REŠENE NALOGE IZ VERJETNOSTI IN STATISTIKE

Martin Raič

Datum zadnje spremembe: 5. april 2024

Kazalo

1. Osnove kombinatorike

Pravilo vsote, pravilo produkta. Variacije, kombinacije in permutacije.

1. Na koliko načinov lahko opremimo dnevno sobo, če imamo na voljo 4 vrste parketa, 3 vrste nelesnih talnih oblog in 5 vrst pohištva?

2. Koliko je:

   a) vseh trimestnih števil?
   b) vseh sodih trimestnih števil?
   c) vseh trimestnih števil s sodo prvo števko?
   d) vseh trimestnih števil s samimi enakimi števkami?
   e) vseh trimestnih števil s samimi različnimi števkami?
   f) vseh trimestnih števil, ki so palindromi?

3. Na koliko načinov lahko iz škatle s petimi različnimi kroglicami vzamemo:

   a) eno kroglico
   b) dve kroglici
   c) tri kroglice

   Pri tem ločite primer, ko kroglice vračamo, in primer, ko jih ne vračamo. Poleg tega ločite primer, ko je vrstni red jemanja pomemben, in primer, ko ni pomemben.

   Pri različici, ko kroglic ne vračamo in vrstni red jemanja ni pomemben, primerjajte rezultata iz točk b) in c).

4. Na koliko načinov lahko na ravno polico razporedimo 3 begonije in 4 fuksije? Pri tem ločite primer, ko razločujemo vse cvetlice, in primer, ko cvetlic iste vrste med seboj ne razločujemo.

Splošneje: iz škatle z n različnimi kroglicami lahko izvlečemo k kroglic na naslednje število načinov:

	vrstni red vlečenja	
	pomemben	ni pomemben
vračamo	${}^{(\mathrm{p})}V_n^k = n^k$	${}^{(\mathrm{p})}C_n^k = \binom{n+k-1}{k}$
ne vračamo	$V_n^k = n(n-1)\ldots(n-k+1)$	$C_n^k = \binom{n}{k} = \dfrac{V_n^k}{k!}$

Velja še $V_n^k = \dfrac{n!}{(n-k)!}$ *in* $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!} = \binom{n}{n-k}$.

5. V posodi je 6 rdečih in 4 modre kroglice, vse kroglice so različne. Na koliko načinov lahko iz posode brez vračanja vzamemo (vrstni red ni pomemben):

   a) 4 rdeče in 2 modri kroglici?

   b) 4 kroglice, a od tega vsaj eno rdečo in vsaj eno modro?

6. Na koliko načinov lahko v vrsto postavimo m miroljubnih in n nasilnih vojakov, če nobena dva nasilna vojaka ne smeta stati skupaj? Vojake med seboj ločimo.

7. Na koliko načinov lahko v ravno vrsto položimo tri brezove, dve leskovi in štiri vrbove šibe, če:

   a) vse šibe razločujemo in ni omejitev?

   b) vse šibe razločujemo ter morajo priti najprej brezove, nato leskove in nazadnje vrbove?

   c) vse šibe razločujemo in morajo biti šibe posamezne vrste skupaj?

   d) šib iste vrste med seboj ne razločujemo in ni omejitev?

8. Na koliko načinov lahko razvrstimo šest otrok (ki jih razločujemo) na vrtiljak s šestimi sedeži (ki jih ločimo le glede na njihovo medsebojno lego)? Kaj pa na vrtiljak z desetimi sedeži? Na vsak sedež gre največ en otrok.

9. [1] 7 moških in 5 žensk se odpravi na taborjenje. Na voljo imajo dva šotora za tri osebe in tri šotore za dve osebi. Vse šotore med seboj ločimo. Na koliko načinov se lahko razporedijo v šotore, če:

   a) ni omejitev?

   b) smejo biti v posameznem šotoru le osebe istega spola?

   c) mora biti v posameznem šotoru najmanj en moški in najmanj ena ženska?

[1] Avtor naloge: Gregor Šega

2. Elementarna verjetnost

Klasična verjetnost, klasična geometrijska verjetnost. Računanje z dogodki.

Klasična verjetnost

Če so vsi izidi enako verjetni, za dogodek A velja:

$$\mathbb{P}(A) = \frac{\textit{število izidov, ki so v } A}{\textit{število vseh izidov}}\,.$$

Temu, da so vse možnosti enako verjetne, pravimo **slepa** *ali* **povsem naključna** *izbira.*
Izbirati dve (splošneje n) stvari na slepo/povsem naključno in **neodvisno** *pa pomeni, da so vse kombinacije možnosti (kjer stvari ločimo) enako verjetne. Z drugimi besedami, to pomeni slepo/povsem naključno izbiro ustreznega urejenega para oz. n-terice.*

1. Vržemo dve neodvisni standardni kocki. Kolikšna je verjetnost, da bo skupno število pik enako 8?

2. Zakonca načrtujeta štiri otroke. Kaj je verjetneje: da bosta oba spola enako zastopana ali da bodo trije enega, eden pa nasprotnega spola? Privzamemo, da sta oba spola pri posameznem rojstvu enako verjetna in da so spoli pri posameznih rojstvih neodvisni.

$$\boxed{\mathbb{P}(A^{\mathrm{c}}) = 1 - \mathbb{P}(A)}$$

3. Vržemo pet neodvisnih standardnih kock. Kolikšna je verjetnost, da bo na vsaj eni kocki padla šestica?

4. V posodi je 5 belih, 4 črne in 3 rdeče kroglice. Iz posode potegnemo tri kroglice. Kolikšna je verjetnost, da bo med njimi po ena kroglica vsake barve, če:

   a) kroglice vračamo?

   b) kroglic ne vračamo?

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)}$$

Dokaz. Ker sta A in $B \setminus A$ nezdružljiva, velja $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Podobno, ker sta $A \cap B$ in $B \setminus A$ nezdružljiva, velja $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$. Ko dobljeni enakosti odštejemo in preoblikujemo, dobimo natančno želeno formulo.

5. Iz posode, v kateri so 3 rdeče, 2 zeleni in 5 belih kroglic, na slepo in brez vračanja izvlečemo dve kroglici. Kolikšna je verjetnost, da je prva rdeča ali pa druga zelena?

Računanje z dogodki

$$A \cup B = B \cup A \qquad A \cap B = B \cap A$$
$$A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cap (B \cap C) = (A \cap B) \cap C$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cup \Omega = \Omega \qquad A \cap \Omega = A$$
$$A \cup \emptyset = A \qquad A \cap \emptyset = \emptyset$$
$$(A \cup B)^{\mathrm{c}} = A^{\mathrm{c}} \cap B^{\mathrm{c}} \qquad (A \cap B)^{\mathrm{c}} = A^{\mathrm{c}} \cup B^{\mathrm{c}}$$
$$A \cup A^{\mathrm{c}} = \Omega \qquad A \cap A^{\mathrm{c}} = \emptyset$$
$$(A^{\mathrm{c}})^{\mathrm{c}} = A$$

6. Poenostavite naslednji izraz z dogodki:

$$(B \cup C) \cap (B \cup C^{\mathrm{c}}) \cap (B^{\mathrm{c}} \cup C)$$

7. Dani so dogodki A, B in C. Matematično zapišite:

   a) dogodek, da se ne zgodi niti A niti B niti C;

   b) dogodek, da se zgodi natanko eden od teh treh dogodkov;

   c) dogodek, da se zgodita vsaj dva od teh treh dogodkov.

   Izračunajte še verjetnosti zgornjih dogodkov, če veste, da je $\mathbb{P}(A) = 0{\cdot}3$, $\mathbb{P}(B) = 0{\cdot}45$, $\mathbb{P}(C) = 0{\cdot}6$, $\mathbb{P}(A \cap B) = 0{\cdot}2$, $\mathbb{P}(A \cap C) = 0{\cdot}2$, $\mathbb{P}(B \cap C) = 0{\cdot}3$ in $\mathbb{P}(A \cap B \cap C) = 0{\cdot}1$.

8. Kolikšna je verjetnost, da v skupini n ljudi obstajata dva, ki imata rojstni dan na isti dan? Prestopna leta zanemarite. Najmanj koliko ljudi mora biti, da je ta verjetnost enaka vsaj 1/2? Zapišite rezultat še za splošno število dni v letu in raziščite asimptotično obnašanje, ko gre le-to proti neskončno.

9. Dan je dobro premešan kup 16 kart, med katerimi so štirje piki. Kolikšna je verjetnost, da sta med prvimi osmimi kartami natanko dva pika?

10. Med 100 izdelki v seriji je 10 okvarjenih. Iz serije na slepo izberemo 10 izdelkov. Če je med njimi več kot en okvarjen, serijo zavrnemo. Kolikšna je verjetnost, da se bo to zgodilo?

11. V posodi je 8 belih, 4 črne in 2 rdeči kroglici. Iz posode brez vračanja potegnemo sedem kroglic. Kolikšna je verjetnost, da bo razmerje barv enako kot v posodi?

12. Pri igri *TikiTaka* Loterije Slovenije prekrižamo od 1 do 10 številk izmed 70. Loterija izžreba 20 številk. Recimo, da se odločimo prekrižati 10 številk. Izračunajte verjetnosti naslednjih dobitkov in jih primerjajte s faktorji izplačil:

- vseh 10 številk je izžrebanih – faktor izplačila: 100.000;
- natanko 5 številk je izžrebanih – faktor izplačila: 2˙5;
- nobena številka ni izžrebana – faktor izplačila: 1.

13. Pri igri Loto na kombinacijskem listku prekrižamo 6 številk izmed 44. Izžreba se 6 rednih številk in še ena dodatna. Kolikšna je verjetnost, da prekrižamo pet redno izžrebanih številk in dodatno?

Če lahko prostor izidov razbijemo na paroma nezdružljive enako verjetne dogodke, jih lahko obravnavamo kot izide: če je dogodek A unija k od n takih dogodkov, je $\mathbb{P}(A) = k/n$.

14. V kupu je 10 kart, od tega dve rdeči, tri zelene in pet belih. Kup dobro premešamo in drugo za drugo brez vračanja vlečemo karte. Izračunajte verjetnosti dogodkov:

    a) da bo prva rdeča karta izvlečena pred prvo zeleno;

    b) da bo prva rdeča karta izvlečena pred zadnjo zeleno.

15. Študenti, ki bodo pisali izpit, se posedejo v tri vrste in tri kolone, tako kot je prikazano spodaj:

Aljaž	Brigita	Cveto
Dragica	Edo	Fani
Gregor	Hana	Iztok

    Asistent povsem naključno izbere tri študente in jih zamenja: prvega premesti na mesto drugega, drugega na mesto tretjega in tretjega na mesto prvega. Kolikšna je verjetnost, da sta Aljaž in Brigita po premestitvi še vedno soseda v isti vrsti?

16. V dobro premešanem kupu 52 kart je 13 src in 13 pikov.

    a) Kolikšna je verjetnost, da so vsa srca med vrhnjimi 27, vsi piki pa med spodnjimi 27 kartami?

    b) Recimo, da so res vsa srca med vrhnjimi 27 in vsi piki med spodnjimi 27 kartami. Kolikšna je pogojna verjetnost, da je 26. karta z vrha srce, 26. karta z dna (tj. 27. karta z vrha) pa pik?

Načelo vključitev in izključitev

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) - \\ - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \cdots - \mathbb{P}(A_{n-1} \cap A_n) + \\ \cdots \\ + (-1)^{n+1}\,\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)\end{aligned}$$

Dokaz z indukcijo. Za $n = 1$ in $n = 2$ velja. Naredimo indukcijski korak z n na $n + 1$. Najprej opazimo:

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) &= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}\big((A_1 \cup A_2 \cup \cdots \cup A_n) \cap A_{n+1}\big) = \\ &= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}\big((A_1 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1})\big)\,.\end{aligned}$$

Po indukcijski predpostavki je:

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n) - \\ - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \cdots - \mathbb{P}(A_{n-1} \cap A_n) + \\ + \mathbf{1}_{A_1 \cap A_2 \cap A_3} + \mathbf{1}_{A_1 \cap A_2 \cap A_4} + \cdots + \mathbf{1}_{A_{n-2} \cap A_{n-1} \cap A_n} - \\ \cdots \\ + (-1)^{n+1}\,\mathbb{P}(A_1 \cap \cdots \cap A_n) + \\ + \mathbb{P}(A_{n+1}) - \\ - \mathbb{P}(A_1 \cap A_{n+1}) - \mathbb{P}(A_2 \cap A_{n+1}) - \cdots + \mathbb{P}(A_n \cap A_{n+1}) - \\ + \mathbb{P}(A_1 \cap A_2 \cap A_{n+1}) + \mathbb{P}(A_1 \cap A_3 \cap A_{n+1}) + \cdots + \mathbb{P}(A_{n-1} \cap A_n \cap A_{n+1}) + \\ \cdots \\ + (-1)^{n+2}\,\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n \cap A_{n+1})\,,\end{aligned}$$

kar je natančno ustrezna desna stran: iz razvoja verjetnosti $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n)$ smo dobili tiste člene, ki ne vsebujejo dogodka A_{n+1}, iz $\mathbb{P}(A_{n+1})$ in razvoja verjetnosti $\mathbb{P}\big((A_1 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1})\big)$ pa tiste člene, ki dogodek A_{n+1} vsebujejo.

17. Na okensko polico povsem naključno razporedimo b begonij, f fuksij in k kalanhoj.

    a) Kolikšna je verjetnost, da so vse begonije skupaj, prav tako pa tudi vse fuksije?

    b) Kolikšna je verjetnost, da niti za begonije niti za fuksije niti za kalanhoje ne velja, da so skupaj?

18. Mama napiše pet različnih pisem in pripravi pet kuvert za ta pisma s samimi različnimi naslovi. Mali Pepček želi pomagati in povsem naključno vtakne pisma v kuverte, v vsako kuverto po eno pismo. Kolikšna je verjetnost, da je vsaj eno pismo v pravi kuverti?

19. Desetkrat vržemo pošteno kocko in pri vsakem metu zabeležimo, koliko pik je padlo. Kolikšna je verjetnost, da bodo na koncu zabeležena vsa možna števila pik (od 1 do 6)?

20. Dano je slučajno zaporedje m števil od 1 do m, vse možnosti so enako verjetne; števila se lahko ponavljajo. Kolikšna je verjetnost, da najdaljše začetno strnjeno strogo naraščajoče podzaporedje vsebuje dano število $l \in \{1, 2, \ldots, m\}$?

    Primeri najdaljših začetnih strnjenih strogo naraščajočih podzaporedij pri $m = 5$:

$$\begin{array}{ll} 1,2,3,4,5: & 1,2,3,4,5 \\ 2,3,5,4,5: & 2,3,5 \\ 2,2,3,5,4: & 2 \end{array}$$

Klasična geometrijska verjetnost

Točka je izbrana **na slepo** *ali* **povsem naključno** *iz množice G, ki je lahko interval, lik, telo ipd., če za vsako merljivo podmnožico $A \subseteq G$ velja:*

$$\mathbb{P}(\textit{točka pripada } A) = \frac{\textit{mera množice } A}{\textit{mera množice } G}\,.$$

Pri tem je mera lahko dolžina, ploščina itd.
Na slepo/povsem naključno in **neodvisno** *izbrati dve točki (splošneje, n točk) pomeni slepo/povsem naključno izbiro njunega urejenega para (oz. n-terice) v ustreznem kartezijskem produktu.*

21. Do šole je štiri minute hoda, vmes pa je semafor, na katerem dve minuti gori zelena, dve minuti pa rdeča luč. Od doma se odpravim pet minut pred začetkom pouka. Kolikšna je verjetnost, da pridem še pravočasno, če se držim predpisov? Kaj pa, če sta na poti dva semaforja? Seveda privzamemo, da je faza semaforja določena povsem naključno (oz. da sta fazi semaforjev določeni povsem naključno in neodvisno).

22. Avtobus odpelje s postaje povsem naključno med 6:55 in 7:05. Študent pa je nagnjen k zamujanju in pride na postajo povsem naključno med 7:00 in 7:07, neodvisno od avtobusa.

    a) Kolikšna je verjetnost, da ujame ta avtobus?
    b) Če želi študent še pravočasno priti na predavanje, mora biti na tem avtobusu najkasneje ob 7:02. Kolikšna je verjetnost, da se to zgodi?

23. Kolikšna je verjetnost, da je povsem naključno izbrana točka v kvadratu bližje robu kot središču kvadrata?

24. *Buffonova*[2] *igla.* Na list papirja z ravnimi vzporednimi črtami, razmaknjenimi za a, na slepo vržemo iglo dolžine b. Kolikšna je verjetnost, da igla seka katero od črt?

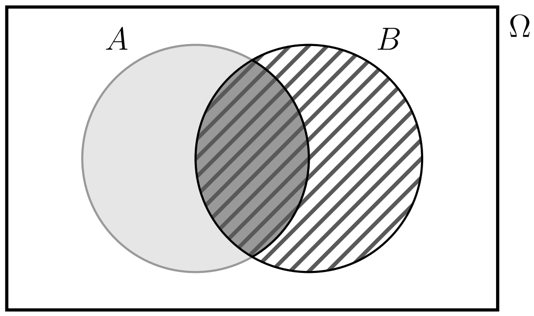
[2] Georges-Louis Leclerc, comte de Buffon (1707–1788), francoski naravoslovec, matematik, kozmolog in filozof; enciklopedist; grof

3. Pogojna verjetnost

Računanje pogojne verjetnosti po definiciji. Izrek o polni verjetnosti, Bayesova formula. Neodvisnost. Zapletenejši primeri pogojne verjetnosti.

Definicija pogojne verjetnosti

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$



Če je dogodek B sestavljen iz samih enako verjetnih izidov, pa je tudi:

$$\mathbb{P}(A \mid B) = \frac{\sharp(A \cap B)}{\sharp(B)}\,.$$

1. Vržemo standardno kocko. Naj bo A dogodek, da padejo vsaj štiri pike, B dogodek, da pade šest pik, L pa dogodek, da pade liho mnogo pik. Izračunajte $\mathbb{P}(A \mid L)$ in $\mathbb{P}(B \mid L)$. Kaj pa, če kocka ni poštena, tako da ena pika pade z verjetnostjo $0{\cdot}3$, izidi z dvema, tremi, štirimi in petimi pikami imajo verjetnost $0{\cdot}15$, šest pik pa pade z verjetnostjo $0{\cdot}1$?

2. Iz dobro premešanega kupa 16 kart, med katerimi so štirje piki, na slepo in brez vračanja izvlečemo štiri karte. Kolikšna je pogojna verjetnost, da je prva med njimi pik, če vemo, da sta med njimi natanko dva pika?

3. V dobro premešanem kupu 52 kart je 13 src in 13 pikov. Recimo, da so vsa srca med vrhnjimi 27 in vsi piki med spodnjimi 27 kartami. Kolikšna je pogojna verjetnost, da je 26. karta z vrha pik, 26. karta z dna (tj. 27. karta z vrha) pa srce?

4. *Bertrandov paradoks.* Dane so tri škatle. V eni sta dva zlata kovanca, v drugi en zlat in en srebrn, v tretji pa dva srebrna. Dovoljeno nam je, da na slepo izberemo en kovanec (tj. vseh šest z enako verjetnostjo). Če uganemo, kakšen je drugi kovanec v škatli, ki smo jo izbrali, dobimo kovanec. Kolikšna je verjetnost, da dobimo kovanec?

   Razmislek: Recimo, da je kovanec zlat. Potem vemo, da je prišel ali iz škatle z dvema zlatima kovancema ali pa iz škatle z enim zlatim in enim srebrnim kovancem. Ker sta obe škatli enako verjetni, je verjetnost, da bomo uganili, enaka $1/2$, ne glede na to, kaj rečemo.

   Je s tem razmislekom vse v redu?

5. *Monty Hallov*[3] *paradoks.* Danih je troje vrat. Za enimi je skrit avto, za preostalimi dvojimi pa buča. Najprej izberemo ena vrata, ki ostanejo zaprta, nakar vodja igre odpre ena izmed vrat, za katerima je buča in ki jih nismo izbrali. Nato nam ponovno ponudi, da izberemo ena izmed še zaprtih vrat. Tisto, kar se skriva za njimi, dobimo.

   Kako naj ravnamo, če želimo dobiti avto? Privzamemo, da so vse možnosti za vrata, za katerimi stoji avto, enako verjetne.

   Razmislek: Recimo, da smo najprej pokazali na prva vrata, vodja igre pa je nato odprl tretja vrata. Prva in druga vrata so še zaprta. Ker so vsa vrata enako verjetna, je pri obojih verjetnost, da bo zadaj avto, enaka $1/2$. Torej je čisto vseeno, kaj storimo.

   Je s tem razmislekom vse v redu?

6. Dana je naslednja različica Monty Hallovega paradoksa: za enimi izmed trojih vrat je skrita nagrada, za preostalimi dvojimi ni ničesar. Igralec najprej pokaže na ena vrata, nakar vodja igre odpre ena izmed vrat, za katerima ni ničesar in na katera igralec ni pokazal. Nato igralec izbere ena izmed še zaprtih vrat in jih odpre (lahko torej ostane pri istem ali pa se premisli). Če se za temi vrati skriva nagrada, jo dobi, sicer ne dobi ničesar.

   Nagrada je za prvimi vrati z verjetnostjo $\frac{1}{2}$, za drugimi vrati z verjetnostjo $\frac{3}{10}$ in za tretjimi z verjetnostjo $\frac{1}{5}$. Nadalje so v tabeli na desni prikazane verjetnosti, da vodja igre odpre j-ta vrata, če je igralec pokazal na i-ta vrata in je za temi vrati nagrada.

$i\backslash j$	1	2	3
1	0	$\frac{1}{4}$	$\frac{3}{4}$
2	$\frac{3}{4}$	0	$\frac{1}{4}$
3	$\frac{1}{4}$	$\frac{3}{4}$	0

   a) Recimo, da igralec najprej pokaže na prva vrata, vodja igre pa odpre druga vrata. Kolikšna je pogojna verjetnost, da igralec dobi nagrado, če ostane pri istem? Kolikšna pa, če se premisli? Kaj od tega se mu bolj splača? Izračunajte še za primer, ko vodja igre odpre tretja vrata (igralec pa še vedno najprej pokaže na prva vrata).

   b) Recimo, da igralec pokaže na prva vrata, nakar glede na to, katera vrata odpre vodja igre, igra tako, kot se mu bolj splača. Kolikšna je verjetnost, da bo dobil nagrado?

   c) Se igralcu splača v prvem koraku pokazati na katera druga vrata?

7. Janez in Peter igrata namizni tenis. V vsaki rundi nekdo zmaga in oba sta enakovredna, ne glede na zgodovino, igrata pa, dokler eden od njiju ne dobi šest rund. Trenutni izid je $4:2$ za Janeza. Kolikšna je verjetnost, da bo Janez dobil cel dvoboj?

[3] Monte Halparin (1921–2017), bolj znan pod odrskim imenom Monty Hall, kanadsko-ameriški radijski in televizijski voditelj, producent in filantrop

Izrek o polni verjetnosti

Dogodki $H_1, H_2, \ldots, H_n$ *oziroma* $H_1, H_2, H_3, \ldots$ *tvorijo* **popoln sistem dogodkov**, *če tvorijo particijo množice* Ω, *tj. vedno se zgodi natanko eden izmed teh dogodkov. Če je še* $\mathbb{P}(H_i) > 0$ *za vse* i, *za poljuben dogodek* A *velja:*

$$\mathbb{P}(A) = \mathbb{P}(H_1)\,\mathbb{P}(A \mid H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A \mid H_2) + \cdots + \mathbb{P}(H_n)\,\mathbb{P}(A \mid H_n)$$

oziroma

$$\mathbb{P}(A) = \mathbb{P}(H_1)\,\mathbb{P}(A \mid H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A \mid H_2) + \mathbb{P}(H_3)\,\mathbb{P}(A \mid H_3) + \cdots$$

Dogodkom H_i *često pravimo* **hipoteze**.

8. V prvi posodi je 5 belih, 3 rdeče in 2 črni kroglici, v drugi posodi pa so tri bele in tri črne kroglice. Iz prve posode v drugo na slepo premestimo eno kroglico, nato pa iz druge na slepo in brez vračanja potegnemo dve kroglici. Kolikšna je verjetnost, da je med njima ena bela in ena črna?

9. Če je $\mathbb{P}(A \mid B) = \mathbb{P}(A \mid C) = p$, ali je nujno tudi $\mathbb{P}(A \mid B \cup C) = p$? Morda potrebujemo kakšen dodaten pogoj? Bi se dalo to posplošiti na več dogodkov?

10. Janez in Peter spet igrata namizni tenis. Spet v vsaki rundi nekdo zmaga, a tokrat Janez dobi posamezno rundo z verjetnostjo $1/3$ (ne glede na zgodovino), igrata pa na dve točki razlike. Kolikšna je zdaj verjetnost, da bo Janez dobil dvoboj? Le-to zdaj računamo od začetka, tj. izida $0 : 0$.

    Opomba. Strogo gledano gre za verjetnost dogodka, da se dvoboj nekoč konča in Janez dobi dvoboj.

11. Pri določenem slučajnem poskusu lahko med drugim pride do opažanja A in do opažanja B. Lahko pride tudi do obeh opažanj, to označimo z $A \cap B$. Poskus ponavljamo in pri vsaki izvedbi pride do opažanja A z verjetnostjo $\mathbb{P}_1(A)$, do opažanja B z verjetnostjo $\mathbb{P}_1(B)$ in do obeh opažanj z verjetnostjo $\mathbb{P}_1(A \cap B)$, ne glede na zgodovino. Privzemimo, da je $\mathbb{P}_1(B) > 0$.

    Poskus ponavljamo, dokler ne pride do opažanja B. Kolikšna je verjetnost, da pri zadnjem poskusu pride tudi do opažanja A?

    Opomba. Strogo gledano gre za verjetnost dogodka, da se izvajanje poskusov nekoč konča in da pri zadnjem poskusu pride do opažanja A.

Pri izreku o polni verjetnosti lahko malo popustimo pri predpostavki, da dogodki $H_1, H_2, \ldots, H_n$ *oziroma* $H_1, H_2, \ldots$ *tvorijo particijo množice* Ω: *dovolj je, da tvorijo* **skorajšnjo** *particijo, kar pomeni, da so paroma nezdružljivi in velja* $\sum_{k=1}^{n} \mathbb{P}(H_k) = 1$ *oziroma* $\sum_{k=1}^{\infty} \mathbb{P}(H_k) = 1$.

12. Iz posode, v kateri je r rdečih, z zelenih in b belih kroglic, na slepo in z vračanjem vlečemo kroglice. Kolikšna je verjetnost, da prvi kroglici, ki ni bela, sledi kroglica iste barve? Privzamemo, da je $r + z > 0$.

    Opomba. Strogo gledano gre za verjetnost dogodka, da nekoč izvlečemo kroglico, ki ni bela, in da ji sledi kroglica iste barve.

13. Mečemo pošten kovanec, pri čemer privzamemo, da je verjetnost, da v posameznem metu pade grb, enaka $1/2$ ne glede na prejšnje mete. Kolikšna je verjetnost, da v prvih n metih *nista* padli dve zaporedni cifri?

Bayesova[4] formula

Če $H_1, H_2, \ldots, H_n$ oziroma $H_1, H_2, H_3, \ldots$ tvorijo popoln sistem dogodkov in če je $\mathbb{P}(H_i) > 0$ za vse i, za poljuben dogodek A s $\mathbb{P}(A) > 0$ velja:

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(H_i)\,\mathbb{P}(A \mid H_i)}{\mathbb{P}(H_1)\,\mathbb{P}(A \mid H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A \mid H_2) + \cdots + \mathbb{P}(H_n)\,\mathbb{P}(A \mid H_n)}$$

oziroma

$$\mathbb{P}(H_i \mid A) = \frac{\mathbb{P}(H_i)\,\mathbb{P}(A \mid H_i)}{\mathbb{P}(H_1)\,\mathbb{P}(A \mid H_1) + \mathbb{P}(H_2)\,\mathbb{P}(A \mid H_2) + \mathbb{P}(H_3)\,\mathbb{P}(A \mid H_3) + \cdots}\,.$$

Brezpogojnim verjetnostim $\mathbb{P}(H_i)$ pravimo **apriorne**, *pogojnim verjetnostim $\mathbb{P}(H_i \mid A)$ pa* **aposteriorne** *verjetnosti hipotez.*

14. Žena pošlja moža na trg po solato, ki jo prodajata dve branjevki, Francka in Micka. Verjetnost, da mož kupi solato pri Francki, je 60%, verjetnost, da kupi pri Micki, pa 40%. Francka ima 10%, Micka pa 20% nagnitih glav solate. Mož prinese domov nagnito glavo solate. Katero branjevko lahko žena bolj upravičeno osumi, da mu je prodala nagnito solato? Privzamemo, da branjevki solato izbirata na slepo.

15. Matičnemu podjetju dobavljajo trije kooperanti: kooperant Alfa Deli dobavlja 20%, kooperant Bobo Deli 50%, kooperant Centro Deli pa 30% vseh delov. Kooperant Alfa Deli ima 5%, Bobo Deli 1%, Centro Deli pa 2% okvarjenih delov. Kontrolor v matičnem podjetju preizkusi povsem naključno izbran del in izkaže se, da je okvarjen, zato zavzdihne: "Oh, že spet ti Alfa Deli!" Kolikšna je verjetnost, da je bil del dobavil kooperant Alfa Deli?

16. V prvi posodi je 6 belih in 4 rdeče kroglice, v drugi pa ena bela in ena rdeča. Najprej na slepo premestimo 3 kroglice iz prve posode v drugo, nato pa iz druge posode potegnemo dve kroglici (na slepo in brez vračanja). Obe sta rdeči. Kolikšna je pogojna verjetnost, da so bile vse tri premeščene kroglice rdeče?

[4] Thomas Bayes (ok. 1701–1761), angleški statistik, filozof in duhovnik

17. Manja ugiba neznano besedo, ki ima štiri soglasnike in en samoglasnik. Črke se ji prikazujejo druga za drugo, vrstni red prikazovanja pa je izbran povsem naključno. Ko so črke enkrat prikazane, ostanejo na zaslonu. V tabeli na desni so prikazane pogojne verjetnosti, da Manja ugane besedo po določenem številu prikazanih soglasnikov in samoglasnikov, če besede prej še ni uganila. Te pogojne verjetnosti veljajo ne glede na to, katere črke so bile prikazane prej.

sogl.	samogl.	ugane
0	0	0·05
0	1	0·1
1	0	0·15
1	1	0·2
2	0	0·4
2	1	0·6
3	0	0·7

   a) Kolikšna je verjetnost, da Manja ugane besedo po treh prikazanih črkah (ne pa tudi prej)?
   b) Recimo, da je Manja uganila besedo po treh prikazanih črkah. Kolikšna je pogojna verjetnost, da so bili to sami soglasniki?

Dogodka A in B sta **neodvisna**, *če velja:*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\ \mathbb{P}(B)$$

Če je $\mathbb{P}(B) > 0$, je to ekvivalentno pogoju, da je $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.
Če je $0 < \mathbb{P}(B) < 1$, pa je to ekvivalentno tudi pogoju, da je $\mathbb{P}(A \mid B) = \mathbb{P}(A \mid B^{\mathrm{c}})$.
Dogodki $A_1, A_2, A_3 \ldots$ so neodvisni, če za poljubne različne indekse $i_1, i_2, \ldots, i_k$ velja:

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\ \mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k})\,.$$

18. Na kupu so štiri karte: pikov kralj, pikova dama, srčev kralj in srčeva dama. Na slepo izvlečemo eno izmed kart. Definirajmo naslednje dogodke:

$$\begin{aligned} A &:= \{\text{izvlekli smo pika}\} \\ B &:= \{\text{izvlekli smo damo}\} \\ C &:= \{\text{izvlekli smo srčevega kralja ali pikovo damo}\} \end{aligned}$$

   Sta dogodka A in B neodvisna? Kaj pa A in C? Kaj pa B in C? Kako pa je z dogodki A, B in C, so neodvisni?

19. Danih je osem kart: as, kralj, dama, fant, 10, 9, 8 in 7. Na slepo izvlečemo eno karto. Definirajmo naslednje dogodke:

$$\begin{aligned} F &:= \{\text{karta je as, kralj, dama ali fant}\} \\ G &:= \{\text{karta je as, kralj, 10 ali 9}\} \\ H &:= \{\text{karta je as, dama, 8 ali 7}\} \end{aligned}$$

   So dogodki F, G in H neodvisni?

20. Vržemo tri kovance. Meti so med seboj neodvisni, verjetnosti, da pade grb, pa niso nujno enake. Naj bo A dogodek, da se na prvem kovancu pojavi grb, B pa dogodek, da se grb pojavi na natanko dveh kovancih.

    a) Recimo, da so vsi trije kovanci pošteni, se pravi, da je verjetnost za grb pri vseh kovancih enaka $1/2$. Sta dogodka A in B neodvisna?

    b) Recimo, da je prvi kovanec pošten, druga dva pa ne: na vsakem od njiju se grb pojavi z verjetnostjo p. Pri katerih p sta A in B neodvisna?

21. Standardno kocko n-krat vržemo, meti so neodvisni. Označimo jih z $1, 2, \ldots, n$. Za $k = 1, 2, \ldots, n$ označimo z A_k dogodek, da je v k-tem metu prvič padla ena pika, z B_l pa označimo dogodek, da je v l-tem metu zadnjič padlo šest pik. Pri katerih k in l sta dogodka A_k in B_l neodvisna?

22. *Simpsonov paradoks.* Dve zdravili so preizkušali na ženskah in moških. Rezultati so naslednji:

zdravljenje	ženske		moški	
	Prvo zdravilo	Drugo zdravilo	Prvo zdravilo	Drugo zdravilo
uspelo	200	10	190	1000
ni uspelo	1800	190	10	1000

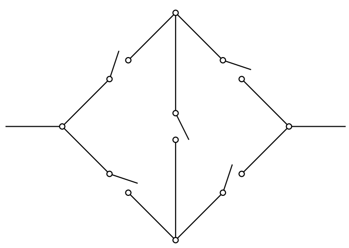
Katero zdravilo je bilo uspešnejše:

- pri ženskah?
- pri moških?
- pri obojih skupaj?

Komentirajte!

Če so $A_1, A_2, \ldots$ in $B_1, B_2, \ldots$ neodvisni dogodki, tudi za poljuben dogodek $A \in \sigma(A_1, A_2, \ldots)$ velja, da so dogodki $A, B_1, B_2, \ldots$ neodvisni. Pri tem je $\sigma(A_1, A_2, \ldots)$ najmanjša σ-algebra, ki vsebuje dogodke $A_1, A_2, \ldots$

23. V vezju, ki ga prikazuje spodnja skica, vsako stikalo prepušča električni tok z verjetnostjo $1/3$, posamezna stikala pa so med seboj neodvisna. Kolikšna je verjetnost, da vezje prepušča tok?



24. Janez, Francelj in Tone gredo streljat zajce. Janez zadene z verjetnostjo 0˙1, Francelj z verjetnostjo 0˙2, Tone pa z verjetnostjo 0˙3, neodvisno drug od drugega.

    a) Vsi pomerijo, ustrelijo in zajec je zadet. Kolikšna je pogojna verjetnost, da ga je Janez zadel?

    b) Ko pridejo do zajca, se izkaže, da ga je zadel natanko eden. Kolikšna je pogojna verjetnost, da je bil to Janez?

25. Andraž, Bojan, Cilka in Darja streljajo v tarčo. Andraž in Bojan streljata z modrimi, Cilka in Darja pa z rdečimi puščicami. Andraž zadene z verjetnostjo 0˙6, Bojan z verjetnostjo 0˙7, Cilka z verjetnostjo 0˙5, Darja pa z verjetnostjo 0˙9. Vsi hkrati pomerijo in ustrelijo, neodvisno drug od drugega. V tarči se znajdeta ena modra in ena rdeča puščica. Kolikšna je pogojna verjetnost, da sta to Bojanova in Darjina?

26. Študent se od 50 izpitnih vprašanj nauči le 30. Za vsako vprašanje, ki se ga nauči, je potem še 30% verjetnosti, da pozabi odgovor, za vsako vprašanje, ki se ga ne nauči, pa je še 10% verjetnosti, da odgovor ugane. Privzamemo, da so dogodki, da študent posamezno vprašanje pozabi oz. ugane odgovor nanj, neodvisni. Na izpitu dobi tri povsem naključno izbrana vprašanja in izpit naredi, če pravilno odgovori na vsaj dve vprašanji. Kolikšna je verjetnost, da bo naredil izpit?

27. Miha se odpravi na obisk k vinogradnikom Janezu, Lojzu in Štefanu. Vsak mu ponudi cviček ali šmarnico. Janez mu ponudi šmarnico z verjetnostjo 60%, Lojz z verjetnostjo 40%, Štefan pa z verjetnostjo 10%. Če mu noben vinogradnik ne ponudi šmarnice, Miho boli glava z verjetnostjo 10%. Če mu šmarnico ponudi natanko en vinogradnik (ne glede na to, kdo), Miho boli glava z verjetnostjo 40%. Če mu šmarnico ponudita natanko dva (ne glede na to, katera dva), Miho boli glava z verjetnostjo 70%. Če pa mu šmarnico ponudijo vsi trije, Miho zagotovo boli glava.

    Naslednji dan Miho boli glava in prijatelj Tone mu pravi: "Janez in Lojz sta ti gotovo dala šmarnico!" Kolikšna je pogojna verjetnost, da ima prav? Privzamemo, da vinogradniki izberejo vrsto vina neodvisno drug od drugega.

28. V kolektivu je n delavcev, med katerimi je tudi Zdravko. Vsi opravijo test za novi koronavirus SARS-Cov-2. Pri okuženih je test pozitiven z verjetnostjo a, pri neokuženih pa z verjetnostjo $1 - b$. Številu a pravimo *občutljivost*, številu b pa *specifičnost* testa. Vsak od delavcev je okužen z verjetnostjo p, neodvisno od vseh ostalih. Privzamemo še, da so testiranja posameznih oseb med seboj neodvisna.

    a) Kolikšna je verjetnost, da je Zdravkov test pozitiven?

    b) Izkaže se, da je bilo pozitivnih natanko k delavcev, a ni znano, kateri so to bili. Kolikšna je pogojna verjetnost, da je Zdravko *okužen*?

29. Pustolovec Albert pride v tujo deželo, kjer ga takoj primejo in vtaknejo v ječo. Po prvi noči, prebiti v ječi, ga obišče kralj in mu ponudi posodo, v kateri je ena rdeča in ena zelena kroglica. Albert na slepo izvleče eno kroglico. Če izvleče zeleno, je

izpuščen, če izvleče rdečo, pa mora prebiti v ječi še eno noč. Naslednji dan ga spet obišče kralj in spet mu ponudi posodo, le da sta tokrat notri dve rdeči in ena zelena kroglica. Spet je Albert izpuščen, če izvleče zeleno kroglico, sicer pa mora ponovno prespati v ječi. Tako se nadaljuje: vsak dan je v posodi ena rdeča kroglica več.

a) Dokažite, da Albert z verjetnostjo ena nekoč pride iz ječe.

b) Ko Alberta izpustijo, mu kralj izroči posodo s kroglicami (n rdečimi in eno zeleno, če je Albert v ječi prespal n-krat). Albert nato sam takoj izvleče eno kroglico. Če je zelena, takoj zapusti deželo, sicer pa izvlečeno rdečo kroglico odvrže in tam prespi (tokrat na svobodi). Nato spet vleče kroglice (tokrat z eno rdečo manj) in če izvleče zeleno, deželo zapusti, sicer pa ponovno prespi. Tako nadaljuje, vsakič z eno rdečo kroglico manj.

   Recimo, da je Albert v tej deželi prespal natanko petkrat. Kolikšna je pogojna verjetnost, da je v ječi prespal trikrat?

30. Iz posode, v kateri je sprva a rdečih in b belih kroglic, vlečemo kroglice. Če izvlečemo rdečo, končamo, če izvlečemo belo, pa jo vrnemo v posodo in dodamo še d novih belih kroglic. Seveda vsakič vlečemo na slepo. Kolikšna je verjetnost, da nikoli ne nehamo vleči?

    Opomba: tovrstni protokoli vlečenja kroglic so znani kot *Pólyeva žara*.[5]

31. Miranda je na nočni zabavi spoznala Ferdinanda. V dneh po zabavi čaka na njegov klic. Verjetnost, da jo Ferdinand prvič pokliče k-ti dan po zabavi, je enaka 3^{-k}. Vsako noč, ki sledi dnevu, ko Ferdinand Mirande ne pokliče, Miranda spozna novega fanta z verjetnostjo $1/10$.

    Recimo, da je Ferdinand poklical Mirando. Kolikšna je pogojna verjetnost, da je, preden jo je prvič poklical, že spoznala novega fanta? Privzamemo, da Miranda fante spoznava le ponoči in da Ferdinand na posamezen dan pokliče Mirando neodvisno od tega, ali je prej spoznala novega fanta ali ne.

[5] György Pólya (1887–1985), madžarski matematik judovskega rodu

4. Slučajne spremenljivke

Pojem porazdelitve, kumulativna porazdelitvena funkcija, porazdelitvena shema diskretno porazdeljene slučajne spremenljivke, porazdelitvena gostota zvezno porazdeljene slučajne spremenljivke. Ugotavljanje in prepoznavanje porazdelitev. Približni obrazci za binomsko porazdelitev. Vrstilne karakteristike. Transformacije (funkcije) slučajnih spremenljivk. Generiranje slučajnih spremenljivk.

Diskretne slučajne spremenljivke

Slučajna spremenljivka je **diskretna**, *če zavzame vrednosti le na števni množici, tj. bodisi končni množici* $\{a_1, a_2, \ldots, a_n\}$ *bodisi števno neskončni množici* $\{a_1, a_2, a_3, \ldots\}$.

Porazdelitev vsake take slučajne spremenljivke X *lahko opišemo s* **porazdelitveno shemo**: *če so vse vrednosti* $a_1, a_2, \ldots$ *različne, pišemo:*

$$X \sim \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix} \qquad \textit{oziroma} \qquad X \sim \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{pmatrix},$$

kar pomeni $\mathbb{P}(X = a_1) = p_1$, $\mathbb{P}(X = a_2) = p_2$ *itd. Velja:*

$$p_1 + p_2 + \cdots + p_n = 1 \qquad \textit{oziroma} \qquad p_1 + p_2 + p_3 + \cdots = 1\,.$$

Slučajna spremenljivka X *je* **diskretno porazdeljena**, *če se njena porazdelitev ujema s porazdelitvijo neke diskretne slučajne spremenljivke. To je natanko tedaj, ko obstaja taka števna množica* S, *da je* $\mathbb{P}(X \in S) = 1$, *in natanko tedaj, ko je* $\sum_x \mathbb{P}(X = x) = 1$ *(v splošnem je namreč le* $\sum_x \mathbb{P}(X = x) \le 1$*).*

1. Vržemo standardno kocko in število pik, ki padejo, označimo z X. Zapišite porazdelitev te slučajne spremenljivke.

Diskretna enakomerna porazdelitev *na* n*-elementni množici* $\{a_1, a_2, \ldots, a_n\}$ *je porazdelitev na slepo/povsem naključno izbranega elementa te množice, tj. porazdelitev s shemo:*

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Označevali jo bomo z $\mathrm{Unif}\{a_1, a_2, \ldots, a_n\}$.

2. Neodvisno vržemo dva poštena kovanca in standardno kocko. Za vsako piko na kocki dobimo en evro, za vsako cifro na kovancu pa dva evra. Slučajna spremenljivka S naj predstavlja skupni znesek v evrih, ki ga dobimo. Zapišite njeno porazdelitev.

3. Slučajna spremenljivka X zavzame vrednosti v množici $\{1, 2, \ldots, 10\}$. Verjetnost, da je X enaka določenemu številu iz te množice, je premo sorazmerna s tem številom. Izračunajte $\mathbb{P}(X > 3)$.

4. Naj bo X število šestic, ki padejo v desetih neodvisnih metih standardne kocke. Zapišite porazdelitev te slučajne spremenljivke.

Bernoullijevo[6] **zaporedje poskusov** *je zaporedje neodvisnih slučajnih poskusov, od katerih lahko vsak uspe ali ne uspe, in sicer vsak poskus uspe z isto verjetnostjo.*
Binomska porazdelitev $\mathrm{Bin}(n,p)$ *je porazdelitev števila uspelih poskusov v Bernoullijevem zaporedju n poskusov, od katerih vsak uspe z verjetnostjo p. Če je $X \sim \mathrm{Bin}(n,p)$, velja:*

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}\,; \quad k = 0, 1, \ldots, n\,.$$

5. Šestkrat vržemo nepošten kovanec, pri katerem grb pade z verjetnostjo $1/3$. Meti so med seboj neodvisni. Kolikšna je verjetnost, da padeta več kot dva grba?

6. Dani sta dve posodi, v vsaki je po n kroglic. Na vsakem koraku iz vsake posode na slepo in neodvisno izberemo po eno kroglico, tako da so izbire neodvisne tudi od prejšnjih izbir. Kroglici nato zamenjamo, tako da imamo v vsaki posodi spet po n kroglic. Kolikšna je verjetnost, da je posamezna kroglica po k korakih spet v isti posodi kot na začetku? Poskusite zapisati z zaključenim izrazom.

   Opomba. Ta "difuzijski model" je leta 1769 opisal švicarski matematik in fizik Daniel Bernoulli (1700–1782), sin matematika Johanna Bernoullija (1667–1748), brata Jakoba Bernoullija (1655–1705).

Aproksimacija točkastih verjetnosti pri binomski porazdelitvi

Naj bo $X \sim \mathrm{Bin}(n,p)$ in $n \to \infty$ ter še $k \in \mathbb{N}_0$. Če gre $p \to 0$, velja **Poissonov**[7] **obrazec**:

$$\mathbb{P}(X = k) \approx \frac{(np)^k e^{-np}}{k!}\,;$$

za majhno relativno napako zahtevamo še $|k - np| \ll \sqrt{n}$.
Če pa je $p, 1-p \gg 1/n$ (ali, ekvivalentno, $\sigma := \sqrt{np(1-p)} \to \infty$), velja **Laplaceova**[8] **lokalna formula**:

$$\mathbb{P}(X = k) \approx \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(k-np)^2/(2\sigma^2)}\,;$$

za majhno relativno napako zahtevamo še $|k - np| \ll \sigma^{4/3}$.

Produkt np je **pričakovano** *število uspelih poskusov, količina σ pa je njegov* **standardni odklon**.

Meja med smotrnostjo uporabe Poissonovega obrazca in Laplaceove lokalne formule je za velike n približno pri $p = 0{\cdot}6/\sqrt[3]{n}$.

[6] Jakob Bernoulli (1655–1705), švicarski matematik

V okviru dometa aproksimacij lahko relativne napake pri aproksimaciji točkastih verjetnosti $\mathbb{P}(X = k)$ navzgor omejimo s količinami naslednjih velikostnih redov:

- pri Poissonovi aproksimaciji: $p + \frac{(k-np)^2}{n}$;
- pri Laplaceovi lokalni formuli: $\dfrac{1}{\sigma} + \dfrac{|k-np|^3}{\sigma^4}$;
- pri Laplaceovi lokalni formuli, če je $k \in \mathbb{Z} + 1/2$: $\dfrac{1 + |k-np|}{\sigma^2} + \dfrac{|k-np|^3}{\sigma^4}$.

Izboljšave aproksimacij (asimptotski razvoj 1. reda):

- pri Poissonovem obrazcu: $\mathbb{P}(X = k) \approx \dfrac{(np)^k}{k!} \exp\left(-np + \dfrac{k - (k-np)^2}{2n}\right) \approx \dfrac{(np)^k}{k!} \exp\left(-np + \dfrac{p - (k-np)^2}{2n}\right)$;
- pri Laplaceovi lokalni formuli: $\mathbb{P}(X = k) \approx \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{x^2}{2} + \dfrac{2p-1}{6\sigma}(3x - x^3)\right)$;
- pri Laplaceovi integralski formuli: $\mathbb{P}(X < k) \approx \dfrac{1}{2} + \Phi\left(x + \dfrac{2p-1}{6\sigma}(x^2 - 1)\right)$ (za $k \in \mathbb{Z} + 1/2$);

Označili smo $x = (k-np)/\sigma$. Iz zgornjih izboljšanih aproksimacij lahko izpeljemo asimptotično obnašanje napake pri Poissonovi aproksimaciji in pri Laplaceovi lokalni formuli, če je $1/n \ll p \ll 1$:

- $\displaystyle \max_k \left|\mathbb{P}(X = k) - \frac{(np)^k e^{-np}}{k!}\right| \sim \frac{p}{2\sigma\sqrt{2\pi}} \max_x \left|1 - x^2\right| e^{-x^2/2} = \frac{p}{2\sigma\sqrt{2\pi}} \doteq 0{\cdot}199\, \frac{p}{\sigma}$;
- $\displaystyle \sum_k \left|\mathbb{P}(X = k) - \frac{(np)^k e^{-np}}{k!}\right| \sim \frac{p}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} \left|1 - x^2\right| e^{-x^2/2}\, \mathrm{d}x = \frac{2p}{\sqrt{2\pi e}} \doteq 0{\cdot}484\, p$;
- $\displaystyle \max_k \left|\mathbb{P}(X = k) - \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(k-np)^2}{2\sigma^2}\right)\right| \sim \frac{1}{6\sigma^2\sqrt{2\pi}} \max_x \left|3x - x^3\right| e^{-x^2/2} = \frac{1}{\sigma^2\sqrt{2\pi}} \sqrt{\frac{3-\sqrt{6}}{6}}\, e^{-(3-\sqrt{6})/2} \doteq \frac{0{\cdot}0918}{\sigma^2}$;
- $\displaystyle \sum_k \left|\mathbb{P}(X = k) - \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(k-np)^2}{2\sigma^2}\right)\right| \sim \frac{1}{6\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \left|3x - x^3\right| e^{-x^2/2}\, \mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}} \, \frac{1 + 4\, e^{-3/2}}{3} \doteq \frac{0{\cdot}252}{\sigma}$.

Asimptotična meja med smotrnostjo uporabe Poissonove in Laplaceove aproksimacije bo torej:

- če gledamo maksimalno absolutno napako: pri $p = \left(\frac{2(3-\sqrt{6})}{3}\, e^{-(3-\sqrt{6})}\right)^{1/3} n^{-1/3} \doteq 0{\cdot}596\, n^{-1/3}$;
- če gledamo vsoto absolutnih napak: pri $p = \left(\frac{e^{1/2} + 4\, e^{-1}}{6}\right)^{2/3} n^{-1/3} \doteq 0{\cdot}647\, n^{-1/3}$.

7. Naj bo X spet število šestic, ki padejo v desetih neodvisnih metih standardne kocke. Preverite, kako natančna sta Poissonov obrazec in Laplaceova lokalna formula pri izračunu $\mathbb{P}(X = 1)$.

8. 50-krat vržemo nepošten kovanec, pri katerem je verjetnost, da pade grb, enaka $0{\cdot}4$. Meti so neodvisni. Kolikšna je verjetnost, da bo padlo natanko 20 grbov? Kolikšna pa je verjetnost, da bo padlo natanko 25 grbov?

   Točen rezultat primerjajte z rezultatoma, dobljenima po Poissonovem obrazcu in po Laplaceovi lokalni formuli.

9. Verjetnost, da uporabnik stranišča potegne vodo, je $0{\cdot}99$. Kolikšna je verjetnost, da se pri 1000 uporabah voda potegne natanko 990-krat?

10. Enota nujne medicinske pomoči pokriva večje število prebivalcev in ima v povprečju 10 obiskov na dan. Ocenite verjetnost, da bo imela v določenem dnevu več kot 15 obiskov. Poenostavljeno privzamemo, da vsak prebivalec potrebuje nujno medicinsko pomoč z isto verjetnostjo in neodvisno od drugih prebivalcev.

---

[7] Siméon Denis Poisson (1781–1840), francoski matematik, geometer in fizik

[8] Pierre-Simon de Laplace (1749–1827), francoski matematik, astronom, fizik in politik

Aproksimacija intervalskih verjetnosti pri binomski porazdelitvi

Če je $X \sim \mathrm{Bin}(n, p)$, $a \le b$ *in je* $p, 1-p \gg 1/n$ *(ali, ekvivalentno,* $\sigma := \sqrt{np(1-p)} \to \infty$*), velja* **Laplaceova integralska formula***:*

$$\mathbb{P}(a < X < b) \approx \mathbb{P}(a \le X \le b) \sim \Phi\left(\frac{b-np}{\sigma}\right) - \Phi\left(\frac{a-np}{\sigma}\right) \; ;$$

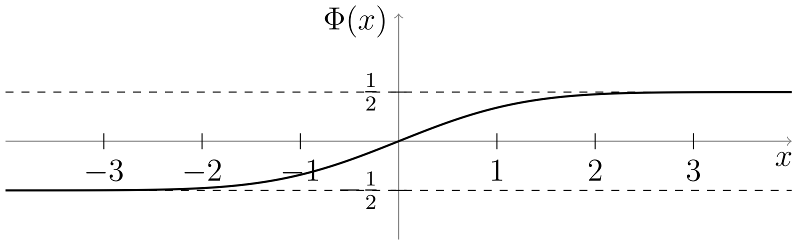
za majhno relativno napako zahtevamo še:

- $|a - np| \ll \sigma^{4/3}$ *ali* $|b - np| \ll \sigma^{4/3}$*;*
- $a, b \in \mathbb{Z} + \frac{1}{2}$ *ali* $b - a \gg 1$.

Funkcija Φ *je* **Gaussov**[9] **verjetnostni integral**:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,\mathrm{d}t$$

in je liha. Graf:



V literaturi so definicije funkcije Φ *različne, zato je treba paziti!*

11. Hotelski kompleks ima 1600 sob. Gostje jih vse naenkrat rezervirajo, a vsaka rezervacija je z verjetnostjo 10% odpovedana. Odpovedi so med seboj neodvisne.

    a) Kolikšna je verjetnost, da bo zasedenih od 1420 do 1450 sob, pri čemer sta obe meji vključeni?
    b) Kolikšna je verjetnost, da bo zasedenih manj kot 1420 sob?
    c) Uprava se odloči prenoviti sobe, a glede na možne odpovedi špekulira, da najbrž ni treba prenoviti vseh. Pripravljena je sprejeti 5% tveganje, da bo treba kakšnega gosta namestiti v neprenovljeno sobo. Najmanj koliko sob morajo prenoviti?
    d) Uprava končno prenovi vse sobe, a spet špekulira, da lahko najbrž sprejme več rezervacij, kot je sob, in je pripravljena sprejeti 5% tveganje, da ne bodo mogli sprejeti vseh gostov. Največ koliko rezervacij lahko sprejmejo?

12. V deželi razsaja epidemija, okuženih je 3% prebivalcev. Obravnavajte različne metode (približnega) izračuna verjetnosti, da sta med 100 neodvisnimi mimoidočimi več kot dva okužena: katere so smiselne, katere hitrejše in katere natančnejše.

[9] Carl Friedrich Gauß (1777–1855), nemški matematik

13. Verjetnost, da je določen izdelek prvovrsten, je 60%. Proizvajalec jih pakira v pošiljke za velegrosista in na pošiljko napiše deklaracijo, da je prvovrstnih izdelkov najmanj 59%. Najmanj koliko izdelkov naj ima pošiljka, če naj deklaracija drži z verjetnostjo najmanj 0˙99? Seveda privzamemo, da so posamezni izdelki med seboj neodvisni.

14. Verjetnost, da je izdelek prvovrsten, je 10%. Najmanj koliko izdelkov približno moramo naročiti, če naj bo med naročenimi izdelki z verjetnostjo najmanj 0˙95 vsaj 100 izdelkov prvovrstnih?

15. Naj bo X število metov standardne kocke, ki jo mečemo, dokler ne pade šestica. Meti so med seboj neodvisni. Zapišite in poimenujte porazdelitev slučajne spremenljivke X.

Geometrijska porazdelitev *je porazdelitev na* $\mathbb{N}$, *pri kateri točkaste verjetnosti tvorijo geometrijsko zaporedje. Natančneje, zapis* $X \sim \mathrm{Geom}(p)$, *kjer je* $0 < p \le 1$, *pomeni:*

$$\mathbb{P}(X = k) = p(1-p)^{k-1}\,; \quad k \in \mathbb{N}\,.$$

Geometrijska porazdelitev je tudi porazdelitev števila poskusov do vključno prvega uspelega, če izvajamo Bernoullijevo zaporedje poskusov, pri katerih vsak uspe z verjetnostjo p.

16. Naj bo X število metov standardne kocke, ki jo mečemo, dokler šestica ne pade desetkrat. Meti so med seboj neodvisni. Zapišite porazdelitev slučajne spremenljivke X.

Negativna binomska (Pascalova)[10] **porazdelitev** $\mathrm{NB}(n, p)$ *je porazdelitev števila poskusov do vključno* n*-tega uspelega, če izvajamo Bernoullijevo zaporedje poskusov, pri katerih vsak uspe z verjetnostjo* p. *Če je* $X \sim \mathrm{NB}(n, p)$, *velja:*

$$\mathbb{P}(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}\,; \quad k = n, n+1, n+2, \ldots$$

POZOR! Marsikje v literaturi je negativna binomska porazdelitev pomaknjena za n v levo, tj.:

$$\mathbb{P}(Y = l) = \binom{n+l-1}{n-1} p^n (1-p)^l\,; \quad l = 0, 1, 2, \ldots$$

To lahko interpretiramo tako, da uspelih poskusov ne štejemo v Y.

[10] Blaise Pascal (1623–1662), francoski matematik, fizik, izumitelj, filozof, moralist in teolog

Tako premaknjena porazdelitev se da lepo posplošiti tudi na necele $n > 0$, tako da definiramo:

$$\mathbb{P}(Y = l) = \frac{n(n+1)\cdots(n+l-1)}{k!} p^n (1-p)^l = \binom{-n}{l} p^n (p-1)^l .$$

Tej porazdelitvi često pravimo **Pólyeva**[11] porazdelitev. Uporablja se v statistiki za modeliranje raznih stvari, kot npr. števila tropskih ciklonov v sezoni ali števila dni, ki jih pacient preživi v bolnišnici.

17. Naj bo X število metov poštenega kovanca, ki ga mečemo, dokler ne pade cifra, takoj za njo pa še grb. Meti so med seboj neodvisni. Zapišite in poimenujte porazdelitev slučajne spremenljivke X.

    Kaj pa, če kovanec ni pošten?

18. V dobro premešanem kupu šestih kart so tri rdeče in tri črne. Iz kupa brez vračanja vlečemo karte, dokler črna karta ne sledi rdeči karti ali pa ne izvlečemo vseh kart. Zapišite porazdelitev števila izvlečenih kart.

19. Naj bo X število metov poštenega kovanca, ki ga mečemo, dokler ne pade tako cifra kot tudi grb. Meti so med seboj neodvisni. Zapišite porazdelitev slučajne spremenljivke X. Je le-ta kaj povezana s kako znano porazdelitvijo?

20. Kovanec, na katerem grb pade z verjetnostjo $p \in (0, 1)$, mečemo, dokler ne dobimo ali m grbov ali pa m cifer, kjer je m dano naravno število. Izračunajte porazdelitev števila potrebnih metov. Privzamemo, da so meti med seboj neodvisni.

21. Pošten kovanec mečemo, dokler ne padeta dve zaporedni cifri, vendar največkrat 7-krat. Meti so med seboj neodvisni. Označimo z X število metov. Zapišite porazdelitev te slučajne spremenljivke. Kaj pa, če umaknemo omejitev, da vržemo največ 7-krat?

22. Med 16 kartami so štirje piki. Na slepo in brez vračanja izvlečemo sedem kart. Naj bo X število pikov med njimi. Zapišite porazdelitev slučajne spremenljivke X.

Hipergeometrijska porazdelitev

Iz posode, v kateri je n kroglic, od tega r rdečih, na slepo in brez vračanja izvlečemo s kroglic. Če z X označimo število rdečih med izvlečenimi, ima ta slučajna spremenljivka hipergeometrijsko porazdelitev: $X \sim \mathrm{Hip}(s, r, n) = \mathrm{Hip}(r, s, n)$. *Velja:*

$$\mathbb{P}(X = k) = \frac{\binom{r}{k}\binom{n-r}{s-k}}{\binom{n}{s}} = \frac{\binom{s}{k}\binom{n-s}{r-k}}{\binom{n}{r}} ; \quad k = 0, 1, 2, \ldots$$

Opomba. Če naredimo limito, ko gre n proti neskončno in r/n proti nekemu fiksnemu številu p, s pa ostane konstanten, je plavzibilno, da postanejo vlečenja med seboj neodvisna: dobimo torej Bernoullijevo zaporedje poskusov z verjetnostjo uspeha p. V kakšnem

[11] György Pólya (1887–1985), madžarski matematik judovskega rodu

smislu limito dobimo, bi bilo sicer treba še precizirati, a tukaj tega ne bomo storili. Opazimo pa, da, brž ko je $r_1, r_2, \ldots$ zaporedje z $\lim_{n\to\infty} r_n/n = p$, za vsak k velja:

$$
\begin{aligned}
&\lim_{n\to\infty} \frac{\binom{s}{k}\binom{n-s}{r_n-k}}{\binom{n}{r_n}} = \\
&= \lim_{n\to\infty} \binom{s}{k} \frac{(n-s)!\,r_n!\,(n-r_n)!}{(r_n-k)!\,(n-r_n-s+k)!\,n!} = \\
&= \lim_{n\to\infty} \binom{s}{k} \frac{r_n(r_n-1)\cdots(r_n-k+1)\cdot(n-r_n)(n-r_n-1)\cdots(n-r_n-s+k+1)}{n(n-1)\cdots(n-s+1)} = \\
&= \lim_{n\to\infty} \binom{s}{k} p^k(1-p)^{s-k}\,.
\end{aligned}
$$

Hipergeometrijska porazdelitev $\mathrm{Hip}(s,r,n)$, ki izhaja iz izvirnega poskusa, torej glede na točkaste verjetnosti konvergira proti binomski porazdelitvi $\mathrm{Bin}(s,p)$, ki izhaja iz limitnega Bernoullijevega zaporedja poskusov.

23. Danih je 12 praznih škatel. Mimo pride Janezek, povsem naključno izbere tri škatle in v vsako vrže po eno kroglico. Mimo pride še Marička, povsem naključno (in neodvisno od Janezka) izbere štiri škatle in prav tako v vsako vrže po eno kroglico. Zapišite in poimenujte porazdelitev števila škatel, ki so ostale prazne.

24. Na nekem izpitu dobi študent dve povsem naključno izbrani vprašanji izmed 10 možnih. Študent se je učil le polovico vseh vprašanj. Vendar pa na vsako vprašanje, ki se ga ni učil, z verjetnostjo 20% ugane odgovor. Glede tega so vprašanja neodvisna, prav tako je študentova zmožnost ugibanja odgovorov neodvisna od izbire izpitnih vprašanj.

    Slučajna spremenljivka U naj pove število vprašanj, ki se jih študent ni učil, je pa uganil odgovor. Zapišite njeno porazdelitev numerično na 4 decimalke natančno.

25. V Pólyevi žari je sprva b belih in r rdečih kroglic. Iz nje na slepo vlečemo kroglice. Vsakič, ko izvlečemo kroglico posamezne barve, jo vrnemo v posodo in dodamo še eno kroglico enake barve. Določite porazdelitev števila belih izvlečenih kroglic po n vlečenjih. Kakšna je ta porazdelitev za $b = r = 1$?

26. Iz posode, v kateri je n kroglic, od tega r rdečih, na slepo in brez vračanja vlečemo kroglice, dokler ne izvlečemo s rdečih ($1 \le s \le r$). Porazdelitvi števila izvlečenih kroglic pravimo *negativna hipergeometrijska porazdelitev*. Določite jo.

27. V Pólyevi žari je sprva a rdečih in b belih kroglic. Iz nje na slepo vlečemo kroglice. Če izvlečemo rdečo, končamo, če izvlečemo belo, pa jo vrnemo v posodo in dodamo še k novih belih kroglic. V 30. nalogi iz 3. razdelka smo dokazali, da z verjetnostjo ena nekoč nehamo vleči. Zapišite porazdelitev števila vlečenj.

28. V skupini je n moških in n žensk. Le-te povsem naključno razporedimo po parih. Določite porazdelitev števila parov, v katerih sta osebi različnega spola.

Kumulativna porazdelitvena funkcija

V splošnem porazdelitev opišemo z verjetnostmi $\mathbb{P}(X \in C)$ za vse merljive množice C. Pri **realnih** *slučajnih spremenljivkah pa zadostuje za C vzeti poltrake $(-\infty, x]$. Tako dobimo* **kumulativno porazdelitveno funkcijo**:

$$F_X(x) = \mathbb{P}(X \le x)\,.$$

29. Narišite grafa kumulativnih porazdelitvenih funkcij slučajnih spremenljivk iz 1. in 2. naloge.

30. Na intervalu $[-3, 3]$ povsem naključno izberemo število. Označimo z D oddaljenost tega števila od intervala $[0, 1]$. Izračunajte kumulativno porazdelitveno funkcijo slučajne spremenljivke D in narišite njen graf.

31. V kvadratu s stranico 2 povsem naključno izberemo točko. Izračunajte kumulativno porazdelitveno funkcijo oddaljenosti izbrane točke od najbližje stranice in narišite njen graf.

Realna slučajna spremenljivka X je porazdeljena **zvezno**, *če obstaja taka integrabilna funkcija $f_X\colon\ \mathbb{R} \to [0, \infty)$, da za poljubna $a \le b$ velja:*

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,\mathrm{d}x\,. \qquad (*)$$

Funkciji f_X pravimo **porazdelitvena gostota**.

Če formula $(*)$ velja za vse realne a in b, $a \le b$, velja tudi za $a = \infty$ in/ali $b = \infty$. Posledično je:

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1\,.$$

Iz formule $(*)$ pa sledi tudi, da je $\mathbb{P}(X = x) = 0$ za vse $x \in \mathbb{R}$.

32. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

$$f_X(x) = \begin{cases} \dfrac{c}{\sqrt{x}} & ;\ 0 < x < 9 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Določite konstanto c ter izračunajte $\mathbb{P}(1 < X < 4)$ in $\mathbb{P}(X > 1)$.

33. Za slučajni spremenljivki iz 30. in 31. naloge določite, ali sta porazdeljeni diskretno in ali sta porazdeljeni zvezno. Za primer, ko je katera od teh slučajnih spremenljivk porazdeljena zvezno, zapišite še porazdelitveno gostoto.

Naj bo F_X kumulativna porazdelitvena funkcija slučajne spremenljivke X.

- *Če je X porazdeljena zvezno, je F_X zvezna in velja $F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t\,.$*
- *Če je F_X zvezna in odsekoma zvezno odvedljiva, je X porazdeljena zvezno in za vse razen za končno mnogo točk x velja $f_X(x) = F'_X(x)$.*

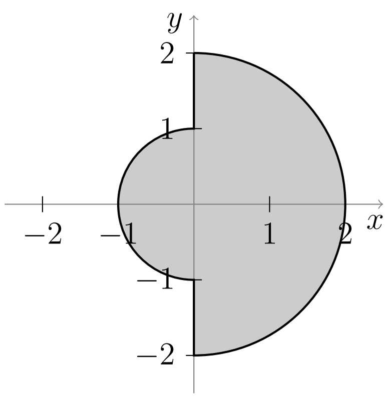
34. Avtobus vozi na 10 minut, na postajo pa pridemo povsem naključno. Slučajna spremenljivka X naj predstavlja čas čakanja na avtobus v minutah. Zapišite kumulativno porazdelitveno funkcijo te slučajne spremenljivke. Nadalje dokažite, da je porazdelitev zvezna, in zapišite še njeno gostoto.

Zvezna enakomerna porazdelitev *na intervalu (a, b) $(a < b)$ je porazdelitev povsem naključne točke iz tega intervala. To je porazdelitev z gostoto:*

$$f(x) = \begin{cases} \dfrac{1}{b-a} & ;\ a < x < b \\ 0 & ;\ \textit{sicer.} \end{cases}$$

35. Rok in Simona se dogovorita za zmenek natanko ob osmih pod starim zvonikom. A v resnici prideta enkrat med 20:00 in 20:10, in sicer z enakomerno porazdelitvijo in neodvisno drug od drugega. Ljubosumni Maks vse od 20:00 opreza za vogalom in čaka, dokler ne prideta obadva. Slučajna spremenljivka M naj predstavlja, koliko časa (v minutah) je čakal Maks. Zapišite kumulativno porazdelitveno funkcijo in gostoto te slučajne spremenljivke. Kolikšna je verjetnost, da je Maks čakal med 5 in 6 minut?

36. Povsem naključno izberemo točko iz lika, ki ga sestavljajo točke, ki ležijo levo od ordinatne osi in so od izhodišča oddaljene največ 1, in točke, ki ležijo desno od ordinatne osi in so od izhodišča oddaljene največ 2 (glej sliko). Naj bo Z oddaljenost izbrane točke od izhodišča. Zapišite kumulativno porazdelitveno funkcijo in gostoto porazdelitve te slučajne spremenljivke.



37. Naj bo $\lambda > 0$. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

$$f_X(x) = \begin{cases} ce^{-\lambda x} & ;\ x > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

Izračunajte konstanto c in določite kumulativno porazdelitveno funkcijo $F_X(x)$. Izračunajte še $\mathbb{P}(1 < X < 2)$.

> **Eksponentna porazdelitev** *je zvezna porazdelitev, skoncentrirana na intervalu* $[0, \infty)$ *in katere gostota na tem intervalu je eksponentna funkcija. Natančneje, porazdelitev* $\mathrm{Exp}(\lambda)$ *ima gostoto:*
>
> $$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x > 0 \\ 0 & ; \textit{sicer} \end{cases}$$

38. Neodvisno in povsem naključno izbiramo naravna števila od 1 do m in jih uvrstimo v zaporedje. Določite porazdelitev dolžine najdaljšega začetnega strnjenega strogo naraščajočega podzaporedja (za primere, ki pojasnijo, kaj se razume kot najdaljše začetno strnjeno strogo naraščajoče podzaporedje, glej 20. nalogo iz 2. razdelka).

39. Slučajna spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0{\cdot}4 & 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}2 \end{pmatrix}$$

Zapišite porazdelitev slučajne spremenljivke $Y = X^2$.

40. Slučajna spremenljivka X je porazdeljena geometrijsko $\mathrm{Geom}(p)$. Zapišite porazdelitev slučajne spremenljivke $Y := \sin(\pi X/2)$.

41. *Nekorektno zastavljen problem*: kakšna je porazdelitev prve signifikantne decimalke naključno izbranega pozitivnega realnega števila (za število π je npr. to 3, za število 2038 je to 2, za število $1/16$ pa je to 6)?

Če želimo dobiti korekten problem, moramo sprejeti določene dodatne predpostavke, npr. kakšno porazdelitev imamo v mislih, ko govorimo o naključnosti števila. Števila, s katerimi se srečujemo, zavzemajo precej velik razpon, od zelo majhnih, kakršen je npr. *Planckov čas* $5{\cdot}39 \cdot 10^{-44}$ s, do zelo velikih, npr. en kilogram vodika vsebuje približno $5{\cdot}975 \cdot 10^{26}$ atomov.

Privzemimo, da gostota porazdelitve v glavnini zaloge vrednosti ostane v glavnini nespremenjena, če slučajno število pomnožimo s faktorjem reda velikosti od $1/10$ do 10. To si lahko predstavljamo tako, da zamenjamo enote (npr. če namesto palcev vzamemo centimetre, se mersko število pomnoži z $2{\cdot}54$). To ustreza predpostavki, da gostota porazdelitve *logaritma* tega števila ostane v glavnini nespremenjena, če temu logaritmu prištejemo število reda velikosti od -1 do 1. To pa je takrat, ko je gostota porazdelitve logaritma v glavnini približno konstantna.

Označimo dano slučajno število z X. Njegova prva decimalka se ohrani, če ga pomnožimo z določeno potenco števila 10. To lahko povemo tudi tako, da njegovemu desetiškemu logaritmu $Y := \log_{10} X$ prištejemo določeno celo število. Število X je z Y natančno določeno, prva decimalka števila X pa je natančno določena že z *necelim* delom števila Y, torej z $U := Y - \lfloor Y \rfloor$.

Iz predpostavke, da je gostota porazdelitve slučajne spremenljivke Y v glavnini, ki se razpenja čez veliko celih števil, sledi, da je slučajna spremenljivka U porazdeljena približno enakomerno na intervalu $[0, 1)$. Približno kakšna je torej porazdelitev iskane prve decimalke?

42. Naj bo $a \in \mathbb{R}$ in $\lambda > 0$. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Dokažite, da je slučajna spremenljivka $Y = (X + a)^2$ zvezno porazdeljena, in zapišite njeno porazdelitveno gostoto.

43. Na razpolago imamo generator slučajnih števil, ki generira enakomerno porazdelitev $\mathrm{Unif}(0, 1)$. Kako bi generirali porazdelitev slučajne spremenljivke X iz 39. naloge? Kaj pa eksponentno porazdelitev $\mathrm{Exp}(\lambda)$?

Naj bo X slučajna spremenljivka, porazdeljena zvezno z gostoto, ki je na nekem intervalu (končnem ali neskončnem) strogo pozitivna, drugje pa enaka nič. Naj bo $0 < \alpha < 1$. Število x_α je **kvantil** *slučajne spremenljivke X za verjetnost α, če velja:*

$$\mathbb{P}(X < x_\alpha) = \mathbb{P}(X \le x_\alpha) = F_X(x_\alpha) = \alpha\,.$$

Kvantili take slučajne spremenljivke za vse verjetnosti iz $(0, 1)$ obstajajo in so enolično določeni.

Kvantilu za verjetnost $1/2$ pravimo **mediana**.
Kvantiloma za verjetnosti $1/3$ in $2/3$ pravimo prvi in drugi **tercil**.
Kvantili za verjetnosti $1/4$, $2/4$ in $3/4$ so **kvartili**.
Kvantili za verjetnosti $0{\cdot}1, 0{\cdot}2, \ldots, 0{\cdot}9$ so **decili**.
Kvantili za verjetnosti $0{\cdot}01, 0{\cdot}02, \ldots, 0{\cdot}99$ pa so **centili** *ali* **percentili**.
Medkvartilni razmik *je razlika* $\mathrm{IQR} = x_{3/4} - x_{1/4}$.

44. Slučajna spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$f_X(x) = \begin{cases} \dfrac{1}{(1+x)^2} & ; x \ge 0 \\ 0 & ; \text{sicer.} \end{cases}$$

Izračunajte vse njene kvantile. Posebej izračunajte še mediano in medkvartilni razmik.

Število x_α je kvantil splošne slučajne spremenljivke X za verjetnost α, če velja:

$$\mathbb{P}(X < x_\alpha) \le \alpha \le \mathbb{P}(X \le x_\alpha)\,.$$

Kvantil za vsako verjetnost iz $(0, 1)$ še vedno obstaja, ni pa več nujno enolično določen.

45. Slučajna spremenljivka X je porazdeljena diskretno po shemi:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0\text{˙}05 & 0\text{˙}15 & 0\text{˙}15 & 0\text{˙}15 & 0\text{˙}15 & 0\text{˙}35 \end{pmatrix}.$$

Določite tretji decil in mediano. Kateri od teh dveh kvantilov je enolično določen?

46. Naj bo X slučajna spremenljivka s kumulativno porazdelitveno funkcijo F.

a) Naj bo q kvantil slučajne spremenljivke X za verjetnost p. Dokažite, da za vsak x veljajo naslednje implikacije:

$$x < q \Longrightarrow F(x) \le p\,, \qquad F(x) < p \Longrightarrow x \ge p\,,$$
$$x > q \Longrightarrow F(x) \ge p\,, \qquad F(x) > p \Longrightarrow x \le p\,.$$

Ali lahko katero od teh implikacij še okrepimo, tako da na levi strani dodamo enačaj in/ali ga na desni odvzamemo?

b) Naj bo $Q\colon (0,1) \to \mathbb{R}$ *kvantilna funkcija* slučajne spremenljivke X, tj. za vsak $p \in (0,1)$ naj bo $Q(p)$ kvantil slučajne spremenljivke X za verjetnost p. Dokažite, da je Q (ne nujno strogo) naraščajoča.

c) Naj bo spet $Q\colon (0,1) \to \mathbb{R}$ kvantilna funkcija slučajne spremenljivke X. Nadalje naj bo slučajna spremenljivka U porazdeljena zvezno enakomerno na intervalu $(0,1)$. Dokažite, da ima slučajna spremenljivka $Q(U)$ enako porazdelitev kot X. To je torej splošni recept za generiranje porazdelitev.

Naj bo $\mu \in \mathbb{R}$ in $\sigma > 0$. **Normalna (Gaussova[12])** *porazdelitev* $\mathrm{N}(\mu,\sigma)$ *je porazdelitev z gostoto:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,.$$

Normalna porazdelitev $\mathrm{N}(\mu,0)$ *je porazdelitev, ki je skoncentrirana v* μ *(*$X \sim \mathrm{N}(\mu,0)$ *pomeni* $\mathbb{P}(X=\mu)=1$*).*

Parametru μ *pravimo* **pričakovana vrednost**, *parametru* σ *pa* **standardni odklon**.

Standardna normalna porazdelitev $\mathrm{N}(0,1)$ *ima potemtakem gostoto:*

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

in kumulativno porazdelitveno funkcijo $F(x) = \frac{1}{2} + \Phi(x)$*.*

47. Slučajna spremenljivka Z je porazdeljena standardno normalno. Izračunajte $\mathbb{P}(Z < 1\text{˙}5)$.

[12] Carl Friedrich Gauß(1777–1855), nemški matematik

Naj bodo:

- $A, B \subseteq \mathbb{R}$ *odprti množici (npr. odprta intervala);*
- $h\colon A \to B$ *taka bijekcija, da je funkcija* $h^{-1}\colon B \to A$ *zvezno odvedljiva;*
- X *slučajna spremenljivka z vrednostmi v* A*, porazdeljena zvezno z gostoto* f_X*.*

Tedaj je slučajna spremenljivka $Y := h(X)$ *porazdeljena zvezno z gostoto:*

$$f_Y(y) = \begin{cases} f_X\big(h^{-1}(y)\big)\big|(h^{-1})'(y)\big| & ; y \in B \\ 0 & ; \textit{sicer.} \end{cases}$$

Slika:

$$\begin{tikzcd} & \Omega \arrow[dl, "X"'] \arrow[dr, "Y"] & \\ A \arrow[rr, "h"] \arrow[dr, "f_X"'] & & B \arrow[dl, "f_Y"] \\ & {[0,\infty)} & \end{tikzcd}$$

48. Slučajna spremenljivka X je porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Kako je porazdeljena slučajna spremenljivka $Y := aX + b$?

Če je $X \sim \mathrm{N}(\mu, \sigma)$*, velja:*

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Laplaceova integralska formula tako ne pomeni nič drugega kot to, da za velike n *in za* p*, ki ni preblizu* 0 *ali* 1*, velja:*

$$\mathrm{Bin}(n, p) \approx \mathrm{N}\big(np, \sqrt{npq}\big)$$

kjer je $q = 1 - p$*.*

49. Slučajna spremenljivka X je porazdeljena normalno $\mathrm{N}(9, 5)$. Izračunajte $\mathbb{P}(X < 0)$.

Porazdelitev gama, *ki jo bomo označevali z* $\mathrm{Gama}(a, \lambda)$, *je zvezna porazdelitev z gostoto:*

$$f(x) = \begin{cases} \dfrac{\lambda^a}{\Gamma(a)}\, x^{a-1}\, e^{-\lambda x} & ;\ x > 0 \\ 0 & ;\ \textit{sicer.} \end{cases}$$

Poseben primer te porazdelitve je eksponentna porazdelitev $\mathrm{Exp}(\lambda) = \mathrm{Gama}(1, \lambda)$.

50. Če je $X \sim \mathrm{Gama}(a, \lambda)$ in $k > 0$, določite porazdelitev slučajne spremenljivke $Y = kX$.

51. Slučajna spremenljivka X je porazdeljena standardno normalno. Določite porazdelitev slučajne spremenljivke $Y = e^X$.

Če je slučajna spremenljivka X **skoncentrirana** *na množici* A, *tj. če je* $\mathbb{P}(X \in A) = 1$ *oziroma* $\mathbb{P}(X \notin A) = 0$ *(kar je npr. res, če je* X *porazdeljena zvezno z gostoto, ki je izven množice* A *enaka* 0*), obstaja tudi slučajna spremenljivka* X', *ki je porazdeljena enako kot* X *in zavzame vrednosti* **izključno** *v množici* A.

Skoncentriranost slučajne spremenljivke na določeni množici je stvar njene porazdelitve. Če gre torej za vprašanje porazdelitve in je določena porazdelitev skoncentrirana na določeni množici A, *lahko za slučajno spremenljivko s to porazdelitvijo vselej privzamemo, da zavzame vrednosti izključno v* A.

52. Slučajna spremenljivka R ima *Rayleighovo*[13] *porazdelitev*, tj. porazdelitev z gostoto:

$$f_R(r) = \begin{cases} ar\, e^{-ar^2/2} & ;\ r > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

(to je porazdelitev razdalje, za katero se po določenem času premakne delec, ki se giblje v skladu z ravninskim Brownovim gibanjem; glej 18. nalogo iz 5. razdelka).

a) Določite porazdelitev slučajne spremenljivke R^2.

b) Slučajnima spremenljivkama R in R^2 določite 95. centil.
Namig: katerega je lažje izračunati?

53. Naj bo $\lambda > 0$ in $a \in \mathbb{R}$. Slučajna spremenljivka X naj bo porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Kako je porazdeljena slučajna spremenljivka $Y := e^{aX}$?

[13] John William Strutt, 3rd Baron Rayleigh (1842–1919), angleški fizik

54. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

$$f_X(x) = \begin{cases} \dfrac{1}{(1+x)^2} & ; x > 0 \\ 0 & ; \text{sicer.} \end{cases}$$

Izračunajte porazdelitev slučajne spremenljivke $Y = \dfrac{X}{1-X}$.

Tehnična opomba. Strogo gledano slučajna spremenljivka X tu ne sme zavzeti vrednosti 1. A njena porazdelitev to dopušča, saj ima tak dogodek verjetnost nič.

55. Naj bo $a, b > 0$ in X slučajna spremenljivka, porazdeljena zvezno z gostoto

$$f_X(x) = \begin{cases} c\, e^{-a^2x^2 - b^2/x^2} & ; x > 0 \\ 0 & ; \text{sicer.} \end{cases}$$

Konstanta c je izbrana tako, da se gostota zintegrira v 1.

a) Dokažite, da je tudi slučajna spremenljivka $Y := aX - b/X$ porazdeljena zvezno, in določite njeno gostoto.

b) Slučajna spremenljivka W je porazdeljena zvezno z gostoto

$$f_W(w) = \frac{f_Y(w) + f_Y(-w)}{2}\,.$$

Poimenujte to porazdelitev.

c) Izračunajte integral $\displaystyle\int_0^\infty e^{-a^2x^2 - b^2/x^2}\,\mathrm{d}x$.

Naj bodo:

- $A \subseteq \mathbb{R}$ *odprta množica;*
- X *slučajna spremenljivka z vrednostmi v* A*, porazdeljena zvezno z gostoto* f_X*;*
- $h\colon\ A \to \mathbb{R}$ *zvezno odvedljiva;*
- $\mathbb{P}\big(h'(X) = 0\big) = 0$.

Tedaj je slučajna spremenljivka $Y = h(X)$ *porazdeljena zvezno z gostoto:*

$$f_Y(y) = \sum_{\substack{x \in A\,; \\ h(x) = y \\ h'(x) \neq 0}} \frac{f_X(x)}{\big|h'(x)\big|}\,.$$

56. Slučajna spremenljivka X je porazdeljena standardno normalno. Določite porazdelitvi slučajnih spremenljivk $Y = X^2$ in $Z = (X-1)^2$.

5. Slučajni vektorji

Skupne (navzkrižne) in robne porazdelitve. Neodvisnost slučajnih spremenljivk. Transformacije slučajnih vektorjev.

Diskretni slučajni vektorji

- *Porazdelitev slučajnega vektorja* (X, Y) *podamo z verjetnostmi* $\mathbb{P}\big((X,Y) = (x,y)\big) = \mathbb{P}(X = x, Y = y)$ *(***skupna** *ali* **navzkrižna porazdelitev** *slučajnih spremenljivk* X *in* Y*).*
- *Porazdelitve komponent imenujemo* **robne porazdelitve***:*

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y)$$
$$\mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y)\,.$$

- X *in* Y *sta neodvisni, brž ko za poljubna* x *in* y *velja* $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$.

1. Iz posode, v kateri so 3 rdeče, 2 modri in 5 belih kroglic, na slepo in brez vračanja izvlečemo tri kroglice. Naj bo R število rdečih, M pa število modrih med njimi. Zapišite porazdelitev slučajnega vektorja (R, M) ter določite in poimenujte še robni porazdelitvi. Sta slučajni spremenljivki R in M neodvisni? Zapišite in poimenujte še porazdelitev slučajne spremenljivke $R + M$.

2. Slučajni spremenljivki R' in M' sta neodvisni in porazdeljeni hipergeometrijsko: $R' \sim \mathrm{Hip}(3, 3, 10)$, $M' \sim \mathrm{Hip}(3, 2, 10)$. Zapišite porazdelitev slučajne spremenljivke $R' + M'$.

3. Slučajni vektor (X, Y) je porazdeljen po naslednji shemi:

	$Y = 1$	$Y = 2$
$X = 1$	$\frac{5}{6} - \frac{1}{6}a - \frac{1}{3}a^2$	$\frac{1}{3}a^2$
$X = 2$	$\frac{1}{6}a$	$\frac{1}{6}$

   a) Za katere vrednosti parametra a je z zgornjo shemo določena porazdelitev slučajnega vektorja?
   b) Pri katerih vrednostih parametra a sta X in Y enako porazdeljeni?
   c) Pri katerih vrednostih parametra a sta X in Y skoraj gotovo enaki?
   d) Pri katerih vrednostih parametra a sta X in Y neodvisni?

4. Slučajni vektor (X, Y) je porazdeljen po naslednji shemi:

	$Y=0$	$Y=1$	$Y=2$
$X=-1$	0˙05	0˙1	
$X=0$	0˙1		
$X=1$	0˙05		

Dopolnite tabelo tako, da bosta slučajni spremenljivki X in Y neodvisni. Poiščite še robni porazdelitvi in porazdelitev razlike $Y-X$.

Diskretni slučajni spremenljivki s končno zalogo vrednosti sta neodvisni natanko tedaj, ko ima matrika njune navzkrižne porazdelitve rang ena.

5. Naj bo $a, b > 0$ in $a + b < 1$. Diskretni slučajni spremenljivki X in Y naj imata skupno porazdelitev, podano po predpisu:

$$\mathbb{P}(X=k, Y=l) = c\,\frac{(k+l)!}{k!\,l!}\,a^k b^l\,; \quad k, l = 0, 1, 2, \ldots\,;$$

sicer je $\mathbb{P}(X=k, Y=l) = 0$. Izrazite c z a in b ter izračunajte robni porazdelitvi.

Diskretne slučajne spremenljivke $X_1, X_2, \ldots, X_n$ so neodvisne, če za poljubne $x_1, x_2, \ldots, x_n$ velja:

$$\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1)\,\mathbb{P}(X_2 = x_2)\cdots\mathbb{P}(X_n = x_n)\,.$$

6. Dane so neodvisne slučajne spremenljivke $X_1, X_2, X_3 \sim \mathrm{Ber}(0˙3)$. Določite porazdelitev njihove vsote $S := X_1 + X_2 + X_3$.

   Splošneje, naj bodo $X_1, \ldots, X_n \sim \mathrm{Ber}(p)$ neodvisne slučajne spremenljivke. Kako je porazdeljena vsota $S := X_1 + \cdots + X_n$?

Posledica. Če sta $S \sim \mathrm{Bin}(m, p)$ in $T \sim \mathrm{Bin}(n, p)$ neodvisni slučajni spremenljivki, je $U := S + T \sim \mathrm{Bin}(m + n, p)$.

Poissonova porazdelitev

Slučajna spremenljivka X ima Poissonovo porazdelitev, kar označimo z $X \sim \mathrm{Pois}(\lambda)$, če velja:

$$\mathbb{P}(X = k) = \frac{\lambda^k\,e^{-\lambda}}{k!}\,; \quad k = 0, 1, 2, \ldots$$

Poissonova porazdelitev je torej limita binomske porazdelitve $\mathrm{Bin}(n, p)$, ko gre $n \to \infty$ in $np \to \lambda$.

7. Naj bosta $S \sim \mathrm{Pois}(\lambda)$ in $T \sim \mathrm{Pois}(\mu)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $U := S + T$?

8. Ministrstvo dobi vsak dan na mizo slučajno število prošenj, ki je porazdeljeno po Poissonovem zakonu $\mathrm{Pois}(\lambda)$. Dnevi so med seboj neodvisni. Ministrstvo vsak dan razreši eno prošnjo, če jo seveda ima, morebitne preostale pa pusti za kasneje. Kolikšna je verjetnost, da ministrstvu po dveh dneh dela preostaneta vsaj še dve nerazrešeni prošnji?

9. V avtobusu, ki ima 32 sedežev, je 30 potnikov. Vsak potnik bo z verjetnostjo $1/30$ na naslednji postaji izstopil, neodvisno od drugih potnikov. Število potnikov, ki bodo na naslednji postaji vstopili, pa je porazdeljeno po Poissonu $\mathrm{Pois}(1)$ in seveda neodvisno od izstopnih namer trenutnih potnikov. Približno izračunajte verjetnost, da ne bodo mogli vsi potniki sedeti, ko bo avtobus speljal s postaje.

   Namig: Dogodek, da izstopita več kot dva potnika in vseeno ne bo dovolj sedežev za vse, je tako malo verjeten, da ga lahko zanemarimo.

10. Dan je dobro premešan kup n kart, med katerimi je $r \geq 1$ rdečih, preostale pa so črne. Za $k = 1, 2, \ldots, r-1$ označimo s ξ_k število črnih kart, ki se nahajajo med k-to in $(k+1)$-to rdečo karto, gledano z vrha navzdol; ξ_0 naj označuje število črnih kart nad prvo rdečo, ξ_r pa število črnih kart pod zadnjo rdečo karto.

    a) Določite porazdelitev slučajnega vektorja $(\xi_0, \xi_1, \ldots, \xi_r)$.
    b) Določite robne porazdelitve.
    c) So slučajne spremenljivke $\xi_0, \xi_1, \ldots, \xi_r$ neodvisne?

11. Dana je povsem naključna permutacija n elementov. Za $k = 1, 2, \ldots, n$ označimo z X_k število ciklov dolžine k. Zapišite porazdelitev slučajnega vektorja $(X_1, X_2, \ldots, X_n)$ ter še porazdelitev slučajne spremenljivke X_1.

Realne slučajne spremenljivke $X_1, X_2, \ldots, X_n$ so neodvisne, brž ko za poljubne $x_1, x_2, \ldots, x_n$ velja:

$$\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1)\,\mathbb{P}(X_2 \leq x_2) \cdots \mathbb{P}(X_n \leq x_n)\,.$$

12. Slučajna spremenljivka X z vrednostmi v naravnih številih naj ima *Zipfovo*[14] *porazdelitev*: za $k = 1, 2, \ldots$ naj velja:

$$\mathbb{P}(X = k) = \frac{1}{\zeta(s)\,k^s}\,,$$

    kjer je $s > 1$, $\zeta(s)$ pa je definirana kot $\zeta(s) := \sum_{k=1}^{\infty} k^{-s}$. Naj bodo $p_1 < p_2 < \cdots$ vsa praštevila ($p_1 = 2$). Definirajmo slučajne spremenljivke α_i s predpisom:

$$X =: \prod_{i=1}^{\infty} p_i^{\alpha_i}\,.$$

[14] George Kengsley Zipf (1902–1950), ameriški jezikoslovec

Spremenljivke α_i so dobro definirane, saj vsako število lahko enolično faktoriziramo na produkt praštevil. Primer: pri $X = 504 = 2^3 \cdot 3^2 \cdot 7$ dobimo $\alpha_1 = 3$, $\alpha_2 = 2$, $\alpha_3 = 0$, $\alpha_4 = 1$ in $\alpha_i = 0$ za vse $i \geq 5$.

a) Poiščite porazdelitve posameznih slučajnih spremenljivke α_i.
Namig: izračunajte $\mathbb{P}(\alpha_i \geq j)$.

b) Poiščite porazdelitve slučajnih vektorjev $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Kaj opazite?

Opomba: Zipfova porazdelitev je model za deleže posameznih besed v določenem jeziku.

13. V Pólyevi žari je sprva b belih in r rdečih kroglic. Iz nje na slepo vlečemo kroglice. Vsakič, ko izvlečemo kroglico posamezne barve, jo vrnemo v posodo in dodamo še d kroglic enake barve. Definirajmo:

$$X_n := \begin{cases} 1 & ; n\text{-ta izvlečena kroglica je rdeča,} \\ 0 & ; n\text{-ta izvlečena kroglica je bela.} \end{cases}$$

a) Določite porazdelitev slučajnega vektorja $(X_1, X_2, \ldots, X_n)$. Kaj opazite?

b) Kolikšna je verjetnost, da bo n-ta izvlečena kroglica rdeča?

14. *Še ena različica Pólyeve žare.* V posodi so črna kroglica in kroglice, označene z naravnimi števili. Kroglice vlečemo iz posode, pri čemer ima vsaka kroglica, ki je v posodi, enako verjetnost, da bo izvlečena (tudi pogojno na zgodovino). Če ima izvlečena kroglica številsko oznako, jo vrnemo in dodamo še eno enako kroglico. Če pa je izvlečena kroglica črna, jo vrnemo v posodo in dodamo kroglico z oznako $m+1$, kjer je m dotlej največja številska oznaka. Na začetku je v posodi poleg črne kroglice le še kroglica z oznako 1.

Primer: Prvih nekaj korakov je lahko videti takole:

● ① → ● ① ② → ● ① ② ① → ● ① ② ① ③ →

● ① ② ① ③ ① → ● ① ② ① ③ ① ③ → ⋯

Izberimo $n \in \mathbb{N}$ in definirajmo X_k kot število kroglic z oznako k po $n-1$ vlečenjih, ko je v posodi vključno s črno natanko $n+1$ kroglic. Določite porazdelitev slučajnega vektorja $(X_1, X_2, \ldots, X_n)$.

15. Pri Bernoullijevem zaporedju poskusov z verjetnostjo uspeha p definirajmo Y_1 kot število poskusov do vključno prvega uspelega, za $i = 2, 3, 4, \ldots$ pa naj bo Y_i število poskusov od nevključno $(i-1)$-tega do vključno i-tega uspelega.

a) Naj bo $m \in \mathbb{N}$. Poiščite skupno porazdelitev slučajnih spremenljivk $Y_1, Y_2, \ldots, Y_m$.

b) Poiščite robne porazdelitve in utemeljite, da so slučajne spremenljivke $Y_1, Y_2, \ldots, Y_m$ neodvisne.

c) Iz rezultata prejšnje točke naredite sklep o ustrezni vsoti neodvisnih slučajnih spremenljivk.

d) Če sta $S \sim \mathrm{NB}(m, p)$ in $T \sim \mathrm{NB}(n, p)$ neodvisni slučajni spremenljivki, kako je porazdeljena njuna vsota?

Dvorazsežni slučajni vektor (X, Y) je porazdeljen **zvezno**, *če obstaja taka integrabilna funkcija $f_{X,Y} \colon \mathbb{R}^2 \to [0, \infty)$, da za poljubne $a \le b$ in $c \le d$ velja:*

$$\mathbb{P}\big(a \le X \le b,\ c \le Y \le d\big) = \int_a^b \int_c^d f_{X,Y}(x, y)\,\mathrm{d}y\,\mathrm{d}x\,.$$

Tedaj tudi za vsako merljivo množico $A \subseteq \mathbb{R}^2$ velja:

$$\mathbb{P}\big((X, Y) \in A)\big) = \iint_A f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y\,.$$

Funkcija $f_{X,Y}$ je porazdelitvena gostota slučajnega vektorja (X, Y), pravimo pa ji tudi **skupna** *ali* **navzkrižna** *porazdelitvena gostota slučajnih spremenljivk X in Y.*

Splošneje, slučajni vektor $\boldsymbol{W}$ z vrednostmi v $\mathbb{R}^n$ je porazdeljen zvezno, če obstaja taka integrabilna funkcija $f_{\boldsymbol{W}} \colon \mathbb{R}^n \to [0, \infty)$, da za poljubno merljivo množico $A \subseteq \mathbb{R}^n$ velja:

$$\mathbb{P}(\boldsymbol{W} \in A) = \int_A f_{\boldsymbol{W}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}\,.$$

Funkcija $f_{\boldsymbol{W}}$ je porazdelitvena gostota slučajnega vektorja $\boldsymbol{W}$. Seveda velja:

$$\int_{\mathbb{R}^n} f_{\boldsymbol{W}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w} = 1\,.$$

Če sta $\boldsymbol{X}$ in $\boldsymbol{Y}$ slučajna vektorja z vrednostmi v $\mathbb{R}^m$ in $\mathbb{R}^n$ in je slučajni vektor $(\boldsymbol{X}, \boldsymbol{Y})$ porazdeljen zvezno z $(m+n)$-razsežno **skupno (navzkrižno)** *gostoto $f_{\boldsymbol{X},\boldsymbol{Y}}$, sta tudi njegovi komponenti $\boldsymbol{X}$ in $\boldsymbol{Y}$ porazdeljeni zvezno, in sicer z* **robnima gostotama***:*

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{\mathbb{R}^n} f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,, \qquad f_{\boldsymbol{Y}}(\boldsymbol{y}) = \int_{\mathbb{R}^m} f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,.$$

Zvezno porazdeljena slučajna vektorja $\boldsymbol{X}$ in $\boldsymbol{Y}$ sta neodvisna natanko tedaj, ko je tudi slučajni vektor $(\boldsymbol{X}, \boldsymbol{Y})$ porazdeljen zvezno z gostoto:

$$f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{x})\, f_{\boldsymbol{Y}}(\boldsymbol{y})\,.$$

16. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$f_{X,Y}(x, y) = \begin{cases} c\,e^{-x} & ; x > 2y > 0 \\ 0 & ; \text{sicer.} \end{cases}$$

a) Poiščite konstanto c.

b) Poiščite robni gostoti f_X in f_Y.

c) Sta slučajni spremenljivki X in Y neodvisni?

d) Izračunajte $\mathbb{P}(2 < Y < 3)$, $\mathbb{P}(X > 3Y)$ in $\mathbb{P}(3X > Y)$.

e) Določite porazdelitev vsote $Z := X + Y$.

17. Slučajni vektor (X, Y, Z) naj bo porazdeljen enakomerno na enotski krogli v $\mathbb{R}^3$. Določite eno- in dvorazsežne robne porazdelitve.

Jacobijeva[15] **matrika** *parcialno odvedljive preslikave* $f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$: $\mathbb{R}^n \to \mathbb{R}^n$ *je matrika iz njenih parcialnih odvodov:*

$$Jf = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}.$$

Jacobijeva determinanta $\det J$ *je determinanta Jacobijeve matrike.*

Naj bodo:

- $A, B \subseteq \mathbb{R}^n$ *odprti množici;*
- $h\colon\ A \to B$ *taka bijekcija, da je preslikava* $h^{-1}\colon\ B \to A$ *parcialno zvezno odvedljiva;*
- $\boldsymbol{X}$ *slučajni vektor z vrednostmi v* A, *porazdeljen zvezno z gostoto* $f_{\boldsymbol{X}}$.

Tedaj je slučajni vektor $\boldsymbol{Y} := h(\boldsymbol{X})$ *porazdeljen zvezno z gostoto:*

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \begin{cases} f_{\boldsymbol{X}}\big(h^{-1}(\boldsymbol{y})\big)\,\big|\det J(h^{-1})(\boldsymbol{y})\big| & ; \boldsymbol{y} \in B \\ 0 & ; \textit{sicer.} \end{cases}$$

[15] Carl Gustav Jacob Jacobi (1804–1851), nemški matematik judovskega rodu

Slika:

$$\begin{tikzcd}
& \Omega \arrow[dl, "\boldsymbol{X}"'] \arrow[dr, "\boldsymbol{Y}"] & \\
A \arrow[rr, "h"] \arrow[dr, "f_{\boldsymbol{X}}"'] & & B \arrow[dl, "f_{\boldsymbol{Y}}"] \\
& {[0, \infty)} &
\end{tikzcd}$$

18. Naj bosta X in Y neodvisni standardni normalni slučajni spremenljivki in naj bodo (R, Θ) polarne koordinate slučajnega vektorja (X, Y), tj. $R = \sqrt{X^2 + Y^2}$ in $\Theta = \arg(X, Y) \in (-\pi, \pi]$.

    a) Določite porazdelitev slučajnega vektorja (R, Θ).
    b) Določite robni porazdelitvi.
    c) V kakšnem medsebojnem odnosu sta slučajni sprememenljivki R in Θ?

19. Naj bosta $X \sim \mathrm{N}(\mu, \sigma)$ in $Y \sim \mathrm{N}(\nu, \tau)$ neodvisni slučajni spremenljivki. Definirajmo $U := X + Y$ in $V := \sigma^2 Y - \tau^2 X$. Določite porazdelitev slučajnega vektorja (U, V) ter porazdelitvi slučajnih spremenljivk U in V.

    Namig: najprej obravnavajte primer, ko je $\mu = \nu = 0$.

20. Avtobus odpelje s postaje ob času, ki je porazdeljen normalno $\mathrm{N}(7{:}00, 3\mathrm{min})$, študent pa je nagnjen k zamujanju in pride na postajo ob času, ki je porazdeljen normalno $\mathrm{N}(7{:}01, 4\,\mathrm{min})$, neodvisno od avtobusa. Kolikšna je verjetnost, da študent še ujame avtobus?

Naj bodo:

- *$A \subseteq \mathbb{R}^n$ odprta množica;*
- *$\boldsymbol{X}$ slučajni vektor z vrednostmi v A, porazdeljen zvezno z gostoto $f_{\boldsymbol{X}}$;*
- *$h\colon\ A \to \mathbb{R}^n$ parcialno zvezno odvedljiva;*
- *$\mathbb{P}\big(\det Jh(\boldsymbol{X}) = 0\big) = 0$.*

Tedaj je slučajni vektor $\boldsymbol{Y} := h(\boldsymbol{X})$ porazdeljen zvezno z gostoto:

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\substack{\boldsymbol{x} \in A\,; \\ h(\boldsymbol{x}) = \boldsymbol{y} \\ \det Jh(\boldsymbol{x}) \neq 0}} \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\big|\det Jh(\boldsymbol{x})\big|}\,.$$

21. Naj bosta X in Y neodvisni slučajni spremenljivki, pri čemer naj bo $X \sim \mathrm{N}(0, 1)$ in $Y \sim \mathrm{N}(1, 1)$. Določite porazdelitev slučajnega vektorja $\left(XY, \frac{Y}{X}\right)$.

    Tehnična opomba. Strogo gledano moramo tu privzeti, da slučajna spremenjivka X ni nikoli enaka nič. A njena porazdelitev to dopušča, saj ima tak dogodek verjetnost nič.

Naj bodo:

- $A \subseteq \mathbb{R}^{m+n}$ *odprta množica;*
- $\boldsymbol{X}$ *in* $\boldsymbol{Y}$ *slučajna vektorja z vrednostmi v* $\mathbb{R}^m$ *in* $\mathbb{R}^n$*, pri čemer naj bločni slučajni vektor* $(\boldsymbol{X}, \boldsymbol{Y})$ *zavzame vrednosti v* A *in naj bo porazdeljen zvezno z gostoto* $f_{\boldsymbol{X},\boldsymbol{Y}}$*;*
- $h\colon\ A \to \mathbb{R}^n$ *parcialno zvezno odvedljiva vektorska funkcija;*
- $\mathbb{P}\big(\det J_{\mathbf{y}} h(\boldsymbol{X}, \boldsymbol{Y}) = 0\big) = 0$*;*
  pri tem je $J_{\mathbf{y}} h$ *Jacobijeva matrika iz parcialnih odvodov vektorske funkcije* $h(\boldsymbol{x}, \boldsymbol{y})$ *po komponentah argumenta* $\boldsymbol{y} \in \mathbb{R}^n$*.*

Tedaj je slučajni vektor $\boldsymbol{Z} := h(\boldsymbol{X}, \boldsymbol{Y})$ *porazdeljen zvezno z gostoto:*

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \int_{\mathbb{R}^m} \sum_{\substack{\boldsymbol{y} \in \mathbb{R}^n\,; \\ (\boldsymbol{x},\boldsymbol{y}) \in A \\ h(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{z} \\ \det J_{\mathbf{y}} h(\boldsymbol{x},\boldsymbol{y}) \neq 0}} \frac{f_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y})}{|\det J_{\mathbf{y}} h(\boldsymbol{x}, \boldsymbol{y})|} \, \mathrm{d}\boldsymbol{x}\,.$$

22. Slučajni spremenljivki X in Y sta neodvisni in porazdeljeni eksponentno $\mathrm{Exp}(\lambda)$. Zapišite porazdelitev njune razlike $Z := X - Y$.

23. Slučajni vektor (X, Y) je porazdeljen zvezno z gostoto:

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{c}{(1 + x + y)^3} & ;\ x, y > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Zapišite porazdelitev slučajne spremenljivke $Z := Y/X$ in izračunajte konstanto c. Izračunajte še $\mathbb{P}(X < 2Y)$.

Tehnična opomba. Strogo gledano moramo tu privzeti, da slučajna spremenjivka X ni nikoli enaka nič. A njena porazdelitev to dopušča, saj ima tak dogodek verjetnost nič.

24. Slučajni spremenljivki X in Y sta neodvisni in porazdeljeni enakomerno na intervalu $[0, 1]$. Zapišite porazdelitev slučajne spremenljivke $Z := XY$.

25. Slučajni vektor (X, Y) ima naslednjo porazdelitveno gostoto:

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{Cx^2y^2}{(1 + x^2 + y^2)(x^2 + y^2)^{5/2}} & ;\ (x, y) \neq (0, 0) \\ 0 & ;\ \text{sicer.} \end{cases}$$

Izračunajte porazdelitveno gostoto slučajne spremenljivke $Z := X^2 + Y^2$ (v celoti, konstante C pa ni treba izračunati).

26. Slučajni vektor (X, Y, Z) je porazdeljen zvezno z gostoto:

$$f_{X,Y,Z}(x,y,z) = \begin{cases} \dfrac{1}{2\pi\sqrt{xy - z^2}}\, e^{-(x+y)/2} & ;\ x, y > 0\,,\ 0 < z < \sqrt{xy} \\ 0 & ;\ \text{sicer.} \end{cases}$$

Določite porazdelitev slučajne spremenljivke $S := \dfrac{XY - Z^2}{X}$.

Tehnična opomba. Strogo gledano moramo tu privzeti, da slučajna spremenjivka X ni nikoli enaka nič. A njena porazdelitev to dopušča, saj ima tak dogodek verjetnost nič.

27. Določene pojave, ki se pojavljajo v času (npr. telefonski klici, nesreče, radioaktivni razpadi), lahko modeliramo s t. i. *homogenim Poissonovim tokom* oz. *Poissonovim procesom štetja* z intenzivnostjo λ, ki je karakteriziran z lastnostma, da je število pojavov, ki se zgodijo v katerem koli časovnem intervalu dolžine t, porazdeljeno po Poissonu $\mathrm{Pois}(\lambda t)$, in da za poljubna disjunktna časovna intervala velja, da je število pojavov, ki se zgodijo v prvem, neodvisno od števila pojavov, ki se zgodijo v drugem časovnem intervalu.

Privzemimo le, da je za vsak $t \ge 0$ število pojavov, ki se zgodijo do vključno časa t, porazdeljeno po Poissonu $\mathrm{Pois}(\lambda t)$. Naj bo T_n slučajna spremenljivka, ki pove čas, ob katerem se zgodi n-ti pojav (čas štejemo od 0 naprej). Določite njeno porazdelitev. Kaj pride pri $n = 1$?

28. Naj bosta $S \sim \mathrm{Gama}(a, \lambda)$ in $T \sim \mathrm{Gama}(b, \lambda)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $U := S + T$?

Posledica. Če so $X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$ neodvisne slučajne spremenljivke, je $S := X_1 + \cdots + X_n \sim \mathrm{Gama}(n, \lambda)$.

Porazdelitev hi hvadrat z n prostostnimi stopnjami, *ki jo označujemo s $\chi^2(n)$, je porazdelitev vsote $Z_1^2 + \cdots + Z_n^2$, kjer so $Z_1, \ldots, Z_n$ neodvisne slučajne spremenljivke, porazdeljene standardno normalno. Ta porazdelitev je pomembna v statistiki, saj med drugim igra ključno vlogo pri ocenjevanju variance. Velja* $\chi^2(n) = \mathrm{Gama}\big(\frac{n}{2}, \frac{1}{2}\big)$.

29. Na zabavo, ki se začne ob določeni uri, je povabljenih n gostov. Vsak malo zamudi. Natančneje, zamuda vsakega je porazdeljena enakomerno na intervalu $[0, 1]$ in zamude so med seboj neodvisne. Slučajna spremenljivka T_k naj predstavlja čas, ko je na zabavo prišel k-ti gost po vrsti (glede na čas prihoda). Določite njeno porazdelitev.

30. Slučajni spremenljivki X in Y sta neodvisni, pri čemer je X porazdeljena normalno $\mathrm{N}(0, \sigma)$, Y pa ima porazdelitev $\mathrm{Gama}(a, \lambda)$.

a) Določite porazdelitev slučajne spremenljivke $T = \dfrac{X}{\sqrt{Y}}$.

b) *Studentova*[16] *porazdelitev* z n prostostnimi stopnjami je porazdelitev slučajne spremenljivke:

$$T = \frac{X}{\sqrt{\dfrac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}}}\,, \qquad (*)$$

kjer so $X, X_1, X_2, \ldots, X_n \sim \mathrm{N}(0, \sigma)$ neodvisne slučajne spremenljivke. Dokažite, da je ta porazdelitev neodvisna od σ, in zapišite njeno gostoto. Kam konvergira ta gostota, ko gre n proti neskončno?

Opomba. Studentova porazdelitev je pomembna v statistiki. Zaenkrat si lahko predstavljamo, da želimo standardizirati opažanje $X \sim \mathrm{N}(0, \sigma)$, pri čemer pa parametra σ ne poznamo; pač pa lahko σ ocenimo tako, da opažanje n-krat neodvisno ponovimo. Imenovalec v $(*)$ je *cenilka* za σ.

31. Slučajni spremenljivki X in Y sta neodvisni s porazdelitvijo gama, in sicer $X \sim \mathrm{Gama}(a, \lambda)$ in $Y \sim \mathrm{Gama}(b, \lambda)$. Določite porazdelitev slučajne spremenljivke $Z = X/(X + Y)$.

32. Slučajni spremenljivki X in Y sta neodvisni s porazdelitvijo gama: $X \sim \mathrm{Gama}(a, \lambda)$ in $Y \sim \mathrm{Gama}(b, \mu)$. Določite porazdelitev slučajne spremenljivke $Q = X/Y$.

Snedecorjeva (Fisherjeva)[17,18] *porazdelitev z m in n prostostnimi stopnjami,* $\mathrm{F}(m, n)$*, je porazdelitev, dobljena na enega od naslednjih dveh ekvivalentnih načinov:*

- *kot porazdelitev kvocienta X/Y, kjer sta X in Y neodvisni ter $X \sim \mathrm{Gama}(m/2, m/2)$ in $Y \sim \mathrm{Gama}(n/2, n/2)$;*
- *kot porazdelitev kvocienta $\dfrac{U/m}{V/n}$, kjer sta U in V neodvisni ter $U \sim \chi^2(m)$ in $V \sim \chi^2(n)$.*

33. Konstruirajte funkcijsko zvezo, za katero obstajata slučajni spremenljivki $F \sim \mathrm{F}(m, n)$ in $B \sim \mathrm{Beta}(m/2, n/2)$, ki sta v tej zvezi.

[16] William Sealy Gosset (1876–1937), angleški statistik, bolj znan pod psevdonimom *Student*
[17] George Waddel Snedecor (1882–1974), ameriški matematik in statistik
[18] Sir Ronald Aymler Fisher (1899–1962), angleški statistik in biolog

Večrazsežna normalna porazdelitev *s* **pričakovano vrednostjo $\boldsymbol{\mu}$** *in* **kovariančno matriko $\boldsymbol{\Sigma}$**, *kjer je $\boldsymbol{\mu}$ vektor v $\mathbb{R}^n$, $\boldsymbol{\Sigma}$ pa je pozitivno semidefinitna matrika v $\mathbb{R}^{n\times n}$, je določena z naslednjimi lastnostmi:*

- *Komponente* **standardne n-razsežne normalne porazdelitve**, *kjer je $\boldsymbol{\mu} = \mathbf{0}$ in $\boldsymbol{\Sigma} = \mathbf{I}_n$ (identiteta na $\mathbb{R}^n$), so neodvisne in porazdeljene (enorazsežno) standardno normalno.*
- *Če je $\boldsymbol{X}$ porazdeljen večrazsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}$ in kovariančno matriko $\boldsymbol{\Sigma}$, $\mathbf{A}$ matrika v $\mathbb{R}^{m\times n}$ in $\boldsymbol{b} \in \mathbb{R}^m$, je tudi $\mathbf{A}\boldsymbol{X}+\boldsymbol{b}$ porazdeljena večrazsežno normalno, in sicer s pričakovano vrednostjo $\mathbf{A}\boldsymbol{\mu}+\boldsymbol{b}$ in kovariančno matriko $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}}$. Med drugim to pomeni, da je v primeru, ko je bločni vektor $\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}$ porazdeljen večrazsežno normalno s pričakovano vrednostjo $\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}$ in kovariančno matriko $\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$, njegova komponenta $\boldsymbol{X}_1$ porazdeljena večrazsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}$ in kovariančno matriko $\boldsymbol{\Sigma}_{11}$.*

Če je $\boldsymbol{\Sigma}$ pozitivno definitna, ima večrazsežna normalna porazdelitev s pričakovano vrednostjo $\boldsymbol{\mu}$ in kovariančno matriko $\boldsymbol{\Sigma}$ gostoto:

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}}\, e^{-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2}\,.$$

V enorazsežnem primeru je torej to porazdelitev $\mathrm{N}\big(\mu, \sqrt{\Sigma}\big)$ (vektor $\boldsymbol{\mu}$ in matrika $\boldsymbol{\Sigma}$ sta skalarja μ in Σ).

34. Naj bo $\begin{bmatrix}X_1\\ X_2\end{bmatrix}$ slučajni vektor, porazdeljen dvorazsežno normalno s pričakovano vrednostjo $\begin{bmatrix}0\\ 0\end{bmatrix}$ in kovariančno matriko $\begin{bmatrix}8 & 2\\ 2 & 5\end{bmatrix}$. Izračunajte $\mathbb{P}(X_1 > 0, X_2 > 0)$.

35. Naj bodo X, Y in Z neodvisne standardne normalne slučajne spremenljivke. Izračunajte $\mathbb{P}(X + Y \geq 0,\ X + Z \geq 0)$.

Če je $\begin{bmatrix}\boldsymbol{X}_1\\ \vdots\\ \boldsymbol{X}_n\end{bmatrix}$ večrazsežni normalni bločni slučajni vektor s pričakovano vrednostjo $\begin{bmatrix}\boldsymbol{\mu}_1\\ \vdots\\ \boldsymbol{\mu}_n\end{bmatrix}$ in kovariančno matriko $\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1n}\\ \vdots & \ddots & \vdots\\ \boldsymbol{\Sigma}_{n1} & \cdots & \boldsymbol{\Sigma}_{nn}\end{bmatrix}$, so komponente $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ neodvisne natanko tedaj, ko za poljubna $i \neq j$ velja $\boldsymbol{\Sigma}_{ij} = 0$. Bloki večrazsežnega bločnega normalnega slučajnega vektorja so torej neodvisni natanko tedaj, ko so paroma neodvisni.

36. Dana naj bosta neodvisna n-razsežna slučajna vektorja $\boldsymbol{X}_1$ in $\boldsymbol{X}_2$, pri čemer naj ima $\boldsymbol{X}_1$ pričakovano vrednost $\boldsymbol{\mu}_1$ in kovariančno matriko $\boldsymbol{\Sigma}_1$, $\boldsymbol{X}_2$ pa naj ima pričakovano vrednost $\boldsymbol{\mu}_2$ in kovariančno matriko $\boldsymbol{\Sigma}_2$. Kako je porazdeljena njuna vsota? Posebej zapišite to ugotovitev še za enorazsežni primer!

37. Naj bosta $\boldsymbol{X}_1$ in $\boldsymbol{X}_2$ slučajna vektorja z vrednostmi v $\mathbb{R}^m$ in $\mathbb{R}^n$, pri čemer naj bo $\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$ porazdeljen $(m+n)$-razsežno normalno s pričakovano vrednostjo $\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ in kovariančno matriko $\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$. Dokažite, da obstaja taka matrika $\mathbf{C} \in \mathbb{R}^{n\times m}$, da je $\boldsymbol{X}_2 + \mathbf{C}\boldsymbol{X}_1$ neodvisen od $\boldsymbol{X}_1$.

6. Pričakovana vrednost in sorodne karakteristike

Pričakovana vrednost, varianca. Linearnost pričakovane vrednosti, metoda indikatorjev. Nekoreliranost. Kovarianca, kovariančna matrika, Pearsonov korelacijski koeficient.

Pričakovana vrednost (matematično upanje) diskretnih slučajnih spremenljivk

$$\mathbb{E}(X) = \sum_x x \, \mathbb{P}(X = x)$$

$$\mathbb{E}\big[h(X)\big] = \sum_x h(x) \, \mathbb{P}(X = x)$$

Če je vrednosti neskončno, pričakovana vrednost obstaja, če je vrsta absolutno konvergentna.

1. Slučajna spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 2 & 3 & 4 \\ 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}1 & 0{\cdot}1 & 0{\cdot}4 \end{pmatrix}.$$

Izračunajte $\mathbb{E}(X)$ in $\mathbb{E}(X^2)$.

2. Matematik je obseden s tem, da mora biti, ko pečica piska, število piskov vedno potenca števila 2. Vsakič, ko število piskov pride do potence števila 2, z verjetnostjo 1/2 pečico ustavi, z verjetnostjo 1/2 pa pusti, da pečica piska naprej. Naj bo X število piskov. Izračunajte $\mathbb{E}(X)$ in $\mathbb{E}\big(\sqrt{X}\,\big)$.

$$\boxed{\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b}$$

3. Slučajna spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} 4900 & 5000 & 5200 \\ 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}6 \end{pmatrix}.$$

Izračunajte $\mathbb{E}(X)$.

Ne le, če je W realna slučajna spremenljivka, tudi če je W poljuben slučajni element (recimo množica), za realno funkcijo h velja formula:

$$\mathbb{E}\big[h(W)\big] = \sum_w h(w)\,\mathbb{P}(W = w)\,.$$

Če je $W = (X, Y)$ dvorazsežni slučajni vektor, tako dobimo:

$$\mathbb{E}\big[h(X, Y)\big] = \sum_x \sum_y h(x, y)\,\mathbb{P}(X = x, Y = y)\,.$$

4. Diskretna porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y = 0$	$Y = 1$	$Y = 3$
$X = 0$	0˙05	0˙1	0˙2
$X = 1$	0˙15	0	0˙15
$X = 4$	0˙2	0˙1	0˙05

   Izračunajte $\mathbb{E}(XY^2)$.

5. Pošten kovanec mečemo, dokler ne dobimo ali m grbov ali m cifer, kjer je $m > 1$ dano celo število. Seveda privzamemo, da so meti med sabo neodvisni. Izračunajte pričakovano število metov.

   Namig: prevedite na vsoto verjetnosti za porazdelitev s parametrom m, povečanim za ena.

Pričakovana vrednost zveznih slučajnih spremenljivk

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\,\mathrm{d}x$$

$$\mathbb{E}\big[h(X)\big] = \int_{-\infty}^{\infty} h(x)\, f_X(x)\,\mathrm{d}x$$

Spet le-ta obstaja, če integral absolutno konvergira.

6. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$, kjer je $\lambda > 0$. Izračunajte $\mathbb{E}(X)$ in *Laplaceovo transformacijo*: za vsak $t \in \mathbb{R}$ izračunajte $\mathbb{E}\big(e^{-tX}\big)$. Za katere t slednje obstaja?

7. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

$$f_X(x) = \begin{cases} \dfrac{1}{(1+x)^2} & ; x \geq 0 \\ 0 & ; \text{sicer.} \end{cases}$$

Pokažite, da $\mathbb{E}(X)$ ne obstaja.

Varianca (disperzija): $\operatorname{var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$.
Standardni odklon: $\sigma(X) = \sqrt{\operatorname{var}(X)}$.

8. Izračunajte varianco in standardni odklon slučajne spremenljivke s porazdelitvijo:

$$\begin{pmatrix} -1 & 0 & 2 & 3 & 4 \\ 0\text{'}1 & 0\text{'}3 & 0\text{'}1 & 0\text{'}1 & 0\text{'}4 \end{pmatrix}.$$

$$\boxed{\operatorname{var}(aX + b) = a^2 \operatorname{var}(X)}$$

9. Izračunajte varianco in standardni odklon slučajne spremenljivke s porazdelitvijo:

$$\begin{pmatrix} 4900 & 5000 & 5200 \\ 0\text{'}1 & 0\text{'}3 & 0\text{'}6 \end{pmatrix}.$$

10. Izračunajte pričakovano vrednost in varianco Poissonove porazdelitve.

11. Izračunajte vse pozitivne momente in varianco porazdelitve gama, tj. če ima slučajna spremenljivka S tako porazdelitev, izračunajte $\mathbb{E}(S^p)$ za vse $p > 0$ in še $\operatorname{var}(S)$.

12. Slučajna spremenljivka Z naj bo porazdeljena standardno normalno. Za vsak $n \in \mathbb{N}$ izračunajte $\mathbb{E}(Z^n)$. *Namig*: simetrija, indukcija.

Posledica. Če je $Z \sim \mathrm{N}(0, 1)$, je $\mathbb{E}(Z) = 0$ in $D(Z) = 1$.

Posledica. Če je $X \sim \mathrm{N}(\mu, \sigma)$, je $\mathbb{E}(X) = \mu$ in $\sigma(X) = \sigma$. Tako se pričakovana vrednost in standardni odklon iz definicije normalne porazdelitve ujemata s pojmoma, definiranima tukaj.

13. Slučajna spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$f_X(x) = \frac{4}{\pi(1 + 4x^2)^2}$$

Izračunajte $\mathbb{E}(1 + 4X^2)$ in $\operatorname{var}(X)$.

$$\boxed{\mathbb{E}\big[h(X,Y)\big] = \iint_{\mathbb{R}^2} h(x,y)\, f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y}$$

14. Slučajni vektor (X,Y) je porazdeljen zvezno z naslednjo gostoto:

$$f_{X,Y}(x,y) = \begin{cases} 2\,e^{-x} & ;\ x > 2y > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Izračunajte $\mathbb{E}\big(e^{Y-X}\big)$.

15. Izračunajte pričakovani kot, pod katerim meteorit pade na planet. Privzemite, da ima planet obliko krogle, da so vse smeri v vesolju, iz katerih prileti meteorit, enako zastopane, pri posamezni smeri pa privzemite tudi enakomerno porazdelitev trajektorij (premic), ki sekajo planet. Prav tako zanemarite ukrivljenje trajektorij zaradi gravitacije.

$$\boxed{\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)}$$

16. Slučajni vektor (X,Y) je porazdeljen zvezno z naslednjo gostoto:

$$f_{X,Y}(x,y) = \frac{c}{(1+x^2+y^2)^3}$$

Izračunajte konstanto c ter še $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{E}(X^2)$, $\mathbb{E}(Y^2)$ in $\mathbb{E}(X^2+Y^2)$.

17. Naj bo slučajna točka (X,Y,Z) porazdeljena enakomerno na enotski krogli $\{(x,y,z); x^2+y^2+z^2<1\}$. Naj bo S sinus zemljepisne širine te točke, tj.:

$$S = \frac{Z}{\sqrt{X^2+Y^2+Z^2}}\,.$$

Izračunajte $\mathbb{E}(S^2)$.

Namig: izmenjajte koordinate.

Tehnična opomba. Slučajni vektor (X,Y,Z) strogo gledano tu ne sme zavzeti vrednosti $(0,0,0)$. Njegova porazdelitev to seveda dopušča, saj za poljuben slučajni vektor (X,Y,Z) s to gostoto velja $\mathbb{P}(X=Y=Z=0)=0$. Natančneje, *obstaja* slučajni vektor (X,Y,Z) z dano gostoto, ki sploh ne zavzame vrednosti $(0,0,0)$.

Indikatorji dogodkov

Indikator *dogodka je slučajna spremenljivka, ki je na danem dogodku enaka* 1*, zunaj njega pa* 0.

Indikator dogodka A *bomo označevali z* $\mathbf{1}_A$.

Indikator dogodka, da je izjava $\mathcal{A}$ *pravilna, bomo označevali z* $\mathbf{1}(\mathcal{A})$.

Velja $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.

18. Danih je šest ploščic z naslednjimi oznakami:

$$\boxed{1} \quad \boxed{1} \quad \boxed{1} \quad \boxed{2} \quad \boxed{3} \quad \boxed{\mathrm{S}}$$

Ploščice se naključno premešajo in obrnejo. Nato jih odkrivamo, dokler ne odkrijemo ploščice $\boxed{\mathrm{S}}$. Označimo z X vsoto vseh odkritih številk. Izračunajte $\mathbb{E}(X)$ in $\mathrm{var}(X)$.

19. Izračunajte pričakovano vrednost in varianco števila negibnih točk v slučajni permutaciji n elementov, porazdeljeni enakomerno.

20. Naj $n, r \in \mathbb{N}$ ter naj bodo $X_1, X_2, \dots, X_n$ med sabo neodvisne in enako porazdeljene celoštevilske slučajne spremenljivke s:

$$\mathbb{P}(X_k = i) = \frac{1}{r}\,; \quad k = 1, 2, \dots, n, \;\; i = 1, 2, \dots, r\,.$$

Označite z Y moč slučajne množice $\{X_1, X_2, \dots, X_n\}$. Izračunajte $\mathbb{E}(Y)$ in $\mathrm{var}(Y)$.

21. Izračunajte pričakovano vrednost in varianco hipergeometrijske porazdelitve.

22. Za okroglo mizo sedi 8 gostov, ki igrajo karte. Vsak ima v rokah eno karto, razdeljeno z dobro premešanega kupa standardnih 52 kart. Vsak lahko vidi svojo karto ter karti svojih sosedov na levi in na desni. Posamezen igralec stavi zlatnik, če ima asa in hkrati nobeden od njegovih sosedov nima asa. Označimo z S število stavljenih zlatnikov. Izračunajte $\mathbb{E}(S)$ in $\mathrm{var}(S)$.

23. Iz posode, v kateri je sprva n kroglic, od tega r rdečih, na slepo in brez vračanja vlečemo kroglice, dokler ne izvlečemo rdeče. Izračunajte pričakovano vrednost in varianco števila izvlečenih kroglic.

24. Urška kupuje čevlje. Obiskati namerava tri trgovine. Če v prvi ne dobi čevljev, ki so ji všeč, po ceni največ 50 evrov, gre naprej v drugo trgovino in če tam ne dobi čevljev, ki so ji všeč, po ceni največ 50 evrov, gre še v tretjo trgovino, kjer kupi čevlje v vsakem primeru.

Cena najugodnejših čevljev, ki so Urški všeč, je v prvi trgovini porazdeljena diskretno enakomerno na množici $\{36, 45, 60\}$, v drugi zvezno z gostoto:

$$f_2(x) = \begin{cases} \frac{60-x}{200} & ; \ 40 < x < 60 \\ 0 & ; \ \text{sicer} \end{cases}$$

in v tretji normalno $\mathrm{N}(54, 10)$. Te tri cene so neodvisne.

Naj bo C cena, po kateri Urška kupi čevlje. Izračunajte $\mathbb{E}(C)$.

25. Izrazite indikator nasprotnega dogodka in indikator preseka dogodkov z indikatorji izvirnih dogodkov. Nato s pomočjo indikatorjev izpeljite načelo vključitev in izključitev.

26. Dokažite, da za poljubno slučajno spremenljivko N z vrednostmi v $\{0, 1, 2, \ldots\}$ velja:
$$\mathbb{E}(N) = \sum_{n=1}^{\infty} \mathbb{P}(N \geq n) = \sum_{n=0}^{\infty} \mathbb{P}(N > n)$$

27. Dana je Pólyeva[19] žara, v kateri je sprva b belih in dve rdeči kroglici. Iz žare drugo za drugo na slepo vlečemo kroglice. Vsakič, ko izvlečemo kroglico določene barve, jo vrnemo v posodo in dodamo še eno novo kroglico iste barve. Vlečemo, dokler ne izvlečemo rdeče kroglice. Izračunajte pričakovano število vlečenj.

28. Izračunajte pričakovano vrednost in varianco geometrijske porazdelitve.

Nekoreliranost

Slučajni spremenljivki X in Y sta nekorelirani, če velja:
$$\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)\,.$$
Poljubni neodvisni slučajni spremenljivki sta nekorelirani, obratno pa ni nujno res: nekoreliranost še ne pomeni neodvisnosti.

29. Diskretna porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 0$	0·3	0	0·3
$X = 1$	0	0·4	0

Dokažite, da sta slučajni spremenljivki X in Y nekorelirani. Sta tudi neodvisni?

[19] György Pólya (1887–1985), madžarski matematik judovskega rodu

Slučajni spremenljivki X in Y sta zagotovo neodvisni v naslednjih treh primerih:

- *če sta nekorelirani in dihotomni, tj. posamezna slučajna spremenljivka lahko zavzame kvečjemu dve vrednosti;*
- *če sta nekorelirani in je njuna navzkrižna porazdelitev dvorazsežna normalna;*
- *če za poljubni omejeni merljivi funkciji g in h velja, da sta slučajni spremenljivki $g(X)$ in $h(Y)$ nekorelirani.*

30. Diskretno porazdeljen slučajni vektor (X, Y) ima naslednjo porazdelitveno tabelo:

	$Y = -1$	$Y = 0$	$Y = 1$
$X = -1$	$\frac{1}{6} - \frac{t^2}{18}$	$\frac{t}{9}$	$\frac{t^2}{18} - \frac{t}{9} + \frac{1}{6}$
$X = 0$	$\frac{t}{9}$	$\frac{1}{3} - \frac{2t}{9}$	$\frac{t}{9}$
$X = 1$	$\frac{t^2}{18} - \frac{t}{9} + \frac{1}{6}$	$\frac{t}{9}$	$\frac{1}{6} - \frac{t^2}{18}$

a) Pri katerih t je z zgornjo tabelo res določena porazdelitev slučajnega vektorja?
b) Za katere vrednosti t sta X in Y nekorelirani?
c) Za katere vrednosti t sta X in Y neodvisni?

Če imata X in Y varianco, je nekoreliranost ekvivalentna zvezi:

$$\operatorname{var}(X + Y) = \operatorname{var}(X) + \operatorname{var}(Y)\,.$$

Pozor! *Zakaj ne velja:*

$$\operatorname{var}(X + X) = \operatorname{var}(X) + \operatorname{var}(X) = 2\operatorname{var}(X)\,?$$

31. Dani sta neodvisni slučajni spremenljivki $X \sim \chi^2(3)$ in $Y \sim \chi^2(5)$. Izračunajte $\operatorname{var}(3X - Y)$ in $\mathbb{E}\big((X - 2Y)^2\big)$.

32. Izračunajte pričakovano vrednost in varianco binomske porazdelitve.

33. Izračunajte pričakovano vrednost in varianco negativne binomske porazdelitve.

34. Naj bo $0 < q < 1/4$ ter naj bosta X in Y neodvisni in enako porazdeljeni slučajni spremenljivki s porazdelitvijo, dano po predpisu:

$$\mathbb{P}(X = n) = \mathbb{P}(Y = n) = \binom{2n}{n} q^n \sqrt{1 - 4q}\,; \quad n = 0, 1, 2, \ldots$$

(izven $\mathbb{N}_0$ pa so vse verjetnosti enake nič).

a) Določite porazdelitev vsote $X + Y$. Kot znano lahko privzamete formulo:

$$\sum_{k=0}^{n} \binom{2k}{k} \binom{2n-2k}{n-k} = 4^n .$$

b) Izračunajte $\mathrm{var}(X)$. *Namig*: pomagajte si s prejšnjo točko.

Da gre res za porazdelitev, je ekvivalentno formuli $(1-4q)^{-1/2} = \sum_{n=0}^{\infty} \binom{2n}{n} q^n$. Zlahka preverimo, da je $\binom{2n}{n} = (-4)^n \binom{-1/2}{n}$. Če uvedemo $x = -4q$, se formula prevede na $(1+x)^{-1/2} = \sum_{n=0}^{\infty} \binom{-1/2}{n} x^n$, kar je Newtonova binomska formula. V resnici gre torej za Pólyevo porazdelitev.

Formula z vsoto binomskih koeficientov je konvolucijska formula za centralne binomske koeficiente in se prevede na obliko $\sum_{k=0}^{n} \binom{-1/2}{k}\binom{-1/2}{n-k} = (-1)^n$. To formulo zlahka preverimo tako, da v identiteti $(1+x)^{-1/2}(1+x)^{-1/2} = (1+x)^{-1}$ preverimo koeficient pri x^n.

Konvolucijska formula za centralne binomske koeficiente pa ima tudi razmeroma eleganten kombinatoričen dokaz. Ta temelji na končnih celoštevilskih zaporedjih ki se na vsakem koraku dvignejo za 1 ali pa spustijo za 1. Recimo takim zaporedjem *ustrezna*. Vsako ustrezno zaporedje lahko opišemo bodisi z vrednostmi $x_0, \ldots, x_N$ bodisi z diferencami $x_i - x_{i-1}$, $i = 1, 2, \ldots, N$. Te lahko izbiramo neodvisno. Če torej predpišemo začetno vrednost x_0, bo v okviru tega vseh ustreznih zaporedij 2^N.

Za zaporedje $x_0, \ldots, x_N$ bomo rekli, da je to zaporedje tipa (m, d), če je $m = \max_i(x_i - x_0)$ in $d = x_k - x_0$. Oglejmo naslednjo transformacijo, pri kateri vse člene od vključno prvega, kjer je dosežen globalni maksimum, znižamo za 2. Če je $m > 0$, ta transformacija zaporedje tipa (m, d), ki je ustrezno, preslika v zaporedje tipa $(m-1, d-2)$, ki je spet ustrezno. Nadalje je dobljena preslikava bijektivna: inverzna transformacija vse člene od nevključno zadnjega, kjer je dosežen globalni maksimum, zviša za 2. Z iteriranjem dobimo bijektivno korespondenco med ustreznimi zaporedji tipa (m, d) in ustreznimi zaporedji tipa $(0, d-2m)$. To velja tudi za $m = 0$ – v tem primeru zaporedja pustimo pri miru.

Naj bo zdaj $N = 2k$ sodo število. Če je $x_0 = x_N = 0$, je maksimum zaporedja lahko enak $0, 1, \ldots, k$. Z zgoraj omenjeno transformacijo se zaporedje transformira v zaporedje z maksimumom nič, zadnji člen pa je lahko $-2k, -2k+2, \ldots, 0$. S tem pa zajamemo vsa možna zaporedja z maksimumom nič. Sklep: zaporedij $x_0, \ldots, x_N$ z $x_0 = 0$ in maksimumom nič je toliko kot zaporedij z $x_0 = x_N = 0$. Ta zaporedja pa lahko opišemo z diferencami, med katerimi jih je k enakih 1, k pa -1. Takih zaporedij je torej $\binom{2k}{k}$.

Oglejmo so zdaj vsa zaporedja $x_0, x_1, \ldots, x_{2n+1}$, ki so ustrezna ter za katera je $x_0 = 0$ in $x_{2n+1} < 0$. Zaradi simetrije je takih zaporedij $\frac{1}{2} 2^{2n+1} = 4^n$. Ta zaporedja pa lahko preštejemo tudi še drugače. Oglejmo si zadnji indeks, kjer je tako zaporedje enako nič. Ta indeks mora biti sod, naj bo npr. $2k$. Ker je to zadnji indeks, kjer je vrednost enaka nič, končni člen pa je manjši od nič, mora biti $x_{2k+1} = -1$ in noben nadaljnji člen ne sme preseči -1. Zaporedju smo torej priredili trojico, ki jo sestavljajo število $k \in \{0, 1, \ldots, n\}$, zaporedje $x_0, \ldots, x_{2k}$ z lastnostjo, da je $x_0 = x_{2k} = 0$, in zaporedje $x_{2k+1}, \ldots, x_{2n+1}$ z lastnostjo, da je $x_{2k+1} = -1$ in $\max_{2k+1 \le j \le 2n+1}(x_j - x_{2k+1}) = 0$. Velja pa tudi obratno – vsako trojico z zgoraj opisanimi lastnostmi lahko sestavimo v zaporedje $x_0, x_1, \ldots, x_{2n+1}$, ki je ustrezno ter za katero je $x_0 = 0$ in $x_{2n+1} < 0$.

Preštejmo zdaj omenjene trojice. Če predpišemo k, je zaporedij $x_0, \ldots, x_{2k}$ z lastnostjo, da je $x_0 = x_{2k} = 0$, $\binom{2k}{k}$, zaporedij $x_{2k+1}, \ldots, x_{2n+1}$ z lastnostjo, da je $x_{2k+1} = -1$ in $\max_{2k+1 \le j \le 2n+1}(x_j - x_{2k+1}) = 0$, pa je $\binom{2n-2k}{n-k}$. Seštejemo po k in rezultat sledi.

35. Naj bosta X in Y neodvisni in enako porazdeljeni slučajni spremenljivki z $\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = \sigma^2$ in $\mathbb{E}(X^4) = \mathbb{E}(Y^4) = \kappa^4$. Izračunajte $\mathrm{var}\big((X+Y)^2\big)$. Posplošite na več spremenljivk!

Kovarianca:

$$\operatorname{cov}(X,Y) := \mathbb{E}\big((X-\mathbb{E}(X))\,(Y-\mathbb{E}(Y))\big) = \\ = \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)$$

Velja $\operatorname{cov}(X,X) = \operatorname{var}(X)$ *in* $\operatorname{cov}(Y,X) = \operatorname{cov}(X,Y)$.
Če so a, b, c *in* d *konstante, velja* $\operatorname{cov}(aX+b, cY+d) = ac\operatorname{cov}(X,Y)$.
Slučajni spremenljivki X *in* Y *sta nekorelirani natanko tedaj, ko je* $\operatorname{cov}(X,Y) = 0$.

Korelacijski koeficient:

$$\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\,\sigma(Y)}$$

Velja $-1 \le \operatorname{corr}(X,Y) \le 1$. *Če so* a, b, c *in* d *konstante ter* $a, c > 0$, *velja* $\operatorname{corr}(aX+b,\ cY+d) = \operatorname{corr}(X,Y)$.

36. Diskretna porazdelitev slučajnega vektorja (X,Y) je podana z naslednjo tabelo:

	$Y=0$	$Y=10$	$Y=30$
$X=0$	0'05	0'1	0'2
$X=10$	0'15	0	0'15
$X=40$	0'2	0'1	0'05

Izračunajte $\operatorname{cov}(X,Y)$ in $\operatorname{corr}(X,Y)$.

37. Slučajni vektor (X,Y) je porazdeljen zvezno z naslednjo gostoto:

$$f_{X,Y}(x,y) = \begin{cases} 3(x^2y + xy^2) & ;\ x,y \in [0,1] \\ 0 & ;\ \text{sicer.} \end{cases}$$

Izračunajte $\operatorname{cov}(X,Y)$ in $\operatorname{corr}(X,Y)$.

38. V posodi je N kroglic, med njimi r rdečih in z zelenih. Na slepo in brez vračanja izvlečemo n kroglic. Označimo z R število rdečih, z Z pa število zelenih med izvlečenimi. Izračunajte $\operatorname{cov}(R,Z)$.

$$\operatorname{cov}(X+Y,Z) = \operatorname{cov}(X,Z) + \operatorname{cov}(Y,Z)\,, \quad \operatorname{cov}(X,Y+Z) = \operatorname{cov}(X,Y) + \operatorname{cov}(X,Z)$$

39. Za okroglo mizo izmenoma sedi n žensk in n moških, kjer je $n \ge 3$. Vsak vrže standardno kocko, meti so neodvisni. Vsaka ženska stavi biser, če vrže šestico, obenem pa nobeden od njenih sosedov ne vrže šestice. Nadalje vsak moški stavi cekin, če vrže enojko, obenem pa nobena od njegovih sosed ne vrže enojke. Naj bo X število stavljenih biserov, Y pa število stavljenih cekinov. Izračunajte $\operatorname{corr}(X,Y)$.

40. *Grayeva koda* je zapis naravnega števila ali ničle, ki ga dobimo iz binarnega zapisa, tako da števko (bit) spremenimo, če je števka levo od nje (torej naslednja pomembnejša) enaka `1`; sicer števko ohranimo. Primer: število 13 ima binarni zapis `1101` in Grayevo kodo `1011`.

    Grayeva koda posameznega števila ima enako mest kot binarni zapis in različni števili imata različni Grayevi kodi (preslikava, ki število preslika v njegovo Grayevo kodo, je torej injektivna). Odlikuje pa se po tem, da se, če število povečamo za 1, vselej spremeni le ena števka.

    Naj bo $n \in \mathbb{N}$ in N naključno izbrano število iz množice $\{0, 1, 2, \ldots, 2^n - 1\}$.

    a) Če binarnemu zapisu ali Grayevi kodi števila N dodamo ustrezno število vodilnih ničel, dobimo slučajno zaporedje n ničel in enic. Utemeljite, da za tako dobljeno slučajno zaporedje, ki pripada binarnemu zapisu ali Grayevi kodi števila N, velja, da se na posameznem mestu ničla ali enica pojavi z verjetnostjo 1/2 in da so mesta med seboj neodvisna.

    b) Naj bo X število enic v binarnem zapisu, Y pa število enic v Grayevi kodi števila N. Izračunajte $\operatorname{corr}(X, Y)$.

Pričakovano vrednost slučajnega vektorja definiramo po komponentah:
za $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ *definiramo* $\boldsymbol{E}(\boldsymbol{X}) := \begin{bmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{bmatrix}$.
Podobno po komponentah definiramo pričakovano vrednost slučajne matrike $\mathbf{M} = [Y_{ij}]_{i,j}$:

$$\mathbf{E}(\mathbf{M}) := [\mathbb{E}(Y_{ij})]_{i,j} \,.$$

Varianci slučajnega vektorja ustreza **kovariančna matrika**:

$$\mathbf{K}(\boldsymbol{X}) := \begin{bmatrix} \operatorname{cov}(X_1, X_1) & \cdots & \operatorname{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_n, X_1) & \cdots & \operatorname{cov}(X_n, X_n) \end{bmatrix} = \mathbf{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}) - \mathbf{E}(\boldsymbol{X})\,\mathbf{E}(\boldsymbol{X})^{\mathrm{T}} \,.$$

Kovarianci med dvema slučajnima vektorjema ustreza splošnejša kovariančna matrika:

$$\mathbf{K}(\boldsymbol{X}, \boldsymbol{Y}) := \begin{bmatrix} \operatorname{cov}(X_1, Y_1) & \cdots & \operatorname{cov}(X_1, Y_n) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}(X_m, Y_1) & \cdots & \operatorname{cov}(X_n, Y_n) \end{bmatrix} = \mathbf{E}(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}) - \mathbf{E}(\boldsymbol{X})\,\mathbf{E}(\boldsymbol{Y})^{\mathrm{T}} \,.$$

41. Slučajni vektor (X, Y, Z) ima naslednjo kovariančno matriko:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix} .$$

Določite parameter a tako, da bosta slučajni spremenljivki $U := aX + 2Y + Z$ in $V := X - Y + aZ$ nekorelirani.

Za poljubno deterministično matriko $\mathbf{A}$ *in poljuben slučajni vektor* $\boldsymbol{X}$ *velja:*

$$\boldsymbol{E}(\mathbf{A}\boldsymbol{X}) = \mathbf{A}\,\boldsymbol{E}(\boldsymbol{X})\,.$$

Posledično za vsak determinističen vektor $\boldsymbol{u}$ *prave dimenzije velja:*

$$\mathbb{E}(\langle \boldsymbol{X}, \boldsymbol{u}\rangle) = \mathbb{E}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{X}) = \boldsymbol{u}^{\mathrm{T}}\,\boldsymbol{E}(\boldsymbol{X}) = \langle \boldsymbol{E}(\boldsymbol{X}), \boldsymbol{u}\rangle\,.$$

Nadalje za poljubni deterministični matriki $\mathbf{A}$ *in* $\mathbf{B}$ *in poljubno slučajno matriko* $\mathbf{M}$ *pravih dimenzij velja:*

$$\mathbf{E}(\mathbf{AM}) = \mathbf{A}\,\mathbf{E}(\mathbf{M}) \quad \textit{in} \quad \mathbf{E}(\mathbf{MB}) = \mathbf{E}(\mathbf{M})\,\mathbf{B}\,.$$

42. Naj bosta $\boldsymbol{X}$ in $\boldsymbol{Y}$ slučajna vektorja.

    a) Za poljubno deterministično matriko $\mathbf{A}$ pravih dimenzij izrazite $\mathbf{K}(\mathbf{A}\boldsymbol{X})$ s $\mathbf{K}(\boldsymbol{X})$.
    b) Za poljubni deterministični matriki $\mathbf{A}$ in $\mathbf{B}$ pravih dimenzij izrazite $\mathbf{K}(\mathbf{A}\boldsymbol{X}, \mathbf{B}\boldsymbol{Y})$ s $\mathbf{K}(\boldsymbol{X}, \boldsymbol{Y})$.
    c) Za poljubna deterministična vektorja $\boldsymbol{u}$ in $\boldsymbol{v}$, ki sta iste dimenzije kot $\boldsymbol{X}$ oziroma $\boldsymbol{Y}$, izrazite $\operatorname{cov}\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle, \langle \boldsymbol{Y}, \boldsymbol{v}\rangle\big)$ s $\mathbf{K}(\boldsymbol{X}, \boldsymbol{Y})$.
    d) Karakterizirajte matrike, ki so kovariančne matrike enega slučajnega vektorja.

43. Slučajni vektor $\boldsymbol{X}$ ima naslednjo kovariančno matriko:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 5 & 2 & 4 \\ 2 & 2 & 2 \\ 4 & 2 & 5 \end{bmatrix}.$$

    V katerih smereh je varianca največja in v katerih najmanjša ter koliko znaša? Z drugimi besedami, za katere enotske vektorje $\boldsymbol{u}$ je varianca slučajne spremenljivke $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ največja in za katere najmanjša?

44. Dokažite, da se pričakovana vrednost in kovariančna matrika iz definicije večrazsežne normalne porazdelitve ujemata s pričakovano vrednostjo in kovariančno matriko, definiranima v tem razdelku.

7. Pogojne porazdelitve in pogojevanje na slučajne spremenljivke

Pogojna porazdelitev diskretne slučajne spremenljivke glede na dogodek in glede na diskretno slučajno spremenljivko. Pogojna verjetnost dogodka glede na poljubno slučajno spremenljivko. Pogojna gostota. Pogojna pričakovana vrednost.

Pogojno porazdelitev *slučajne spremenljivke X glede na dogodek B opišemo s pogojnimi verjetnostmi $\mathbb{P}(X \in C \mid B)$, kjer C preteče vse merljive množice. Če je X diskretna, lahko njeno pogojno porazdelitev opišemo s* **pogojno porazdelitveno shemo**:

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ \mathbb{P}(X = a_1 \mid B) & \mathbb{P}(X = a_2 \mid B) & \cdots \end{pmatrix}.$$

1. Vržemo tri poštene in neodvisne kovance in jih razporedimo v vrsto. Naj bo X število grbov pri prvih dveh kovancih. Nato pride Pepček, povsem naključno izbere dva različna kovanca in ju zamenja. Zdaj je na obeh prvih kovancih grb. Zapišite pogojno porazdelitev slučajne spremenljivke X glede na omenjeno opažanje.

2. Slučajna spremenljivka X je porazdeljena geometrijsko Geom(1/3). Določite njeno pogojno porazdelitev glede na dogodek $\{X < 5\}$.

3. Porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	$\frac{1}{12}$	$\frac{1}{6}$	0
$X = 1$	$\frac{1}{4}$	0	$\frac{1}{12}$
$X = 2$	0	$\frac{1}{6}$	$\frac{1}{4}$

Določite pogojno porazdelitev slučajne spremenljivke X glede na $Y = 2$ ter Y glede na $X = 0$ in glede na $Y \geq X$.

Za vsako realno slučajno spremenljivko in vsak dogodek B s pozitivno verjetnostjo lahko pogojno porazdelitev slučajne spremenljivke X glede na B opišemo s **pogojno kumulativno porazdelitveno funkcijo**:

$$F_{X|B}(x) = \mathbb{P}(X \le x \mid B)\,.$$

Če je pogojna porazdelitev zvezna, obstaja tudi **pogojna porazdelitvena gostota***:*

$$f_{X|B}(x) = F'_{X|B}(x)\,.$$

Brž ko je X zvezno porazdeljena, je tudi njena pogojna porazdelitev zvezna – glede na vsak dogodek s pozitivno verjetnostjo.
Če je X porazdeljena zvezna z gostoto f_X in $\mathbb{P}(X \in C) > 0$, je:

$$f_{X|X\in C}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbb{P}(X \in C)} & ; x \in C \\ 0 & ; \textit{sicer.} \end{cases}$$

Podobno velja tudi za zvezne slučajne vektorje.

4. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Za vsak $a \ge 0$ določite pogojno porazdelitev slučajne spremenljivke X glede na dogodek $\{X \ge a\}$.

5. Slučajni spremenljivki X in Y sta neodvisni in porazdeljeni enakomerno na intervalu $[0, 1]$. Določite pogojno porazdelitev slučajne spremenljivke X glede na dogodek $\{X < Y\}$.

6. Računalnik povsem naključno izbere realno število X med 0 in 1. Tine ga pogleda in Tone ga z verjetnostjo $1/2$ vpraša, ali je to število manjše od $2/3$, z verjetnostjo $1/2$ pa, ali je večje od $1/3$ (izbira vprašanja je neodvisna od izbranega števila). Odgovor je pritrdilen. Zapišite pogojno porazdelitveno gostoto izbranega števila X glede na dani odgovor (kaj točno je Tone vprašal Tineta, ne vemo).

Pogojna pričakovana vrednost *slučajne spremenljivke* X *glede na dogodek* B *je pričakovana vrednost, ki pripada ustrezni pogojni porazdelitvi, in ga označimo z* $\mathbb{E}(X \mid B)$. *Tako velja:*

$$\mathbb{E}(X \mid B) = \sum_x x\,\mathbb{P}(X = x \mid B)$$

in splošneje:

$$\mathbb{E}\big(h(X) \mid B\big) = \sum_x h(x)\,\mathbb{P}(X = x \mid B)\,.$$

Pogojna pričakovana vrednost ima vse lastnosti običajne pričakovane vrednosti, npr. linearnost.
Podobno definiramo tudi **pogojno varianco**. *Velja:*

$$\operatorname{var}(X \mid B) = \mathbb{E}\Big[\big(X - \mathbb{E}(X \mid B)\big)^2 \;\Big|\; B\Big] = \mathbb{E}\big(X^2 \mid B\big) - \big(\mathbb{E}(X \mid B)\big)^2\,.$$

7. Za slučajno spremenljivko X iz 1. naloge in dogodek B, da je po Pepčkovi zamenjavi na obeh prvih kovancih grb, izračunajte $\mathbb{E}(X \mid B)$ in $\operatorname{var}(X \mid B)$. Prav tako izračunajte $\mathbb{E}(X \mid D)$ in $\operatorname{var}(X \mid D)$ za slučajno spremenljivko X iz 6. naloge in dogodek D, da Tine odgovori pritrdilno.

Za vsako slučajno spremenljivko X *in vsak dogodek* B *velja:*

$$\mathbb{E}(X \mid B) = \frac{\mathbb{E}(XZ)}{\mathbb{P}(B)} = \frac{\mathbb{E}(XZ)}{\mathbb{E}(Z)}\,,$$

kjer je slučajna spremenljivka Z **indikator** *dogodka* B, *tj. enaka* 1 *na dogodku* B *in* 0 *zunaj njega.*

8. Naj bosta X in Y neodvisni slučajni spremenljivki, pri čemer naj bo X porazdeljena enakomerno na $[0,1]$, Y pa enakomerno na $\{0,1,2\}$. Izračunajte $\mathbb{E}(XY \mid 2X > Y)$.

Za vsako slučajno spremenljivko X *s pričakovano vrednostjo in vsak popoln sistem dogodkov* $H_1, H_2, H_3, \ldots$ *velja* **izrek o polni pričakovani vrednosti**:

$$\mathbb{E}(X) = \mathbb{P}(H_1)\,\mathbb{E}(X \mid H_1) + \mathbb{P}(H_2)\,\mathbb{E}(X \mid H_2) + \mathbb{P}(H_3)\,\mathbb{E}(X \mid H_3) + \cdots$$

9. V prvi posodi je 6 belih in 4 rdeče kroglice, v drugi pa 3 bele in 6 rdečih. Najprej na slepo premestimo eno kroglico iz prve posode v drugo, nato pa iz druge posode na slepo in brez vračanja izvlečemo tri kroglice. Izračunajte pričakovano število belih med njimi.

10. Pošten kovanec mečemo, dokler ne padeta dve cifri zapored. Meti so med seboj neodvisni. Izračunajte pričakovano število vseh metov.

Pogojno verjetnost *dogodka A* **glede na diskretno slučajno spremenljivko** *Y lahko definiramo bodisi kot funkcijo, ki y slika v $\mathbb{P}(A \mid Y = y)$, bodisi kot slučajno spremenljivko, ki jo označimo s $\mathbb{P}(A \mid Y)$: na dogodku $\{Y = y\}$ s pozitivno verjetnostjo definiramo:*

$$\mathbb{P}(A \mid Y) := \mathbb{P}(A \mid Y = y)\,,$$

na dogodku $\{Y = y\}$ z verjetnostjo nič pa vrednost izberemo poljubno, a konstantno na celem dogodku. Tako dobimo slučajno spremenljivko, ki je funkcija slučajne spremenljivke Y, določena pa je **skoraj gotovo**: *poljubni izbiri se z verjetnostjo ena ujemata.*
Tako definirana slučajna spremenljivka je odvisna samo od informacije, ki jo nudi Y (natančneje σ-algebre, generirane z Y): če je g merljiva bijektivna funkcija, je $\mathbb{P}(A \mid g(Y)) = \mathbb{P}(A \mid Y)$ (natančneje, vsaka izbira, ki je dobra za levo stran, je dobra tudi za desno stran).

11. Pošteno kocko mečemo, dokler ne pade šestica, meti so neodvisni. Izračunajte pogojno verjetnost, da pri tem ne pade nobena enojka, glede na število vseh metov.

12. Na kupu je 16 dobro premešanih kart, in sicer po štirje asi, kralji, dame in fanti. Naj bo A dogodek, da je prva karta as, z Y pa označimo število kraljev med prvimi štirimi kartami.

    a) Določite pogojno verjetnost dogodka A glede na Y.

    b) Izračunajte $\mathbb{E}\big(\mathbb{P}(A \mid Y)\big)$. Ali kaj opazite?

    c) Dokažite, da za *vsako* diskretno slučajno spremenljivko Y in *vsak* dogodek A velja:
    $$\mathbb{E}\big(\mathbb{P}(A \mid Y)\big) = \mathbb{P}(A)\,.$$

13. Pošteno kocko spet mečemo, dokler ne pade šestica, meti so neodvisni. Kolikšna je verjetnost, da pri tem ne pade nobena enojka?

Pogojno porazdelitev *slučajne spremenljivke X glede na diskretno slučajno spremenljivko Y opišemo s pogojnimi verjetnostmi $\mathbb{P}(X \in C \mid Y)$, kjer C preteče vse merljive množice. Tako dobimo slučajno verjetnostno mero. Če je X diskretna, lahko seveda njeno pogojno porazdelitev opišemo s pogojno porazdelitveno shemo:*

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ \mathbb{P}(X = a_1 \mid Y) & \mathbb{P}(X = a_2 \mid Y) & \cdots \end{pmatrix}.$$

Za poljubno realno slučajno spremenljivko X lahko definiramo pogojno kumulativno porazdelitveno funkcijo:

$$F_{X|Y}(x) = \mathbb{P}(X \le x \mid Y) \quad \textit{ali tudi} \quad F_{X|Y}(x \mid y) = \mathbb{P}(X \le x \mid Y = y),$$

za zvezno porazdeljene pa pogojno gostoto: $f_{X|Y}(x)$ je slučajna funkcija spremenljivke x, $f_{X|Y}(x \mid y)$ pa je (deterministična) funkcija spremenljivk x in y. Lahko definiramo tudi pogojno pričakovano vrednost:

$$\mathbb{E}(X \mid Y) = g(Y), \quad \textit{kjer je} \quad g(y) = \mathbb{E}(X \mid Y = y).$$

Funkciji g pravimo **regresijska funkcija**. *Prav tako lahko definiramo pogojno varianco* $\mathrm{var}(X \mid Y)$ *in podobno.*

14. Iz posode, v kateri so najprej ena rdeča, dve zeleni in sedem belih kroglic, na slepo in brez vračanja vlečemo kroglice, dokler ne izvlečemo rdeče. Naj bo X število zelenih, Y pa število vseh izvlečenih kroglic.

    a) Zapišite pogojno porazdelitev slučajne spremenljivke X glede na Y.
    b) Izračunajte $\mathbb{E}(X \mid Y)$ in $\mathbb{E}\big(\mathbb{E}(X \mid Y)\big)$. Ali kaj opazite?
    c) Dokažite, da za vsako slučajno spremenljivko X in diskretno slučajno spremenljivko Y velja $\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = \mathbb{E}(X)$.

15. Dani sta neodvisni slučajni spremenljivki $X \sim \mathrm{Pois}(\lambda)$ in $Y \sim \mathrm{Pois}(\mu)$, kjer je $\lambda, \mu > 0$. Določite pogojno porazdelitev slučajne spremenljivke X glede na $S := X + Y$.

16. Podobno kot v prejšnji nalogi naj bodo $X_1, X_2, \ldots, X_n$ neodvisne slučajne spremenljivke z $X_i \sim \mathrm{Pois}(\lambda_i)$. Označimo $S := X_1 + X_2 + \cdots + X_n$. Izračunajte pogojno porazdelitev slučajnega vektorja $(X_1, X_2, \ldots, X_n)$ glede na S.

17. Naj bosta X in Y slučajni spremenljivki, pri čemer naj bo Y porazdeljena binomsko $\mathrm{Bin}(2, 1/2)$ ter še $\mathbb{E}(X \mid Y) = Y$ in $\mathrm{var}(X \mid Y) = Y + 1$. Izračunajte $\mathbb{E}(X)$ in $\mathrm{var}(X)$.

18. Naj bo $X_1, X_2, \ldots$ zaporedje slučajnih spremenljivk, pri čemer naj bo $X_1 = 1$, za $n = 1, 2, \ldots$ pa naj velja:

$$\mathbb{P}(X_{n+1} = X_n \mid X_n) = \frac{n + 1 - X_n}{n + 1} \quad \text{in} \quad \mathbb{P}(X_{n+1} = X_n + 1 \mid X_n) = \frac{X_n}{n + 1}.$$

Izračunajte $\mathbb{E}(X_n)$ in $\operatorname{var}(X_n)$.

Namig: pri varianci si pomagajte s slučajnimi spremenljivkami $Y_n := X_n(X_n + 1)$.

19. Naj bo $0 < a, b < 1$. Slučajna spremenljivka X je porazdeljena geometrijsko $\mathrm{Geom}(a)$ in pogojno na X je slučajna spremenljivka Y porazdeljena negativno binomsko $\mathrm{NB}(X, b)$. Določite brezpogojno porazdelitev slučajne spremenljivke Y.

20. Pošteno kocko spet mečemo, dokler ne pade šestica, meti so neodvisni.

   a) Določite *pogojno* porazdelitev števila vseh enojk glede na število vseh metov.

   b) Določite *brezpogojno* porazdelitev števila vseh enojk.

21. Naj bodo $X_1, X_2, \ldots$ neodvisne enako porazdeljene Bernoullijeve slučajne spremenljivke, tj.:

$$X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

in naj bo $N \sim \mathrm{Pois}(\lambda)$ neodvisna od slučajnih spremenljivk X_i. Označimo:

$$S = X_1 + X_2 + \cdots + X_N \,.$$

   a) Določite *pogojno* porazdelitev slučajne spremenljivke S glede na N.

   b) Določite *brezpogojno* porazdelitev slučajne spremenljivke S.

   c) Dokažite, da sta slučajni spremenljivki S in $T := N - S$ neodvisni.

   d) Dokažite, da, če spremenimo porazdelitev slučajne spremenljivke N, ni več nujno, da sta S in T neodvisni.

Opomba. Transformaciji, pri katerih iz slučajne spremenlivke N nastane slučajna spremenljivka S, pravimo *redčenje* (angl. *thinning*).

Če je Y porazdeljena diskretno, X pa pogojno na Y porazdeljena zvezno z gostoto $f_{X|Y}$, je tudi brezpogojna porazdelitev slučajne spremenljivke X zvezna z gostoto:

$$f_X(x) = \mathbb{E}\big[f_{X|Y}(x)\big] \,.$$

22. Naj bodo $X_1, X_2, \ldots$ neodvisne slučajne spremenljivke, porazdeljene eksponentno $\mathrm{Exp}(\lambda)$, N pa naj bo neodvisna od prej omenjenih slučajnih spremenljivk in porazdeljena geometrijsko $\mathrm{Geom}(p)$. Zapišite porazdelitveno gostoto slučajne spremenljivke:

$$S = X_1 + X_2 + \cdots + X_N \,.$$

Če sta X in Y slučajni spremenljivki in $\mathbb{P}(Y = y) > 0$, se pogojna porazdelitev slučajne spremenljivke $h(X, Y)$ glede na $Y = y$ ujema s pogojno porazdelitvijo slučajne spremenljivke $h(X, y)$ glede na ta dogodek. Če je X diskretna, torej velja:

$$\mathbb{E}\big[h(X, Y) \mid Y = y\big] = \sum_x h(x, y)\,\mathbb{P}(X = x \mid Y = y)\,.$$

23. Naj bo slučajni vektor (X, Y) porazdeljen tako kot v 3. nalogi. Izračunajte $\mathbb{E}(X \mid Y = 2)$, $\mathbb{E}(X^2 \mid Y = 2)$ in $\mathbb{E}(X^2 Y^2 \mid Y = 2)$.

Če je X slučajna spremenljivka s pričakovano vrednostjo in Y diskretna slučajna spremenljivka, velja:

$$\mathbb{E}\big[X\,g(Y) \bigm| Y\big] = \mathbb{E}(X \mid Y)\,g(Y)$$

in posledično:

$$\mathbb{E}\big[X\,g(Y)\big] = \mathbb{E}\big[\mathbb{E}(X \mid Y)\,g(Y)\big]\,.$$

24. Pošteno kocko spet mečemo, dokler ne pade šestica, meti so neodvisni. Izračunajte pričakovani delež enojk med vsemi meti.

Za vsako slučajno spremenljivko X s pričakovano vrednostjo in **vsako** *slučajno spremenljivko Y obstaja slučajna spremenljivka Z, ki je funkcija slučajne spremenljivke Y in za katero je $\mathbb{E}\big[Z\,g(Y)\big] = \mathbb{E}\big[X\,g(Y)\big]$ za vsako merljivo funkcijo g, za katero desna stran obstaja. Slučajna spremenljivka Z je določena do skoraj gotovega ujemanja natančno. Slučajni spremenljivki Z pravimo* **pogojna pričakovana vrednost glede na slučajno spremenljivko** *Y in pišemo $Z = \mathbb{E}(X \mid Y)$. Velja tudi:*

$$\mathbb{E}\big[X\,g(Y) \bigm| Y\big] = \mathbb{E}(X \mid Y)\,g(Y)\,.$$

Regresijska funkcija *je definirana kot $\mathbb{E}(X \mid Y = y) = h(y)$, kjer je $h(Y) = \mathbb{E}(X \mid Y)$. Ni pa nujno natančno določena.*

Če sta X in Y neodvisni, je $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$.

25. Naj bodo X, Y in Z slučajne spremenljivke ter $a \in \mathbb{R}$. Privzemimo naslednje:
    - X, $Y - aX$ in $Z - aY$ so neodvisne.
    - $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z) = 0$.
    - $\mathbb{E}(X^2) = \mathbb{E}(Y^2) = \mathbb{E}(Z^2) = 1$.

    Izračunajte $\mathbb{E}(Z \mid X, Y)$ in $\mathbb{E}(YZ \mid X)$.

Pogojna verjetnost *dogodka A glede na slučajno spremenljivko Y je definirana kot pogojna pričakovana vrednost njegovega indikatorja Z: $\mathbb{P}(A \mid Y) = \mathbb{E}(Z \mid Y)$.*
Pogojna porazdelitev *slučajne spremenljivke X glede na Y je nabor pogojnih verjetnosti $\mathbb{P}(X \in C \mid Y)$, kjer C preteče vse merljive množice. Lahko jo gledamo tudi kot slučajno preslikavo $C \mapsto \mathbb{P}(X \in C \mid Y)$, ki pa mora biti povsod verjetnostna mera. V splošnem ni nujno, da pogojna porazdelitev obstaja, a pogojne porazdelitve realnih (in mnogih drugih) slučajnih spremenljivk vedno obstajajo.*
Na podlagi pogojne porazdelitve lahko definiramo pogojno kumulativno porazdelitveno funkcijo, pogojno gostoto, pogojno pričakovano vrednost, pogojno varianco itd. Definicija pogojne pričakovane vrednosti poljubne funkcije slučajne spremenljivke na podlagi pogojne porazdelitve se ujema s prvotno definicijo.

26. Naj bodo $X_1, X_2, \ldots$ in N tako kot v 22. nalogi. Zapišite pogojno porazdelitev slučajne spremenljivke N glede na S.

27. Slučajna spremenljivka U naj bo porazdeljena zvezno enakomerno na $(0,1)$ in pogojno na U naj bo slučajna spremenljivka X porazdeljena binomsko $\mathrm{Bin}(n, U)$. Določite porazdelitev slučajne spremenljivke X in še pogojno porazdelitev slučajne spremenljivke U glede na X.

    Namig: Ker je X porazdeljena diskretno, je dovolj izračunati pogojne porazdelitve slučajne spremenljivke U glede na dogodke $\{X = k\}$, kjer je $k = 0, 1, \ldots, n$; le-te pa lahko dobimo na podlagi pogojnih pričakovanih vrednosti $\mathbb{E}\big[h(U) \bigm| X = k\big]$, kjer je h merljiva *preizkusna funkcija*.

Opomba. Ta naloga sodi v *Bayesovo statistiko*: enakomerna porazdelitev je *apriorna* porazdelitev slučajne spremenljivke U, želimo pa izračunati njeno *aposteriorno* porazdelitev glede na opažanje X.

28. Naj bo $0 < q < 1$. Slučajna spremenljivka N naj ima logaritemsko porazdelitev, tj. zavzame vrednosti v $\mathbb{N}$ in velja:

$$\mathbb{P}(N = n) = -\frac{1}{\ln(1-q)} \frac{q^n}{n}\,; \quad n = 1, 2, 3, \ldots$$

    Nadalje naj ima slučajna spremenljivka X pogojno na N Paretovo[20] porazdelitev z gostoto:

$$f_{X \mid N}(x \mid n) = \begin{cases} \dfrac{n}{x^{n+1}} & ; x > 1 \\ 0 & ; \text{sicer.} \end{cases}$$

    Določite brezpogojno porazdelitev slučajne spremenljivke X in še pogojno porazdelitev slučajne spremenljivke N glede na X.

[20] Vilfredo Federico Damaso Pareto (1848–1923), italijanski inženir, sociolog, ekonomist, politolog in filozof

29. Za slučajni vektor (X, Y) iz 3. naloge zapišite porazdelitev slučajnih spremenljivk $\mathbb{E}(X \mid Y)$ in $\operatorname{var}(X \mid Y)$, nato pa izračunajte še $\operatorname{var}\big(\mathbb{E}(X \mid Y)\big)$ in $\mathbb{E}\big(\operatorname{var}(X \mid Y)\big)$. Rezultata primerjajte z $\operatorname{var}(X)$. Kaj opazite?

30. Naj obstaja $\mathbb{E}(X^2)$. Dokažite zvezo:

$$\operatorname{var}(X) = \operatorname{var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\operatorname{var}(X \mid Y)\big)\,.$$

    Opomba: prvemu členu pravimo *pojasnjena*, drugemu členu pa *nepojasnjena* ali *rezidualna* varianca. Z izrazom pojasnjena je mišljena pojasnjenost z odvisnostjo slučajne spremenljivke X od Y (tj. slučajna spremenljivka X ima za različne vrednosti Y različne pogojne porazdelitve, z njimi pa lahko različne pogojne pričakovane vrednosti).

31. Naj bo $Y \sim \operatorname{Exp}(\lambda)$, pogojno na Y pa naj bo $X \sim \operatorname{Pois}(Y)$. Izračunajte $\mathbb{E}(X^2)$.

32. Naj bo spet $\lambda > 0$, $Y \sim \operatorname{Exp}(\lambda)$ in pogojno na Y naj bo $X \sim \operatorname{Pois}(Y)$. Za katere $a \in \mathbb{R}$ obstaja $\mathbb{E}(a^X)$ in koliko je to enako?

Če je Y **poljubna** *slučajna spremenljivka, X pa pogojno na Y porazdeljena zvezno z gostoto $f_{X|Y}$, je tudi brezpogojna porazdelitev slučajne spremenljivke X zvezna z gostoto $f_X(x) = \mathbb{E}\big[f_{X|Y}(x)\big]$.*

33. Dani sta slučajni spremenljivki N in T. Slučajna spremenljivka N je porazdeljena geometrijsko $\operatorname{Geom}(p)$ in pogojno na N naj ima T porazdelitev $\operatorname{Gama}(N, \lambda)$. Določite brezpogojno porazdelitev slučajne spremenljivke T.

Slučajni vektor (X, Y) je zvezno porazdeljen natanko tedaj, ko je hkrati zvezno porazdeljena slučajna spremenljivka Y in pogojna porazdelitev slučajne spremenljivke X glede Y skoraj gotovo zvezna. Za ustrezne (pogojne) gostote tedaj velja zveza:

$$f_{X,Y}(x, y) = f_Y(y)\, f_{X|Y}(x \mid y)\,.$$

Pogojno porazdelitev slučajne spremenljivke $h(X, Y)$ glede na Y dobimo tako, da pogojno porazdelitev slučajne spremenljivke X glede na Y preslikamo s funkcijo h_Y, kjer je $h_y(x) = h(x, y)$.

Verjetnostna in funkcijska neodvisnost

Slučajni spremenljivki X in Y sta (v smislu verjetnosti) neodvisni natanko tedaj, ko se pogojna porazdelitev slučajne spremenljivke X glede na Y izraža tako, da Y v njej ne nastopa (funkcijska neodvisnost).

34. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$f_{X,Y}(x, y) = \begin{cases} 2\,e^{-x} & ;\ x > 2y > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Določite pogojno porazdelitev slučajne spremenljivke X glede na Y in Y glede na X. Poiščite še kakšno netrivialno slučajno spremenljivko, neodvisno od X, in še kakšno neodvisno od Y.

35. Slučajna spremenljivka X je porazdeljena standardno normalno $\mathrm{N}(0, 1)$, pogojna porazdelitev slučajne spremenljivke Y glede na X pa je normalna $\mathrm{N}(0, 1/|X|)$. Določite porazdelitev slučajne spremenljivke Y in še pogojno porazdelitev slučajne spremenljivke X^2 glede na Y.

36. Naj bo $0 < a < b$. Slučajna spremenljivka X ima porazdelitev $\mathrm{Gama}(2, 1)$, tj. ima gostoto:

$$f_X(x) = \begin{cases} x\,e^{-x} & ;\ x > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Pogojno na $X = x$ je slučajna spremenljivka Y porazdeljena enakomerno na intervalu (ax, bx). Določite brezpogojno porazdelitev slučajne spremenljivke Y in izračunajte $\mathbb{E}(1/Y)$.

Namig: uporabite pogojno pričakovano vrednost.

37. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$f_{X,Y}(x, y) = \begin{cases} 3(x^2 y + x y^2) & ;\ x, y \in [0, 1] \\ 0 & ;\ \text{sicer.} \end{cases}$$

Izračunajte $\mathbb{E}(X \mid Y)$ in $\mathbb{E}(XY \mid Y)$.

38. Naj bo $0 < a < b$. Slučajna spremenljivka X ima porazdelitev $\mathrm{Gama}(2, 1)$, tj. ima gostoto:

$$f_X(x) = \begin{cases} x\,e^{-x} & ;\ x > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Pogojno na $X = x$ je slučajna spremenljivka Y porazdeljena enakomerno na intervalu (ax, bx).

a) Določite brezpogojno porazdelitev slučajne spremenljivke Y.

b) Določite, za katere $t \in \mathbb{R}$ obstaja $\mathbb{E}(X^t/Y)$; če obstaja, ga tudi izračunajte.
Namig: uporabite pogojno matematično upanje.

39. Naj bosta (X, Y) in (Z, W) neodvisna slučajna vektorja z enako gostoto:

$$f_{X,Y}(x, y) = \begin{cases} e^{-x} \cdot \dfrac{1}{\sqrt{2\pi x}}\, e^{-\frac{(y-x)^2}{2x}} & ; x > 0 \\ 0 & ; \text{sicer.} \end{cases}$$

Določite porazdelitveno gostoto slučajnega vektorja $(X + Z, Y + W)$.

Namig: pogojujte na X in Z.

Dvorazsežna normalna porazdelitev $\mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, *kjer je* $\sigma_1, \sigma_2 > 0$ *in* $-1 < \rho < 1$, *je porazdelitev z gostoto:*

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{(x-\mu_1)^2}{2(1-\rho^2)\sigma_1^2} + \frac{\rho(x-\mu_1)(y-\mu_2)}{(1-\rho^2)\sigma_1\sigma_2} - \frac{(y-\mu_2)^2}{2(1-\rho^2)\sigma_2^2}}\,.$$

Če ima slučajni vektor (X, Y) *to porazdelitev, je* $X \sim \mathrm{N}(\mu_1, \sigma_1)$ *in* $Y \sim \mathrm{N}(\mu_2, \sigma_2)$.

40. a) Naj ima slučajni vektor (X_1, X_2) dvorazsežno normalno porazdelitev $\mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Določite pogojno porazdelitev slučajne spremenljivke X_2 glede na X_1.

b) Če za rezultate prvega in drugega kolokvija iz verjetnosti in statistike za računalničarje privzamemo model z dvorazsežno normalno porazdelitvijo, iz rezultatov 963 parov kolokvijev iz let od 1997 do 2010 dobimo naslednje ocene parametrov:

$$\mu_1 \doteq 59{\cdot}2\,, \mu_2 \doteq 58{\cdot}0\,, \sigma_1 \doteq 20{\cdot}1\,, \sigma_2 \doteq 25{\cdot}3\,, \rho \doteq 0{\cdot}291\,.$$

Na podlagi privzetega modela ocenite verjetnost, da bo kandidat, ki na prvem kolokviju zbere 25 točk, kolokvije naredil, tj. na obeh zbral skupaj vsaj 100 točk.

8. Rodovne, momentno-rodovne in karakteristične funkcije

Rodovne funkcije: osnove, konvolucijski izrek, procesi razvejanja. Momentno-rodovne funkcije, kumulante, asimetrija, sploščenost. Neenakosti. Karakteristične funkcije: osnove, povezava z rodovnimi funkcijami, konvolucijski izrek, inverzna formula.

Rodovna funkcija

Rodovna funkcija slučajne spremenljivke:

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \end{pmatrix}$$

je definirana po predpisu:

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + p_3 s^3 + \cdots = \mathbb{E}\big[s^X\big]$$

in zagotovo obstaja za vse kompleksne s z $|s| \le 1$. Velja:

$$p_k = \frac{G_X^{(k)}(0)}{k!}\,,$$

torej rodovna funkcija natančno določa porazdelitev take slučajne spremenljivke.

Če za definicijo rodovne funkcije vzamemo $G_X(s) := \mathbb{E}\big[s^X\big]$, jo lahko definiramo tudi za splošnejše slučajne spremenljivke, vendar pa moramo paziti na njeno definicijsko območje (ki ga omejuje tako potenciranje kot tudi pričakovana vrednost): lahko se zgodi, da bo definirana le za $s = 1$. Pomembna posebna primera:

- Če je $X \ge 0$, je $G_X(s)$ definirana za $0 \le s \le 1$ ali še splošneje za vse kompleksne s z $|s| \le 1$, ki ne pripadajo $[-1, 0)$.
- Če je X celoštevilska, je $G_X(s)$ definirana za vse kompleksne s z $|s| = 1$.

Velja $G_X(1) = \lim_{s\uparrow 1} G_X^{(r)}(s) = 1$. Če se omejimo na slučajne spremenljivke z vrednostmi v $\mathbb{N}_0$, se r-ti **faktorski moment** *slučajne spremenljivke X izraža s formulo:*

$$f_r := \mathbb{E}\big[X(X-1)(X-2)\cdots(X-r+1)\big] = \lim_{s\uparrow 1} G^{(r)}(s)\,,$$

Pri tem faktorski moment obstaja natanko tedaj, ko obstaja ustrezna leva limita r-tega odvoda.

Brž ko je rodovna funkcija definirana za kakšen $s > 1$, vsi faktorski momenti obstajajo in velja kar $f_r = G^{(r)}(1)$.

1. Izračunajte rodovno funkcijo slučajne spremenljivke X, porazdeljene po Poissonu $\mathrm{Pois}(\lambda)$. Izračunajte še $\mathbb{E}(X)$, $\mathrm{var}(X)$ in vse faktorske momente.

2. Izračunajte rodovno funkcijo geometrijske porazdelitve $\mathrm{Geom}(p)$ ter še pričakovano vrednost, varianco in vse faktorske momente.

3. Naj bo $0 < q < 1$. Slučajna spremenljivka X z vrednostmi v naravnih številih ima porazdelitev, podano po predpisu:

$$\mathbb{P}(X = k) = \frac{cq^k}{k} \quad ; \quad k = 1, 2, 3, \ldots$$

Izračunajte konstanto c ter rodovno funkcijo in faktorske momente te porazdelitve.

Namig: $\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$

4. Rodovna funkcija slučajne spremenljivke X je podana s predpisom:

$$G_X(s) = a\, e^{s+s^2} .$$

Določite konstanto a in izračunajte $\mathbb{P}(X = 2)$.

5. *Ehrenfestov model* opisuje prehajanje molekul iz enega predela dvodelne posode v drugega. V obeh predelih je skupaj m molekul. Na posameznem koraku se povsem naključno izbere kroglica (torej vsaka z enako verjetnostjo, ne glede na to, v katerem predelu je), nakar se ta kroglica prestavi v drug predel posode, ostale pa ostanejo tam, kjer so. Za $n = 0, 1, 2, \ldots$ naj bo X_n število molekul v prvem predelu posode po n korakih. Označimo rodovno funkcijo te slučajne spremenljivke z G_n.

   a) Smiselno izrazite G_{n+1} z G_n: uporabite operatorje, ki delujejo na funkcijah.

   b) Kako mora biti porazdeljena slučajna spremenljivka X_0, da imajo vse slučajne spremenljivke $X_0, X_1, \ldots$ enako porazdelitev?

6. Naj bo Π_n slučajna permutacija n elementov, pri čemer so vse permutacije enako verjetne. Naj bo X_n število negibnih točk v Π_n.

   a) Izračunajte razmerje $\dfrac{\mathbb{P}(X_n = j)}{\mathbb{P}(X_{n+1} = j + 1)}$.

   b) Naj bo G_n rodovna funkcija slučajne spremenljivke X_n. Izrazite $G_n(s)$ kot polinom v spremenljivki $s - 1$.

   Namig: na podlagi razmerja iz prejšnje točke izpeljite primerno zvezo med G_n in G_{n+1}.

   c) Za vse $k = 0, 1, 2, \ldots, n$ izračunajte $\mathbb{P}(X_n = k)$. Dovolj je, da zapišete izraz, ki za izračun zahteva $O(n)$ operacij, pri čemer so dovoljene operacije seštevanje, odštevanje, množenje, deljenje in faktoriela.

$$\boxed{\textit{Če sta } X \textit{ in } Y \textit{ neodvisni, je } G_{X+Y} = G_X\, G_Y.}$$

7. Naj bosta $X \sim \mathrm{Geom}(1/2)$ in $Y \sim \mathrm{Geom}(1/3)$ neodvisni slučajni spremenljivki. Določite porazdelitev njune vsote.

8. Slučajna spremenljivka X je porazdeljena diskretno po predpisu:

$$\mathbb{P}(X=0)=\frac{9}{16}\,,\quad \mathbb{P}(X=1)=\frac{9}{32}\,,\quad \mathbb{P}(X=k)=\frac{5}{2^{k+4}}\,;\quad k=2,3,4,\ldots$$

od nje neodvisna slučajna spremenljivka Y pa ima geometrijsko porazdelitev $\mathrm{Geom}(8/9)$. Določite porazdelitev slučajne spremenljivke $X+2Y$.

9. Naj bosta X in Y neodvisni slučajni spremenljivki in $Z=X+Y$ njuna vsota.

   a) Recimo, da je $X\sim\mathrm{Geom}(4/5)$ in $Z\sim\mathrm{Geom}(2/3)$. Določite porazdelitev slučajne spremenljivke Y.

   b) Naj bo $0<a,b<1$. Določite potreben in zadosten pogoj za obstoj neodvisnih slučajnih spremenljivk X in Y, pri katerih je $X\sim\mathrm{Geom}(a)$ in $X+Y\sim\mathrm{Geom}(b)$.

10. Naj bosta X in Y neodvisni in enako porazdeljeni slučajni spremenljivki z vrednostmi v $\mathbb{N}_0$. Določite porazdelitev, če veste, da je:

$$\mathbb{P}(X+Y=n)=(1-q)q^n\,;\quad n=0,1,2,\ldots$$

kjer je $0<q<1$.

11. Naj bo $X\sim\mathrm{Bin}(n,p)$. S pomočjo rodovne funkcije izračunajte $\mathbb{E}\left(\dfrac{1}{1+X}\right)$.

12. Janez ima dva otroka, ki še nimata svojih otrok. Za vsakega od njiju je porazdelitev števila otrok, ki jih bo imel, enaka:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Privzamemo še, da sta števili otrok Janezovih otrok neodvisni. Določite porazdelitev števila njegovih vnukov.

13. Nika še nima otrok. Porazdelitev števila njenih bodočih otrok je enaka:

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix},$$

porazdelitev števila otrok vsakega eventuelnega Nikinega otroka pa je enaka:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Privzamemo še, da so števila otrok Nikinih otrok neodvisna. Določite pričakovano število Nikinih vnukov in še verjetnost, da Nika ostane brez vnukov.

14. Naravna števila sklenejo, da bodo šla v kitajsko restavracijo, kjer je neskončno okroglih miz, ki so oštevilčene z $1, 2, \ldots$ Za vsako mizo lahko sedi poljubno mnogo gostov. Vratar v restavracijo spusti le N začetnih naravnih števil, kjer je $N + 1 \sim \mathrm{Geom}(p)$ in $0 < p < 1$. Natančneje, velja:

$$\mathbb{P}(N = n) = p(1-p)^n \,; \quad n = 0, 1, 2, \ldots$$

(če je $N = 0$, vratar v restavracijo ne spusti nikogar). Naravna števila prihajajo v restavracijo po vrsti in se posedejo po naslednjih pravilih: 1 se usede za mizo 1. Ko pride število n (če ga vratar spusti), se z verjetnostjo $1/n$ usede za prazno mizo z najnižjo možno številko, sicer pa se z verjetnostjo $1/n$ usede levo od nekega predhodnega števila, neodvisno od N in položajev predhodnih števil. Določite porazdelitev števila zasedenih (tj. nepraznih) miz.

Če so $X_1, X_2, \ldots$ neodvisne in enako porazdeljene slučajne spremenljivke z rodovno funkcijo G in je N slučajna spremenljivka z vrednostmi v $\mathbb{N}_0$, neodvisna od slučajnih spremenljivk X_i in z rodovno funkcijo H, ima vsota $Z := X_1 + X_2 + \ldots + X_N$ rodovno funkcijo $G_Z(s) = H\big(G(s)\big)$.

15. Naj bodo $X_1, X_2, X_3, \ldots$ in N neodvisne slučajne spremenljivke, pri čemer naj bo $X_i \sim \mathrm{Geom}(a)$ za vse i in $N \sim \mathrm{Geom}(b)$. Določite porazdelitev vsote $X_1 + X_2 + \ldots + X_N$.

16. Slučajne spremenljivke $X_1, X_2, \ldots$ so porazdeljene geometrijsko $\mathrm{Geom}(2/3)$, slučajna spremenljivka N pa geometrijsko $\mathrm{Geom}(3/4)$. Vse omenjene slučajne spremenljivke so neodvisne. Zapišite porazdelitev slučajne spremenljivke $Z = X_1 + X_2 + \cdots + X_{2N}$.

17. Naj bo $0 < p < 1$. Slučajne spremenljivke $X_1, X_2, \ldots$ naj bodo neodvisne, vse s porazdelitvijo, podano po predpisu:

$$\mathbb{P}(X_n = k) = \frac{(1-p)^k}{k \ln(1/p)} \,; \quad k = 1, 2, 3, \ldots$$

Nadalje naj bo N od njih neodvisna slučajna spremenljivka s Poissonovo porazdelitvijo $\mathrm{Pois}(-a \ln p)$, kjer je $a > 0$. Izračunajte porazdelitev slučajne spremenljivke $Z = X_1 + X_2 + \cdots + X_N$.

18. Naj bodo $X_1, X_2, \ldots$ in N neodvisne slučajne spremenljivke in $\lambda > 0$. Slučajne spremenljivke $X_1, X_2, \ldots$ naj bodo enako porazdeljene, slučajna spremenljivka N pa naj ima Poissonovo porazdelitev $\mathrm{Pois}(\lambda)$. Nadalje je znano, da je porazdelitev slučajne spremenljivke $Z = X_1 + X_2 + \cdots + X_N$ podana s predpisom

$$\mathbb{P}(Z = k) = \frac{1}{2^{k+1}} \,; \quad k = 0, 1, 2, \ldots$$

a) Določite porazdelitev slučajnih spremenljivk $X_1, X_2, \ldots$ za $\lambda = 1$.

b) Za katere λ sploh obstajajo takšne slučajne spremenljivke?

19. Maks še nima otrok. Porazdelitev število njegovih otrok in števila otrok posameznega njegovega potomca je enaka:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/6 & 1/3 & 1/2 \end{pmatrix} .$$

Privzamemo še, da so števila otrok vsakega posameznika neodvisna. Določite verjetnost, da bo Maksovo potomstvo nekoč izumrlo.

Procesi razvejanja

Vzemimo proces razvejanja, pri katerem so števila neposrednih potomcev vsakega posameznika neodvisna in enako porazdeljena z rodovno funkcijo G_1*. Verjetnost, da tak proces izumre, je enaka* $\min\{s \in [0,1] \; ; \; G_1(s) = s\}$*.*

Če je $\mathbb{E}(X) < 1$*, proces izumre z verjetnostjo ena.*

Proces izumre z verjetnostjo nič natanko tedaj, ko je skoraj gotovo $X \ge 1$*.*

20. Dan naj bo proces razvejanja kot zgoraj, ki skoraj gotovo izumre. Naj bo Z_1 število vseh potomcev v prvi generaciji, Z pa število vseh predstavnikov tega procesa. Označimo z G_1 rodovno funkcijo števila vseh potomcev v prvi generaciji, z G pa označimo število vseh predstavnikov procesa.

    a) Izpeljite smiselno zvezo med G in G_1.

    b) Vzemimo posebni primer, ko ima posameznik z verjetnostjo $1/2$ dva potomca, z verjetnostjo $1/2$ pa je brez potomcev. Utemeljite, da proces v tem primeru skoraj gotovo izumre, in izračunajte porazdelitev števila vseh njegovih predstavnikov.

21. V igri *Book of Ra* vsaka igra lahko z verjetnostjo p generira $m \ge 2$ nagradnih iger, sicer pa ne generira nobene igre. Vsaka nagradna igra lahko spet z verjetnostjo p generira m nagradnih iger, te pa lahko spet z verjetnostjo p generirajo m nagradnih iger in tako naprej. Privzemimo, da so vse igre med sabo neodvisne. Predpostavite še, da je $mp < 1$. Naj bo Z število iger, ki jih bo igralec odigral za eno stavo.

    a) Utemeljite, da bo igralec za eno stavo skoraj gotovo odigral le končno število iger.

    b) Utemeljite, da $\mathbb{E}(Z)$ in $\operatorname{var}(Z)$ obstajata, ter ju izračunajte.
    Namig: upoštevajte rezultat prejšnje naloge.

Rodovno funkcijo slučajne spremenljivke T *z vrednostmi v* $\{0, 1, 2, \ldots\} \cup \{\infty\}$ *definiramo kot:*

$$F(s) := \sum_{n=0}^{\infty} \mathbb{P}(T = n) \, s^n .$$

Tako je $F(1) = \mathbb{P}(T < \infty)$*.*

Markovska[21] **veriga** *v diskretnem času na števnem prostoru stanj A je zaporedje slučajnih spremenljivk $S_0, S_1, \ldots$ z vrednostmi v A, ki slikajo iz prostora, opremljenega z več verjetnostnimi merami: za vsako stanje $a \in A$ je na njem definirana verjetnostna mera $\mathbb{P}_a$. Pri tem mora veljati:*

- $\mathbb{P}_a(S_0 = a) = 1$;
- $\mathbb{P}_{a_0}(S_{n+1} = a_{n+1} \mid S_1 = a_1, \ldots, S_n = a_n) = \mathbb{P}_{a_n}(S_1 = a_{n+1})$.

Tako je markovska veriga določena že z verjetnostmi $p_{ab} := \mathbb{P}_a(X_1 = b)$, ki sestavljajo **prehodno matriko** $\mathbf{P} = [p_{ab}]_{a,b}$.

Za markovsko verigo $S_0, S_1, \ldots$ in stanje $a \in S$ označimo $T_a := \min\{n \in \mathbb{N} \,;\, S_n = a\}$; če je $S_n \neq a$ za vse $n \in \mathbb{N}$, definiramo $T_a := \infty$.

Za stanji a in b naj bo $F_{a \to b}$ rodovna funkcija slučajne spremenljivke T_b pri verjetnostni meri $\mathbb{P}_a$.

Enostavni slučajni sprehod *je markovska veriga na $\mathbb{Z}$, določena z verjetnostma $p_{a,a+1} = p$ in $p_{a,a-1} = 1 - p$. Dogovorimo se, naj bo $\mathbb{P}(A \mid p) := \mathbb{P}_0(A)$ in $F_r(s \mid p) := F_{0 \to r}$; označimo še $F := F_0$.*
Pri verjetnostni meri $\mathbb{P}(\cdot \mid p)$ je veriga enako porazdeljena kot zaporedje delnih vsot neodvisnih slučajnih spremenljivk $X_k \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}$. Lahko torej vzamemo, da je skoraj gotovo $S_n = X_1 + X_2 + \ldots + X_n$.

22. Za enostavni slučajni sprehod:

    a) za $r \in \mathbb{N}$ izrazite F_r z F_1 (*namig*: glejte slučajne spremenljivke $T_2 - T_1$, $T_3 - T_2, \ldots$);

    b) določite porazdelitev časov T_0 in T_1 (vključno z verjetnostma $\mathbb{P}(T_0 < \infty)$ in $\mathbb{P}(T_1 < \infty)$) (*namig*: najprej izračunajte za T_1, nato za T_0);

    c) izračunajte $\mathbb{E}(T_0 \mid T_0 < \infty)$ in $\mathbb{E}(T_1 \mid T_1 < \infty)$.

[21] Andrej Andrejevič Markov (1856–1922), ruski matematik

Moment *reda r je pričakovana vrednost r-te potence:* $z_r := \mathbb{E}(X^r)$.
Centralni moment *reda r je pričakovana vrednost r-te potence centrirane slučajne spremenljivke:* $m_r := \mathbb{E}\Big[\big(X - \mathbb{E}(X)\big)^r\Big]$. *Drugi centralni moment je torej ravno varianca.*
Momentno-rodovna funkcija *je eksponentna rodovna funkcija momentov:*

$$M_X(t) = \mathbb{E}\big(e^{tX}\big) = 1 + z_1\, t + \frac{z_2}{2!}\, t^2 + \frac{z_3}{3!}\, t^3 + \cdots$$

Tudi če so vsi momenti definirani, momentno-rodovna funkcija ni nujno definirana za vse t (lahko se zgodi, da je definirana le za $t = 0$).
Če je X celoštevilska, je tudi:

$$M_X(t) = G_X(e^t)\,.$$

Če sta X in Y neodvisni, je $M_{X+Y} = M_X\, M_Y$.

23. Naj bo $a < b$. Določite momentno-rodovno funkcijo enakomerne porazdelitve na intervalu $[a, b]$. Kje je definirana?

24. Naj bo $\lambda > 0$. Določite momentno-rodovno funkcijo in prve štiri momente Poissonove porazdelitve $\mathrm{Pois}(\lambda)$. Kje je definirana?

25. Naj bo $\lambda > 0$. Določite vse momente in momentno-rodovno funkcijo eksponentne porazdelitve na intervalu $\mathrm{Exp}(\lambda)$. Kje je definirana? Izračunajte še tretji centralni moment.

26. Izračunajte momentno-rodovno funkcijo in vse momente standardne normalne porazdelitve.

27. Izračunajte momentno-rodovno funkcijo splošne normalne porazdelitve. Na podlagi tega sklepajte o porazdelitvi vsote neodvisnih normalno porazdeljenih slučajnih spremenljivk.

28. Izračunajte momentno-rodovno funkcijo porazdelitve gama. Kje je le-ta definirana? Na podlagi tega sklepajte o porazdelitvi vsote določenih neodvisnih slučajnih spremenljivk s to porazdelitvijo.

Kumulante *so odvodi logaritma momentno-rodovne funkcije v izhodišču:*

$$\kappa_r(X) := (\ln M_X)^{(r)}(0)\,.$$

29. Izračunajte vse kumulante normalne porazdelitve $\mathrm{N}(\mu, \sigma)$.

30. a) Koliko je enaka prva kumulanta?

b) Dokažite, da so višje kumulante invariantne za translacije: za $r \geq 2$ velja $\kappa_r(X + a) = \kappa_r(X)$.

c) Izrazite $\kappa_r(aX)$ s $\kappa_r(X)$.

Asimetrija:

$$A(X) = \frac{\kappa_3(X)}{\big(\mathrm{var}(X)\big)^{3/2}}$$

Če sta $a > 0$ in b konstanti, je $A(aX + b) = A(X)$.

Presežna sploščenost (kurtozis):

$$K(X) = \frac{\kappa_4(X)}{\big(\mathrm{var}(X)\big)^2}$$

Če sta $a \neq 0$ in b konstanti, je $K(aX + b) = K(X)$.

31. Izrazite drugo, tretjo in četrto kumulanto ter še asimetrijo in presežno sploščenost s centralnimi momenti. *Namig*: pomagajte si z razvojem v potenčne vrste.

32. Naj bo $a < b$. Izračunajte prve štiri kumulante ter asimetrijo in presežno sploščenost enakomerne porazdelitve na intervalu $[a, b]$.

33. Naj bo $\lambda > 0$. Izračunajte drugi, tretji in četrti centralni moment Poissonove porazdelitve $\mathrm{Pois}(\lambda)$, poleg tega pa še njeno asimetrijo in presežno sploščenost.

Neenačba Markova[22]: *če je $X \geq 0$, za vsak $x > 0$ velja ocena:*

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x}\,.$$

Dostikrat se namesto slučajne spremenljivke X splača vzeti kakšno njeno funkcijo. Tako lahko dobimo **neenačbo Čebiševa**[23]:

$$\mathbb{P}\big(|X - \mathbb{E}(X)| \geq t\big) \leq \frac{\mathrm{var}(X)}{t^2}\,.$$

Dobre ocene pa često dobimo tudi iz momentno-rodovne funkcije:

$$\mathbb{P}(X \geq x) \leq e^{-tx}\, M_X(t)\,.$$

34. V tovarni proizvedejo 1600 izdelkov, a je vsak okvarjen z verjetnostjo 0˙1. Posamezni izdelki so med seboj neodvisni. Navzgor ocenite verjetnost, da bo pokvarjenih več kot 250 izdelkov:

[22] Andrej Andrejevič Markov (1856–1922), ruski matematik

[23] Pafnutij Lvovič Čebišev (1821–1894), ruski matematik

a) s pomočjo neenačbe Markova;

b) s pomočjo neenačbe Čebiševa z upoštevanjem simetrije;

c) s pomočjo momentno-rodovne funkcije.

Primerjajte še z aproksimacijo po Laplaceovi integralski formuli!

35. Slučajne spremenljivke $X_1, X_2, \ldots,$ naj bodo neodvisne in naj imajo standardno *Laplaceovo porazdelitev*, tj. porazdelitev z gostoto:

$$f(x) = \frac{1}{2} e^{-|x|} .$$

Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Ocenite $\mathbb{P}(S_5 > 5)$ in $\mathbb{P}(S_{20} > 20)$, in sicer z uporabo neenakosti Čebiševa in simetrije ter z uporabo momentno-rodovne funkcije.

Karakteristična funkcija

$$\phi_X(t) = \mathbb{E}\big(e^{itX}\big)$$

Definirana je za vsak $t \in \mathbb{R}$.

Velja tudi $\phi_X(t) = M_X(it)$. Brž ko je momentno-rodovna funkcija definirana za kakšno neničelno realno število, lahko karakteristično funkcijo izračunamo že iz vrednosti momentno-rodovne funkcije na realnih številih, in sicer tako, da jo ustrezno analitično razširimo.

Če je X celoštevilska, je tudi:

$$\phi_X(t) = G_X(e^{it}) .$$

36. Izračunajte karakteristično funkcijo slučajne spremenljivke, ki je porazdeljena geometrijsko $\mathrm{Geom}(p)$.

37. Izračunajte karakteristično funkcijo slučajne spremenljivke, ki je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$.

38. Izračunajte karakteristično funkcijo slučajne spremenljivke, ki je porazdeljena zvezno z gostoto $f_X(x) = \frac{1}{2}e^{-|x|}$.

Če sta X in Y neodvisni, je $\phi_{X+Y} = \phi_X \, \phi_Y$.

39. Karakteristična funkcija binomske porazdelitve.

40. Karakteristična funkcija negativne binomske porazdelitve.

41. Karakteristična funkcija porazdelitve gama.

42. Rekonstruirajte porazdelitev slučajne spremenljivke s karakteristično funkcijo:

$$\phi_X(t) = \frac{\cos t + 2\cos^2 t + 2i \sin t \cos t}{3}$$

Inverzna formula

Če je slučajna spremenljivka X porazdeljena zvezno z gostoto f, ki je v dani točki x odvedljiva, velja formula:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) e^{-itx} \, \mathrm{d}t$$

kjer je ϕ karakteristična funkcija slučajne spremenljivke X (pod pogojem, da zgornji integral obstaja).

43. Slučajna spremenljivka X ima naslednjo karakteristično funkcijo:

$$\phi_X(t) = \begin{cases} 1 - |x| & ; |x| \le 1 \\ 0 & ; \text{sicer.} \end{cases}$$

Rekonstruirajte njeno porazdelitev.

44. Rekonstruirajte porazdelitev slučajne spremenljivke s karakteristično funkcijo $\phi_X(t) = e^{-|t|}$.

45. Neodvisne slučajne spremenljivke $X_1, \ldots, X_n$ imajo standardno *Cauchyjevo*[24] porazdelitev, tj. zvezno porazdelitev z gostoto:

$$p(x) = \frac{1}{\pi(1 + x^2)}$$

Določite porazdelitev slučajne spremenljivke:

$$\bar{X} := \frac{X_1 + \cdots + X_n}{n}$$

[24] baron Augustin-Louis Cauchy (1789–1857), francoski matematik

9. Limitni izreki

Konvergenca slučajnih spremenljivk. Šibki in krepki zakon velikih števil. Centralni limitni izrek.

Prva Borel–Cantellijeva lema

Če za dogodke $A_1, A_2, \ldots$ velja $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, se skoraj gotovo zgodi kvečjemu končno mnogo teh dogodkov.

1. [25] Mečemo pošten kovanec, meti so neodvisni. Naj bo L_n dolžina najdaljšega sosledja samih cifer do vključno n-tega meta. Dokažite, da skoraj gotovo velja:

$$\lim_{n\to\infty} \frac{L_n}{\log_2 n} = 1\,.$$

Konvergenca slučajnih spremenljivk

Zaporedje slučajnih spremenljivk $X_1, X_2, \ldots$ konvergira proti slučajni spremenljivki X:

- **skoraj gotovo** *($X_n \xrightarrow[n\to\infty]{\text{s.g.}} X$), če velja $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$;*
- **v verjetnosti** *($X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$), če za vsak $\varepsilon > 0$ velja $\lim_{n\to\infty} \mathbb{P}(|X_n - X| < \varepsilon) = 1$.*
- **v porazdelitvi** *ali tudi* **šibko** *($X_n \xrightarrow[n\to\infty]{\text{d}} X$), če velja ena izmed naslednjih dveh ekvivalentnih trditev:*
  - *Za vsak x, kjer je F_X zvezna, velja $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.*
  - *Za vsako zvezno in omejeno funkcijo h velja $\lim_{n\to\infty} \mathbb{E}\big(h(X_n)\big) = \mathbb{E}\big(h(X)\big)$.*

Iz skoraj gotove konvergence sledi konvergenca v verjetnosti, iz nje pa šibka konvergenca.
Šibko konvergenco lahko definiramo zgolj za porazdelitve.

2. *Standardni slučajni sprehod* je zaporedje slučajnih spremenljivk $S_n = X_1 + \cdots + X_n$, kjer so $X_1, X_2, \ldots$ neodvisne in:

$$X_i \sim \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

Ali zaporedje S_n/n konvergira proti nič v verjetnosti? Kaj pa šibko? Pa skoraj gotovo?

[25] Vir: R. Durrett: Probability: Theory and Examples, druga izdaja, Duxbury Press, 1995.

Krepki zakon velikih števil Kolmogorova[26]. *Če so slučajne spremenljivke $X_1, X_2, \ldots$ neodvisne in enako porazdeljene ter če obstaja $\mu = \mathbb{E}(X_n)$ (tj. $\mathbb{E}\big(|X_n|\big) < \infty$), velja:*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{s.g.}} \mu$$

Druga Borel[27]–Cantellijeva[28] lema

Če za **neodvisne** *dogodke $A_1, A_2, \ldots$ velja $\displaystyle\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, se skoraj gotovo zgodi neskončno mnogo teh dogodkov.*

3. [29] Naj bodo $X_2, X_3, X_4, \ldots$ neodvisne slučajne spremenljivke z:

$$X_n \sim \begin{pmatrix} -n & 0 & n \\ \frac{1}{2n\ln n} & 1 - \frac{1}{n\ln n} & \frac{1}{2n\ln n} \end{pmatrix}.$$

Dokažite, da za to zaporedje velja šibki, ne pa tudi krepki zakon velikih števil: če označimo $S_n = X_1 + X_2 + \cdots + X_n$, zaporedje S_n/n v verjetnosti konvergira proti nič, vendar skoraj gotovo divergira.

4. Naj bo spet S_n standardni slučajni sprehod. Pokažite, da se le-ta skoraj gotovo vrne v izhodišče, tj. da z verjetnostjo 1 obstaja tak $n \in \mathbb{N}$, da je $S_n = 0$.

5. Naj bo S_n *nesimetrični* slučajni sprehod, torej $S_n = X_1 + \cdots + X_n$, kjer so $X_1, X_2, \ldots$ neodvisne in:

$$X_i \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}.$$

Za vsak $k \in \mathbb{Z}$ izračunajte verjetnost, da sprehod (še) kdaj obišče stanje k.

6. Naj bo spet S_n standardni slučajni sprehod. Definirajmo slučajne spremenljivke T_n po predpisu:

$$T_n = \begin{cases} \dfrac{S_n}{n} & ; S_n \neq 0 \\ 2 & ; S_n = 0 \end{cases}$$

Pokažite, da zaporedje T_n konvergira proti nič v verjetnosti, vendar pa je skoraj gotovo divergentno.

[26]Andrej Nikolajevič Kolmogorov (1903–1987), ruski matematik
[27]Félix Édouard Justin Émile Borel (1871–1956), francoski matematik
[28]Francesco Paolo Cantelli (1875–1966), italijanski matematik
[29]Vir: G. R. Grimmett, D. R. Stirzacker: Probability and Random Processes. Problems and Solutions. Clarenson Press, Oxford, 1997.

7. Dokažite, da iz konvergence v porazdelitvi ne sledi nujno konvergenca v verjetnosti, pač pa to sledi, če gre za konvergenco proti konstantni slučajni spremenljivki.

8. Dokažite, da za vsako zaporedje porazdelitev, ki šibko konvergira proti dani porazdelitvi, obstaja verjetnostni prostor, na njem pa zaporedje slučajnih spremenljivk z ustreznimi porazdelitvami, ki skoraj gotovo konvergira proti slučajni spremenljivki z limitno porazdelitvijo.

9. Dan je naslednji prototip trditve:

   Če X_n konvergira proti X in Y_n konvergira proti Y, tudi $X_n + Y_n$ konvergira proti $X + Y$.

   Raziščite, kako je z njegovo veljavnostjo pri skoraj gotovi konvergenci ter konvergenci v verjetnosti in porazdelitvi.

10. Dokažite *izrek Sluckega*[30]: če X_n konvergira proti X in Y_n konvergira proti konstanti c (oboje v porazdelitvi), tudi $X_n + Y_n$ konvergira proti $X + c$.

Centralni limitni izrek
(klasična formulacija)

Naj bodo $X_1, X_2, \ldots$ neodvisne in enako porazdeljene slučajne spremenljivke z $\mathbb{E}(X_i^2) < \infty$ ter $\mathbb{E}(X_i) = \mu_1$ in $\mathrm{var}(X_i) = \sigma_1^2$ ($\sigma_1 > 0$). Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Tedaj je, če je n velik, približno $S_n \sim \mathrm{N}\big(\mu_n, \sigma_n\big)$, kjer je:

$$\mu_n = \mathbb{E}(S_n) = n\mu_1 \quad \text{in} \quad \sigma_n = \sqrt{\mathrm{var}(S_n)} = \sigma_1\sqrt{n}\,.$$

To pomeni, da je:

$$\mathbb{P}(a < S_n < b),\ \mathbb{P}(a \le S_n \le b) \approx \Phi\left(\frac{b - \mu_n}{\sigma_n}\right) - \Phi\left(\frac{a - \mu_n}{\sigma_n}\right),$$

kjer napaka pri fiksni porazdelitvi seštevancev X_i konvergira proti nič, ko gre n proti neskončno, in sicer enakomerno po a in b.

11. Standardno kocko vržemo 105-krat, meti so neodvisni. Ocenite verjetnost, da bo skupaj padlo manj kot 350 pik.

12. Slučajne spremenljivke $X_1, \ldots, X_{100}$ so neodvisne in porazdeljene diskretno po naslednji shemi:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0{\cdot}5 & 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}1 \end{pmatrix}$$

    Ocenite $\mathbb{P}(170 < X_1 + \cdots + X_{100} < 210)$ in utemeljite odgovor.

[30] Evgenij Evgenjevič Slucki (1880–1948), ruski matematik, statistik in ekonomist

13. Loterija izda srečko, ki stane 1 evro. Srečka z verjetnostjo 5% zadene 10 evrov, z verjetnostjo 45% zadene 1 evro (v tem primeru torej kupec dobi povrnjen vložek), z verjetnostjo 50% pa ne zadene nič. Kupimo 50 srečk. Kolikšna je verjetnost, da imamo dobiček? Privzamemo, da so srečke neodvisne.

Centralni limitni izrek za trikotne sheme

Oglejmo si shemo slučajnih spremenljivk:

$$\begin{array}{llll} X_1^{(1)} & & & \\ X_1^{(2)}, & X_2^{(2)} & & \\ X_1^{(3)}, & X_2^{(3)}, & X_3^{(3)} & \\ & & & \ddots \end{array}$$

kjer so za vsak $n \in \mathbb{N}$ slučajne spremenljivke $X_1^{(n)}, \ldots, X_n^{(n)}$ neodvisne in enako porazdeljene. Označimo:

- $\mu_1^{(n)} := \mathbb{E}\big(X_i^{(n)}\big), \quad \sigma_1^{(n)} := \sigma\big(X_i^{(n)}\big), \quad \gamma_1^{(n)} := \left[\mathbb{E}\Big(\big|X_i^{(n)} - \mu_1^{(n)}\big|^3\Big)\right]^{1/3}$
- $S_n := X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}$
- $\mu_n := \mathbb{E}(S_n) = n\mu_1^{(n)}, \quad \sigma_n := \sigma(S_n) = \sigma_1^{(n)}\sqrt{n}$

Tedaj standardizirane vsote $(S_n - \mu_n)/\sigma_n$ šibko konvergirajo proti standardni normalni porazdelitvi, brž ko je:

$$n \gg \left(\frac{\gamma_1^{(n)}}{\sigma_1^{(n)}}\right)^6 .$$

Tedaj gre **absolutna** *napaka pri normalni aproksimaciji intervalskih verjetnosti $\mathbb{P}(a_n < S_n < b_n)$ in $\mathbb{P}(a_n \le S_n \le b_n)$ proti nič, in sicer ne glede na zaporedji a_n in b_n.*

Da se tudi oceniti napaka, ki jo naredimo pri normalni aproksimaciji. Ocen je veliko, med najenostavnejšimi pa je ocena na podlagi tretjega absolutnega momenta, znana kot **Berry**[31]**–Esseenov**[32] **izrek**. Če so $X_1, X_2, \ldots, X_n$ neodvisne in enako porazdeljene z $\mathbb{E}(X_i) = \mu_1$, $\sigma_1 = \sigma(X_i)$ in $\gamma_1 = \left[\mathbb{E}(|X_i - \mathbb{E}(X_i)|^3)\right]^{1/3}$, vsota $S_n = X_1 + X_2 + \cdots + X_n$ zadošča oceni:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(S_n \le x) - \Phi\left(\frac{x - n\mu_1}{\sigma_1\sqrt{n}}\right) - \frac{1}{2} \right| \le \frac{0{\cdot}4690}{\sqrt{n}} \, \frac{\gamma_1^3}{\sigma_1^3} .$$

(Berry, 1941; Esseen, 1942; . . . ; Ševcova, 2013).

I. G. Ševcova: Ob absolutnyh konstantah v neravenstve Berri–Esseena i jego strukturnyh i neravnomernyh utočnjenijah. *Informatika i jeje primenjenija* **7**/1 (2013), 124–125.

14. Slučajna spremenljivka X ima porazdelitev hi kvadrat s 100 prostostnimi stopnjami. Približno izračunajte $\mathbb{P}(X > 110)$ in $\mathbb{P}(90 < X < 110)$. Odgovor utemeljite!

[31] Andrew Campbell Berry (1906–1998), ameriški matematik

[32] Carl-Gustav Esseen (1918–2001), švedski matematik

15. Slučajne spremenljivke $X_1, X_2, \ldots, X_n$ naj bodo neodvisne in naj imajo standardno *Laplaceovo porazdelitev*, tj. porazdelitev z gostoto:

$$f(x) = \tfrac{1}{2}\, e^{-|x|}\,.$$

Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Približno izračunajte $\mathbb{P}(S_5 > 5)$ in $\mathbb{P}(S_{20} > 20)$.

Relativna napaka v centralnem limitnem izreku za trikotne sheme

Spet naj bo dana trikotna shema slučajnih spremenljivk $X_k^{(n)}$. *Zadosten pogoj, da gre* **relativna** *napaka pri normalni aproksimaciji intervalskih verjetnosti* $\mathbb{P}(a_n \le S_n \le b_n)$ *proti nič, je:*

$$n \gg \left(\frac{\gamma_1^{(n)}}{\sigma_1^{(n)}}\right)^6 \quad \textit{in} \quad n \gg \left(\frac{\gamma_1^{(n)}}{\sigma_1^{(n)}} \min\left\{\frac{|a_n - \mu_n|}{\sigma_n}, \frac{|b_n - \mu_n|}{\sigma_n}\right\}\right)^6 .$$

Drugi pogoj je ekvivalenten pogoju $\min\bigl\{|a_n - \mu_n|, |b_n - \mu_n|\bigr\} \ll n^{2/3}\,\dfrac{\bigl(\sigma_1^{(n)}\bigr)^2}{\gamma_1^{(n)}}$.

16. Slučajne spremenljivke $U_1, \ldots, U_{100}$ so porazdeljene enakomerno na intervalu $[0,1]$, slučajne spremenljivke $V_1, \ldots, V_{100}$ pa diskretno po shemi:

$$\begin{pmatrix} 1 & 2 \\ 2/3 & 1/3 \end{pmatrix}.$$

Vse omenjene slučajne spremenljivke so med seboj neodvisne.

a) Definirajmo $X_i := U_i V_i$. Izračunajte $\mathbb{E}(X_i)$ in $\operatorname{var}(X_i)$.

b) Naj bo $S := X_1 + X_2 + \cdots + X_{100}$. Približno izračunajte $\mathbb{P}(S < 60)$.

17. Slučajne spremenljivke $X_1, X_2, \ldots, X_{100}$ so neodvisne in porazdeljene diskretno po shemi:

$$\begin{pmatrix} 0 & 1 & 2 \\ p & 1-2p & p \end{pmatrix}.$$

Označimo $S := X_1 + X_2 + \cdots + X_{100}$. Približno določite vrednost parametra p, pri kateri je $\mathbb{P}(S < 90) = 0{\cdot}05$.

18. Slučajne spremenljivke $X_1, X_2, \ldots$ so neodvisne s porazdelitvijo:

$$\begin{pmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$$

Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Približno od katerega n naprej je $\mathbb{P}(|S_n| > 100) > 0{\cdot}05$?

19. Potapljač se potaplja, dokler ne nabere 80 biserov. Pri vsakem potopu nabere največ en biser, pa še tega le z verjetnostjo 20%. Potopi so med seboj neodvisni.

a) Kolikšno je pričakovano število potopov?

b) Ocenite verjetnost, da se bo moral potopiti več kot 450-krat. Odgovor utemeljite (možni sta dve uporabi centralnega limitnega izreka)!

Centralni limitni izrek ima veliko posplošitev. Pomembna posplošitev je na primer, ko seštevanci niso nujno enako porazdeljeni. Površno povedano, če je S vsota veliko neodvisnih slučajnih spremenljivk z dovolj lepimi porazdelitvami, od katerih nobena posebej ne izstopa, je približno $S \sim \mathrm{N}(\mu, \sigma)$, kjer je $\mu = \mathbb{E}(S)$ in $\sigma^2 = \mathrm{var}(S)$, torej je:

$$\mathbb{P}(a < S < b),\ \mathbb{P}(a \le S \le b) \approx \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Če je $S = X_1 + X_2 + \cdots + X_n$ in so $X_1, \ldots, X_n$ neodvisne, je seveda:

$$\begin{aligned} \mu &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) \\ \sigma^2 &= \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n)\,. \end{aligned}$$

Tudi v tem primeru se da napaka pri normalni aproksimaciji eksplicitno oceniti. Velja:

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}(S \le x) - \Phi\left(\frac{x-\mu}{\sigma}\right) - \frac{1}{2} \right| \le \frac{0{\cdot}5583}{\sigma^3} \sum_{k=1}^{n} \mathbb{E}\big(|X_k - \mathbb{E}(X_k)|^3\big)\,.$$

(Esseen, 1945; …; Ševcova, 2013). Glej:

I. G. Ševcova: Ob absolutnyh konstantah v neravenstve Berri–Esseena i jego strukturnyh i neravnomernyh utočnjenijah. *Informatika i jeje primenjenija* **7**/1 (2013), 124–125.

Centralni limitni izrek za trikotne sheme in za različno porazdeljene seštevance

Dana naj bo trikotna shema slučajnih spremeljivk:

$$\begin{array}{llll} X_1^{(1)} & & & \\ X_1^{(2)}, & X_2^{(2)} & & \\ X_1^{(3)}, & X_2^{(3)}, & X_3^{(3)} & \\ & & & \ddots \end{array}$$

kjer so za vsak $n \in \mathbb{N}$ *slučajne spremenljivke* $X_1^{(n)}, \ldots, X_n^{(n)}$ *neodvisne, ne pa tudi nujno enako porazdeljene. Označimo:*

- $S_n := X_1^{(n)} + X_2^{(n)} + \cdots + X_n^{(n)}$
- $\mu_n := \mathbb{E}(S_n) = n\mu_1^{(n)}, \quad \sigma_n := \sigma(S_n) = \sigma_1^{(n)}\sqrt{n},$

  $\gamma_n := \left[\sum_{i=1}^{n} \mathbb{E}\Big(\big|X_i^{(n)} - \mathbb{E}\big(X_i^{(n)}\big)\big|^3\Big)\right]^{1/3}$

Tedaj standardizirane vsote $(S_n - \mu_n)/\sigma_n$ *šibko konvergirajo proti standardni normalni porazdelitvi, brž ko je:*

$$\lim_{n\to\infty} \frac{\gamma_n}{\sigma_n} = 0\,.$$

Tedaj gre **absolutna** *napaka pri normalni aproksimaciji intervalskih verjetnosti* $\mathbb{P}(a_n < S_n < b_n)$ *in* $\mathbb{P}(a_n \le S_n \le b_n)$ *proti nič, in sicer ne glede na zaporedji* a_n *in* b_n*.* **Relativna** *napaka pa gre proti nič, brž ko poleg zgornjega velja še:*

$$\lim_{n\to\infty} \frac{\gamma_n}{\sigma_n} \min\left\{\frac{a_n - \mu_n}{\sigma_n}, \frac{b_n - \mu_n}{\sigma_n}\right\} = 0\,.$$

Količini γ_n^3/σ_n^3 *pravimo* **razmerje Ljapunova** *za eksponent* 3*.*

20. Naj bodo $X_1, X_2, \ldots$ neodvisne in naj bo $X_k \sim \mathrm{Exp}(1/k)$. Označimo $S_n := X_1 + X_2 + \cdots + X_n$.

    a) Dokažite, da standardizirane vsote $(S_n - \mathbb{E}(S_n))/\sigma(S_n)$ šibko konvergirajo proti standardni normalni porazdelitvi.

    b) Ocenite $\mathbb{P}(S_{100} < 6000)$.

21. Slučajne spremenljivke $X_1, X_2, \ldots, X_{100}$ so neodvisne in porazdeljene diskretno po shemi:

$$\begin{pmatrix} -1 & 1 \\ 1/3 & 2/3 \end{pmatrix}.$$

    Označimo $S = 10X_1 + X_2 + X_3 + \cdots + X_{100}$. Približno izračunajte $\mathbb{P}(S < 41)$.

[33] Aleksandr Mihajlovič Ljapunov (1857–1918), ruski matematik in mehanik

22. Ljubiteljski kolesar Bernard vsak dan od ponedeljka do sobote z verjetnostjo 1/2 prevozi 20, z verjetnostjo 1/2 pa 25 km. V nedeljo pa z verjetnostjo 3/4 počiva, z verjetnostjo 1/4 pa naredi izlet, na katerem prevozi 100 km. Privzamemo, da so vse Bernardove odločitve neodvisne.

    a) Čim natančneje izračunajte verjetnost, da v 12 tednih prevozi več kot 2000 km.

    b) Približno po koliko tednih postane verjetnost, da prevozi več kot 20.000 km, večja od 95%?

23. Naj bodo $X_1, X_2, X_3, \dots$ neodvisne slučajne spremenljivke, porazdeljene diskretno po predpisih:
$$X_n \sim \begin{pmatrix} -n & 0 & n \\ \frac{1}{2n^2} & 1 - \frac{1}{n^2} & \frac{1}{2n^2} \end{pmatrix}.$$
Pokažite, da za njihove delne vsote $S_n := X_1 + X_2 + \cdots + X_n$ centralni limitni izrek ne velja, čeprav imajo slučajne spremenljivke X_n enake pričakovane vrednosti in variance.

Centralni limitni izrek ima posplošitve celo na situacije, ko med sumandi obstaja določena vrsta odvisnosti. Med drugim ga lahko uporabimo za hipergeometrijsko porazdelitev.

24. Naj bo $N \sim \mathrm{Hip}(s, r, n)$.

    a) Izračunajte $\mathbb{E}(N)$ in $\mathrm{var}(N)$.

    b) Za $n = 500$, $r = 200$ in $s = 100$ približno izračunajte $\mathbb{P}(N > 45)$.

25. Slučajne spremenljivke $X, Y_1, Y_2, \dots, Y_{100}$ so neodvisne, pri čemer je $\mathbb{P}(X = 1) = 2/3$, $\mathbb{P}(X = 2) = 1/3$, $\mathbb{E}(Y_i) = 3$ in $\mathrm{var}(Y_i) = 100$. Slučajne spremenljivke $Y_1, \dots, Y_{100}$ so tudi enako porazdeljene. Označimo $S = X(Y_1 + Y_2 + \cdots + Y_{100})$. Približno izračunajte $\mathbb{P}(400 < S < 500)$.

10. Zadostne in postranske statistike

Zadostnost statistik. Fisher–Neymanov faktorizacijski izrek. Minimalne zadostne statistike. Eksponentne družine porazdelitev. Postranske statistike, uporaba Basujevega izreka.

Statistika T, ki izhaja iz opažanja X, tj. $T = \tau(X)$, je **zadostna** *za model, v katerem je verjetnost odvisna od parametra θ, če je pogojna porazdelitev opažanja X glede na T neodvisna od θ (pri klasični statistiki je to neodvisnost v funkcijskem smislu).*

1. Dan je statistični model, pri katerem opazimo X in Y, ki sta neodvisni statistični spremenljivki, pri čemer je $X \sim \mathrm{Pois}(p\lambda)$ in $Y \sim \mathrm{Pois}((1-p)\lambda)$. Pri tem je $\lambda > 0$ in $p \in (0,1)$. Je $X + Y$ v tem modelu zadostna statistika:

   a) če je λ znan, p pa neznan (torej parameter modela)?

   b) če je p znan, λ pa neznan (torej parameter modela)?

2. Dan je statistični model, pri katerem opazimo slučajni vektor (X, Y), ki je porazdeljen zvezno z dvorazsežno gostoto:

$$f_{X,Y}(x, y; \lambda) = \begin{cases} \lambda^2 e^{-\lambda x} & ; x > y > 0 \\ 0 & ; \text{sicer,} \end{cases}$$

   pri čemer je $\lambda > 0$ parameter modela. Je X ali Y v tem modelu zadostna statistika?

Fisher–Neymanov faktorizacijski izrek

Če je $f_X(x;\theta)$ verjetnostna funkcija ali pa gostota opažanja X, je statistika $T = \tau(X)$ zadostna za parameter θ natanko tedaj, ko obstajata taki funkciji ρ in g, da je:

$$f_X(x;\theta) = \rho(x)\, g\big(\tau(x);\theta\big)$$

za vse x in Θ.

3. Dan je statistični model, pri katerem opazimo $X \sim \mathrm{N}(\mu, \sigma^2)$, pri čemer je $\mu \in \mathbb{R}$ in $\sigma > 0$. Je $|X|$ v tem modelu zadostna statistika:

   a) če sta μ in σ oba neznana, torej je model parametriziran z (μ, σ)?

   b) če je znano, da je $\mu = 0$, σ pa je parameter modela?

   c) če je znano, da je $\mu = 1$, σ pa je parameter modela?

   d) če je $\mu = a\sigma^2$, kjer je a znan, σ pa je parameter modela?

4. Dokažite, da je pri Bernoullijevem zaporedju poskusov, od katerih vsak uspe z verjetnostjo θ, število uspelih poskusov zadostna statistika.

5. Poiščite kakšno enorazsežno zadostno statistiko za družino Poissonovih porazdelitev $\mathrm{Pois}(\lambda)$, ki izhaja iz opažanja neodvisnih realizacij $X_1, \ldots, X_n$ te porazdelitve.

6. Denimo, da zaporedje poskusov tvori *markovsko*[34] *verigo* z začetno verjetnostjo uspeha θ in prehodno matriko:

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\theta}{2} & \frac{\theta}{2} \\ \frac{1-\theta}{2} & \frac{1+\theta}{2} \end{bmatrix} .$$

To pomeni, da, če definiramo $X_k := 1$, če k-ti poskus uspe, in $X_k := 0$, če ne uspe, velja:

$$\mathbb{P}_\theta(X_1 = 1) = \theta \,,$$
$$\mathbb{P}_\theta(X_{k+1} = x_{k+1} \mid X_1 = x_1, \ldots, X_k = x_k) = p_{x_k x_{k+1}} \,; \quad k = 1, 2, \ldots, n-1 \,.$$

a) Dokažite, da vsak poskus uspe z verjetnostjo θ.

b) Dokažite, da število uspelih poskusov ni zadostna statistika.

c) Poiščite kakšno zadostno statistiko, katere razsežnost ni odvisna od velikosti vzorca.

Zadostna statistika S je **minimalna***, če je skoraj gotovo funkcija vsake zadostne statistike. Natančneje, za vsako zadostno statistiko T morata obstajati taka merljiva funkcija h in tak dogodek A, ki ima pri vseh vrednostih parametrov verjetnost ena, da na A velja $S = h(M)$.*

7. Opažanje X je porazdeljeno diskretno po shemi:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ a & a^2 & 2a^2 & 4a & 3a^2 & b \end{pmatrix}$$

a) Izrazite b z a.

b) Pri katerih vrednostih parametra a je z zgornjo shemo določena porazdelitev slučajne spremenljivke?

c) Poiščite minimalno zadostno statistiko za model, pri katerem a zavzame vse vrednosti iz prejšnje točke.

d) Poiščite minimalno zadostno statistiko za model, pri katerem a zavzame le robni vrednosti iz prejšnje točke.

[34] Andrej Andrejevič Markov (1856–1922), ruski matematik

Eksponentna družina *porazdelitev je tista, pri kateri se da verjetnostna funkcija ali gostota zapisati v obliki:*

$$f_X(x) = \rho(x)\, g(\gamma_1, \ldots, \gamma_m)\, e^{h_1(x)\,\gamma_1 + h_2(x)\,\gamma_2 + \cdots + h_m(x)\,\gamma_m}\,.$$

Pri zgoraj opisani družini je $\big(h_1(X), \ldots, h_m(X)\big)$ *zadostna statistika za to družino. Pravimo ji* **pripadajoča (zadostna) statistika**.
Pripadajočemu naboru $\gamma_1, \ldots, \gamma_m$ *pravimo* **naravni nabor parametrov** *za ta zapis, množico v* $\mathbb{R}^m$*, ki jo preteče, pa imenujemo* **naravni parametrični prostor**.

8. Dokažite, da je družina Bernoullijevih porazdelitev:

$$\begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}$$

eksponentna. Zapisu, ki ga najdete, določite naravni parametrični prostor. Poskusite to storiti tako, da bo število naravnih parametrov čim manjše. Nato enako storite še za binomsko porazdelitev $\mathrm{Bin}(n, p)$.

Če naravni parametrični prostor, ki pripada zapisu eksponentne družine, vsebuje afino neodvisno množico (denimo oglišča neizrojenega simpleksa) ali, ekvivalentno, če so parametri $1, \gamma_1, \gamma_2, \ldots, \gamma_m$ *linearno neodvisni, je pripadajoča zadostna statistika minimalna.*

9. Pokažite, da je vsota opaženih vrednosti minimalna zadostna statistika za družino Poissonovih porazdelitev, če opažanje sestoji iz neodvisnih realizacij $X_1, \ldots, X_n$ te porazdelitve.

Če porazdelitev statistične spremenljivke X *pripada eksponentni družini, ki ima za določen naraven nabor parametrov pripadajočo zadostno statistiko* $\boldsymbol{h}(X) = \big(h_1(X), \ldots, h_m(X)\big)$*, eksponentno družino tvori tudi porazdelitev slučajnega vektorja* $(X_1, \ldots, X_n)$ *neodvisnih kopij spremenljivke* X *(torej porazdelitev nabora* n *neodvisnih realizacij). Parametrični prostor je isti, pripadajoča zadostna statistika pa je* $\boldsymbol{h}(X_1) + \boldsymbol{h}(X_2) + \cdots + \boldsymbol{h}(X_n)$*. Tako lahko tudi iskanje minimalne zadostne statistike pri naboru več realizacij prevedemo na iskanje le-te pri eni sami realizaciji.*

V vseh nadaljnjih nalogah tega razdelka je na populaciji definirana statistična spremenljivka X s porazdelitvijo iz dane družine, na voljo pa imamo nabor neodvisnih realizacij $X_1, X_2, \ldots, X_n$ te statistične spremenljivke.

10. Poiščite minimalno zadostno statistiko za:

a) normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, kjer je μ neznan, σ pa znan;

b) normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, kjer je σ neznan, μ pa znan;

c) normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, kjer sta oba parametra neznana;

d) normalno porazdelitev $\mathrm{N}\big(a, \sqrt{a}\big)$;

e) normalno porazdelitev $\mathrm{N}(a, a)$.

11. Poiščite minimalno zadostno statistiko za porazdelitev gama, kjer sta oba parametra neznana.

> *Statistika U je* **postranska** *(angl. ancillary), če njena porazdelitev ni odvisna od parametra.*
> *Postranske statistike so med drugim pomembne pri preizkušanju statističnega modela, v katerem so postranske, proti širšemu modelu (več o tem v 13. razdelku).*

12. Denimo, da statistični model predvideva normalno porazdelitev $\mathrm{N}(1, 1)$ ali pa $\mathrm{N}(-1, 1)$. Poiščite kakšno netrivialno postransko statistiko.

13. Poiščite postransko statistiko za normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, katere zaloga vrednosti ima čimvečjo dimenzijo:

a) če je μ neznan, σ pa znan;

b) če je σ znan, μ pa neznan;

c) če sta μ in σ oba neznana.

> *Brž ko naravni parametrični prostor, ki pripada zapisu eksponentne družine, vsebuje kakšno odprto množico v $\mathbb{R}^m$, je pripadajoča zadostna statistika $\big(h_1(X), \ldots, h_m(X)\big)$ neodvisna od katere koli postranske statistike pri vseh vrednostih parametrov.*

14. Iz postranskih statistik iz 13. naloge potegnite zaključke o neodvisnosti pomembnih statistik pri Gaussovih modelih.

15. Za preizkušanje domneve, da je slučajna spremenljivka porazdeljena normalno $\mathrm{N}(\mu, \sigma)$, kjer je σ dan, μ pa neznan, lahko uporabimo postransko statistiko $\sum_{i=1}^{n}(X_i - \bar{X})^2$. Določite njeno porazdelitev.

16. Za preizkušanje domneve, da je slučajna spremenljivka porazdeljena normalno $\mathrm{N}(\mu, \sigma)$, kjer je μ dan, σ pa neznan, lahko uporabimo postransko statistiko:

$$U := \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^{n}(X_i - \mu)^2}} \,.$$

Vendar pa navadno uporabimo *Studentovo*[35] *statistiko*:

$$T := \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \sqrt{n(n-1)}\,.$$

Dokažite, da sta statistiki T in U v determinističnì bijektivni korespondenci in zato ekvivalentni (če smo natančni, se moramo v resnici omejiti na dogodek z verjetnostjo ena). Določite še porazdelitev statistike T.

[35]William Sealy Gosset (1876–1937), angleški statistik, bolj znan pod psevdonimom *Student*

11. Točkasto ocenjevanje in vzorčenje

Doslednost in nepristranskost. Srednja kvadratična napaka. Pridobivanje cenilk: metoda momentov in metoda največjega verjetja, nepristranske cenilke z enakomerno najmanjšo varianco.

Cenilka $\hat{q}$ karakteristike q je **nepristranska**, *če je vselej $\mathbb{E}(\hat{q}) = q$.*
Pristranskost *cenilke $\hat{q}$ definiramo kot:*

$$B(\hat{q}) := \mathbb{E}(\hat{q}) - q\,.$$

Cenilka je **dosledna**, *če za vsak $\varepsilon > 0$ velja:*

$$\lim_{n\to\infty} \mathbb{P}(|\hat{q} - q| < \varepsilon) = 1\,.$$

Zadosten pogoj za doslednost je, da velja $\lim_{n\to\infty} \mathrm{MSE}(\hat{q}) = 0$, kjer je $\mathrm{MSE}(\hat{q})$ **srednja kvadratična napaka**, *definirana po predpisu:*

$$\mathrm{MSE}(\hat{q}) := \mathbb{E}\big[(\hat{q} - q)^2\big] = \mathrm{var}(\hat{q}) + \big(B(\hat{q})\big)^2\,.$$

Manjša kot je srednja kvadratična napaka, učinkovitejša je cenilka.

1. Iz populacije velikosti N vzamemo vzorec velikosti n. Na populaciji je definirana statistična spremenljivka X, njene vrednosti na populaciji označimo z $x_1, x_2, \ldots, x_N$, na vzorcu pa z $X_1, X_2, \ldots, X_n$. Tedaj lahko definiramo *populacijsko povprečje*:

   $$\mu := \mathbb{E}(X) = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

   in *vzorčno povprečje*:

   $$\bar{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}\,.$$

   Dokažite, da je $\bar{X}$ nepristranska cenilka za μ, in izračunajte srednjo kvadratično napako. Je to dosledna cenilka? Ločite dve možnosti, in sicer, da gre za enostavni slučajni vzorec s ponavljanjem in brez ponavljanja.

2. Če opažene vrednosti $X_1, \ldots, X_n$ uredimo po velikosti, dobimo *vrstilne statistike*:

   $$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}\,.$$

   a) Privzemimo, da so vrednosti $X_1, \ldots, X_n$ nastale kot neodvisne realizacije zvezne enakomerne porazdelitve na $(0, 1)$. Izračunajte pričakovane vrednosti vseh vrstilnih statistik.

   b) Privzemimo zdaj, da so vrednosti $X_1, \ldots, X_n$ nastale kot neodvisne realizacije zvezne enakomerne porazdelitve na poljubnem intervalu. Poiščite nepristransko cenilko za kvantil porazdelitve za verjetnost p, ki temelji na danih dveh vrstilnih statistikah $X_{(i)}$ in $X_{(j)}$, kjer sta i in j različna. Seveda mora biti funkcija statistik $X_{(i)}$ in $X_{(j)}$ neodvisna od intervala.

c) Če predpišemo i in j, za katera p je cenilka enaka kar $X_{(i)}$ oz. $X_{(j)}$? Se to ujema z ustreznima *vzorčnima* kvantiloma (ki ju definiramo kot kvantila porazdelitve povsem naključno izbrane opažene vrednosti)? Kako bi s pomočjo tega poljudno opisali cenilko za kvantil za poljubno verjetnost?

d) Pri konstrukciji cenilke za q_p pri danem p lahko i in j poljubno izberemo. Kateri vrstilni statistiki pa je smiselno izbrati, če p ni preblizu 0 ali 1?

e) Izračunajte ustrezno oceno prvega kvartila porazdelitve, za katero privzamemo, da je porazdeljena zvezno enakomerno, pri vzorčnih vrednostih:

$$3,\ 6,\ 14,\ 16,\ 17,\ 17,\ 18,\ 20\,.$$

3. Pri *stratificiranem vzorčenju* populacijo razdelimo na več podpopulacij – *stratumov* – in iz vsakega vzamemo vzorec predpisane velikosti. Recimo, da je populacija velika in da so deleži posameznih stratumov v celotni populaciji enaki $p_1, p_2, \ldots, p_r$.

   Na populaciji imamo spet definirano statistično spremenljivko X. Na i-tem stratumu naj ima X povprečje μ_i in varianco σ_i^2.

   a) Označimo z $X_1, X_2, \ldots, X_r$ zožitve populacijske spremenljivke X na posamezne stratume. Dokažite, da za poljubno funkcijo f velja:

   $$\mathbb{E}\big[f(X)\big] = p_1\,\mathbb{E}\big[f(X_1)\big] + p_2\,\mathbb{E}\big[f(X_2)\big] + \cdots + p_r\,\mathbb{E}\big[f(X_r)\big]\,.$$

   Pravimo, da je porazdelitev statistične spremenljivke X *mešanica* porazdelitev statističnih spremenljivk $X_1, X_2, \ldots, X_r$.

   b) Naj bo μ povprečje, σ^2 pa varianca statistične spremenljivke na celi populaciji. Izrazite μ in σ^2 z μ_i in σ_i^2.

   c) Iz vsakega stratuma vzamemo enostavni slučajni vzorec – iz i-tega stratuma vzorec velikosti n_i. Označimo z $\bar{X}_1, \ldots, \bar{X}_r$ vzorčna povprečja na teh stratumih. Zapišite nepristransko cenilko za μ, ki temelji na teh vzorčnih povprečjih, in izračunajte njeno varianco za primer, ko so delni vzorci majhni v primerjavi s stratumi.

   d) Dokažite, da je le-ta pri *proporcionalnem stratificiranem vzorčenju*, tj. $n_i = np_i$, kjer je n velikost celotnega vzorca, manjša ali enaka varianci cenilke, ki bi jo dobili iz enostavnega slučajnega vzorca brez stratifikacije.

   e) Recimo, da vzamemo dovolj velik vzorec in da poznamo variance znotraj stratumov (v praksi te variance ocenimo bodisi iz preteklih raziskav bodisi iz manjših *pilotnih vzorcev*). Približno kako velike vzorce moramo vzeti iz posameznih stratumov pri določeni velikosti celotnega vzorca, da bo varianca cenilke $\bar{X}$ najmanjša?

4. Niso vedno vse enote v vzorcu brez ponavljanja z enako verjetnostjo. Vzorčni načrt je lahko zasnovan celo tako, da niso nujno vsi vzorci enako veliki. Predstavimo vzorec s slučajno množico S in za različne enote $i_1, i_2, \ldots, i_k$ označimo:

   $$\pi_{i_1 i_2 \cdots i_k} := \mathbb{P}(i_1 \in S, i_2 \in S, \ldots, i_k \in S)\,.$$

Homogene linearne statistike spremenljivke X na tako predstavljenih vzorcih so oblike:

$$\sum_{i \in S} a_i x_i \,,$$

kjer je x_i vrednost spremenljivke X na i-ti enoti.

a) Recimo, da želimo oceniti populacijsko povprečje dane spremenljivke s homogeno linearno statistiko. Pokažite, da obstaja največ ena izbira koeficientov a_i, pri katerih je ta statistika nepristranska (kdaj obstaja?). Pravimo ji *Horvitz*[36]*–Thompsonova*[37] *cenilka.*

b) Kaj je Horvitz–Thompsonova cenilka na enostavnih slučajnih vzorcih brez ponavljanja?

c) Denimo, da iz populacije velikosti 10 izberemo enostavni slučajni vzorec velikosti 3 s ponavljanjem (in ga nato zapišemo kot množico). To pomeni, da izvedemo tri runde in v vsaki izberemo eno enoto, vsakič na slepo in neodvisno, pri čemer enote pustimo notri; na koncu vzamemo v vzorec enote, ki so bile vsaj enkrat izbrane v vzorec. Izračunajte vrednost Horvitz–Thompsonove cenilke za naslednje primere:

   - izbrane so tri različne enote, vrednosti statistične spremenljivke na njih pa so 1, 2 in 3;
   - v prvih dveh rundah je izbrana ista enota, na kateri je vrednost spremenljivke enaka 1, v tretji rundi pa je izbrana enota, nakateri je vrednost spremenljivke enaka 4;
   - v prvih dveh rundah sta izbrani različni enoti in na obeh je vrednost spremenljivke enaka 1, v tretji rundi pa je izbrana enota, na kateri je vrednost spremenljivke enaka 4;
   - izbrane so tri različne enote, vrednosti statistične spremenljivke na njih pa so 21, 22 in 23.

   Komentirajte!

d) Poiščite potreben in zadosten pogoj, da Horvitz–Thompsonova cenilka komutira s translacijami.

e) Oglejmo si naslednjo modifikacijo stratificiranega vzorčenja: populacijo razdelimo na K stratumov velikosti $N_1, N_2, \ldots, N_K$. Nato najprej vzamemo enostavni slučajni vzorec iz k stratumov, na vsakem izbranem stratumu pa pogojno na to vzamemo enostaven slučajni vzorec: če je i-ti stratum izbran, na njem vzamemo enostavni slučajni vzorec velikosti n_i. Privzamemo, da so ti delni vzorci pogojno na izbrane stratume med seboj neodvisni. Nato vse delne vzorce združimo v enoten vzorec. Poiščite potreben in zadosten pogoj za to, da Horvitz–Thompsonova cenilka komutira s translacijami.

f) Izračunajte varianco Horvitz–Thompsonove cenilke.

[36] Daniel G. Horvitz, ameriški matematik
[37] Donovan J. Thompson (1919–1991), ameriški biostatistik

g) Za primer, ko Horvitz–Thompsonova cenilka komutira s translacijami, izrazite njeno varianco s centriranimi vrednostmi $x_i - \mu$, kjer je $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ populacijsko povprečje.

V nalogah od 5. do 11. privzamemo, da opažanje sestoji iz neodvisnih realizacij $X_1, X_2, \ldots, X_n$ iste statistične spremenljivke X.

5. Je vzorčno povprečje dosledna cenilka parametra a pri zvezni porazdelitvi z gostoto:

$$p(x) = \frac{1}{\pi\big(1 + (x-a)^2\big)}\ ?$$

Metoda momentov *temelji na tem, da za cenilke populacijskih momentov* $z_k = \mathbb{E}(X^k)$ *vzamemo:*

$$\hat{z}_k = \frac{X_1^k + X_2^k + \cdots + X_n^k}{n}$$

Če želimo oceniti karakteristiko q*, jo najprej izrazimo kot funkcijo momentov:* $q = g(z_1, \ldots, z_r)$*. Njena cenilka po metodi momentov je ista funkcija vzorčnih momentov:* $\hat{q} = g(\hat{z}_1, \ldots, \hat{z}_r)$*. Cenilke* $\hat{z}_k$ *so nepristranske in dosledne, zato so tudi cenilke* $\hat{q}_k$*, dobljene po metodi momentov, vedno dosledne. Niso pa nujno nepristranske.*

6. Statistična spremenljivka X je porazdeljena enakomerno na $[0, a]$.

   a) Poiščite cenilko za a, ki temelji na prvem momentu.

   b) Na podlagi cenilke iz prejšnje točke ocenite a iz vzorca:

   $$1,\ 2,\ 1,\ 3,\ 10\,.$$

   Kaj opazite?

   c) Dokažite, da za prejšnji vzorec dobimo smiseln rezultat, če vzamemo kakšen višji moment.

7. Statistična spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(1/\alpha)$, kjer je $\alpha > 0$ neznan parameter. Po metodi momentov poiščite cenilko za α. Je cenilka nepristranska? Je dosledna?

8. Statistična spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 1 & 2 \\ 1-2a-b & a & b & a \end{pmatrix}$$

Po metodi momentov poiščite cenilki za parametra a in b. Ocenite ju iz naslednjega vzorca:

$$-1,\ 0,\ 1,\ 1,\ 0,\ 0,\ -1,\ 2,\ 1,\ 2$$

9. Statistična spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$f_X(x) = \frac{1}{2\alpha} \exp\left(-\frac{|x|}{\alpha}\right)$$

Po metodi momentov poiščite cenilko za neznani parameter α.

10. Statistična spremenljivka X naj ima končno varianco σ^2.

   a) Po metodi momentov poiščite cenilko za σ^2 in pokažite, da je le-ta vedno pristranska.

   b) Označimo cenilko iz prejšnje točke z s_0^2. Pokažite, da obstaja tak faktor k', odvisen le od velikosti vzorca, da je $s^2 := k' s_0^2$ nepristranska cenilka za σ^2.

   c) Privzemimo, da ima X končen četrti moment. Izračunajte srednji kvadratični napaki cenilk s_0^2 in s^2.

   d) Naj bo X porazdeljena normalno. Pokažite, da obstaja tak faktor k^*, odvisen le od velikosti vzorca, da je $k^* s_0^2$ najučinkovitejša izmed cenilk oblike $k s_0^2$, $k > 0$. Izračunajte še $\mathrm{MSE}(k^* s_0^2)$.

11. Statistična spremenljivka X je porazdeljena po Poissonu $\mathrm{Pois}(\lambda)$. Dokažite, da sta $\bar{X}$ in:

$$s^2 := \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

obe nepristranski cenilki za λ. Katera ima manjšo varianco?

Metoda največjega verjetja

Naj bo $f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ verjetnostna funkcija ali gostota porazdelitve opažanja $\boldsymbol{X}$, ki je odvisna še od parametra θ. Cenilka za θ po metodi največjega verjetja (maksimalne zanesljivosti) pri opažanju $\boldsymbol{X}$ je tisti θ, pri katerem je **verjetje** *$L(\theta) := f_{\boldsymbol{X}}(\boldsymbol{X};\theta)$ maksimalno.*
Za iskanje maksimuma navadno rešujemo enačbo:

$$\frac{\mathrm{d} \ln L}{\mathrm{d}\theta} = 0\,,$$

če je parameter θ enorazsežen, in ustrezen sistem enačb s parcialnimi odvodi, če je večrazsežen.

Opomba. Pričakovana vrednost logaritma verjetja $\mathbb{E}_{\theta_0}\big[\ln L(\theta)\big]$ je maksimalna ravno pri $\theta = \theta_0$ in maksimum je dosežen le pri tistih θ, za katere je $\mathbb{P}_\theta = \mathbb{P}_{\theta_0}$.

12. Denimo, da zaporedje poskusov tako kot v 6. nalogi iz 10. razdelka tvori markovsko verigo z začetno verjetnostjo uspeha θ in prehodno matriko:

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\theta}{2} & \frac{\theta}{2} \\ \frac{1-\theta}{2} & \frac{1+\theta}{2} \end{bmatrix}.$$

Recimo, da smo opazili, da je prvi poskus uspel, drugi ni uspel, tretji pa je spet uspel. Ocenite θ po metodi največjega verjetja.

Če imamo na voljo opažene vrednosti $X_1, \ldots, X_n$, ki so nastale kot neodvisne realizacije porazdelitve z verjetnostno funkcijo ali gostoto $f(x;\theta)$, je verjetje kar produkt verjetij za posamezne spremenljivke:

$$L(\theta) = L_1(\theta)\, L_2(\theta) \cdots L_n(\theta)\,,$$

kjer je $L_i(\theta) = f(X_i;\theta)$. V tem primeru se verjetje često splača logaritmirati, saj je potem:

$$\ln L(\theta) = \ln L_1(\theta) + \ln L_2(\theta) + \cdots + \ln L_n(\theta)\,.$$

V nalogah od 13. do 15. spet privzamemo, da opažanje sestoji iz neodvisnih realizacij $X_1, X_2, \ldots, X_n$ iste statistične spremenljivke X.

13. Statistična spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(1/\alpha)$, kjer je $\alpha > 0$ neznan parameter. Po metodi največjega verjetja poiščite cenilko za α. Dobimo isto cenilko kot po metodi momentov (glej 7. nalogo)?

14. Statistična spremenljivka X ima *Paretovo*[38] *porazdelitev*, ki je zvezna z gostoto:

$$f_X(x;a) = \begin{cases} \dfrac{c}{x^a} & ; x > 1 \\ 0 & ; \text{sicer.} \end{cases}$$

   a) Določite tiste vrednosti parametra a, pri katerih je lahko z zgornjo formulo definirana verjetnostna gostota. V primerih, ko to možno, ustrezno izrazite c z a.

   b) Ocenite a po metodi največjega verjetja.

15. Statistična spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 1 & 2 \\ 1-2a-b & a & b & a \end{pmatrix}$$

   Po metodi največjega verjetja iz opaženih vrednosti:

$$-1, 0, 1, 1, 0, 0, -1, 2, 1, 2$$

   ocenite parametra a in b. Dobimo isto oceno kot po metodi momentov?

[38] Vilfredo Federico Damaso Pareto (1848–1923), italijanski inženir, sociolog, ekonomist, politolog in filozof

Metoda največjega verjetja in zadostnost

Pri metodi največjega verjetja lahko namesto z izvirnim opažanjem delamo tudi z zadostno statistiko. Premislek velja tudi v obratno smer: opažanje lahko nadomestimo s hipotetičnim opažanjem z več informacije, pri katerem pa je izvirno opažanje zadostna statistika.
Karakterizacija zadostnosti prek Fisher–Neymanovega faktorizacijskega izreka se elegantno zapiše z verjetjem: statistika T je za model, parametriziran s θ, zadostna natanko tedaj, ko se da verjetje zapisati v obliki:

$$L = W\, g(T;\theta)\,,$$

kjer je W neka statistika.

16. *Hardy*[39]*–Weinbergov*[40] *model* iz matematične biologije določa zvezo med verjetnostmi genotipov, če sta v populaciji dva *alela* (različici istoležnega gena), recimo R in r. Tedaj so možni genotipi RR, Rr in rr. Po modelu so verjetnosti teh genotipov enake $(1-\theta)^2$, $2\theta(1-\theta)$ in θ^2 (ker je θ delež alela r v populaciji).

    V enostavnem slučajnem vzorcu, zajetem iz velike populacije, je 20 osebkov tipa RR, 30 tipa Rr in 50 tipa rr. Ocenite θ po metodi največjega verjetja.

17. Statistična spremenljivka X je porazdeljena zvezno z gostoto:

$$f_X(x;\lambda) = \begin{cases} e^{-\lambda x} & ; x \ge -\frac{\ln\lambda}{\lambda} \\ 0 & ; \text{sicer,} \end{cases}$$

    kjer je $\lambda > 0$ neznan parameter. Po metodi največjega verjetja poiščite cenilko za λ, ki temelji na enem samem opažanju te statistične spremenljivke.

18. Statistična spremenljivka X je porazdeljena enakomerno $\mathrm{Unif}(0,a)$. Opazimo neodvisne realizacije $X_1, X_2, \ldots, X_n$ te statistične spremenljivke.

    a) Naj bo M cenilka za a, dobljena po metodi momentov, V pa cenilka, dobljena po metodi največjega verjetja. Katera od cenilk je nepristranska? Katera je učinkovitejša?

    b) Modificirajte cenilko V, tako da bo nepristranska. Je nova cenilka učinkovitejša?

    c) Ali obstaja cenilka, ki je še učinkovitejša od cenilke iz prejšnje točke?

[39] Godfrey Harold Hardy (1877–1947), angleški matematik

[40] Wilhelm Weinberg (1862–1937), nemški splošni zdravnik, ginekolog, porodničar, raziskovalec dednosti, statistik in genealog

Nepristranska cenilka z enakomerno najmanjšo možno varianco

V eksponentni družini, ki se da zapisati tako, da pripadajoči naravni parametrični prostor vsebuje odprto množico, ima nepristranska cenilka za dano karakteristiko q enakomerno najmanjšo možno varianco natanko tedaj, ko se izraža s pripadajočo zadostno statistiko $T = \big(h_1(X), \ldots, h_m(X)\big)$. Taka cenilka je (do skoraj gotove enakosti) enolična in obstaja, brž ko obstaja sploh kakšna nepristranska cenilka Z za q. Nepristransko cenilko z enakomerno najmanjšo možno varianco v tem primeru dobimo kot $\mathbb{E}(Z \mid T)$ (ki je opazljiva, ker je T zadostna).

19. Vsak poskus določene vrste uspe z verjetnostjo $p \in (0, 1)$. Poiščite nepristransko cenilko za p^2 z enakomerno najmanjšo možno varianco, ki temelji na izvedbi treh neodvisnih poskusov te vrste.

    Kaj pa, če izvedemo n poskusov?

20. Statistična spremenljivka X je porazdeljena diskretno po shemi:

$$\begin{pmatrix} 0 & 1 & 2 \\ \dfrac{1}{1+3\theta+2\theta^2} & \dfrac{3\theta}{1+3\theta+2\theta^2} & \dfrac{2\theta^2}{1+3\theta+2\theta^2} \end{pmatrix},$$

    kjer je $\theta > 0$ neznan parameter.

    a) Zapišite to kot naravno enoparametrično eksponentno družino porazdelitev. *Namig*: pomagajte si s funkcijami $\rho_i(x) = \begin{cases} 1 & ; x = i \\ 0 & ; \text{sicer} \end{cases}$.
    b) Poiščite nepristransko cenilko za $1/(1 + \theta)$ z enakomerno najmanjšo možno varianco, ki temelji na eni sami realizaciji te statistične spremenljivke.

V 21. in 22. nalogi spet privzamemo, da opažanje sestoji iz neodvisnih realizacij $X_1, X_2, \ldots, X_n$ iste statistične spremenljivke X.

21. Statistična spremenljivka je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$.

    a) Poiščite nepristransko cenilko za λ z enakomerno najmanjšo možno varianco.
    b) Poiščite večkratnik cenilke iz prejšnje točke z najmanjšo srednjo kvadratično napako.

22. Statistična spremenljivka je porazdeljena geometrijsko $\mathrm{Geom}(p)$. Poiščite nepristransko cenilko za p z enakomerno najmanjšo možno varianco.

    Namig: najprej raziščite primer, ko opazimo le eno realizacijo.

23. Statistična spremenljivka X je porazdeljena po Poissonu $\mathrm{Pois}(\lambda)$. Želeli bi oceniti λ^2, opazimo pa eno samo realizacijo (X).

    a) Poiščite cenilko po metodi največjega verjetja. Je le-ta nepristranska?

b) Poiščite nepristransko cenilko z enakomerno najmanjšo možno varianco.

c) Recimo, da nastavimo cenilko v obliki $X^2 - aX$. Če natančnost cenilke merimo s srednjo kvadratično napako, ali obstaja najnatančnejša cenilka te oblike? Katere izbire pa so smiselne?

Pomoč: prvi štirje momenti Poissonove porazdelitve so:

$\mathbb{E}(X) = \lambda\,,\ \mathbb{E}(X^2) = \lambda^2 + \lambda\,,\ \mathbb{E}(X^3) = \lambda^3 + 3\lambda^2 + \lambda\,,\ \mathbb{E}(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda\,.$

24. Statistična spremenljivka je porazdeljena po Poissonu $\mathrm{Pois}(\lambda)$, opazimo pa eno samo realizacijo (X).

a) Poiščite nepristransko cenilko za $e^{-3\lambda}$ z enakomerno najmanjšo možno varianco. Je videti smiselna?

b) Dokažite, da ima cenilka po metodi največjega verjetja enakomerno manjšo srednjo kvadratično napako.

c) Kaj pa, če opazimo vzorec neodvisnih opažanj $X_1, X_2, \ldots, X_n$? Kako se nepristranska cenilka asimptotično obnaša?

12. Intervali zaupanja

Normalna porazdelitev z znanim σ. Normalna porazdelitev z obema neznanima parametroma, iščemo μ ali σ. Asimptotični intervali zaupanja za ne-Gaussove porazdelitve, Bernoullijevo zaporedje poskusov.

Interval zaupanja $(q_{\min}, q_{\max})$ *za karakteristiko* q *pri* **stopnji zaupanja** β *je določen z neenačbo:*

$$\mathbb{P}\big(q_{\min} < q < q_{\max}\big) \geq \beta$$

ki mora veljati za vse verjetnostne mere $\mathbb{P}$ *iz našega statističnega modela,* $q_{\min}$ *in* $q_{\max}$ *pa morata biti opazljivi. Če je res* $q_{\min} < q < q_{\max}$*, pravimo, da pride do* **pokritosti**. *Če se da, interval izberemo tako, da je* β *natančna spodnja meja za verjetnost pokritosti. Tipično vzamemo* $\beta = 0\dot{}90$, $0\dot{}95$ *ali* $0\dot{}99$.
Konstrukcija intervalov zaupanja navadno temelji na pojmu **pivotne funkcije** $T(\boldsymbol{X}, q)$, *kjer je* $\boldsymbol{X}$ *opažanje (recimo vzorec): interval zaupanja za* q *je* $\{q\ ;\ t_{\min} < T(\boldsymbol{X}, q) < t_{\max}\}$ *ali kaj podobnega.* T *je pivotna funkcija, če je porazdelitev slučajne spremenljivke* $T(\boldsymbol{X}, q)$ *konstantna (ko spreminjamo parametre modela in se s tem spreminja* q*, prav tako pa tudi porazdelitev opažanja* $\boldsymbol{X}$*). Tega sicer ni možno vedno doseči, a je dovolj, če dosežemo primerno približno. Obenem pa se mora pri fiksni vrednosti opažanja* $\boldsymbol{X}$ *statistika* T *čimbolj spreminjati z drugim argumentom.*
Če zaokrožujemo, spodnjo mejo vedno zaokrožimo navzdol, zgornjo pa navzgor.

1. Za $n = 1$ in $n = 2$ ter poljuben $\beta \in (0, 1)$ konstruirajte interval zaupanja za parameter θ (z verjetnostjo pokritosti vsaj β, kjer opazimo $S \sim \mathrm{Bin}(n, \theta)$, z naslednjimi lastnostmi:
   - *monotonost*: krajišči sta naraščajoči funkciji opažanja X;
   - *simetrija*: interval zaupanja za $S = n - k$ je interval zaupanja za $S = k$, prezrcaljen okoli $1/2$;
   - *minimalnost*: interval je minimalen med vsemi intervali zaupanja z zgornjima lastnostma.

2. Statistična spremenljivka ima diskretno porazdelitev, podano s shemo:

$$\begin{pmatrix} 0 & 1 & 2 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & \dfrac{\theta(1-\theta)}{\theta^2 - \theta + 1} & \dfrac{(1-\theta)^2}{\theta^2 - \theta + 1} \end{pmatrix}.$$

   Poiščite minimalni *levostranski* 95% interval zaupanja, ki bo temeljil na enem samem opažanju te spremenljivke. Natančneje, če opazimo k, naj bo ta interval oblike $I = [0, b_k)$ ali $[0, 1]$ (za $b_k = 1$).

3. V filmu *Rain Man* iz leta 1988 glavni junak filma, ki ima avtistično motnjo, noče leteti z nobeno letalsko družbo razen z avstralsko družbo Qantas: edino ta namreč še ni imela smrtne nesreče z reaktivnim letalom.

Vzemimo Bernoullijevo zaporedje poskusov, v katerem se posamezen poskus ponesreči z verjetnostjo θ. Pri predpisani stopnji zaupanja β si oglejmo naslednji interval zaupanja za θ, ki temelji na izvedbi n takih poskusov: če se kateri izmed poskusov ponesreči, vzamemo kar $[0, 1]$, če pa se nobeden izmed poskusov ne ponesreči, vzamemo interval $[0, \theta^*)$ (lahko si mislimo, da se glavni junak filma ravna natančno po tej konstrukciji). Kolikšna je najmanjša možna vrednost za θ^*?

4. Spet vzemimo zaporedje Bernoullijevo zaporedje poskusov, pri katerem se vsak ponesreči z verjetnostjo θ. Naj bo N število posrečenih poskusov pred prvim ponesrečenim. Pri dani stopnji zaupanja β konstruirajte minimalni levostranski interval zaupanja za θ, tj. pri $N = n$ naj bo le-ta oblike $[0, b_n)$ ali $[0, b_n]$, kjer je $b_0 > b_1 > \ldots$ Primerjajte s prejšnjo nalogo!

5. Konstruirajte minimalni levostranski interval zaupanja za parameter θ na podlagi opažanja $S \sim \mathrm{Bin}(n, \theta)$, tj. za $S = k$ naj bo interval oblike $[0, b_k)$ ali $[0, b_k]$, kjer je $b_0 < b_1 < \cdots < b_n$.

   Recimo sedaj, da je S število ponesrečenih poskusov med prvimi n poskusi v Bernoullijevem zaporedju. Oglejmo si dogodek, da se prvih $n-1$ poskusov posreči, n-ti pa ponesreči. Kakšen interval zaupanja za θ dobimo na tem dogodku? Kakšnega pa, če uporabimo konstrukcijo iz prejšnje naloge? Razložite!

Če je I_1 množica zaupanja pri stopnji zaupanja $1 - \alpha_1$, I_2 pa množica zaupanja pri stopnji zaupanja $1 - \alpha_2$, je $I_1 \cap I_2$ množica zaupanja pri stopnji zaupanja $1 - \alpha_1 - \alpha_2$. Na podlagi te ugotovitve lahko iz enostranskih intervalov zaupanja dobimo dvostranske. Tako iz prejšnje konstrukcije dobimo naslednji dvostranski interval zaupanja:

Clopper[41]–Pearsonov[42] interval zaupanja za neznano verjetnost

Izvedemo n neodvisnih poskusov, pri katerih vsak uspe z verjetnostjo θ. Opazimo, da je uspelo natanko S poskusov. Tedaj za interval zaupanja za θ postavimo:

$$\begin{array}{cl} \big[0,\ B_{(1-\beta)/2}(1, n)\big) & ;\ S = 0 \\ \big(B_{(1+\beta)/2}(S, n-S+1),\ B_{(1-\beta)/2}(S+1, n-S)\big) & ;\ S = 1, 2, \ldots, n-1 \\ \big(B_{(1+\beta)/2}(n, 1),\ 1\big] & ;\ S = n \end{array},$$

kjer je $B_p(r, s)$ kvantil porazdelitve $\mathrm{Beta}(r, s)$ *za verjetnost $1 - p$. Velja (glej 33. nalogo iz 5. razdelka):*

$$B_p(r, s) = \frac{r\, F_p(2r, 2s)}{r\, F_p(2r, 2s) + s} = \frac{r}{r + s\, F_{1-p}(2s, 2r)},$$

kjer je $F_p(a, b)$ kvantil Snedecorjeve (Fisherjeve)[43,44] porazdelitve z a in b prostostnimi stopnjami za verjetnost $1 - p$.

Iz 5. naloge sledi, da Clopper–Pearsonov interval zagotavlja verjetnost pokritosti vsaj β, vendar pa ni vedno minimalen. Zadošča pa monotonosti in simetriji.

[41] C. J. Clopper

6. Za primer, ko izvedemo dva poskusa, primerjajte Clopper–Pearsonov interval zaupanja za neznano verjetnost s tistim iz 1. naloge. Privzemite, da je $\beta \geq 4/9$.

7. Opazimo vrednosti $X_1, \ldots, X_n$, ki so nastale neodvisne realizacije diskretne porazdelitve:
$$\begin{pmatrix} 0 & a & 3a \\ 1/3 & 1/2 & 1/6 \end{pmatrix},$$
kjer je $a > 0$ neznan parameter. Za velike n poiščite asimptotični 95% interval zaupanja za a oblike $[0, a_{\max}]$, kjer mora biti meja $a_{\max}$ opazljiva.

Sredina pri normalni porazdelitvi z znanim standardnim odklonom

Opazimo vrednosti $X_1, X_2, \ldots, X_n$, ki so nastale kot neodvisne realizacije statistične spremenljivke X, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$. Ocenjujemo parameter μ pri čemer je σ znan. Tedaj ne glede na μ velja:

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim \mathrm{N}(0, 1),$$

kjer je $\bar{X}$ vzorčno povprečje, definirano spodaj. Izračunamo torej:

$$\begin{aligned} \bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\ c &= z_{(1-\beta)/2} = \Phi^{-1}(\beta/2) \\ \Delta &= \frac{c\sigma}{\sqrt{n}} \end{aligned}$$

Interval zaupanja: $\bar{X} - \Delta < \mu < \bar{X} + \Delta$.

8. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, 5)$. Vrednosti na vzorcu so:

101, 91, 93, 103, 91, 101, 103, 95, 95

Določite 95% interval zaupanja za μ.

[42]Karl Pearson (1857–1936), angleški matematik in biostatistik
[43]George Waddel Snedecor (1882–1974), ameriški matematik in statistik
[44]Sir Ronald Aymler Fisher (1899–1962), angleški statistik in biolog

Sredina pri normalni porazdelitvi z neznanim standardnim odklonom

Spet opazimo vrednosti $X_1, X_2, \ldots, X_n$*, ki so nastale kot neodvisne realizacije statistične spremenljivke* X*, porazdeljene normalno* $\mathrm{N}(\mu, \sigma)$*. Ocenjujemo parameter* μ*, pri čemer* σ *zdaj ni znan. Tedaj ne glede na* μ *in* σ *velja:*

$$\frac{\bar{X} - \mu}{s} \sqrt{n} \sim \mathrm{Student}(n-1)\,,$$

kjer je s *popravljeni vzorčni standardni odklon, definiran spodaj. Izračunamo torej:*

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c &= t_{(1-\beta)/2}(n-1) \\
s &= \sqrt{\frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}} \\
\Delta &= \frac{cs}{\sqrt{n}}\,,
\end{aligned}$$

kjer je $t_p(df)$ *kvantil Studentove*[45] *porazdelitve z* df *prostostnimi stopnjami za verjetnost* $1-p$*.*
Interval zaupanja: $\bar{X} - \Delta < \mu < \bar{X} + \Delta$*.*
Za velike vzorce konstrukcija asimptotično deluje tudi pri ne-Gaussovih spremenljivkah. V tem primeru lahko kvantil Studentove porazdelitve $t_{(1-\beta)/2}(n-1)$ *zamenjamo s kvantilom normalne porazdelitve* $z_{(1-\beta)/2} = \Phi^{-1}(\beta/2)$*.*

9. Isto kot prejšnja naloga, le da σ zdaj ni znan.

Asimptotični interval zaupanja za pričakovano vrednost pri ne-Gaussovih spremenljivkah

Naj bo X *nekonstantna statistična spremenljivka s pričakovano vrednostjo* μ *in varianco* σ*. Opazimo vrednosti* $X_1, X_2, \ldots, X_n$*, ki so nastale kot neodvisne realizacije te statistične spremenljivke. Tedaj vemo, da sta* $s_0 = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}$ *in* s *dosledni cenilki za* σ *(glej 10. nalogo iz 11. razdelka), torej po izreku Sluckega*[46] *za deljenje velja:*

$$\frac{\bar{X} - \mu}{S} \sqrt{n}\,,\ \frac{\bar{X} - \mu}{s_0} \sqrt{n} \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)\,.$$

Konstrukcija asimptotičnega intervala zaupanja torej sovpada s konstrukcijo eksaktnega intervala zaupanja za Gaussove spremenljivke pri neznanem σ*, ki ga lahko ocenimo z* s_0 *ali* s*.*

[45] William Sealy Gosset (1876–1937), angleški statistik, bolj znan pod psevdonimom *Student*

10. Za 860 žensk poizvemo, koliko otrok imajo. Dobimo naslednjo frekvenčno porazdelitev:

število otrok	0	1	2	3	4	5	6	7	8	9	10
število žensk	227	168	320	96	29	11	4	2	2	0	1

Poiščite 95% interval zaupanja za povprečno število otrok na žensko na celotni populaciji.

Opomba. Podatki so sicer izmišljeni, so pa ukrojeni po popisu Slovenije iz leta 2002 (števila žensk so deljena s 1000 in zaokrožena, izmišljen je tudi konec tabele).

11. Izvedemo n neodvisnih poskusov, pri katerih vsak uspe z verjetnostjo θ. Opazimo, da je uspelo S poskusov. Konstruirajte asimptotični interval zaupanja za θ, pri čemer za oceno standardnega odklona uporabite s_0.

Wilsonov[47] interval zaupanja za neznano verjetnost

To je interval, ki ga dobimo neposredno iz Laplaceove integralske formule brez popravka s polovico. Ob oznakah iz prejšnje naloge dobimo:

$$\hat{\theta} - c\sqrt{\frac{\theta(1-\theta)}{n}} < \theta < \hat{\theta} + c\sqrt{\frac{\theta(1-\theta)}{n}}\,.$$

V zgornji formuli interval zaupanja za θ še ni eksplicitno izražen z opažanjem S (oz. $\hat{\theta}$). Za ta namen je potrebno rešiti kvadratno enačbo. Dobimo:

$$\tilde{\theta} - \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}(1-\hat{\theta}) + \frac{c^2}{4}} < \theta < \tilde{\theta} + \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}(1-\hat{\theta}) + \frac{c^2}{4}}\,,$$

kjer je $\tilde{n} = n + c^2$ in $\tilde{\theta} = \dfrac{S + \frac{c^2}{2}}{\tilde{n}}$. Če kot spodnje krajišče dobimo 0, ga vključimo v interval, prav tako vključimo zgornje krajišče 1.

Ta interval ima tipično boljšo pokritost kot Waldov. Ima dobro pokritost **v povprečju** *in podobno kot pri Waldovem intervalu velja, da se, ko θ preteče interval $[a, b]$ $(0 < a \le b < 1)$, minimalna verjetnost pokritosti bliža nominalni β, ko gre n proti neskončno. Za $n \ge 20$ in $0{\cdot}1 \le p \le 0{\cdot}9$ je pri $\beta = 0{\cdot}95$ verjetnost pokritosti vsaj $0{\cdot}92$.*

[46] Evgenij Evgenjevič Slucki (1880–1948), ruski matematik, statistik in ekonomist

[47] Edwin Bidwell Wilson (1879–1964), ameriški matematik

Wilsonov interval zaupanja s popravkom za zveznost

Ta interval dobimo neposredno iz Laplaceove integralske formule s popravkom s polovico. Ob oznakah iz prejšnje naloge dobimo:

$$\tilde{\theta}_- - \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}_-(1-\hat{\theta}_-)+\frac{c^2}{4}} < \theta < \tilde{\theta}_+ + \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}_+(1-\hat{\theta}_+)+\frac{c^2}{4}}\,,$$

kjer je $\hat{\theta}_- = \dfrac{S-\frac{1}{2}}{\hat{n}}$, $\tilde{\theta}_- = \dfrac{S-\frac{1}{2}}{\tilde{n}}$, $\hat{\theta}_+ = \dfrac{S+\frac{1}{2}}{\hat{n}}$ *in* $\tilde{\theta}_+ = \dfrac{S+\frac{1}{2}}{\tilde{n}}$.
Če je $S = 0$*, za spodnje krajišče postavimo* 0 *in ga vključimo.*
Če je $S = n$*, za zgornje krajišče postavimo* 1 *in ga vključimo.*
Numerični izračuni kažejo, da različica s popravkom vedno doseže nominalno pokritost. V povprečju pa pride do prepokritosti (ki je višjega velikostnega reda kot povprečna prepokritost pri različici brez popravka).

Agresti[48]–Coullov[49] interval zaupanja za neznano verjetnost

Je zelo blizu Wilsonovemu, a je lažje izračunljiv. Če upoštevamo oznake iz Wilsonovega intervala, je različica brez popravka za zveznost:

$$\tilde{\theta} - c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}} < \theta < \tilde{\theta} + c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}}\,.$$

Tu že za $n \ge 10$ *velja, da je pri* $0{\dot{}}1 \le p \le 0{\dot{}}9$ *in* $\beta = 0{\dot{}}95$ *verjetnost pokritosti vsaj* $0{\dot{}}92$.
Različica s popravkom za zveznost:

$$\tilde{\theta} - c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}} - \frac{1}{2n} < \theta < \tilde{\theta} + c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}} + \frac{1}{2n}\,.$$

Tudi tu numerični izračuni kažejo, da različica s popravkom vedno doseže nominalno pokritost.

12. Zanima nas, kolikšnemu deležu študentov je izmed zvrsti filma najbolj všeč romantični film. Anketiramo 20 študentov in romantični film je najbolj všeč štirim. Določite vse prej omenjene intervale zaupanja za delež vseh študentov, ki jim je najbolj všeč romantični film, pri 90% stopnji zaupanja.

13. V 100 metih kocke je 20-krat padla šestica. Določite vse prej omenjene intervale zaupanja za verjetnost padca šestice pri 95% stopnji zaupanja.

[48]Alan Agresti (1947), ameriški statistik
[49]Brent A. Coull, ameriški biostatistik

14. Pri 10000 metih kovanca je padlo 5048 grbov. Določite vse prej omenjene 90% intervale zaupanja za verjetnost, da pade grb.

V nadaljevanju razdelka privzamemo, da opažanje sestoji iz neodvisnih realizacij $X_1, X_2, \ldots, X_n$ iste statistične spremenljivke X

Standardni odklon pri normalni porazdelitvi

Naj bo $X \sim \mathrm{N}(\mu, \sigma)$. Ocenjujemo parameter σ, pri čemer μ ni znan. Tedaj je:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)\,.$$

Izračunamo torej:

$$\begin{aligned}\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n}\\ c_1 &= \chi^2_{(1+\beta)/2}(n-1)\\ c_2 &= \chi^2_{(1-\beta)/2}(n-1)\\ s^2 &= \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}\,,\end{aligned}$$

kjer je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost $1 - p$.
Interval zaupanja:

$$s\sqrt{\frac{n-1}{c_2}} < \sigma < s\sqrt{\frac{n-1}{c_1}}$$

oziroma:

$$\sqrt{\frac{1}{c_2}\sum_{i=1}^{n}(X_i - \bar{X})^2} < \sigma < \sqrt{\frac{1}{c_1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

15. Isti podatki kot pri 8. nalogi, le da ocenjujemo σ.

16. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Opažene vrednosti so:

124, 129, 126, 122, 124

Ocenite σ po občutku. Ali pride v 90% interval zaupanja?

17. Telesna teža v skupini 75 učencev ima naslednjo frekvenčno porazdelitev:

teža [kg]	39	40	41	42	43	44	45	46	47	48	49	50
št. učencev	1	3	5	8	8	7	9	8	6	6	4	3

teža [kg]	51	52	53	54	59
št. učencev	2	2	1	1	1

Poiščite 99% interval zaupanja za μ in za σ (za vsakega posebej, pri čemer privzemite, da drugi parameter ni znan).

18. Naj ima X pričakovano vrednost μ, standardni odklon σ in končen četrti moment. Želeli bi oceniti σ. Modificirajte neopazljivo spremenljivko:

$$\sum_{i=1}^{n}(X_i - \mu)^2$$

v opazljivo funkcijo karakteristike σ, ki ima, če je v argumentu prava vrednost karakteristike σ, asimptotično normalno porazdelitev. Na podlagi tega konstruirajte asimptotični interval zaupanja za σ.

Asimptotični interval zaupanja za populacijski standardni odklon pri ne-Gaussovih spremenljivkah

$$\hat{\sigma}^2 - c\,\frac{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}{\sqrt{n}} < \sigma^2 < \hat{\sigma}^2 + c\,\frac{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}{\sqrt{n}}\,,$$

kjer je $c = z_{(1-\beta)/2} = \Phi^{-1}(\beta/2)$.

19. **MANJKA!**

20. Poiščite 95% interval zaupanja za standardni odklon števila otrok na žensko za podatke iz 10. naloge:

število otrok	0	1	2	3	4	5	6	7	8	9	10
število žensk	227	168	320	96	29	11	4	2	2	0	1

13. Preizkusi značilnosti

p-vrednosti. Preizkus na podlagi razmerja verjetij. Nerandomizirani preizkus neznane verjetnosti. Neyman–Pearsonova lema. Wilksov izrek. Z-preizkus neznane verjetnosti. Z- in T-preizkus sredine. T-preizkus razlike sredin (za parne in neodvisne vzorce). Analiza variance z enojno klasifikacijo. Preizkušanje variance. Preizkus skladnosti s fiksno porazdelitvijo in z družino porazdelitev. Preizkus z znaki. Inverzijski preizkus.

Želeli bi preizkusiti, ali so opaženi podatki v vzorcu v skladu z ničelno domnevo H_0 o porazdelitvi ali pa so morda bolj v skladu z alternativno domnevo H_1. Pri preizkusih značilnosti bodisi zavrnemo ničelno domnevo bodisi pravimo, da odstopanja niso statistično dovolj **značilna**, *da bi jo zavrnili.*
Postopku, po katerem se odločimo, ali bomo ničelno domnevo zavrnili ali ne, pravimo **preizkus**. *Preizkus ustreza* **stopnji značilnosti** *oz.* **tveganja** α, *če za vsako verjetnostno mero $\mathbb{P}$, za katero velja H_0, velja:*

$$\mathbb{P}(H_0 \text{ zavrnemo}) \le \alpha\,.$$

Če se da, preizkus načrtujemo tako, da je α natančna zgornja meja, z drugimi besedami, da je stopnja značilnosti **eksaktna**.
Če ničelno domnevo zavrnemo pri $\alpha = 0{\cdot}05$, pravimo, da so odstopanja statistično **značilna**. *Če se to zgodi pri pragu $0{\cdot}01$, pa pravimo, da so statistično* **zelo značilna**.
Moč preizkusa *je verjetnost, da H_0 zavrnemo, gledana na H_1.*

Odločanje o tem, ali bomo domnevo zavrnili ali ne, navadno temelji na **preizkusni statistiki**, *tj. opazljivi spremenljivki z vrednostmi v* $\mathbb{R}$. *Domnevo zavrnemo, če preizkusna statistika pade v* **kritično območje**, *ki ga navadno označimo s* K_α.
Lahko si predstavljamo, da preizkusna statistika T *meri, koliko opažanje "ustreza" ničelni domnevi, tj. večja kot je vrednost* T, *"ustreznejše" je opažanje. Tako domnevo zavrnemo, če je* $T \le c$ *ali pa* $T < c$. *Pragu* c *pravimo* **kritična vrednost**. *Da bi dosegli eksaktno stopnjo značilnosti, lahko preizkus tudi* **randomiziramo**: *če je* $T < c$, *domnevo zavrnemo, če je* $T > c$, *je ne zavrnemo, če je* $T = c$, *pa jo zavrnemo z določeno verjetnostjo.*
Lahko si pomagamo tudi s p*-vrednostmi: definirajmo* $p(t) := \bar{\mathbb{P}}(T' \le t)$, *kjer smo s* $\bar{\mathbb{P}}(A)$ *označili supremum vseh verjetnosti dogodka* A *pri ničelni domnevi,* T' *pa je kopija preizkusne statistike* T *(pomožna slučajna spremenljivka, porazdeljena tako kot* T*). Domnevo zavrnemo, brž ko je* $p(T) \le \alpha$. *Če to ni res, lahko ravnamo na dva načina: ali je ne zavrnemo ali pa* **randomiziramo**: *definiramo še* $p^-(t) := \bar{\mathbb{P}}(T' < t)$. *Če je* $p^-(T) > \alpha$, *domneve nikakor ne zavrnemo, sicer pa jo zavrnemo s pogojno verjetnostjo:*

$$\frac{\alpha - p^-(T)}{p(T) - p^-(T)} .$$

Preizkus na podlagi razmerja verjetij

Naj bo statistični model parametriziran s $\theta \in \Theta$. *Preizkušamo ničelno domnevo, da je* $\theta \in \Theta_{H_0}$, *proti alternativni domnevi, da je* $\theta \in \Theta_{H_1} := \Theta \setminus \Theta_{H_0}$. *Preizkus na podlagi razmerja verjetij temelji na preizkusni statistiki:*

$$\Lambda = \frac{\sup_{\theta \in \Theta_{H_0}} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} ,$$

kjer je $L(\theta)$ *verjetje, tj. verjetnostna funkcija ali gostota, evaluirana na opažanju. Večje kot je razmerje verjetij, ustreznejši je vzorec. Z drugimi besedami, ničelno domnevo zavrnemo, če je razmerje verjetij* **premajhno**. *Lahko vzamemo tudi:*

$$\Lambda' = \frac{\sup_{\theta \in \Theta_{H_0}} L(\theta)}{\sup_{\theta \in \Theta_{H_1}} L(\theta)} .$$

1. Loterija za neko vrsto srečke trdi, da jih je vsaj pol dobitnih. Kupili smo osem srečk in le dve sta bili dobitni. Lahko pri stopnji značilnosti $\alpha = 0{\cdot}05$ začnemo sumiti, da loterija laže? Seveda privzamemo, da so posamezne kupljene srečke med seboj neodvisne.

2. Pri 20 metih kovanca je padlo 5 grbov. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da je verjetnost, da pade grb, enaka $1/2$, proti alternativni domnevi, da

je različna od 1/2. Kaj pa, če bi vzeli $\alpha = 0{\cdot}01$? Uporabite preizkus na podlagi razmerja verjetij.

3. Na neki fakulteti študira 70% žensk in 30% moških. Posebno priznanje za izjemne študijske dosežke je bilo podeljeno 5 ženskam in 7 moškim. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite ničelno domnevo, da so izjemni študijski dosežki (po merilih komisije) neodvisni od spola, proti alternativni domnevi, da so od spola odvisni. Privzemite, da se spoli posameznikov, ki dobijo nagrado, obnašajo kot Bernoullijevo zaporedje poskusov.

Neznana verjetnost (enostavnejši, nerandomiziran preizkus)

Naj bo θ verjetnost, da se zgodi določen dogodek. Pri stopnji značilnosti α preizkušamo ničelno domnevo H_0, ki pravi, da je $\theta = \theta_0$. Izvedemo n poskusov in dani dogodek se zgodi pri natanko S poskusih. Ničelno domnevo zavrnemo, če je $\rho(S) \le \alpha$, kjer je funkcija ρ odvisna od alternativne domneve:

- *pri H_1^+: $\theta > \theta_0$ je $\rho(k) = \mathbb{P}(S' \ge k)$;*
- *pri H_1^-: $\theta < \theta_0$ je $\rho(k) = \mathbb{P}(S' \le k)$;*
- *pri $H_1^\pm$: $\theta \ne \theta_0$ je $\rho(k) = 2\min\{\mathbb{P}(S' \le k), \mathbb{P}(S' \ge k)\}$.*

Tu je $S' \sim \mathrm{Bin}(n, \theta_0)$.

4. Uporabite zgornjo različico preizkusa na podatkih iz prejšnje naloge. Izvedite preizkus še proti alternativni domnevi, da so izjemni študijski dosežki pristranski v korist moških. Komentirajte!

5. Prireditelj neke igre na srečo navaja, da je verjetnost, da bo v posamezni igri izžreban dobitek, enaka 1/50. To želimo preizkusiti na osnovi opažanja, po kolikšnem številu iger je dobitek prvič izžreban. Natančno opišite izvedbo ustreznega randomiziranega preizkusa za stopnjo značilnosti $\alpha = 0{\cdot}1$. Alternativna domneva naj bo, da je verjetnost, da bo dobitek izžreban, manjša od 1/50.

Neyman[50]–Pearsonova[51] lema

Če ima parameter le dve možni vrednosti (tj. statistični model obsega le dve verjetnostni meri), je preizkus na podlagi razmerja verjetij najmočnejši.

6. Življenjska doba originalne žarnice je porazdeljena eksponentno s pričakovano vrednostjo 500 ur, življenjska doba ponaredka pa je porazdeljena eksponentno s pričakovano vrednostjo 100 ur. Na podlagi opažene življenjske dobe ene žarnice preizkušamo ničelno domnevo, da je originalna, proti alternativni domnevi, da je ponaredek. Konstruirajte najmočnejši preizkus pri stopnji značilnosti $\alpha = 0{\cdot}05$. Kolikšna je njegova moč?

[50] Jerzy Sp lawa-Neyman (1894–1891), poljski matematik in statistik

[51] Karl Pearson (1857–1936), angleški matematik in biostatistik

7. Tine je zgeneriral 100 slučajnih števil, ki so neodvisna in porazdeljena normalno $\mathrm{N}(0,1)$, Tone pa je zgeneriral 10.000 takih števil. Oba povesta povprečje števil, ki sta jih dobila.

   a) Zapišite porazdelitev Tinetovega in Tonetovega povprečja.

   b) Zapomnimo si enega izmed povprečij, ne pa tudi, čigavo je. Konstruirajte najmočnejši preizkus, ki na podlagi tega opažanja pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusi ničelno domnevo, da je opaženo povprečje Tonetovo, proti alternativni domnevi, da je Tinetovo. Kolikšna je njegova moč?

Asimptotično obnašanje preizkusa na podlagi razmerja verjetij (Wilksov[52] izrek)

Denimo, da opazimo n vrednosti, ki so nastale kot neodvisne realizacije statistične spremenljivke X s porazdelitvijo iz eksponentne družine, parametrizirane z naborom naravnih parametrov $\gamma_1, \ldots, \gamma_m$, ki preteče odprto množico $\Theta \subseteq \mathbb{R}^m$. Preizkušamo ničelno domnevo:

$$H_0\colon\ (\gamma_1, \ldots, \gamma_m) \in \Theta_{H_0}\,,$$

kjer je $\Theta_{H_0} \subseteq \Theta$ vložena podmnogoterost razreda $\mathcal{C}^{(2)}$ (glej spodaj). Z drugimi besedami, preizkušamo **ožji model** *Θ_{H_0} proti* **širšemu modelu** *Θ. Pri veljavnosti H_0 naj imajo komponente $h_1(X), \ldots, h_m(X)$ pripadajoče zadostne statistike končne druge momente in neizrojeno kovariančno matriko. Označimo z Λ razmerje verjetij. Tedaj pri H_0 velja:*

$$-2\ln\Lambda \xrightarrow[n\to\infty]{\mathrm{d}} \chi^2(\dim\Theta - \dim\Theta_{H_0})\,.$$

Podmnogoterost razreda $\mathcal{C}^{(s)}$ lahko definiramo na vsaj dva ekvivalentna načina.

Prva definicija. Množica $\Theta_{H_0} \subseteq \Theta$ je vložena podmnogoterost razreda $\mathcal{C}^{(k)}$ dimenzije $m - s$ $(s = 0, 1, \ldots, m)$, če obstaja tako odprto pokritje množice Θ_{H_0}, da je v okviru vsakega elementa tega pokritja množica Θ_{H_0} določena z enačbami $q_1 = z_1, \ldots, q_s = z_s$ za primerne karakteristike $q_1, \ldots, q_s$ $(s \le m)$, ki so funkcije naravnih parametrov $\gamma_1, \ldots, \gamma_m$. Zahtevamo še, da je vektorska funkcija, ki naravnim parametrom priredi vektor karakteristik, k-krat parcialno zvezno odvedljiva in da ima matrika njenih prvih parcialnih odvodov maksimalni rang.

Druga definicija. Množica $\Theta_{H_0} \subseteq \Theta$ je vložena podmnogoterost razreda $\mathcal{C}^{(s)}$ dimenzije r $(r = 0, 1, \ldots, m)$, če obstaja tako odprto pokritje množice Θ_{H_0}, da je v okviru vsakega elementa tega pokritja množica Θ_{H_0} slika vektorske funkcije $g\colon V \to \mathbb{R}^m$, kjer je $V \subseteq \mathbb{R}^r$ odprta množica, g pa k-krat parcialno zvezno odvedljiva topološka vložitev, pri kateri ima matrika prvih parcialnih odvodov povsod maksimalni rang. To torej pomeni, da lahko lokalno Θ_{H_0} primerno parametriziramo z r parametri.

8. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Na voljo imamo vzorec $X_1, X_2, \ldots, X_n$, kjer so vse spremenljivke v vzorcu neodvisne in imajo predpisano porazdelitev.

   Opazimo $n = 100$, $\sum_{i=1}^n X_i = 37$, $\sum_{i=1}^n X_i^2 = 75$. Pri stopnji značilnosti $\alpha = 0{\cdot}01$ na podlagi razmerja verjetij preizkusite domnevo, da je $\mu = \sigma^2 \neq 0$, proti alternativni domnevi, da to ni res.

[52] Samuel Stanley Wilks (1906–1964), ameriški matematik

Neznana verjetnost pri veliko poskusih

Naj bo θ verjetnost, da se zgodi določen dogodek uspeha. Preizkušamo ničelno domnevo H_0, ki pravi, da je $\theta = \theta_0$. Izvedemo n poskusov in pri dani dogodek se zgodi pri natanko S poskusih. Tokrat privzamemo, da je število poskusov dovolj veliko: za minimalno sprejemljivo natančnost preizkusa je potrebno, da je $n\theta_0 \geq 5$ in $n(1-\theta_0) \geq 5$. Glede na alternativno domnevo H_1 ničelno domnevo zavrnemo:

- *pri $H_1^{\pm}$: $\theta \neq \theta_0$: če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\left|\dfrac{S}{n} - \theta_0\right| - \dfrac{1}{2n}\right) \geq z_{\alpha/2}$;
- *pri H_1^{+}: $\theta > \theta_0$: če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\dfrac{S}{n} - \theta_0 - \dfrac{1}{2n}\right) \geq z_{\alpha}$;
- *pri H_1^{-}: $\theta < \theta_0$: če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\dfrac{S}{n} - \theta_0 + \dfrac{1}{2n}\right) \leq -z_{\alpha}$.

Z $z_p = \Phi^{-1}\left(\frac{1}{2} - p\right)$ smo označili kvantil standardne normalne porazdelitve za verjetnost $1-p$.

Z-preizkus

Z-preizkus temelji na standardni normalni porazdelitvi. Natančneje, Z-preizkus na preizkusni statistiki Z s popravkom δ ima tri različice:

- **Dvostranska različica** *zavrne ničelno domnevo, če je* $|Z| - \delta \geq z_{\alpha/2}$;
- **Enostranska različica v desno** *zavrne ničelno domnevo, če je* $Z - \delta \geq z_{\alpha}$;
- **Enostranska različica v levo** *zavrne ničelno domnevo, če je* $Z + \delta \leq -z_{\alpha}$.

9. Tovarna jamči, da je delež drugorazrednih izdelkov enak 20%.

   a) V vzorcu 100 izdelkov jih je 24 drugorazrednih. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da je verjetnost, da je izdelek drugorazreden, enaka $0{\cdot}2$, proti alternativni domnevi, da je večja od $0{\cdot}2$.

   b) Kaj pa, če bi bilo drugorazrednih 72 izdelkov izmed 300?

   c) Kaj, če bi v slednjem primeru vzeli stopnjo značilnosti $0{\cdot}01$?

10. 10000-krat vržemo kovanec in 5090-krat je padel grb. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da je kovanec pošten, proti alternativni domnevi, da ni pošten. Kaj pa, če bi za alternativno domnevo vzeli, da grb pade z večjo verjetnostjo kot cifra?

V vseh nadaljnjih nalogah tega razdelka imamo za vsako statistično spremenljivko X, definirano na populaciji, na voljo vzorec, na katerem ima ustrezna spremenljivka vrednosti $X_1, X_2, \ldots, X_n$, pri čemer so vrednosti neodvisne in porazdeljene enako kot X.

Sredina pri normalni porazdelitvi z znanim standardnim odklonom

Opazimo vrednosti $X_1, X_2, \ldots, X_n$, ki so nastale kot neodvisne realizacije statistične spremenljivke X, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, kjer je σ znan. Preizkušamo ničelno domnevo $H_0\colon\ \mu = \mu_0$. Ker pri H_0 velja:

$$Z := \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \sim \mathrm{N}(0, 1)\,,$$

izvedemo Z-preizkus brez popravka na preizkusni statistiki Z, in sicer:

- **dvostransko različico**, *če H_1 trdi, da je $\mu \neq \mu_0$;*
- **enostransko različico v desno**, *če H_1 trdi, da je $\mu > \mu_0$;*
- **enostransko različico v levo**, *če H_1 trdi, da je $\mu < \mu_0$;*

11. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, 5)$, dajo naslednje vrednosti:

    101, 91, 93, 103, 91, 101, 103, 95, 95

    Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite ničelno domnevo, da je $\mu = 100$, proti alternativni domnevi, da je $\mu \neq 100$. Kaj pa, če bi za alternativno domnevo vzeli $\mu < 100$ ali $\mu > 100$?

12. Podobno kot v prejšnji nalogi na podlagi 9 neodvisnih meritev količine, porazdeljene normalno $\mathrm{N}(\mu, 5)$, preizkušamo ničelno domnevo, da je $\mu = 100$, proti alternativni, da je $\mu \neq 100$. Izračunajte moč preizkusa pri $\mu = 99{\cdot}9$, $\mu = 97$ in $\mu = 90$.

Sredina pri normalni porazdelitvi
z neznanim standardnim odklonom

Spet opazimo vrednosti $X_1, X_2, \ldots, X_n$, ki so nastale kot neodvisne realizacije statistične spremenljivke X, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, le da tokrat σ ni znan. Preizkušamo ničelno domnevo $H_0\colon\ \mu = \mu_0$. Ker pri H_0 velja:

$$T := \frac{\bar{X} - \mu_0}{s} \sqrt{n} \sim \mathrm{Student}(n-1)\,,$$

kjer je s definiran tako kot pri konstrukciji intervala zaupanja za ta primer, ničelno domnevo zavrnemo:

- *proti $H_1^{\pm}\colon\ \mu \neq \mu_0$, če je $|T| \geq t_{\alpha/2}(n-1)$;*
- *proti $H_1^{+}\colon\ \mu > \mu_0$, če je $T \geq t_{\alpha}(n-1)$;*
- *proti $H_1^{-}\colon\ \mu < \mu_0$, če je $T \leq -t_{\alpha}(n-1)$.*

Tu je $t_p(df)$ kvantil Studentove[53] porazdelitve z df prostostnimi stopnjami za verjetnost $1-p$.

[53] William Sealy Gosset (1876–1937), angleški statistik, bolj znan pod psevdonimom *Student*

T-preizkus

Prej opisani preizkus je primer T-preizkusa. Ta temelji na Studentovi porazdelitvi. Natančneje, T-preizkus na preizkusni statistiki T z df prostostnimi stopnjami ima tri različice:

- **Dvostranska različica** *zavrne ničelno domnevo, če je* $|T| \geq t_{\alpha/2}(df)$;
- **Enostranska različica v desno** *zavrne ničelno domnevo, če je* $T \geq t_{\alpha}(df)$;
- **Enostranska različica v levo** *zavrne ničelno domnevo, če je* $T \leq -t_{\alpha}(df)$.

Tu je $t_p(df)$ kvantil Studentove porazdelitve z df prostostnimi stopnjami za verjetnost $1-p$.

13. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$109,\ 107,\ 95,\ 121,\ 107,\ 97,\ 111,\ 121,\ 95$$

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite ničelno domnevo, da je $\mu = 100$, proti alternativni domnevi, da je $\mu \neq 100$. Kaj pa, če bi vedeli, da je $\sigma = 10$?

14. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$46,\ 51,\ 48,\ 46,\ 52,\ 47,\ 51,\ 44,\ 47,\ 48$$

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite ničelno domnevo, da je $\mu = 50$, proti alternativni domnevi, da je $\mu \neq 50$. Kaj pa, če bi za alternativno domnevo vzeli, da je $\mu < 50$?

Preizkušanje enakosti sredin za parne vzorce

Večkrat hkrati izmerimo dve statistični spremenljivki. Opazimo torej pare $(X_1, Y_1), \ldots, (X_n, Y_n)$, za katere privzamemo, da so neodvisni in da je $(X_i, Y_i) \sim \mathrm{N}(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$.

Preizkušamo ničelno domnevo $H_0\colon\ \mu_X = \mu_Y$, alternativna domneva pa je lahko:

- $H_1^{\pm}\colon\ \mu_X \neq \mu_Y$;
- $H_1^{X}\colon\ \mu_X > \mu_Y$;
- $H_1^{Y}\colon\ \mu_Y > \mu_X$.

Preizkus izvedemo tako, da ustrezno preizkusimo sredino razlike $X - Y$ glede na vrednost 0.

15. Na desetih osebah so preizkušali učinek neke diete proti debelosti. Osebe so stehtali pred začetkom in po koncu diete. Podatki so naslednji:

Pred dieto	125	131	126	117	114
Po dieti	121	118	119	121	113

Pred dieto	134	123	135	100	117
Po dieti	118	111	130	97	118

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite ničelno domnevo, da dieta nima učinka, proti alternativni domnevi, da ima shujševalni učinek. Privzeti smete, da je vektor telesne teže pred in po dieti porazdeljen dvorazsežno normalno.

Preizkušanje enakosti sredin za neodvisne vzorce (Welchev[54] preizkus)

Opazimo vrednosti $X_1, X_2, \ldots, X_m$, ki so nastale kot realizacije statistične spremenljivke X, porazdeljene normalno $\mathrm{N}(\mu_X, \sigma_X)$, in vrednosti $Y_1, Y_2, \ldots, Y_n$, ki so nastale kot realizacije statistične spremenljivke Y, porazdeljene normalno $\mathrm{N}(\mu_Y, \sigma_Y)$. Za vse realizacije privzamemo, da so neodvisne. Če tovrstno opažanje temelji na jemanju vzorcev, sta lahko X in Y definirani na različnih populacijah.
Preizkušamo domnevo $H_0\colon\ \mu_X = \mu_Y$. Ker pri H_0 približno velja:

$$T := \frac{\bar{X} - \bar{Y}}{\widehat{\mathrm{SE}}} \sim \mathrm{Student}(df)\,,$$

kjer je:

$$\begin{aligned}
\widehat{\mathrm{SE}} &= \sqrt{\frac{s_X^2}{m} + \frac{s_Y^2}{n}}\,,\\
s_X^2 &= \frac{(X_1 - \bar{X})^2 + \cdots + (X_m - \bar{X})^2}{m-1}\,,\\
s_Y^2 &= \frac{(Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2}{n-1}\,,\\
df &= \frac{\widehat{\mathrm{SE}}^4}{\frac{s_X^4}{m^2(m-1)} + \frac{s_Y^4}{n^2(n-1)}}\,,
\end{aligned}$$

izvedemo T-preizkus na preizkusni statistiki T z df prostostnimi stopnjami, in sicer:

- **dvostransko različico**, *če H_1 trdi, da je $\mu_X \neq \mu_Y$;*
- **enostransko različico v desno**, *če H_1 trdi, da je $\mu_X > \mu_Y$;*
- **enostransko različico v levo**, *če H_1 trdi, da je $\mu_X < \mu_Y$.*

16. V spodnji tabeli so prikazane porodne teže deklic in dečkov, ki so se rodili 18. decembra 1997 v porodnišnici *Mater Mothers' Hospital* v Brisbanu v Avstraliji:[55]

[54] Bernard Lewis Welch (1911–1989), angleški statistik

[55] Vir: `http://jse.amstat.org/jse_data_archive.htm`, podatki `babyboom`, ogled 13. 6. 2021.

Deklice:	3837, 3334, 2208, 1745, 2576, 3208, 3746, 3523.
Dečki:	3554, 3838, 3625, 2846, 3166, 3520, 3380, 3294, 3521, 2902.

Pri stopnji tveganja 0˙05 preizkusite domnevo, da so deklice in dečki v povprečju enako težki, proti alternativni domnevi, ki to zanika.

17. Vzorčne vrednosti statistične spremenljivke $X \sim \mathrm{N}(\mu_X, \sigma_X)$ iz prve populacije pridejo:

$$102,\ 96,\ 103,\ 98,\ 105,\ 97,\ 103,\ 98,\ 100,\ 98,\ 99,\ 101$$

vzorčne vrednosti statistične spremenljivke $Y \sim \mathrm{N}(\mu_Y, \sigma_Y)$ iz druge populacije pa pridejo:

$$95,\ 97,\ 95,\ 99,\ 95,\ 95$$

Pri stopnji značilnosti $\alpha = 0{\dot{}}01$ preizkusite domnevo, da je $\mu_X = \mu_Y$, proti alternativni domnevi, da je $\mu_X > \mu_Y$.

Enakost sredin več normalnih statističnih spremenljivk: analiza variance (ANOVA) z enojno klasifikacijo

Danih je k populacij, na vsaki je definirana statistična spremenljivka, naj bodo to $X_1 \sim \mathrm{N}(\mu_1, \sigma)$, ..., $X_k \sim \mathrm{N}(\mu_k, \sigma)$. Iz vsake populacije vzamemo vzorec, pri čemer so vse enote vzorcev med seboj neodvisne. Vrednosti na vzorcu iz i-te populacije označimo z $X_{i1}, \ldots, X_{in_i}$. Preizkušamo domnevo $H_0\colon \mu_1 = \mu_2 = \ldots = \mu_k$, alternativna domneva H_1 pa je nasprotje H_0. Izračunajmo:

$$\bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}\,, \quad n := \sum_{i=1}^{k} n_i\,, \quad \bar{X} := \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij} = \sum_{i=1}^{k} \frac{n_i}{n} \bar{X}_i\,,$$

$$S_{\mathrm{B}}^2 := \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2\,, \quad S_{\mathrm{W}}^2 := \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2\,.$$

Če velja H_0, sta S_{B}^2 in S_{W}^2 neodvisni ter $S_{\mathrm{B}}^2/\sigma^2 \sim \chi^2(k-1)$ in $S_{\mathrm{W}}^2/\sigma^2 \sim \chi^2(n-k)$, zato je:

$$F := \frac{S_{\mathrm{B}}^2/(k-1)}{S_{\mathrm{W}}^2/(n-k)} \sim \mathrm{F}(k-1, n-k)$$

kjer je $\mathrm{F}(k-1, n-k)$ Snedecorjeva (Fisherjeva)[56,57] porazdelitev. V skladu s tem ničelno domnevo zavrnemo, če je $F \geq F_\alpha(k-1, n-k)$.

[56] George Waddel Snedecor (1882–1974), ameriški matematik in statistik

[57] Sir Ronald Aymler Fisher (1899–1962), angleški statistik in biolog

F-preizkus

Prej opisani preizkus je primer F-preizkusa. Ta temelji na Snedecorjevi porazdelitvi. Natančneje, F-preizkus na preizkusni statistiki F z (df_1, df_2) prostostnimi stopnjami ima tri različice:

- **Dvostranska različica**: *H_0 zavrnemo, če je $F \le F_{\alpha/2}(df_1, df_2)$ ali $F \ge F_{\alpha/2}(df_1, df_2)$;*
- **Enostranska različica v desno**: *H_0 zavrnemo, če je $F \ge F_\alpha(df_1, df_2)$;*
- **Enostranska različica v levo**: *H_0 zavrnemo, če je $F \le F_{1-\alpha}(df_1, df_2)$.*

Tu je $F_p(df_1, df_2)$ kvantil Snedecorjeve porazdelitve z (df_1, df_2) prostostnimi stopnjami za verjetnost $1 - p$. Velja $F_{1-p}(df_1, df_2) = F_p(df_2, df_1)$.

18. Pacientom, ki so jim dajali določena zdravila, so merili neki parameter. Meritve so dale naslednje vrednosti:

Aspirin:	3, 5, 3, 5
Tilenol:	2, 2, 4, 4
Placebo:	2, 1, 2

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da je vrednost parametra neodvisna od tega, ali pacient jemlje katero izmed obeh zdravil ali pa sploh nobenega.

Standardni odklon pri normalni porazdelitvi

Opazimo vrednosti $X_1, X_2, \dots, X_m$, ki so nastale kot realizacije statistične spremenljivke X, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$. Preizkušamo ničelno domnevo $H_0\colon\ \sigma = \sigma_0$. Ker pri H_0 velja:

$$\chi^2 := (n-1)\,\frac{s^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$$

ničelno domnevo zavrnemo:

- *proti $H_1^{\pm}\colon\ \sigma \ne \sigma_0$, če je $\chi^2 \le \chi^2_{1-\alpha/2}(n-1)$ ali $\chi^2 \ge \chi^2_{\alpha/2}(n-1)$;*
- *proti $H_1^{+}\colon\ \sigma > \sigma_0$, če je $\chi^2 \ge \chi^2_{\alpha}(n-1)$;*
- *proti $H_1^{-}\colon\ \sigma < \sigma_0$, če je $\chi^2 \le \chi^2_{1-\alpha}(n-1)$.*

Tu je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost $1 - p$.

Preizkus hi kvadrat

Prej opisani preizkus je primer preizkusa hi kvadrat. Ta temelji na porazdelitvi hi kvadrat. Natančneje, preizkus hi kvadrat na preizkusni statistiki χ^2 z df prostostnimi stopnjami ima tri različice:

- **Enostranska različica v desno** *zavrne ničelno domnevo, če je* $\chi^2 \geq \chi^2_\alpha(df)$;
- **Enostranska različica v levo** *zavrne ničelno domnevo, če je* $\chi^2 \leq \chi^2_{1-\alpha}(df)$;
- **Dvostranska različica** *zavrne ničelno domnevo, če je* $\chi^2 \leq \chi^2_{1-\alpha/2}(df)$ *ali* $\chi^2 \geq \chi^2_{\alpha/2}(df)$.

Tu je $\chi^2_p(df)$ *kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost* $1-p$.

19. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$99,\ 90,\ 108,\ 111,\ 97,\ 93,\ 90,\ 106,\ 104,\ 102$$

Pri $\alpha = 0{\cdot}05$ preizkusite:

a) ničelno domnevo, da je $\sigma = 5$, proti alternativni domnevi, da je $\sigma \neq 5$;

b) ničelno domnevo, da je $\sigma = 10$, proti alternativni domnevi, da je $\sigma < 10$.

Pearsonov[58] preizkus skladnosti s fiksno porazdelitvijo

Preizkušamo, ali je porazdelitev dane statistične spremenljivke enaka:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & \cdots & p_r \end{pmatrix} \qquad (r \geq 3)\,.$$

Pri tem so lahko $a_1, \ldots, a_r$ dejanske vrednosti spremenljivke ali pa le razredi, v katere pade. Opazimo n neodvisnih realizacij te porazdelitve. Iz opaženih vrednosti/razredov naredimo frekvenčno porazdelitev:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ N_1 & N_2 & \cdots & N_r \end{pmatrix}$$

Izračunamo **pričakovane frekvence** $\tilde{N}_i := np_i$*. Ker tedaj pri veljavnosti ničelne domneve približno velja:*

$$\chi^2 := \sum_{i=1}^{r} \frac{(N_i - \tilde{N}_i)^2}{\tilde{N}_i} \sim \chi^2(r-1)$$

domnevo o porazdelitvi preizkusimo s preizkusom hi kvadrat na preizkusni statistiki χ^2 z $r-1$ prostostnimi stopnjami, in sicer enostransko v desno.
Za naše potrebe je preizkus dovolj natančen, če je $r \geq 3$ in $\tilde{N}_i \geq 5$ za vse i. Če dobimo $\tilde{N}_i < 5$, lahko razrede združimo. Za $r = 2$ pa lahko uporabimo kar dvostranski preizkus uspeha poskusa.

20. Pri kvizu Lepo je biti milijonar od 22. novembra do 28. decembra 2003 je bil 21-krat pravilen odgovor A, 42-krat B, 77-krat C in 116-krat D. Pri stopnji značilnosti $\alpha = 0{\cdot}01$ preizkusite ničelno domnevo, da so odgovori enakomerno porazdeljeni, proti alternativni domnevi, da niso.

    Če izvzamemo prvih pet vprašanj, je bil A pravilen 21-krat, B 37-krat, C 53-krat in D 25-krat. Naredite isti preizkus.

21. V vzorcu so 2 osebka tipa RR, 5 tipa Rr in 4 tipa rr. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da je v populaciji 25% osebkov tipa RR, 50% tipa Rr in 25% tipa rr. Kaj pa, če bi bilo v vzorcu 20 osebkov tipa RR, 50 tipa Rr in 50 tipa rr?

22. V spodnji tabeli so za volitve v Državni zbor dne 24. 4. 2022 prikazani rezultati predvolilne ankete agencije Parsifal v dneh od 20. do 22. aprila ($n = 724$), vzporednih volitev, ki jih je izvedla agencija Mediana ($n = 15216$) in dokončni rezultati. Prikazani so samo deleži tistih strank, ki so prišle v Državni zbor:

[58] Karl Pearson (1857–1936), angleški matematik in biostatistik

	GS	SDS	NSi	SD	Levica
Anketa	28˙0%	26˙5%	7˙3%	8˙1%	6˙2%
Vzporedne volitve	35˙8%	22˙5%	6˙8%	6˙6%	4˙4%
Dokončni rezultati	34˙45%	23˙48%	6˙86%	6˙69%	4˙46%

Pri stopnjah značilnosti $\alpha = 0˙05$ in $\alpha = 0˙01$ preizkusite domnevo, da so bili volivci na anketi oz. vzporednih volitvah opredeljeni enako kot na pravih volitvah.

23. Pri stopnji značilnosti $\alpha = 0˙05$ preizkusite domnevo, da ima statistična spremenljivka X *standardno* Laplaceovo porazdelitev, tj. zvezno z gostoto:

$$f_X(x) = \frac{1}{2}\, e^{-|x|}\,,$$

če ima vzorec naslednjo frekvenčno porazdelitev:

pod -1	od -1 do 0	od 0 do 1	nad 1
32	20	21	27

24. Podatki imajo naslednjo frekvenčno porazdelitev:

pod 10	10–20	20–30	30–40	nad 40
5	20	35	25	15

Pri stopnji značilnosti $\alpha = 0˙05$ preizkusite domnevo, da podatki izhajajo iz populacije, porazdeljene normalno $\mathrm{N}(25, 10)$.

Preizkus skladnosti z družino porazdelitev Pearsonovega tipa

Preizkušamo, ali porazdelitev dane statistične spremenljivke pripada družini:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1(\theta_1, \ldots, \theta_k) & p_2(\theta_1, \ldots, \theta_k) & \cdots & p_r(\theta_1, \ldots, \theta_k) \end{pmatrix} \qquad (r \geq 3)\,,$$

kjer nabor $(\theta_1, \ldots, \theta_k)$ *preteče odprto množico v* $\mathbb{R}^k$*, vektorska funkcija* $(p_1, \ldots, p_r)$ *pa je dvakrat zvezno odvedljiva vložitev. Spet opazimo* n *neodvisnih realizacij te porazdelitve in naredimo frekvenčno porazdelitev:*

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ N_1 & N_2 & \cdots & N_r \end{pmatrix}$$

Izračunamo **pričakovane frekvence**
$\tilde{N}_i := np_i(\hat{\theta}_1, \ldots, \hat{\theta}_k)$*, kjer so* $\hat{\theta}_1, \ldots, \hat{\theta}_k$ *ocene za* $\theta_1, \ldots, \theta_k$ *po metodi največjega verjetja. Tedaj pri veljavnosti ničelne domneve približno velja:*

$$\chi^2 := \sum_{i=1}^{r} \frac{(N_i - \tilde{N}_i)^2}{\tilde{N}_i} \sim \chi^2(r - 1 - k)\,.$$

Torej izvedemo preizkus hi kvadrat z $df = r - 1 - k$*, in sicer enostransko različico v desno. Preizkus je dovolj natančen, če je* $\tilde{N}_i \geq 5$ *za vse* i*.*

25. V enostavnem slučajnem vzorcu, zajetem iz velike populacije, je 20 osebkov tipa RR, 30 tipa Rr in 50 tipa rr.

    a) S Pearsonovim preizkusom skladnosti pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da populacija ustreza Hardy[59]–Weinbergovemu[60] modelu, tj. da so deleži osebkov tipa RR, Rr in rr enaki $(1-\theta)^2$, $2\theta(1-\theta)$ in θ^2 (glej tudi 16. nalogo iz 11. razdelka).

    b) Zavedati se je treba, da je Pearsonov preizkus skladnosti zgolj približen – asimptotično natančen pri velikih vzorcih. Alternativa temu preizkusu je lahko preizkus na podlagi Wilksovega izreka (ob dodatni predpostavki, da je vsaka vrednost zajeta s strogo pozitivno verjetnostjo). Na danih podatkih uporabite še ta preizkus in primerjajte!

26. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da je statistična spremenljivka X porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 1 & 2 \\ 1 - 2a - b & a & b & a \end{pmatrix},$$

[59] Godfrey Harold Hardy (1877–1947), angleški matematik
[60] Wilhelm Weinberg (1862–1937), nemški splošni zdravnik, ginekolog, porodničar, raziskovalec dednosti, statistik in genealog

če je v vzorcu 20 enot z vrednostjo -1, 30 enot z vrednostjo 0, 30 enot z vrednostjo 1 in 20 enot z vrednostjo 2.

27. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da ima statistična spremenljivka X *splošno usredinjeno* Laplaceovo porazdelitev, tj. zvezno z gostoto:

$$f_X(x) = \frac{\lambda}{2}\, e^{-\lambda|x|}\,; \quad \lambda > 0\,,$$

če ima vzorec isto frekvenčno porazdelitev kot v 23. nalogi.

Preizkus z znaki

Večkrat hkrati izmerimo urejenostni statistični spremenljivki. Natančneje, opazimo pare $(X_1, Y_1), \ldots, (X_n, Y_n)$, ki so neodvisni in enako porazdeljeni. Predstavljajmo si, da so prisotni vplivi, ki večajo X na račun Y, in vplivi, ki večajo Y na račun X.

Ničelna domneva H_0 pravi, da so ti vplivi **uravnovešeni***:*
$\mathbb{P}(X > Y) = \mathbb{P}(Y > X)$. *Alternativna domneva pa je lahko:*

- H_1^X*, da vplivi, ki večajo X na račun Y, prevladujejo nad vplivi, ki delajo nasprotno:* $\mathbb{P}(X > Y) > \mathbb{P}(Y > X)$.
- H_1^Y*, da vplivi, ki večajo X na račun Y, prevladujejo nad vplivi, ki delajo nasprotno:* $\mathbb{P}(Y > X) > \mathbb{P}(X > Y)$.
- $H_1^{\pm}$*, da velja bodisi H_1^X bodisi H_1^Y.*

Naj bo:

- S_X *označimo število meritev, pri katerih je $X > Y$;*
- S_Y *število meritev, pri katerih je $Y > X$;*
- $\tilde{n} = S_X + S_Y$ *število meritev, pri katerih je $X \neq Y$ (primere, kjer pride enako, torej preprosto izločimo).*

Ničelno domnevo zavrnemo, če je $p(S_X, S_Y) \le \alpha$. Funkcija p (p-vrednost) je odvisna od ničelne domneve in ustreza preizkusu uspeha poskusa pri ničelni domnevi, da je le-ta enaka $1/2$:

- *pri H_1^X postavimo* $p(k,l) = \mathbb{P}(S' \ge k) = \mathbb{P}(S' \le l)$;
- *pri H_1^Y postavimo* $p(k,l) = \mathbb{P}(S' \ge l) = \mathbb{P}(S' \le k)$;
- *pri $H_1^{\pm}$ postavimo* $p(k,l) = 2\min\{\mathbb{P}(S' \ge k), \mathbb{P}(S' \ge l)\} =$
  $= 2\min\{\mathbb{P}(S' \le k), \mathbb{P}(S' \le l)\}$.

Tu je $S' \sim \mathrm{Bin}(\tilde{n}, 1/2)$.

28. Meritve količin X in Y so zapisane v naslednji tabeli:

X_i	28	14	16	16	31	17	13	14	12	13
Y_i	26	29	31	18	37	10	19	33	23	45

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da sta dogodka $\{X > Y\}$ in $\{Y > X\}$ enako verjetna, proti alternativni domnevi, da je dogodek $\{Y > X\}$ verjetnejši. Nato naredite ustrezni enostranski preizkus povprečij.

29. Meritve količin X in Y so zapisane v naslednji tabeli:

X_i	25	28	30	23	28	26	29	23	33	21	33	28
Y_i	35	27	29	21	18	25	28	27	31	19	32	26

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da sta dogodka $\{X > Y\}$ in $\{Y > X\}$ enako verjetna, proti alternativni domnevi, da je dogodek $\{X > Y\}$ verjetnejši. Nato naredite ustrezni enostranski preizkus povprečij.

30. Meritve količin X in Y so zapisane v naslednji tabeli:

X_i	30	24	22	28	26	19	25	31	36	21	25	26	29	29	19	18
Y_i	28	121	21	25	25	17	122	129	34	20	22	23	126	26	18	17

a) Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da sta dogodka $\{X > Y\}$ in $Y > X$ enako verjetna, proti alternativni domnevi, da sta njuni verjetnosti različni.

b) Pri isti stopnji značilnosti preizkusite ničelno domnevo iz prejšnje točke še proti obema enostranskima domnevama.

c) Naredite še ustrezna enostranska preizkusa povprečij.

Preizkus z znaki za veliko število meritev

Če je $\tilde{n} = S_X + S_Y \geq 10$, lahko preizkus z znaki nadomestimo s približnim preizkusom, pri katerem ničelno domnevo zavrnemo:

- *proti H_1^X, če je* $\dfrac{S_X - S_Y - 1}{\sqrt{\tilde{n}}} \geq z_\alpha$;
- *proti H_1^Y, če je* $\dfrac{S_X - S_Y + 1}{\sqrt{\tilde{n}}} \leq -z_\alpha$.
- *proti $H_1^\pm$, če je* $\dfrac{|S_X - S_Y| - 1}{\sqrt{\tilde{n}}} \geq z_{\alpha/2}$;

Z drugimi besedami, izvedemo Z-preizkus na preizkusni statistiki $\dfrac{S_X - S_Y}{\sqrt{\tilde{n}}}$ *s popravkom* $\dfrac{1}{\sqrt{\tilde{n}}}$.

31. 50 ljudi so pred ogledom in po ogledu filma povpraševali, kako se počutijo: zelo slabo, slabo, srednje, dobro ali zelo dobro. Rezultati so naslednji:

pred	po
srednje	srednje
dobro	zelo dobro
srednje	zelo dobro
dobro	srednje
srednje	zelo dobro
dobro	dobro
srednje	dobro
dobro	dobro
dobro	zelo dobro
zelo dobro	zelo dobro
dobro	zelo dobro
zelo dobro	dobro
dobro	srednje
zelo dobro	srednje
srednje	dobro
srednje	dobro
dobro	zelo dobro
srednje	dobro
dobro	zelo dobro
zelo dobro	dobro
zelo dobro	zelo dobro
zelo dobro	dobro
slabo	dobro
dobro	srednje
srednje	zelo dobro

pred	po
dobro	zelo dobro
dobro	dobro
zelo dobro	zelo dobro
dobro	dobro
srednje	zelo slabo
srednje	zelo dobro
zelo dobro	srednje
dobro	dobro
dobro	dobro
srednje	slabo
slabo	srednje
srednje	srednje
zelo slabo	slabo
slabo	srednje
slabo	srednje
slabo	zelo dobro
zelo slabo	srednje
srednje	slabo
srednje	slabo
zelo slabo	srednje
srednje	dobro
slabo	zelo dobro
slabo	slabo
slabo	slabo
zelo slabo	srednje

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite ničelno domnevo, da ogled filma ne spremeni počutja, proti alternativni domnevi, da ga spremeni.

32. V menzi so zamenjali kuharja. Izbranih 50 odjemalcev so pred menjavo in po menjavi vprašali, kako so zadovoljni s kakovostjo menze. Anketiranci so lahko izbirali med naslednjimi možnostmi:

    1: zelo nezadovoljen
    2: nezadovoljen
    3: niti zadovoljen niti nezadovoljen
    4: zadovoljen
    5: zelo zadovoljen

    Rezultati obeh anket so naslednji:

prej\potem	1	2	3	4	5
1	4	11	1	0	0
2	1	3	9	2	0
3	1	0	5	6	0
4	0	0	1	3	1
5	0	0	0	1	1

Pri stopnji značilnosti $\alpha = 0\text{'}01$ preizkusite domnevo, da se kakovost hrane ni spremenila, proti alternativni domnevi, da se je spremenila.

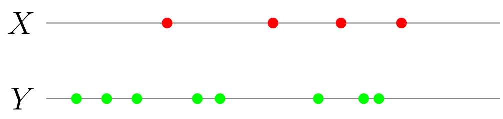
Stohastična primerjava porazdelitev

Porazdelitev slučajne spremenljivke X je **stohastično večja** *od porazdelitve slučajne spremenljivke Y, če je $F_X \le F_Y$.*

Porazdelitev slučajne spremenljivke X je **stohastično manjša** *od porazdelitve slučajne spremenljivke Y, če je $F_X \ge F_Y$.*

Porazdelitev slučajne spremenljivke X je stohastično **strogo** *večja (manjša) od porazdelitve slučajne spremenljivke Y, če je stohastično večja (manjša) in porazdelitvi nista enaki.*

33. Slučajni spremenljivki X in Y sta porazdeljeni enakomerno na točkah, kot je označeno spodaj:



Narišite grafa kumulativnih porazdelitvenih funkcij in opredelite, ali je katera od porazdelitev stohastično večja od druge.

Inverzijski (Wilcoxon[61]–Mann[62]–Whitneyjev[63]) preizkus

Preizkušamo, ali sta **urejenostni** *statistični spremenljivki X in Y enako porazdeljeni. Pri statistični spremenljivki X opazimo $X_1, \ldots, X_m$, pri Y pa $Y_1, \ldots, Y_n$. Privzamemo, da so vse opažene vrednosti med seboj neodvisne. Če opažanje temelji na jemanju vzorcev, sta lahko X in Y definirani na različnih populacijah.*

Opažene vrednosti združimo in jih uredimo po velikosti, recimo od namanjše do največje. Naj bodo $R_1^X, \ldots, R_m^X$ mesta (rangi), ki pripadajo opažanjem spremenljivke X. Privzemimo, da sta vzorca dovolj velika.

Tudi tu ločimo enostransko in dvostransko različico preizkusa. Obravnavali bomo torej tri alternativne domneve:

- *H_1^X, da je X stohastično strogo večja od Y;*
- *H_1^Y, da je X stohastično strogo manjša od Y;*
- *$H_1^{\pm}$, da velja ena od zgornjih dveh možnosti.*

Če je število meritev dovolj veliko, ničelno domnevo zavrnemo:

- *proti H_1^X, če je* $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i^X - m(m+n+1) - 1\right) \geq z_\alpha$;
- *proti H_1^Y, če je* $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i^X - m(m+n+1) + 1\right) \leq -z_\alpha$;
- *proti $H_1^{\pm}$, če je* $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(\left|2\sum_{i=1}^m R_i^X - m(m+n+1)\right| - 1\right) \geq z_{\alpha/2}$.

Z drugimi besedami, izvedemo Z-preizkus na preizkusni statistiki $Z_X := \sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i^X - m(m+n+1)\right)$ *s popravkom* $\sqrt{\dfrac{3}{mn(m+n+1)}}$.

POZOR! Če bi dvostranska alternativna domneva preprosto trdila, da je spremenljivka na prvi skupini drugače porazdeljena kot na drugi, preizkus ne bi bil več dosleden!

34. Tekmovalci dveh ekip, "modrih" in "oranžnih", so se pomerili v teku. Vrstni red tekmovalcev je naslednji:

$$M,\ M,\ O,\ M,\ M,\ O,\ M,\ M,\ O,\ M,\ O,\ O,\ O,\ M,\ O,\ O,\ O,\ O,\ M,\ O$$

(tj. prvi, ki je prispel na cilj, je bil član "modrih", drugi prav tako, tretji je bil član "oranžnih" itd.). Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite domnevo, da so modri enako dobri kot oranžni, proti alternativni domnevi, da je med njimi razlika.

[61] Frank Wilcoxon (1892–1965), ameriški kemik in statistik

[62] Henry Berthold Mann, rojen kot Heinrich Mann (1905–2000), avstrijski matematik judovskega rodu, deloval v ZDA

[63] Donald Ransom Whitney (1915–2007) ameriški statistik

35. Statistični spremenljivki X in Y imata možne vrednosti $a < b < c$. Recimo, da je populacija razdeljena na tri skupine in da so s skupino natančno določene vrednosti obeh spremenljivk, tako kot kaže tabela:

skupina	X	Y
prva	a	b
druga	b	c
tretja	c	a

Recimo, da vzamemo dovolj velik vzorec in da so vse skupine enako zastopane. Kaj bo pokazal preizkus z znaki in kaj bo pokazal inverzijski preizkus?

36. Dijaki so se pomerili v teku na 60 metrov. Določeni so izjavili, da so se prej pripravljali, določeni pa, da ne. Rezultati tistih, ki so se pripravljali, so:

$$7{\cdot}6\,,\ 7{\cdot}6\,,\ 7{\cdot}7\,,\ 7{\cdot}8\,,\ 7{\cdot}8\,,\ 8{\cdot}0\,,\ 8{\cdot}1\,,\ 8{\cdot}2\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 9{\cdot}3\,,$$

rezultati tistih, ki se niso pripravljali, pa so:

$$7{\cdot}9\,,\ 8{\cdot}2\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}4\,,\ 8{\cdot}7\,,\ 8{\cdot}8\,.$$

Z inverzijskim preizkusom preizkusite ničelno domnevo, da tisti, ki se pripravljajo, tečejo enako kot tisti, ki se ne pripravljajo, proti alternativni domnevi, da tisti, ki se pripravljajo, tečejo bolje od tistih, ki se ne pripravljajo. Kaj pa pravi T-preizkus? Stopnja značilnosti naj bo obakrat $\alpha = 0{\cdot}05$.

37. 186 odraslih Britancev je odgovorilo na vprašanje, kakšna se jim je zdela premierka Theresa May v odstopu. Rezultati po spolu so zbrani v naslednji tabeli:

mnenje	ženske	moški
odlična	3	0
dobra	21	13
povprečna	28	20
slaba	18	20
grozna	22	41

Pri stopnji značilnosti $\alpha = 0{\cdot}01$ preizkusite domnevo, da imajo moški in ženske o premierki enako dobro mnenje, proti alternativni domnevi, ki to zanika. Kateri spol je imel na vzorcu boljše mnenje o premierki?

14. Povezanost dveh številskih spremenljivk

Interval zaupanja za korelacijski koeficient. Preizkušanje nekoreliranosti. Kontingenčni preizkus. Enostavna linearna regresija.

V celotnem razdelku je privzeto, da opažanje sestoji iz neodvisnih realizacij $(X_1, Y_1), \ldots, (X_n, Y_n)$ istega para statističnih spremenljivk (X, Y).

Interval zaupanja za korelacijski koeficient

Za izračun korelacijskega koeficienta $\rho = \operatorname{corr}(X, Y)$ *najprej izračunamo* **vzorčni korelacijski koeficient** R:

$$\begin{aligned}
C_x^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\
C_y^2 &= \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 \\
C_{xy} &= \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^n X_i Y_i - n\bar{X}\bar{Y} \\
R &= \frac{C_{xy}}{C_x C_y}
\end{aligned}$$

in ga normaliziramo:

$$Z := \operatorname{Arth} R = \frac{1}{2} \ln \frac{1+R}{1-R}$$

Približen interval zaupanja za ρ:

$$\operatorname{th}\left(Z - \frac{c}{\sqrt{n-3}}\right) \le \rho \le \operatorname{th}\left(Z + \frac{c}{\sqrt{n-3}}\right)$$

kjer je $c = z_{(1-\beta)/2}$ *in* $\operatorname{th}(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$.

1. Meritve krvnega pritiska so zbrane v naslednji tabeli:

sistolični	130	120	120	125	125	125	105	130	130	135
diastolični	80	80	85	80	75	75	75	80	85	70

sistolični	130	125	140	130	120
diastolični	75	80	90	80	85

Poiščite 95% interval zaupanja za korelacijski koeficient med sistoličnim in diastoličnim pritiskom.

Preizkušanje nekoreliranosti (Gaussov[64] model)

Če velja ničelna domneva H_0, da sta X in Y nekorelirani, je približno:

$$T := \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \sim \mathrm{Student}(n-2)$$

kjer je R vzorčni korelacijski koeficient. Tako nekoreliranost preizkušamo s T-preizkusom na preizkusni statistiki T z $n-2$ prostostnimi stopnjami, in sicer z:

- *dvostransko različico, če H_1 trdi, da sta X in Y korelirani;*
- *enostransko različico v desno, če H_1 trdi, da sta X in Y pozitivno korelirani;*
- *enostransko različico v levo, če H_1 trdi, da sta X in Y negativno korelirani.*

2. Za meritve krvnega pritiska iz 1. naloge pri stopnji značilnosti $\alpha = 0{\cdot}1$ preizkusite ničelno domnevo, da sta sistolični in diastolični pritisk nekorelirana, proti alternativni domnevi, da sta korelirana.

Kontingenčni preizkus neodvisnosti

Preizkušamo, ali sta spremenljivki X in Y neodvisni. Recimo, da pri opaženih vrednostih kombinacija $X = a_i, Y = b_j$ nastopa N_{ij}-krat ($i = 1, \ldots, r$, $j = 1, \ldots, s$). Ker pri neodvisnosti približno velja:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(N_{ij}-\tilde N_{ij})^2}{\tilde N_{ij}} \sim \chi^2\big((r-1)(s-1)\big)$$

kjer je:

$$N_{i.} = \sum_{j=1}^{s} N_{ij}\,, \quad N_{.j} = \sum_{i=1}^{r} N_{ij}\,, \quad n = \sum_{i=1}^{r}\sum_{j=1}^{s} N_{ij}\,, \quad \tilde N_{ij} = \frac{N_{i.}N_{.j}}{n}$$

neodvisnost preizkusimo s preizkusom hi kvadrat z $(r-1)(s-1)$ prostostnimi stopnjami na preizkusni statistiki χ^2, in sicer uporabimo enostransko različico v desno. Za naše potrebe je preizkus dovolj natančen, če je $\tilde N_{ij} \geq 5$ za vse i in j.

3. Na vzorcu 62 oseb dobimo naslednjo navzkrižno frekvenčno porazdelitev barve oči in las:

[64] Carl Friedrich Gauß (1777–1855), nemški matematik

oči\lasje	rdeči, blond	rjavi, črni	Skupaj
modre	12	1	13
zelene	14	9	23
rjave	4	22	26
Skupaj	30	32	62

Pri stopnji značilnosti $\alpha = 0{\cdot}01$ preizkusite domnevo, da sta barva las in barva oči neodvisni.

Kontingenčni preizkus neodvisnosti za dihotomni spremenljivki

Če sta spremenljivki X in Y dihotomni in so navzkrižne frekvence podane s tabelo:

A	B
C	D

,

velja:

$$\chi^2 = \frac{(A+B+C+D)(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)} .$$

4. Rezultati ankete z dvema vprašanjema 'Ali verjamete v horoskop?' in 'Ali verjamete v NLP-je?' so zbrani v naslednji tabeli:

Horoskop\NLP	vsaj malo	ne	Skupaj
vsaj malo	6	8	14
ne	7	10	17
Skupaj	13	18	31

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ preizkusite neodvisnost verovanja ljudi v horoskop in v NLP-je.

5. Na neki spletni strani je 1990 ljudi glasovalo, kateri film bo najverjetneje dobil oskarja. Anketirance so razdelili v tri razrede glede na starost. Rezultati ankete so naslednji:

	do 20	od 20 do 40	nad 40
Gospodar prstanov	350	250	180
Skrivnostna reka	80	100	100
Seabiscuit	70	90	130
Zgubljeno s prevodom	50	80	110
Gospodar in bojevnik	200	150	50

S kontingenčnim preizkusom preizkusite domnevo, da je mnenje o oskarjih neodvisno od starosti. Stopnja značilnosti naj bo 0˙05.

15. Linearna regresija

Ocenjevanje in napovedovanje pri enostavni linearni regresiji. Splošna linearna regresija.

Enostavna linearna regresija

Enostavna linearna regresija je statistični model, ki se ravna po predlogi:

$$Y = a + bX + \varepsilon\,,$$

kjer je ε slučajna spremenljivka, ki ima pogojno na X matematično upanje 0 in določeno varianco σ^2; tej slučajni spremenljivki pravimo **šum**. *To torej pomeni, da želimo skozi podatke potegniti premico – pravimo ji* **regresijska premica**.
Opazimo podatke (X_1, Y_1), (X_2, Y_2), $\ldots$, (X_n, Y_n), za katere privzamemo, da se ravnajo po tej predlogi, tj. da velja:

$$Y_i = a + bX + \varepsilon_i\,,$$

*kjer so $\varepsilon_1, \ldots, \varepsilon_n$ nekorelirane ter imajo matematično upanje nič in enako varianco σ^2 (***homoskedastičnost***).*
Parametra a in b točkasto ocenimo po **metodi najmanjših kvadratov***: njuni cenilki $\hat{a}$ in $\hat{b}$ sta tisti vrednosti, pri katerih je vsota kvadratov* **ostankov** *ali* **rezidualov***:*

$$\hat{\varepsilon}_i = Y_i - \hat{a} - \hat{b}X_i$$

najmanjša. Cenilki lahko izračunamo po formuli:

$$\hat{b} = \frac{C_{xy}}{C_{xx}}\,, \quad \hat{a} = \bar{Y} - \hat{b}\bar{X}\,,$$

kjer je:

$$C_{xx} = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2\,,$$
$$C_{xy} = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}\,.$$

Pomembno je, da linearna regresija ni simetrična – spremenljivki X in Y imata različni vlogi. Pravimo, da delamo regresijo Y glede na X. Spremenljivka X je **pojasnjevalna***, spremenljivka Y pa* **odvisna**. *Če njuni vlogi zamenjamo, tipično dobimo drugo regresijsko premico.*

1. Meritve količin X in Y dajo naslednje vrednosti:

X_i	1	2	3	3	6
Y_i	1	3	2	3	4

Določite regresijsko premico Y glede na X in X glede na Y.

Intervalsko ocenjevanje

Če privzamemo, da so šumi ε_i porazdeljeni normalno, lahko konstruiramo intervala zaupanja za a in b pri stopnji tveganja α v obliki:

$$\hat{a} - \Delta_a < a < \hat{a} + \Delta_a\,, \qquad \hat{b} - \Delta_b < b < \hat{b} + \Delta_b\,,$$

kjer je:

$$\Delta_a = t_{\alpha/2}(n-2)\, S \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{C_{xx}}}\,, \qquad \Delta_b = t_{\alpha/2}(n-2)\, \frac{S}{\sqrt{C_{xx}}}$$

in

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2\,.$$

2. Za opažene vrednosti iz prejšnje naloge poiščite 95% intervala zaupanja za regresijska koeficienta.

Napovedovanje

Točkasta napoved za Y pri danem X je seveda $\hat{Y} = \hat{a} + \hat{b}X$. Če spet privzamemo, da so šumi ε_i porazdeljeni normalno, lahko podamo tudi **napovedni interval** *za Y pri* **stopnji tveganja** *α, in sicer:*

$$\hat{Y} - \Delta \le Y \le \hat{Y} + \Delta$$

kjer je:

$$\Delta = t_{\alpha/2}(n-2)\, S \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{C_{xx}}}$$

3. Za opažene vrednosti iz 1. naloge napovejte Y pri $X = 10$ in poiščite 95% napovedni interval.

4. Včasih zelo razširjene žarnice na žarilno nitko se zdaj zamenjujejo s svetili druge vrste. Kako svetla je žarnica, se je tipično opisovalo kar z njeno močjo (P) v wattih. V splošnem pa moč sama po sebi ne določa svetilnosti (Φ) v lumnih: pomembna je tudi vrsta svetila. Pri žarnicah na žarilno nitko v grobem velja naslednja tabela:[65]

[65] Vir: Häberle, Häberle, Jöckel, Krall, Schiemann, Schmitt, Tkotz: *Tabellenbuch Elektrotechnik.* 25. Auflage. Verlag Europa–Lehrmittel, Haan–Gruiten, 2013.

moč [W]	svetilnost [lm]
25	230
40	430
60	730
100	1380
150	2220
200	3150
300	5000
500	8400
1000	18800

a) Ocenite koeficienta enostavne linearne regresije pri modelu, ki ima P za pojasnjevalno, Φ pa za odvisno spremenljivko. Ocenite svetilnost 75-wattne in 15-wattne žarnice.

b) Modelirajte to kot odvisnost oblike $\Phi = kP^\alpha$ (z določenimi napakami). Prevedite na enostavno linearno regresijo, pri čemer vzemite moč za pojasnjevalno, svetilnost pa za odvisno spremenljivko. Ocenite koeficienta k in α ter spet ocenite svetilnost 75-wattne in 15-wattne žarnice.

5. V spodnji tabeli so podane povprečne temperature v Ljubljani za zadnjih 36 let:[66]

Leto	Temp.
1985	9·85
1986	10·78
1987	10·14
1988	11·28
1989	10·98
1990	11·32
1991	10·66
1992	11·38
1993	11·04
1994	12·37
1995	11·44
1996	10·45

Leto	Temp.
1997	11·00
1998	11·99
1999	11·21
2000	12·30
2001	12·42
2002	11·82
2003	11·98
2004	11·13
2005	10·95
2006	11·36
2007	12·77
2008	11·78

Leto	Temp.
2009	11·91
2010	11·18
2011	11·64
2012	12·51
2013	11·69
2014	12·96
2015	12·53
2016	12·26
2017	11·99
2018	12·79
2019	12·71
2020	12·47

a) Ocenite, kako hitro se podnebje segreva.

b) Je segrevanje statistično značilno?

c) Napovejte povprečno temperaturo leta 2050 točkasto in intervalsko s 95% gotovostjo.

[66]Vir: https://data.giss.nasa.gov/gistemp/station_data_v4_globe/, ogled 11. 5. 2021.

Splošna linearna regresija

Splošna linearna regresija je naslednja posplošitev enostavne linearne regresije:

$$Y = b_1 f_1(X) + b_2 f_2(X) + \cdots + b_k f_k(X) + \varepsilon\,.$$

Parametre točkasto ocenimo po **metodi najmanjših kvadratov***: njihove cenilke $\hat{b}_1, \ldots, \hat{b}_k$ so določene tako, da je vsota kvadratov rezidualov:*

$$\hat{\varepsilon}_i = Y_i - b_1 f_1(X_i) - b_2 f_2(X_i) - \cdots - b_k f_k(X_i)$$

minimalna. Ustrezna nepristranska napoved vrednosti Y pri danem X pa je seveda $\hat{Y} = \hat{b}_1 f_1(X) + \hat{b}_2 f_2(X) + \cdots \hat{b}_k f_k(X)$.

6. Za podatke iz 1. naloge zdaj privzamemo model $Y = bX + \varepsilon$. Ocenite b ter točkasto napovejte vrednost spremenljivke Y pri $X = 10$.

REŠITVE

1. Osnove kombinatorike

1. $(4+3)\cdot 5 = 35$.

2. a) 900, b) 450, c) 400, d) 9, e) 648, f) 90.

3. Označimo s k število kroglic, ki jih vzamemo iz posode. Tedaj so rezultati zbrani v naslednji tabeli:

k	vrstni red vlečenja			
	pomemben		ni pomemben	
	vračamo	ne vračamo	vračamo	ne vračamo
1	5	5	5	5
2	25	20	15	10
3	125	60	35	10

Pri različici, ko kroglic ne vračamo in vrstni red ni pomemben, sta rezultata za $k=2$ in $k=3$ enaka, ker lahko namesto tega, katere kroglice smo vzeli, gledamo, katere kroglice so ostale v posodi.

Različico, ko vrstni red ni pomemben, kroglice pa vračamo, v splošnem primeru, ko iz posode z n kroglicami vzamemo k kroglic, izračunamo tako, da jemanja ponazorimo z razporeditvijo $n-1$ rdečih in k modrih puščic v vrsto, pri čemer puščice ločimo le po barvi. Vsaka rdeča puščica predstavlja pregrajo med dvema kroglicama v škatli in vsaka modra puščica predstavlja jemanje ustrezne kroglice. Tako npr. razporeditev $MRRMMMRR$ pomeni, da iz posode s 5 kroglicami vzamemo tri kroglice, in sicer prvo kroglico enkrat, tretjo dvakrat, ostalih pa ne vzamemo. Vsaka razporeditev puščic ustreza natanko enemu jemanju kroglic. Tako dobimo, da je vseh jemanj $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$ (glej naslednjo nalogo).

4. Če vse razločujemo: $(3+4)! = 5040$.

Če cvetlic iste vrste ne razločujemo: $\displaystyle\binom{7}{3} = 35$.

5. a) $\displaystyle\binom{6}{4}\binom{4}{2} = 90$,

b) $\displaystyle\binom{6}{1}\binom{4}{3} + \binom{6}{2}\binom{4}{2} + \binom{6}{3}\binom{4}{1} = \binom{10}{4} - \binom{6}{4} - \binom{4}{4} = 194$.

6. Razporeditev lahko določimo tako, da najprej v vrsto postavimo miroljubne vojake. To lahko storimo na $m!$ načinov. Nato mednje vrinemo nasilne vojake: zanje imamo $m+1$ mest. Vrinemo torej lahko na $(m+1)m\cdots(m-n+2)$ načinov. Možnih načinov je torej $m!\,(m+1)m\cdots(m-n+2)$, kar je enako 0, če je $n \geq n+2$, sicer pa je enako:

$$\frac{m!\,(m+1)!}{(m-n+1)!} = \binom{m+1}{n} m!\,n!\,.$$

7. a) $9! = 362880$, b) $3!\,2!\,4! = 288$, c) $288 \cdot 3! = 1728$, d) $\dfrac{9!}{3!\,2!\,4!} = 1260$.

8. $5! = 120$; $9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 15120$.

9. a) $\dfrac{12!}{3!\,3!\,2!\,2!\,2!} = 1633200$,

b) $2 \cdot 3 \cdot \dfrac{7!}{3!\,2!\,2!} \cdot \dfrac{5!}{3!\,2!} = 12600$,

c) $\dfrac{7!}{2!\,2!} \cdot 5! = 151200$.

2. Elementarna verjetnost

1. $\dfrac{5}{36} \doteq 0{\dot{}}139$.

2. Verjetneje je, da so trije enega, eden pa nasprotnega spola: verjetnost tega dogodka je $8/16 = 0{\dot{}}5$, verjetnost enake zastopanosti pa je $6/16 = 0{\dot{}}375$.

3. $1 - \dfrac{5^5}{6^5} \doteq 0{\dot{}}598$.

4. a) $\dfrac{5}{12} \cdot \dfrac{4}{12} \cdot \dfrac{3}{12} \cdot 6 \doteq 0{\dot{}}208$, b) $\dfrac{5 \cdot 4 \cdot 3}{\binom{12}{3}} \doteq 0{\dot{}}273$.

5. Označimo z R_i dogodek, da je bila i-ta izvlečena kroglica rdeča, z Z_i, da je bila zelena, z R_i pa, da je bila rdeča. Izračunati je treba $\mathbb{P}(R_1 \cup Z_2)$.

Prvi način. Velja:

$$\mathbb{P}(R_1 \cup Z_2) = \mathbb{P}(R_1) + \mathbb{P}(Z_2) - \mathbb{P}(R_1 \cap Z_2)\,.$$

Velja $\mathbb{P}(R_1) = 3/10$. Nadalje je:

$$\mathbb{P}(Z_2) = \mathbb{P}(R_1 \cap Z_2) + \mathbb{P}(Z_1 \cap Z_2) + \mathbb{P}(B_1 \cap Z_2) = \frac{3 \cdot 2 + 2 \cdot 1 + 5 \cdot 2}{10 \cdot 9} = \frac{1}{5}\,.$$

Opazimo, da je to enako $\mathbb{P}(Z_1)$. To je zato, ker sta obe poziciji pri vlečenju enakovredni. Končno je:

$$\mathbb{P}(R_1 \cup Z_2) = \frac{3}{10} + \frac{1}{5} - \frac{3 \cdot 2}{10 \cdot 9} = \frac{13}{30} \doteq 0{\dot{}}433\,.$$

Drugi način. Velja:

$$\begin{aligned}\mathbb{P}(R_1 \cup Z_2) &= \mathbb{P}(R_1) + \mathbb{P}(Z_2 \setminus R_1) = \\ &= \mathbb{P}(R_1) + \mathbb{P}(Z_1 \cap Z_2) + \mathbb{P}(B_1 \cap Z_2) = \\ &= \frac{3}{10} + \frac{2 \cdot 1}{10 \cdot 9} + \frac{3 \cdot 2}{10 \cdot 9} = \\ &= \frac{13}{30}\,,\end{aligned}$$

kar je isto kot prej.

6. $B \cap C$

7. a) $\mathbb{P}(A^c \cap B^c \cap C^c) = 0{\dot{}}25$,
b) $\mathbb{P}\big((A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C)\big) = 0{\dot{}}25$,
c) $\mathbb{P}\big((A \cap B) \cup (A \cap C) \cup (B \cap C)\big) = 0{\dot{}}5$.

8. Označimo iskani dogodek z A. Tedaj velja:

$$\mathbb{P}(A) = 1 - \frac{364 \cdot 363 \cdot \cdots \cdot (366 - n)}{365^{n-1}} .$$

Tabela prvih nekaj verjetnosti:

n	P	n	P	n	P
1	0	26	0'598	51	0'974
2	0'00274	27	0'627	52	0'978
3	0'0082	28	0'654	53	0'981
4	0'0164	29	0'681	54	0'984
5	0'0271	30	0'706	55	0'986
6	0'0405	31	0'73	56	0'988
7	0'0562	32	0'753	57	0'9901
8	0'0743	33	0'775	58	0'9917
9	0'0946	34	0'795	59	0'993
10	0'117	35	0'814	60	0'9941
11	0'141	36	0'832	61	0'9951
12	0'167	37	0'849	62	0'9959
13	0'194	38	0'864	63	0'9966
14	0'223	39	0'878	64	0'9972
15	0'253	40	0'891	65	0'9977
16	0'284	41	0'903	66	0'9981
17	0'315	42	0'914	67	0'9984
18	0'347	43	0'924	68	0'9987
19	0'379	44	0'933	69	0'999
20	0'411	45	0'941	70	0'9992
21	0'444	46	0'948	71	0'9993
22	0'476	47	0'955	72	0'9995
23	0'507	48	0'961	73	0'9996
24	0'538	49	0'966	74	0'9996
25	0'569	50	0'97	75	0'9997

Že pri 23 ljudeh verjetnost prvič preseže 1/2.

Splošneje, če je d dni v letu, velja:

$$\begin{aligned}\mathbb{P}(A) &= 1 - \frac{d(d-1)(d-2)\cdots(d-n+1)}{d^n} = \\ &= 1 - \left(1 - \frac{1}{d}\right)\left(1 - \frac{2}{d}\right)\cdots\left(1 - \frac{n-1}{d}\right) .\end{aligned}$$

Oglejmo si asimptotično obnašanje te verjetnosti, ko gre $d \to \infty$ (n pa se lahko spreminja z d). Najprej opazimo, da je dogodek A unija dogodkov A_{ij}, da imata i-ti in j-ti človek oba rojstni dan na isti dan ($i \neq j$). Vsi ti dogodki imajo verjetnosti

$1/d$. Dogodka A_{ij} in A_{ji} sovpadata, sicer pa je plavzibilno, da so ti dogodki, če je n majhen v primerjavi z d, približno paroma nezdružljivi, zato je:

$$\mathbb{P}(A) = \mathbb{P}\Bigg(\bigcup_{1\le i<j\le n} A_{ij}\Bigg) \approx \sum_{1\le i<j\le n} \mathbb{P}(A_{ij}) = \sum_{1\le i<j\le n} \frac{1}{d} = \frac{n^2-n}{2d} \approx \frac{n^2}{2d}\,.$$

To je enako vsaj $1/2$ pri $n \ge \sqrt{d}$. Iz $\sqrt{365} \doteq 19{\dot{}}10$ dobimo, da je pri $d = 365$ to izpolnjeno pri $n \ge 20$. Dobili smo torej malo prenizek prag.

Natančnejši argument pa nam da analiza. Z aproksimacijo $1 - x \approx e^{-x}$ in oceno logaritma ostankov se da izpeljati:

$$\begin{aligned} \mathbb{P}(A^{\mathrm{c}}) &\sim e^{-n(n-1)/(2d)} &&, \text{ brž ko je } n \ll d^{2/3}\,;\\ \mathbb{P}(A) &\sim 1 - e^{-n(n-1)/(2d)} &&, \text{ brž ko gre } d \to \infty\,;\\ \mathbb{P}(A) &\sim \frac{n(n-1)}{2d} &&, \text{ brž ko je } n \ll \sqrt{d}\,. \end{aligned}$$

Če aproksimiramo $\mathbb{P}(A) \approx 1 - e^{-n^2/(2d)}$, je slednje enako vsaj $1/2$ pri $n \ge \sqrt{2n \ln 2}$. Iz $\sqrt{2 \cdot 365 \cdot \ln 2} \doteq 22{\dot{}}49$ dobimo, da je pri $d = 365$ to izpolnjeno pri $n \ge 23$, kar je pravi prag.

9. $\dfrac{\binom{4}{2}\binom{12}{6}}{\binom{16}{8}} = \dfrac{\left[\binom{8}{2}\right]^2}{\binom{16}{4}} \doteq 0{\dot{}}431.$

10. $1 - \dfrac{\binom{90}{10} + \binom{10}{1}\binom{90}{9}}{\binom{100}{10}} \doteq 0{\dot{}}262.$

11. $\dfrac{\binom{8}{4}\binom{4}{2}\binom{2}{1}}{\binom{14}{7}} \doteq 0{\dot{}}245.$

12. Verjetnosti dobitkov lahko računamo na dva načina. Pri prvem načinu si predstavljamo, da so prekrižane številke fiksne, nakar gledamo vsa možna žrebanja (*pogled igralca*). Lahko pa si predstavljamo tudi, da so fiksne izžrebane številke, nakar gledamo vsa možna križanja (pogled Loterije). Ker se namreč igralčeva izbira in žrebanje številk izvedeta neodvisno, ni pomembno, kaj je bilo prej. Dobimo:

- Za vseh 10 številk: $\dfrac{\binom{60}{10}}{\binom{70}{20}} = \dfrac{\binom{20}{10}}{\binom{70}{10}} \doteq 4{\dot{}}657 \cdot 10^{-7} \doteq \dfrac{1}{2.147.181}$;
- Za 5 številk: $\dfrac{\binom{10}{5}\binom{60}{15}}{\binom{70}{20}} = \dfrac{\binom{20}{5}\binom{50}{5}}{\binom{70}{10}} \doteq 0{\dot{}}0828 \doteq \dfrac{1}{12{\dot{}}1}$;
- Za nič številk: $\dfrac{\binom{60}{20}}{\binom{70}{20}} = \dfrac{\binom{50}{10}}{\binom{70}{10}} \doteq 0{\dot{}}0259 \doteq \dfrac{1}{38{\dot{}}6}$.

13. Nalogo lahko rešimo na vsaj tri načine.

Prvi način: postavimo se v kožo igralca, ki ve, katere številke je prekrižal, ne ve pa, katere so bile izžrebane. Če najprej izberemo redno izžrebane številke in nato dodatno, dobimo, da je verjetnost danega dogodka enaka:

$$\frac{\binom{6}{5}\binom{38}{1}\cdot 1}{\binom{44}{6}\cdot 38}\,.$$

Drugi način: prav tako se postavimo v kožo igralca, le da najprej izberemo dodatno in šele nato redno izžrebane. Iskana verjetnost se sedaj izrazi v obliki:

$$\frac{6\cdot\binom{5}{5}\binom{38}{1}}{44\cdot\binom{43}{6}}\,.$$

Tretji način: postavimo se v kožo Loterije, ki ve, katere številke so bile izžrebane, ne ve pa, katere številke je prekrižal igralec. Iskana verjetnost se sedaj izrazi v obliki:

$$\frac{\binom{6}{5}\binom{1}{1}\binom{43}{0}}{\binom{44}{6}}\,.$$

Preverimo lahko, da vse tri izražave dajo vrednost:

$$\frac{3}{3529526}\approx\frac{1}{1.176.509}\doteq 8{\cdot}50\cdot 10^{-7}\,.$$

14. Izide – razporeditve kart v kupu – lahko opišemo na več načinov. Smiselno je izbrati opis, iz katerega bo razvidno, da lahko enako število izidov združimo v enako verjetne dogodke, s pomočjo katerih lahko opišemo željeni dogodek. V našem primeru si je smiselno predstavljati, da 10! možnih permutacij generiramo v več fazah:

(1) V vrsto razporedimo vse rdeče in zelene karte. Če vse karte ločimo, imamo za to 5! možnosti.

(2) Med razporejene rdeče in zelene karte drugo za drugo vrivamo bele: najprej prvo belo karto razporedimo na 6 možnih pozicij, od skrajne leve do skrajne desne. Nato na 7 možnih pozicij razporedimo drugo belo karto, nato še tretjo, četrto in peto. Pri razporejanju imamo torej $6\cdot 7\cdot 8\cdot 9\cdot 10$ možnosti.

(3) Z leve proti desni poberemo vse karte, tako da se najbolj desna karta znajde na dnu, najbolj leva pa na vrhu kupa.

Na ta način vsaka razporeditev rdečih in zelenih kart predstavlja dogodek in vsak tak dogodek vsebuje $6\cdot 7\cdot 8\cdot 9\cdot 10$ izidov – permutacij. Ker so vsi izidi enako verjetni, so enako verjetni tudi vsi opisani dogodki. Tako lahko rečemo, da lahko na bele karte pozabimo.

Fazo (1) se splača razdeliti še na nadaljnje podfaze, a pri vsaki točki drugače.

a) Dogodek, da je prva rdeča karta izvlečena pred prvo zeleno, ustreza dogodku, da je v fazi (1) na skrajni levi rdeča karta. Zamislimo si, da za rdeče in zelene karte pripravimo mesta in da najprej izberemo rdečo ali zeleno karto, ki bo na prvem mestu. Za to imamo 5 možnosti. Nato drugo za drugo razporedimo še ostale karte. Vsaka razporeditev karte na prvo mesto predstavlja dogodek in vsak tak dogodek zajema $4 \cdot 3 \cdot 2 \cdot 1$ razporeditev rdečih in zelenih kart. Tako dobimo, da so tudi vsi ti dogodki enako verjetni. Dogodek, da je prva rdeča karta izvlečena pred prvo zeleno, zajema dva dogodka od petih enako verjetnih, zato je njegova verjetnost enaka $2/5$.

b) Dogodek, da je prva rdeča karta izvlečena pred zadnjo zeleno, je nasproten dogodku, da sta v fazi (1) na skrajni desni dve rdeči karti. Tu se splača fazo (1) razdeliti na naslednji dve podfazi:

(1.1) Pripravimo pet mest za rdeče in zelene karte in določimo, na katerih mestih bosta rdeči karti. Za to imamo $\binom{5}{2} = 10$ možnosti. Seveda bodo na preostalih treh mestih zelene karte.

(1.2) Določimo, katere konkretne rdeče oz. zelene karte bodo na ustreznih mestih. Za to imamo $2!\,3!$ možnosti.

Spet vsaka določitev barv kart posameznim mestom predstavlja dogodek in vsak tak dogodek vsebuje po $2!\,3!$ razporeditev rdečih in zelenih kart iz faze (1). Zato so tudi vsi ti dogodki enako verjetni. Dogodek, da sta na skrajnih dveh desnih mestih rdeči karti, je tako eden izmed 10 enako verjetnih. Verjetnost, da bo prva rdeča karta izvečena pred zadnjo zeleno, je torej enaka $1 - 1/10 = 9/10$.

15. Verjetnostni prostor lahko razdelimo na $9 \cdot 8 \cdot 7 = 504$ enako verjetne izide – glede na to, katere študente asistent izbere za premestitev in v katerem vrstnem redu. Vsak tak izid lahko torej ponazorimo z zaporedjem črk oblike xyz. Tako zaporedje si lahko predstavljamo kot povelje, naj gre x tja, kjer je bil prej y, y tja, kjer je bil prej z, in z tja, kjer je bil prej x.

Preštejmo ugodne izide. Pri Aljažu in Brigiti ločimo, kateri od njiju je bil premeščen in kateri ne.

- Vsak izid, kjer ni premeščen niti Aljaž niti Brigita, je ugoden. Teh izidov je $7 \cdot 6 \cdot 5 = 210$.
- Če je Aljaž premeščen, Brigita pa ne, mora biti Aljaž premeščen na Cvetovo mesto. Taki so vsi izidi oblike ACW, WAC ali CWA, kjer črka W pomeni kogar koli, ki ni Aljaž, Brigita ali Cveto. Teh izidov je $6 \cdot 3 = 18$.
- Če je Brigita premeščena, Aljaž pa ne, ni ugodnih izidov.
- Če sta premeščena oba, najprej opazimo, da morata ostati v svoji vrsti. Edini ugodni izidi so ABC, BCA in CAB.

Iskana verjetnost je tako enaka $\dfrac{210 + 18 + 3}{504} = \dfrac{231}{504} = \dfrac{11}{24} \doteq 0{'}458$.

Opomba. Po trije in trije izidi, kot smo jih definirali, določajo enako premestitev. Tako bi lahko za izide vzeli tudi kar premestitve. Teh je $\dfrac{9 \cdot 8 \cdot 7}{3} = 2 \cdot \dbinom{9}{3} = 168$.

16. *Prvi način.* Kup razdelimo na vrhnjih 25, srednji 2 in spodnjih 25 kart. Če se zgodi dani dogodek, je število pikov in src v posameznem delu kupa natančno določeno že s številom pikov in src v srednjem delu z dvema kartama. Tam pa imamo naslednje možnosti:

- brez src in pikov – verjetnost tega dogodka je:

$$\frac{\binom{25}{13}^2}{\binom{52}{13}\binom{39}{13}} = \frac{\binom{26}{2}}{\binom{52}{2}} \cdot \frac{\binom{24}{12}}{\binom{50}{25}}\,;$$

- natanko eno srce in brez pikov ali pa natanko en pik in brez src – verjetnost tega dogodka je:

$$\frac{2 \cdot \binom{25}{12}\binom{2}{1}\binom{25}{13}}{\binom{52}{13}\binom{39}{13}} = 2 \cdot \frac{13 \cdot 26}{\binom{52}{2}} \cdot \frac{\binom{25}{13}}{\binom{50}{25}} = 2 \cdot \frac{13 \cdot 26}{\binom{52}{2}} \cdot \frac{\binom{25}{12}}{\binom{50}{25}}\,;$$

- eno srce in en pik – verjetnost tega dogodka je:

$$\frac{2 \cdot \binom{25}{12}^2}{\binom{52}{13}\binom{39}{13}} = \frac{13 \cdot 13}{\binom{52}{2}} \cdot \frac{\binom{26}{13}}{\binom{50}{25}}\,;$$

- dve srci ali dva pika – verjetnost tega dogodka je:

$$\frac{2 \cdot \binom{25}{11}\binom{25}{13}}{\binom{52}{13}\binom{39}{13}} = 2 \cdot \frac{\binom{13}{2}}{\binom{52}{2}} \cdot \frac{\binom{26}{14}}{\binom{50}{25}} = 2 \cdot \frac{\binom{13}{2}}{\binom{52}{2}} \cdot \frac{\binom{26}{12}}{\binom{50}{25}}\,.$$

Ker se te možnosti med seboj izključujejo, je verjetnost danega dogodka vsota zgornjih verjetnosti, ki je enaka:

$$\frac{61}{7} \cdot \frac{(25!)^2 \cdot 26!}{(12!)^2 \cdot 52!} \doteq 4{\cdot}57 \cdot 10^{-8}\,.$$

Drugi način. Srcem pripišemo zgornjo, pikom pa spodnjo polovico kupa, nakar ločimo možnosti, ali katero srce oziroma pik prestopi pripisano polovico kupa:

- vsa srca so v zgornji, vsi piki pa v spodnji polovici – verjetnost tega dogodka je:

$$\frac{\binom{26}{13}^2}{\binom{52}{13}\binom{39}{13}} = \frac{\binom{26}{13}}{\binom{52}{26}}\,;$$

- srce prestopi, pik pa ne, ali pa obratno – verjetnost tega dogodka je:

$$2 \cdot \frac{\binom{26}{12}\binom{25}{13}}{\binom{52}{13}\binom{39}{13}} = \frac{13}{52} \cdot \frac{\binom{26}{14}}{\binom{51}{26}} = \frac{13}{52} \cdot \frac{\binom{26}{12}}{\binom{51}{25}}\,;$$

- prestopi tako eno srce kot tudi en pik – verjetnost tega dogodka je:

$$\frac{\binom{25}{12}^2}{\binom{52}{13}\binom{39}{13}} = \frac{13 \cdot 13}{\binom{52}{2}} \cdot \frac{\binom{26}{13}}{\binom{50}{25}}\,.$$

Tudi ti dogodki se med seboj izključujejo in njihova vsota pride ista kot prej.

Tretji način. Ločimo glede na število src med srednjima dvema kartama:

- brez srca – verjetnost tega dogodka je

$$\frac{\binom{25}{13}\binom{27}{13}}{\binom{52}{13}\binom{39}{13}} = \frac{\binom{26}{12}}{\binom{52}{25}} = \frac{\binom{26}{14}}{\binom{52}{27}};$$

- natanko eno srce – verjetnost tega dogodka je

$$\frac{2 \cdot \binom{25}{12}\binom{26}{13}}{\binom{52}{13}\binom{39}{13}} = 2 \cdot \frac{13}{52} \cdot \frac{\binom{26}{13}}{\binom{51}{25}} = 2 \cdot \frac{13}{52} \cdot \frac{\binom{26}{13}}{\binom{51}{26}};$$

- dve srci – verjetnost tega dogodka je

$$\frac{\binom{25}{11}\binom{25}{13}}{\binom{52}{13}\binom{39}{13}} = \frac{\binom{13}{2}}{\binom{52}{2}} \cdot \frac{\binom{26}{14}}{\binom{50}{25}} = \frac{\binom{13}{2}}{\binom{52}{2}} \cdot \frac{\binom{26}{12}}{\binom{50}{25}}.$$

Tudi ti dogodki se med seboj izključujejo in njihova vsota pride ista kot prej.

17. a) Če vse cvetlice med seboj ločimo, je vseh možnih razporeditev $(b + f + k)!$ in vse so enako verjetne. Prešteti moramo še vse razporeditve, pri katerih so begonije in fuksije skupaj. Le-te lahko obravnavamo kot dva bloka; skupaj s kalanhojami je to $k + 2$ enot. Nato pa moramo še razporediti begonije in fuksije znotraj bloka. Dobimo, da je možnih razporeditev $(k + 2)!\, b!\, f!$. Iskana verjetnost je torej enaka:

$$\frac{(k+2)!\, b!\, f!}{(b+f+k)!}.$$

b) Označimo z B, F in K dogodke, da so begonije, fuksije oziroma kalanhoje skupaj. Po načelu vključitev in izključitev je iskana verjetnost enaka:

$$\begin{aligned}\mathbb{P}(B^{\mathrm{c}} \cap F^{\mathrm{c}} \cap K^{\mathrm{c}}) &= 1 - \mathbb{P}(B) - \mathbb{P}(F) - \mathbb{P}(K) \\ &\quad + \mathbb{P}(B \cap F) + \mathbb{P}(B \cap K) + \mathbb{P}(F \cap K) - \mathbb{P}(B \cap F \cap F) \\ &= \frac{1}{(b+f+k)!}\Big[(b+f+k)! \\ &\quad - (f+k+1)!\, b! - (b+k+1)!\, f! - (b+f+1)!\, k! \\ &\quad + (k+2)!\, b!\, f! + (f+2)!\, b!\, k! + (b+2)!\, f!\, k! - 3!\, b!\, f!\, k!\Big].\end{aligned}$$

18. 76/120. V splošnem, ko je kuvert n, je rezultat:

$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + \frac{(-1)^{n-1}}{n!} = 1 - \frac{1}{n!}\left\langle \frac{n!}{e} \right\rangle = \frac{1}{n!}\left\langle n!\left(1 - \frac{1}{e}\right)\right\rangle$$

19. Za $k = 1, 2, \ldots, 6$ označimo z A_k dogodek, da število k *ni* zabeleženo, z A pa označimo dogodek, da so zabeležena vsa števila pik. Tedaj je:

$$A = A_1^c \cap A_2^c \cap A_3^c \cap A_4^c \cap A_5^c \cap A_6^c = \left(A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6\right)^c .$$

Za katere koli različne indekse $i_1, i_2, \ldots, i_k$ je $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}$ dogodek, da so vsa ustrezna števila pik nezabeležena, torej da je v vsakem metu padlo število pik, različno od $i_1, i_2, \ldots, i_k$. Torej je:

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \frac{(6-k)^{10}}{6^{10}} .$$

Po načelu vključitev in izključitev je:

$$\begin{aligned}\mathbb{P}(A) &= 1 - \sum_{k=1}^{6} (-1)^{k-1} \binom{6}{k} \frac{(6-k)^{10}}{6^{10}} = \sum_{k=0}^{6} (-1)^k \binom{6}{k} \frac{(6-k)^{10}}{6^{10}} = \\ &= \frac{16435440}{6^{10}} = \frac{38045}{139968} \doteq 0{\cdot}272 .\end{aligned}$$

20. *Prvi način.* Naj bo A_l dogodek, da najdaljše začetno strnjeno strogo naraščajoče podzaporedje vsebuje l. Ta dogodek je unija paroma nezdružljivih dogodkov $A_{l1}, A_{l2}, \ldots, A_{ll}$, kjer je A_{lk} dogodek, da prvih k členov izbranega zaporedja tvori strogo naraščajoče podzaporedje, pri čemer je njegov zadnji člen enak l. Z drugimi besedami, prvih $k-1$ členov izbranega zaporedja tvori naraščajoče zaporedje števil od 1 do $l-1$, zadnji člen pa je enak l. Če naraščajoča zaporedja identificiramo z množicami, dobimo:

$$\mathbb{P}(A_{lk}) = \frac{\binom{l-1}{k-1}}{m^k} ,$$

torej je:

$$\begin{aligned}\mathbb{P}(A_l) &= \sum_{k=1}^{l} \mathbb{P}(A_{lk}) = \\ &= \sum_{k=1}^{l} \binom{l-1}{k-1} \frac{1}{m^k} = \\ &= \frac{1}{m} \sum_{i=0}^{l-1} \binom{l-1}{i} \left(\frac{1}{m}\right)^i = \\ &= \frac{1}{m}\left(1 + \frac{1}{m}\right)^{l-1} = \\ &= \frac{(m+1)^{l-1}}{m^l} .\end{aligned}$$

Drugi način. Opazimo, da je to, ali najdaljše začetno strnjeno strogo naraščajoče zaporedje vsebuje l, določeno že s prvimi l členi zaporedja. Zaporedjem l števil od 1 do m, za katera velja, da najdaljše začetno strnjeno strogo naraščajoče podzaporedje vsebuje l, recimo *dobra* zaporedja. Le-ta bomo prešteli s konstrukcijo ustrezne bijekcije.

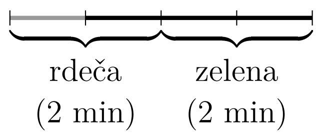
Dobra zaporedja so v očitni bijektivni korespondenci z nabori $(k; i_1, i_2, \ldots, i_l)$, kjer je $k \in \{1, 2, \ldots, l\}$, $i_1, \ldots, i_l \in \{1, 2, \ldots, m\}$ in $i_1 < i_2 < \cdots < i_k = l$. Označimo množico takšnih naborov z N_l. Vsakemu naboru iz N_l priredimo funkcijo $f\colon \{1, 2, \ldots, l-1\} \to \{0, 1, \ldots, m\}$ na naslednji način:

- Postavimo $f(i_1) := f(i_2) := \cdots := f(i_{k-1}) := 0$.
- Pišimo $\{1, 2, \ldots, l-1\} \setminus \{i_1, i_2, \ldots, i_{k-1}\} = \{j_1, j_2, \ldots, j_{l-k}\}$, pri čemer naj bo $j_1 < j_2 < \cdots < j_{l-k}$.
- Postavimo $f(j_1) := i_{k+1},\ f(j_2) := i_{k+2},\ \ldots,\ f(j_{l-k}) := i_l$.

Pokažimo, da je ta predpis bijekcija iz N_l v množico vseh funkcij iz $\{1, 2, \ldots, l-1\}$ v $\{0, 1, \ldots, m\}$, tj. da za vsako funkcijo $f\colon \{1, 2, \ldots, l-1\} \to \{0, 1, \ldots, m\}$ obstaja natanko en nabor iz N_l, ki se preslika vanjo. Če je $(k; i_1, i_2, \ldots, i_l)$ tak nabor, najprej opazimo, da f v 0 preslika natanko števila $i_1, i_2, \ldots, i_{k-1}$. Velja torej $f^{-1}(\{0\}) = \{i_1, i_2, \ldots, i_{k-1}\}$. Element k torej dobimo kot moč množice $f^{-1}(\{0\})$, povečano za 1. Nadalje števila $i_1, i_2, \ldots, i_{k-1}$ dobimo kot prasliko $f^{-1}(\{0\})$, urejeno v strogo naraščajoče zaporedje. Ko spet pišemo $\{1, 2, \ldots, l-1\} \setminus \{i_1, i_2, \ldots, i_{k-1}\} = \{j_1, j_2, \ldots, j_{l-k}\}$, kjer je $j_1 < j_2 < \cdots < j_{l-k}$, pa dobimo še indekse $i_{k+1} = f(j_1)$, $i_{k+2} = f(j_2), \ldots, i_l = f(j_{l-k})$. Tako je nabor natančno določen, pripada N_l in se preslika v funkcijo f. Dani predpis je torej res bijekcija.

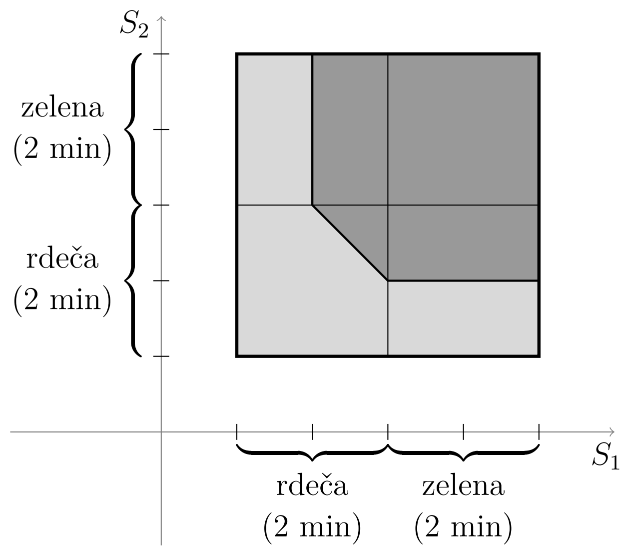
Ustreznih funkcij f je seveda $(m+1)^{l-1}$ in tolikšna je tudi moč množice N_l, torej je toliko tudi dobrih zaporedij. S tem je zgoraj dobljeni rezultat dokazan še bolj neposredno.

21. Za en semafor lahko verjetnostni prostor ponazorimo z naslednjo sliko:



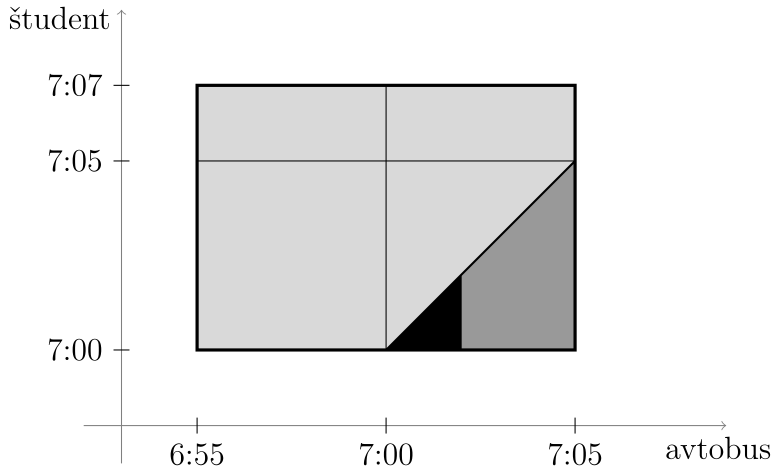
in verjetnost iskanega dogodka je $\dfrac{3}{4} = 0{'}75$.

Za dva semaforja pa je slika naslednja:

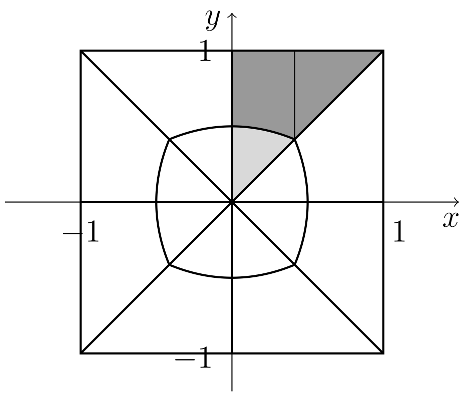


in verjetnost iskanega dogodka je $\dfrac{17}{32} \doteq 0{\cdot}531$.

22. a) $\dfrac{5}{28} \doteq 0{\cdot}179$, b) $\dfrac{1}{35} \doteq 0{\cdot}0286$. Slika:



23. Dani kvadrat postavimo v kooordinatni sistem (x, y), tako da se središče kvadrata ujema z izhodiščem koordinatnega sistema, da ima stranice vzporedne s koordinatnim križem in da je njegova stranica enaka 2. Nato ga razdelimo na osem trikotnikov, med katerimi ima vsak eno oglišče v središču kvadrata, drugo se ujema z enim od oglišč kvadrata, tretje pa se ujema z enim od presečišč roba kvadrata s koordinatnim križem, ki je najbližje prej omenjenemu oglišču kvadrata (glej sliko). Zaradi simetrije je dovolj, če se omejimo na en sam trikotnik.



V zgornjem desnem trikotniku bo tedaj točka bližje robu kot središču kvadrata (temnosivo) natanko tedaj, ko bo $1-y < \sqrt{x^2+y^2}$. Po ureditvi (in ob upoštevanju lege točke) dobimo, da je to natanko tedaj, ko je $y > (1-x^2)/2$ – v resnici gre za geometrijsko definicijo parabole. Mejni del parabole gre od točke $(0, 1/2)$ do točke (a, a), kjer je $a = 1/(1+\sqrt{2}) = \sqrt{2}-1$.

Ploščina celega trikotnika je $1/2$, ploščina ugodnega območja (temnosivo) pa je enaka:

$$\begin{aligned}\frac{(1-a)^2}{2} + \int_0^a \left(1 - \frac{1-x^2}{2}\right) \mathrm{d}x = \frac{(1-a)^2}{2} + \int_0^a \frac{1+x^2}{2}\,\mathrm{d}x &= \\ = \frac{3-3a+3a^2+a^3}{6}& \\ = \frac{8-4\sqrt{2}}{6}&\,.\end{aligned}$$

Iskana verjetnost je torej:

$$\frac{8-4\sqrt{2}}{3} \doteq 0{\cdot}781\,.$$

24. Recimo, da so črte vodoravne. Označimo s h razdaljo med točko na najbolj spodnjem delu igle in prvo črto, ki je nad to točko ($0 \le h < a$). Nadalje naj bo φ kot med iglo in pravokotnico na črte ($0 \le \varphi \le \pi/2$). Tedaj igla seka katero izmed črt natanko tedaj, ko je $h \le b\cos\varphi$.

Slepo metanje igle je smiselno interpretirati tako, da sta razdalja h in kot φ izbrana na slepo in neodvisno iz ustreznih intervalov. V tem primeru dobimo:

$$\mathbb{P} = \frac{2}{\pi a}\int_0^{\pi/2} \min\{a, b\cos\varphi\}\,\mathrm{d}\varphi = \begin{cases} \dfrac{2}{\pi}\left[\arccos\dfrac{a}{b} + \dfrac{b-\sqrt{b^2-a^2}}{a}\right] & ;\ a \le b \\ \dfrac{2b}{\pi a} & ;\ a \ge b \end{cases}$$

3. Pogojna verjetnost

1. Če je kocka poštena, velja $\mathbb{P}(A \mid L) = 1/3$ in $\mathbb{P}(B \mid L) = 0$. Če pa je nepoštena na način, opisan v nalogi, pride $\mathbb{P}(A \mid L) = 0{\cdot}25$ in še vedno $\mathbb{P}(B \mid L) = 0$.

2. Označimo z A dogodek, da je prva karta pik, z B pa dogodek, da sta med izvlečenimi kartami natanko dva pika.

Prvi način. Predstavljamo si, da so mesta kart posamezne barve fiksna, da pa jih slučajno vlečemo. S preprostim premislekom dobimo brezpogojni verjetnosti obeh dogodkov pa tudi pogojno verjetnost $\mathbb{P}(B \mid A)$:

$$\begin{aligned}
\mathbb{P}(A) &= \frac{1}{4}\,,\\
\mathbb{P}(B) &= \frac{\binom{4}{2}\binom{12}{2}}{\binom{16}{4}} = \binom{4}{2} \cdot \frac{4}{16} \cdot \frac{3}{15} \cdot \frac{12}{14} \cdot \frac{11}{13} = \frac{9 \cdot 11}{5 \cdot 7 \cdot 13} = \frac{99}{455}\,,\\
\mathbb{P}(B \mid A) &= \frac{\binom{3}{1}\binom{12}{2}}{\binom{15}{3}} = 3 \cdot \frac{3}{15} \cdot \frac{12}{14} \cdot \frac{11}{13} = \frac{2 \cdot 9 \cdot 11}{5 \cdot 7 \cdot 13} = \frac{198}{455}\,.
\end{aligned}$$

Sledi:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B \mid A) = \frac{99}{910}\,, \qquad \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{2}\,.$$

Drugi način. Predstavljamo si, da so mesta izvlečenih kart fiksna, barve pa se slučajno razporejajo. Vseh možnih izidov je $\binom{16}{4}$. Dogodek B vsebuje $\binom{4}{2}\binom{12}{2}$ izidov, dogodek $A \cap B$ pa $\binom{1}{1}\binom{3}{1}\binom{12}{2}$ izidov. Sledi:

$$\mathbb{P}(A \mid B) = \frac{\binom{1}{1}\binom{3}{1}\binom{12}{2}}{\binom{4}{2}\binom{12}{2}} = \frac{1}{2}\,.$$

Tretji način. Za izide v našem verjetnostnem prostoru postavimo kar konfiguracije, pri katerih opišemo, katere izvlečene karte so piki in katere ne. Tedaj vsi izidi sicer *niso* enako verjetni, pač pa so enako verjetni izidi, ki sestavljajo dogodek B. Teh je $\binom{4}{2} = 6$, trije med njimi so v dogodku A. Zato je $\mathbb{P}(A \mid B) = 1/2$.

3. Rešujemo podobno kot pri drugem načinu rešitve 16. naloge iz 2. razdelka, pri čemer za izide vzamemo vse možne razporeditve mest v kupu za srca in pike (ločimo torej, ali določeno mesto v kupu pripada srcu, piku ali čemu drugemu). Razporeditve, pri katerih so vsa srca med vrhnjimi 27 in vsi piki med spodnjimi 27 kartami, preštejemo tako, da srcem pripišemo zgornjo, pikom pa spodnjo polovico kupa, nakar ločimo možnosti, ali katero srce oziroma pik prestopi pripisano polovico kupa:

- Razporeditev, pri katerih so vsa srca so v zgornji, vsi piki pa v spodnji polovici, je $\binom{26}{13}^2$.
- Razporeditev, pri katerih srce prestopi, pik pa ne, ali pa obratno, je $2\binom{26}{12}\binom{25}{13}$.

- Razporeditve, pri katerih prestopi tako eno srce kot tudi en pik, so (v okviru osnovne zahteve) natančno tiste, pri katerih je 26. karta z dna srce, 26. z vrha pa pik. Le-teh je $\binom{25}{12}^2$.

Iskana pogojna verjetnost je torej enaka:

$$\frac{\binom{25}{12}^2}{\binom{26}{13}^2 + 2 \cdot \binom{26}{12}\binom{25}{13} + \binom{25}{12}^2} = \frac{1}{\left(\frac{26}{13}\right)^2 + 2 \cdot \frac{26}{14} + 1} = \frac{7}{61} \doteq 0{\cdot}115\,.$$

4. Razmislek je napačen: brezpogojni verjetnosti sta enaki, ne pa tudi pogojni verjetnosti glede na dogodek, da smo izvlekli zlat kovanec.

5. Imamo dve možnosti: ali ostanemo pri prvotni izbiri ali pa se premislimo. Ker so (na začetku) vse možnosti za vrata, za katerimi se skriva avto, enako verjetne, velja, da, če se ne premislimo, dobimo avto z verjetnostjo 1/3. Bolj formalno lahko napišemo:

$$\mathbb{P}_{\text{ostanemo pri prvotni izbiri}}(\text{dobimo avto}) = \frac{1}{3}$$

(strategija je zapisana spodaj pri $\mathbb{P}$ in ne kot pogoj – to je še vedno brezpogojna verjetnost). Nadalje opazimo, da, če se premislimo, dobimo ravno nasprotno, kot če ostanemo pri prvotni izbiri. Torej je:

$$\mathbb{P}_{\text{se premislimo}}(\text{dobimo avto}) = \frac{2}{3}\,.$$

Bolje se je torej premisliti – vsaj če se že na začetku igre odločimo, ali se bomo premislili ali ne, ne glede na to, katera vrata bo odprl vodja igre.

Razmislek iz besedila naloge je torej napačen. Da vidimo, zakaj, zapišimo vse možne izide. Brez škode za splošnost lahko privzamemo, da sprva izberemo prva vrata. Če se za njimi skriva avto, ima vodja igre izbiro, katera vrata odpre, sicer pa izbire nima. Izide lahko torej zapišemo kot AOB, ABO, BAO in BOA: črka A pomeni avto za (še zaprtimi) vrati, črka B bučo za zaprtimi vrati, črka O pa odprta vrata.

Razmislek iz besedila naloge je napačen, ker so v njem pomešane pogojne in brezpogojne verjetnosti. Privzemimo najprej, da vodja igre v primeru, ko ima možnost izbire, izbira povsem naključno – da torej nobenih vrat ne favorizira. Tedaj verjetnosti izidov ustrezajo shemi:

$$\begin{pmatrix} AOB & ABO & BAO & BOA \\ 1/6 & 1/6 & 1/3 & 1/3 \end{pmatrix}.$$

Brezpogojni verjetnosti sta res enaki:

$$\mathbb{P}(\text{avto za prvimi vrati}) = \mathbb{P}(\text{avto za drugimi vrati}) = \frac{1}{3}\,,$$

to pa ne drži za pogojni verjetnosti:

$$\begin{aligned}\mathbb{P}\big(\text{avto za prvimi vrati} \bigm| \text{tretja vrata odprta}\big) &= \frac{\mathbb{P}(\{\text{ABO}\})}{\mathbb{P}(\{\text{ABO}, \text{BAO}\})} = \frac{1}{3}\,,\\ \mathbb{P}\big(\text{avto za drugimi vrati} \bigm| \text{tretja vrata odprta}\big) &= \frac{\mathbb{P}(\{\text{BAO}\})}{\mathbb{P}(\{\text{ABO}, \text{BAO}\})} = \frac{2}{3}\,.\end{aligned}$$

Pogojni verjetnosti sta drugačni od brezpogojnih: brž ko je vodja igre odprl tretja vrata, nam je s tem dal dodatno informacijo. Še drugače: ne moremo uporabiti formule za pogojno verjetnost pri enako verjetnih izidih, ker izida ABO in BAO nista enako verjetna.

Dogodka, da vodja igre odpre tretja vrata, ne smemo zamešati z dogodkom, da je za tretjimi vrati buča: prvo je dogodek $\{ABO, BAO\}$, drugo pa je dogodek $\{AOB, ABO, BAO\}$. Tudi pogojni verjetnosti:

$$\begin{aligned}&\mathbb{P}\big(\text{avto za prvimi vrati} \bigm| \text{buča za tretjimi vrati}\big)\\ &\qquad= \mathbb{P}\big(\text{avto za drugimi vrati} \bigm| \text{buča za tretjimi vrati}\big) = \frac{1}{2}\end{aligned}$$

se ne ujemata s pogojnima verjetnostma glede na dogodek, da vodja igre odpre tretja vrata. Pogojevanje na dogodek, da je za tretjimi vrati buča, ustreza alternativnemu protokolu igre, po katerem bi vodja igre najprej odprl vrata, ki bi jih izbral igralec, nakar bi igralec izbral ena od še neodprtih vrat in dobil tisto, kar se skriva za njimi.

Omeniti velja še, da odprtje tretjih vrat ne spremeni verjetnosti, da je avto za prvimi vrati:

$$\mathbb{P}\big(\text{avto za prvimi vrati} \bigm| \text{tretja vrata odprta}\big) = \frac{1}{3} = \mathbb{P}\big(\text{avto za prvimi vrati}\big)\,.$$

Seveda pa odprtje tretjih vrat spremeni verjetnosti za druga in tretja vrata: za druga jo dvigne z $1/3$ na $2/3$, za tretja pa jo spusti z $1/3$ na 0.

Informacija, da je vodja igre odprl prva vrata, pa lahko spremeni tudi verjetnost, da je avto za prvimi vrati, če vodja igra favorizira katera od vrat. Če recimo favorizira tretja vrata v smislu, da jih odpre vsakič, ko jih le lahko, verjetnosti izidov ustrezajo shemi:

$$\begin{pmatrix} AOB & ABO & BAO & BOA \\ 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}$$

in velja:

$$\begin{aligned}&\mathbb{P}\big(\text{avto za prvimi vrati} \bigm| \text{tretja vrata odprta}\big)\\ &\qquad= \mathbb{P}\big(\text{avto za drugimi vrati} \bigm| \text{tretja vrata odprta}\big) = \frac{1}{2}\,,\end{aligned}$$

medtem ko je:

$$\begin{aligned}\mathbb{P}\big(\text{avto za prvimi vrati} \bigm| \text{druga vrata odprta}\big) &= 0\,,\\ \mathbb{P}\big(\text{avto za tretjimi vrati} \bigm| \text{druga vrata odprta}\big) &= 1\,.\end{aligned}$$

V primeru, ko vodja igre favorizira tretja vrata, se dogodek, da vodja igra odpre tretja vrata, in dogodek, da je za tretjimi vrati buča, razlikujeta le za dogodek z verjetnostjo nič.

6. a) Vodja igre lahko odpre druga vrata, če je nagrada za prvimi ali za tretjimi vrati. V prvem primeru vodja igre odpre druga vrata s pogojno verjetnostjo $1/4$ in igralec dobi nagrado. V drugem primeru pa vodja igre z gotovostjo odpre druga vrata in igralec ne dobi nagrade. Ustrezna pogojna verjetnost je torej enaka:

$$\frac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{5}} = \frac{5}{13}\,.$$

Ko vodja igre odpre vrata, je vedno za enimi še neodprtimi vrati nagrada, za drugimi pa ne. Če se igralec premisli, je torej ustrezna pogojna verjetnost enaka $\frac{8}{13}$, kar je več. Torej se igralcu splača premisliti.

Če vodja igre odpre tretja vrata in igralec ostane pri istem, pa je ustrezna pogojna verjetnost enaka:

$$\frac{\frac{1}{3}\cdot\frac{3}{4}}{\frac{1}{2}\cdot\frac{3}{4}+\frac{3}{10}} = \frac{5}{9}\,,$$

če se igralec premisli, pa je ustrezna pogojna verjetnost enaka $\frac{4}{9}$. Če vodja igre odpre tretja vrata, se torej igralcu splača ostati pri istem.

b) Ločimo naslednje možnosti:

- Če je nagrada za prvimi vrati in vodja igre odpre druga vrata, se bo igralec premislil, odprl bo tretja vrata in ne bo dobil nagrade.
- Če je nagrada za prvimi vrati in vodja igre odpre tretja vrata, se igralec ne bo premislil in bo dobil nagrado.
- Če je nagrada za drugimi vrati, bo vodja igre odprl tretja vrata, igralec se ne bo premislil in ne bo dobil nagrade.
- Če je nagrada za tretjimi vrati, bo vodja igre dobil druga vrata, igralec se bo premislil in bo dobil nagrado.

Iskana verjetnost je tako enaka:

$$\frac{1}{2}\cdot\frac{3}{4}+\frac{1}{5} = \frac{23}{40}\,,$$

kar je več kot v primeru, če ne gleda na to, katera vrata odpre vodja igre (tedaj je verjetnost, da dobi nagrado, enaka $\frac{1}{2}$, če se premisli ali ne).

c) Igralcu se bolj splača v prvem koraku pokazati na tretja vrata, potem pa se premisliti ne glede na to, katera vrata odpre vodja igre. V tem primeru bo namreč dobil nagrado, brž ko le-ta ne bo za tretjimi vrati, to pa se zgodi z verjetnostjo $\frac{4}{5}$, kar je več kot $\frac{23}{40}$.

7. Končni izid je lahko:

- 6 : 2, če v nadaljnjih dveh rundah zmaga Janez;
- 6 : 3, če v nadaljnjih dveh rundah enkrat zmaga Janez in enkrat Peter, v tretji rundi pa zmaga Janez;
- 6 : 4, če v nadaljnjih treh rundah enkrat zmaga Janez in dvakrat Peter, v četrti rundi pa zmaga Janez;
- 6 : 5, če v nadaljnjih štirih rundah enkrat zmaga Janez in trikrat Peter, v peti rundi pa zmaga Janez.

Iskana verjetnost je zato enaka:

$$\left(\frac{1}{2}\right)^2 + 2 \cdot \left(\frac{1}{2}\right)^3 + 3 \cdot \left(\frac{1}{2}\right)^4 + 4 \cdot \left(\frac{1}{2}\right)^5 = \frac{13}{16} = 0{\dot{}}8125\,.$$

Še hitreje pridemo do rezultata, če vzamemo nasprotni dogodek (izida 4 : 6 in 5 : 6).

8. $\frac{5}{10} \cdot \frac{4 \cdot 3}{\binom{7}{2}} + \frac{3}{10} \cdot \frac{3 \cdot 3}{\binom{7}{2}} + \frac{2}{10} \cdot \frac{4 \cdot 3}{\binom{7}{2}} = \frac{37}{70} \doteq 0{\dot{}}529.$

9. V splošnem to ni res: za protiprimer lahko postavimo verjetnostni prostor $\Omega = \{0, 1, 2, 3\}$, v katerem so vsi izidi enako verjetni, ter dogodke $A = \{0, 1\}$, $B = \{0, 2\}$ in $C = \{0, 3\}$: velja $\mathbb{P}(A \mid B) = \mathbb{P}(A \mid C) = 1/2$, toda $\mathbb{P}(A \mid B \cup C) = 1/3$.

Trditev pa drži ob dodatni predpostavki, da sta dogodka B in C nezdružljiva. Splošneje velja, da, če so $H_1, H_2, \ldots, H_n$ oziroma $H_1, H_2, H_3, \ldots$ paroma nezdružljivi dogodki z unijo H in $\mathbb{P}(A \mid H_i) = p$ za vse i, velja tudi $\mathbb{P}(A \mid H) = p$. Dogodek $A \cap H$ je namreč disjunktna unija dogodkov $A \cap H_1, A \cap H_2, \ldots, A \cap H_n$ oziroma $A \cap H_1, A \cap H_2, A \cap H_3, \ldots$, zato je:

$$\begin{aligned}
\mathbb{P}(A \cap H) &= \mathbb{P}(A \cap H_1) + \mathbb{P}(A \cap H_2) + \cdots + \mathbb{P}(A \cap H_n) = \\
&= p\,\mathbb{P}(H_1) + p\,\mathbb{P}(H_2) + \cdots + p\,\mathbb{P}(H_n) = \\
&= p\,\mathbb{P}(H)
\end{aligned}$$

oziroma

$$\begin{aligned}
\mathbb{P}(A \cap H) &= \mathbb{P}(A \cap H_1) + \mathbb{P}(A \cap H_2) + \mathbb{P}(A \cap H_3) + \cdots = \\
&= p\,\mathbb{P}(H_1) + p\,\mathbb{P}(H_2) + p\,\mathbb{P}(H_3) + \cdots = \\
&= p\,\mathbb{P}(H)\,,
\end{aligned}$$

od koder sledi, da je $\mathbb{P}(A \mid H) = p$.

10. Naj bo A dogodek, da Janez dobi dvoboj. Njegovo verjetnost lahko izračunamo na vsaj dva načina.

Prvi način: neposredno, kot vsoto geometrijske vrste:

$$\mathbb{P}(A) = \frac{1}{9} + \frac{4}{9} \cdot \frac{1}{9} + \frac{4}{9} \cdot \frac{4}{9} \cdot \frac{1}{9} + \cdots = \frac{1}{9} \cdot \frac{1}{1 - \frac{4}{9}} = \frac{5}{9}\,.$$

Drugi način: s pogojevanjem na rezultat po prvih dveh rundah. Glede na ta rezultat definirajmo dogodke $H_{2:0}$, $H_{1:1}$ in $H_{0:2}$. Le-ti tvorijo popoln sistem dogodkov. Velja:

$$\begin{aligned} \mathbb{P}(H_{2:0}) &= \frac{1}{9}\,, & \mathbb{P}(A \mid H_{2:0}) &= 1\,, \\ \mathbb{P}(H_{1:1}) &= \frac{4}{9}\,, & \mathbb{P}(A \mid H_{1:1}) &= p\,, \\ \mathbb{P}(H_{0:2}) &= \frac{4}{9}\,, & \mathbb{P}(A \mid H_{0:2}) &= 0\,. \end{aligned}$$

Če označimo $p := \mathbb{P}(A)$, po izreku o polni verjetnosti velja:

$$\begin{aligned} p = \mathbb{P}(A) &= \mathbb{P}(H_{2:0})\,\mathbb{P}(A \mid H_{2:0}) + \mathbb{P}(H_{1:1})\,\mathbb{P}(A \mid H_{1:1}) + \mathbb{P}(H_{0:2})\,\mathbb{P}(A \mid H_{0:2}) = \\ &= \frac{1}{9} + \frac{4p}{9}\,. \end{aligned}$$

Rešimo enačbo in dobimo $p = 1/5$.

Opomba. Pogojni verjetnosti $\mathbb{P}(A \mid H_{2:0})$ in $\mathbb{P}(A \mid H_{0:2})$ sta res očitni, verjetnost $\mathbb{P}(A \mid H_{1:1})$ pa je le videti očitna, njena utemeljitev pa zahteva malo premisleka. Prejšnja naloga pokaže, da se da izpeljati iz pogojnih verjetnosti:

$$\begin{aligned} &\mathbb{P}\big(A \mid \text{prvo rundo dobi Janez, drugo pa Peter}\big) = p\,, \\ &\mathbb{P}\big(A \mid \text{prvo rundo dobi Peter, drugo pa Janez}\big) = p\,, \end{aligned}$$

a tudi ti dve pogojni verjetnosti nista očitni. Za posamezno zgodovino smo namreč privzeli samo pogojni verjetnosti dogodkov, da Janez ali Peter dobi naslednjo rundo: privzeli smo, da sta ti dve verjetnosti neodvisni od zgodovine. Iz tega najprej izpeljemo, da to velja tudi za dogodke, ki se nanašajo na točno določene poteke nadaljnjih končno mnogo zaporednih rund, npr. da naslednjo rundo dobi Peter, nato pa Janez dvakrat zmaga. Če sta J_k in P_k dogodka, da Janez oziroma Peter v k-ti rundi zmaga, in Z poljubna zgodovina dolžine n, tako velja:

$$\begin{aligned} &\mathbb{P}(P_{n+1} \cap J_{n+2} \cap J_{n+3} \mid Z) = \\ &\qquad = \mathbb{P}(P_{n+1} \mid Z)\,\mathbb{P}(J_{n+2} \cap J_{n+3} \mid Z \cap P_{n+1}) = \\ &\qquad = \mathbb{P}(P_{n+1} \mid Z)\,\mathbb{P}(J_{n+2} \mid Z \cap P_{n+1})\,\mathbb{P}(J_{n+3} \mid Z \cap P_{n+1} \cap J_{n+2}) = \\ &\qquad = \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{1}{3}\,. \end{aligned}$$

Nadalje opisana neodvisnost od zgodovine velja tudi za disjunktne števne unije takih potekov. Ena od teh unij pa za vsako zgodovino, ki se konča z izenačenim izidom, ustreza dogodku, da Janez zmaga. Omenjeni pogojni verjetnosti sledita.

Če bi obravnavali dogodek, ki se ne da izraziti samo s končnimi poteki, npr. limitno nihanje razmerja med rundami, ki jih je dobil posamezen igralec, pa bi bila teoretična utemeljitev še bistveno težja. Za take primere je na voljo *Dynkinova lema*, znana tudi kot *izrek* π–λ.

11. Tu ima smisel definirati dva verjetnostna prostora: prvi se nanaša na *posamezen* poskus, na njem sta definirana dogodka A in B in ima verjetnostno mero $\mathbb{P}_1$. Drugi pa se nanaša na celotno zaporedje poskusov: na njem so definirani dogodki A_n in B_n, da je v n-tem poskusu prišlo do opažanja A oziroma B, verjetnostno mero na njem pa označimo s $\mathbb{P}$. Tedaj velja:

$$\begin{aligned}\mathbb{P}(A_n \mid Z_1, Z_2, \ldots, Z_{n-1}) &= \mathbb{P}_1(A)\,,\\ \mathbb{P}(B_n \mid Z_1, Z_2, \ldots, Z_{n-1}) &= \mathbb{P}_1(B)\,,\\ \mathbb{P}(A_n \cap B_n \mid Z_1, Z_2, \ldots, Z_{n-1}) &= \mathbb{P}_1(A \cap B)\,,\end{aligned}$$

kjer je lahko za vsak $k = 1, \ldots, n-1$ dogodek Z_k kateri koli dogodek s strogo pozitivno verjetnostjo, ki se nanaša zgolj na k-ti poskus. Dogodki, ki se nanašajo zgolj na k-ti poskus, tvorijo σ-algebro, ki vsebuje dogodka A_k in B_k. Med drugim lahko torej za Z_k vzamemo tudi gotovi dogodek, torej je tudi:

$$\mathbb{P}(A_n) = \mathbb{P}_1(A)\,, \qquad \mathbb{P}(B_n) = \mathbb{P}_1(B) \qquad \text{in} \qquad \mathbb{P}(A_n \cap B_n) = \mathbb{P}_1(A \cap B)\,.$$

Želena verjetnost je verjetnost dogodka:

$$A_B := (A_1 \cap B_1) \cup (B_1^{\mathrm{c}} \cap A_2 \cap B_2) \cup (B_1^{\mathrm{c}} \cap B_2^{\mathrm{c}} \cap A_3 \cap B_3) \cup \cdots$$

Nadaljujemo lahko na vsaj tri načine.

Prvi način: neposredno. Glede na to, da gre pri zgornjem zapisu dogodka A_B za disjunktno unijo, se njegova verjetnost izraža kot vrsta:

$$\begin{aligned}\mathbb{P}(A_B) &= \mathbb{P}_1(A \cap B) + \mathbb{P}_1(B_1^{\mathrm{c}})\,\mathbb{P}_1(A \cap B) + \big(\mathbb{P}_1(B_1^{\mathrm{c}})\big)^2\,\mathbb{P}_1(A \cap B) + \cdots = \\ &= \frac{\mathbb{P}_1(A \cap B)}{1 - \mathbb{P}_1(B^{\mathrm{c}})} = \\ &= \mathbb{P}_1(A \mid B)\,.\end{aligned}$$

Drugi način: s pogojevanjem na dogodek B_1. Po izreku o polni verjetnosti velja:

$$\mathbb{P}(A_B) = \mathbb{P}(B_1)\,\mathbb{P}(A_B \mid B_1) + \mathbb{P}(B_1^{\mathrm{c}})\,\mathbb{P}(A_B \mid B_1^{\mathrm{c}})\,.$$

Na dogodku B_1 se dogodek A_B ujema z dogodkom A_1, zato je $\mathbb{P}(A_B \mid B_1) = \mathbb{P}(A_1 \mid B_1) = \mathbb{P}_1(A \mid B)$. Če se zgodi B_1^{c}, pa se vse začne na novo: s podobnim premislekom kot pri prejšnji nalogi se da izpeljati, da je $\mathbb{P}(A_B \mid B_1) = \mathbb{P}(A_B)$. Torej je:

$$\mathbb{P}(A_B) = \mathbb{P}(B_1)\,\mathbb{P}_1(A \mid B) + \big(1 - \mathbb{P}(B_1)\big)\,\mathbb{P}(A_B)\,.$$

Rešimo enačbo in dobimo $\mathbb{P}(A) = \mathbb{P}_1(A \mid B)$.

Tretji način. Za $n = 1, 2, 3, \ldots$ naj bo:

$$H_n := B_1^{\mathrm{c}} \cap B_2^{\mathrm{c}} \cap \cdots \cap B_{n-1}^{\mathrm{c}} \cap B_n$$

dogodek, da je v n-tem poskusu prišlo do opažanja B, prej pa ne. Nadalje naj bo $H = H_1 \cup H_2 \cup \cdots$ dogodek, da se izvajanje poskusov nekoč konča. Ker gre za disjunktno unijo, je:

$$\begin{aligned}
\mathbb{P}(H) &= \mathbb{P}(H_1) + \mathbb{P}(H_2) + \cdots = \\
&= \mathbb{P}_1(B) + \big(1 - \mathbb{P}_1(B)\big)\,\mathbb{P}(B) + \big(1 - \mathbb{P}_1(B)\big)^2\,\mathbb{P}(B) + \cdots = \\
&= \frac{\mathbb{P}_1(B)}{1 - \big(1 - \mathbb{P}_1(B)\big)} = \\
&= 1\,.
\end{aligned}$$

Torej je $\mathbb{P}(A_B) = \mathbb{P}(A_B \mid H)$. Nadalje za vsak n velja:

$$\begin{aligned}
\mathbb{P}(A_B \mid H_n) &= \frac{\mathbb{P}(A_B \cap H_n)}{\mathbb{P}(H_n)} = \\
&= \frac{\mathbb{P}(B_1^{\mathrm{c}} \cap B_2^{\mathrm{c}} \cap \cdots \cap B_{n-1}^{\mathrm{c}} \cap A_n \cap B_n)}{\mathbb{P}(B_1^{\mathrm{c}} \cap B_2^{\mathrm{c}} \cap \cdots \cap B_{n-1}^{\mathrm{c}} \cap B_n)} = \\
&= \frac{\big(\mathbb{P}_1(B^{\mathrm{c}})\big)^{n-1}\,\mathbb{P}_1(A \cap B)}{\big(\mathbb{P}_1(B^{\mathrm{c}})\big)^{n-1}\,\mathbb{P}_1(B)} = \\
&= \mathbb{P}_1(A \cap B)\,.
\end{aligned}$$

Iz rešitve 9. naloge dobimo, da je tudi $\mathbb{P}(A_B) = \mathbb{P}(A_B \mid H) = \mathbb{P}_1(A \cap B)$.

Opomba. Ta naloga nudi še eno možno rešitev prejšnje naloge. Za n-ti poskus namreč vzamemo $(2n-1)$-to in $(2n)$-to rundo. Na verjetnostnem prostoru, ki se nanaša na posamezen poskus, definiramo dogodek J, da v obeh rundah zmaga Janez, in dogodek I, da v obeh rundah zmaga isti igralec. Dogodek A, da Janez dobi dvoboj, se tedaj ujema z dogodkom, da se pri prvem poskusu, ko se zgodi dogodek I, zgodi tudi dogodek J. Očitno je $J \subseteq I$ in iz rezultata te naloge dobimo:

$$\mathbb{P}(A) = \mathbb{P}_1(J \mid I) = \frac{\mathbb{P}_1(J)}{\mathbb{P}_1(I)} = \frac{\frac{1}{9}}{\frac{5}{9}} = \frac{1}{5}\,.$$

12. Naj bo A dogodek, da nekoč izvlečemo kroglico, ki ni bela, in da ji sledi kroglica iste barve. Nadalje za $n = 1, 2, 3, \ldots$ definirajmo dogodek R_n, da v n-tem vlečenju izvlečemo rdečo kroglico, pred tem pa smo izvlekli same bele, Z_n pa naj bo dogodek, da v n-tem vlečenju izvlečemo zeleno kroglico, pred tem pa smo izvlekli same bele. Očitno je:

$$\mathbb{P}(A \mid R_n) = \frac{r}{r+z+b} \quad \text{in} \quad \mathbb{P}(A \mid Z_n) = \frac{z}{r+z+b}\,.$$

Če je $R := R_1 \cup R_2 \cup \cdots$ dogodek, da je prva izvlečena kroglica, ki ni bela, rdeča, $Z := Z_1 \cup Z_2 \cup \cdots$ pa dogodek, da je prva izvlečena kroglica, ki ni bela, zelena, iz rešitve 9. naloge sledi, da je tudi

$$\mathbb{P}(A \mid R) = \frac{r}{r+z+b} \quad \text{in} \quad \mathbb{P}(A \mid Z) = \frac{z}{r+z+b}\,.$$

Iz prejšnje naloge pa sledi, da je:

$$\mathbb{P}(R) = \frac{r}{r+z} \quad \text{in} \quad \mathbb{P}(Z) = \frac{z}{r+z}$$

in še, da je $\mathbb{P}(R \cup Z) = 1$. Po izreku o polni verjetnosti za skorajšnje particije je:

$$\mathbb{P}(A) = \mathbb{P}(R)\,\mathbb{P}(A \mid R) + \mathbb{P}(Z)\,\mathbb{P}(A \mid Z) = \frac{r^2 + z^2}{(r+z)(r+z+b)}\,.$$

13. Označimo iskano verjetnost s p_n. Očitno je $p_0 = 1$ in $p_1 = 3/4$, za $n = 2$ pa lahko pogojujemo na naslednji popoln sistem dogodkov:

$$\begin{aligned} H_{\mathrm{G}} &:= \{\text{v prvem metu je padel grb}\}\,, \\ H_{\mathrm{CG}} &:= \{\text{v prvem metu je padla cifra, v drugem pa grb}\}\,, \\ H_{\mathrm{CC}} &:= \{\text{v prvem in drugem metu je padla cifra}\}\,. \end{aligned}$$

Označimo z A_n dogodek, da v prvih n metih nista padli dve zaporedni cifri, se pravi, da je $\mathbb{P}(A_n) = p_n$. Oglejmo si še pogojne verjetnosti glede na prej definirani popoln sistem dogodkov. Očitno je $\mathbb{P}(A_n \mid H_{\mathrm{CC}}) = 0$. Pri H_{G} in H_{CG} pa se vse začne na novo, le da imamo en oziroma dva meta manj. Podobno kot v 10. nalogi lahko utemeljimo, da je $\mathbb{P}(A_n \mid H_{\mathrm{G}}) = p_{n-1}$ in $\mathbb{P}(A_n \mid H_{\mathrm{CG}}) = p_{n-2}$. Po izreku o polni verjetnosti je:

$$p_n = \tfrac{1}{2} p_{n-1} + \tfrac{1}{4} p_{n-2}\,.$$

Glede na to, da smo p_1 in p_2 že izračunali, so verjetnosti p_n zdaj enolično podane in se lahko izračunajo rekurzivno. Zveza je podobna tisti za Fibonaccijevo zaporedje $F_1, F_2, F_3, F_4, F_5, F_6 \ldots = 1, 1, 2, 3, 5, 8, \ldots$ Če postavimo $a_n := 2^n p_n$, to zaporedje zadošča zvezi $a_n = a_{n-1} + a_{n-2}$, ki je natančno ista kot za Fibonaccijevo zaporedje. Izračunamo $a_1 = 2, a_2 = 3$ in tako dobimo $a_n = F_{n-2}$ oziroma $p_n = \frac{F_{n+2}}{2^n}$, eksplicitna formula pa je:

$$p_n = \frac{4}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{4} \right)^{n+2} - \left(\frac{1-\sqrt{5}}{4} \right)^{n+2} \right].$$

Opomba. Pogojni verjetnosti $\mathbb{P}(A_n \mid H_{\mathrm{G}}) = p_{n-1}$ in $\mathbb{P}(A_n \mid H_{\mathrm{CG}}) = p_{n-2}$ nista očitni, saj smo za posamezno zgodovino privzeli samo pogojni verjetnosti dogodkov, da v naslednjem metu pade cifra oziroma grb: privzeli smo, da sta ti dve verjetnosti neodvisni od zgodovine. Iz tega najprej izpeljemo, da to velja tudi poljubno predpisano zaporedje nadaljnjih končno mnogo metov, npr. da v naslednjem metu pade cifra, nato grb, nato pa spet cifra. To pa velja tudi za števne disjunktne unije takih potekov, taka unija pa je tudi dogodek, da v nadaljnjih $n-1$ oziroma $n-2$ metih nista padli dve zaporedni cifri. To utemelji omenjeni pogojni verjetnosti.

Opomba. Naloga izpelje, da obstaja natanko F_{n+2} zaporedij n cifer in grbov, med katerimi ni dveh zaporednih cifer. Izpeljavo pa lahko nadgradimo v konstrukcijo bijekcije med množico takih zaporedij in množico $\mathcal{B}_n := \{0, 1, 2, \ldots, F_{n+1} - 1\}$: zaporedju cifer in grbov najprej priredimo zaporedje $(e_1, e_2, \ldots, e_n)$, pri čemer je

$e_k = 1$, če je na k-tem mestu cifra, in $x_k = 0$, če je na k-tem mestu grb. Označimo z $\mathcal{A}_n$ množico vseh zaporedij n ničel in enic, ki nimajo zaporednih enic. Zdaj pa vsakemu zaporedju $(e_1, e_2, \ldots, e_n) \in \mathcal{A}_n$ priredimo vsoto $\sum_{k=1}^n e_k F_{k+1}$. Pokažimo, da ta vsota res pripada množici $\mathcal{B}_n$ in da za vsak $x \in \mathcal{B}_n$ obstaja natanko eno zaporedje $(e_1, e_2, \ldots, e_n) \in \mathcal{A}_n$, za katerega je $x = \sum_{k=1}^n e_k F_{k+1}$.

To bomo dokazali z indukcijo, pri čemer bosta baza indukcije primera $n = 1$ in $n = 2$. Velja $\mathcal{A}_1 = \{0, 1\}$. Ker je $F_2 = 1$, se 0 preslika v 0 in 1 v 1, to pa sta natanko elementa množice $\mathcal{B}_1$, saj je $F_3 = 2$. Nadalje je $\mathcal{A}_2 = \{(0,0), (1,0), (0,1)\}$ in ti elementi se preslikajo v 0, 1 oziroma 2. To pa so natanko elementi množice $\mathcal{B}_2$, saj je $F_4 = 3$.

Naredimo zdaj indukcijski korak z $n-2$ in $n-1$ na n. Najprej pokažimo, da za $(e_1, e_2, \ldots, e_n) \in \mathcal{A}_n$ velja $\sum_{k=1}^n e_k F_{k+1} \in \mathcal{B}_n$ ali ekvivalentno $\sum_{k=1}^n e_k F_{k+1} < F_{n+2}$. Če je $e_n = 0$, je po indukcijski predpostavki $\sum_{k=1}^n e_k F_{k+1} = \sum_{k=1}^{n-1} e_k F_{k+1} < F_{n+1} < F_{n+2}$. Če pa je $e_n = 1$, mora biti $e_{n-1} = 0$, potem pa spet iz indukcijske predpostavke dobimo, da je $\sum_{k=1}^n e_k F_{k+1} = \sum_{k=1}^{n-2} e_k F_{k+1} + F_{n+1} < F_n + F_{n+1} = F_{n+2}$.

Končno naredimo indukcijski korak še za trditev, da $x \in \mathcal{B}_n$ obstaja natanko eno zaporedje $(e_1, e_2, \ldots, e_n) \in \mathcal{A}_n$, za katerega je $x = \sum_{k=1}^n e_k F_{k+1}$. Če je $x < F_{n+1}$, mora biti $e_n = 0$, potem pa je ekvivalentno dokazati, da obstaja natanko eno zaporedje $(e_1, e_2, \ldots, e_{n-1}) \in \mathcal{A}_{n-1}$, za katerega je $x = \sum_{k=1}^{n-1} e_k F_{k+1}$. To pa velja po indukcijski predpostavki, saj je $x \in \mathcal{B}_{n-1}$. Če pa je $x \ge F_{n+1}$, iz prejšnje že dokazane trditve sledi, da mora biti $e_n = 1$, potem pa mora biti $e_{n-1} = 0$. V tem primeru je torej ekvivalentno dokazati, da obstaja natanko eno zaporedje $(e_1, e_2, \ldots, e_{n-2}) \in \mathcal{A}_{n-2}$, za katerega je $x - F_{n+1} = \sum_{k=1}^{n-1} e_k F_{k+1}$. A v tem primeru je $0 \le x - F_{n+1} < F_{n+2} - F_{n+1} = F_n$, torej je $x - F_{n+1} \in \mathcal{B}_{n-2}$ in rezultat spet dobimo iz indukcijske predpostavke.

14. Če s H_F označimo dogodek, da mož kupi solato pri Francki, s H_M dogodek, da kupi pri Micki, z G pa dogodek, da je kupil nagnito solato, velja:

$$\mathbb{P}(H_F \mid G) = \frac{0\text{'}6 \cdot 0\text{'}1}{0\text{'}6 \cdot 0\text{'}1 + 0\text{'}4 \cdot 0\text{'}2} \doteq 0\text{'}429\,,$$
$$\mathbb{P}(H_M \mid G) = \frac{0\text{'}4 \cdot 0\text{'}2}{0\text{'}6 \cdot 0\text{'}1 + 0\text{'}4 \cdot 0\text{'}2} \doteq 0\text{'}571\,.$$

Torej žena bolj upravičeno sumi Micko.

15. $\dfrac{0\text{'}2 \cdot 0\text{'}05}{0\text{'}2 \cdot 0\text{'}05 + 0\text{'}5 \cdot 0\text{'}01 + 0\text{'}3 \cdot 0\text{'}02} \doteq 0\text{'}476.$

16. $$\frac{\dfrac{\binom{4}{3}}{\binom{10}{3}} \cdot \dfrac{4}{5} \cdot \dfrac{3}{4}}{\dfrac{\binom{4}{3}}{\binom{10}{3}} \cdot \dfrac{4}{5} \cdot \dfrac{3}{4} + \dfrac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} \cdot \dfrac{3}{5} \cdot \dfrac{2}{4} + \dfrac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} \cdot \dfrac{2}{5} \cdot \dfrac{1}{4}} = 0\text{'}125.$$

17. Za $i = 1, 2, 3$ naj bo H_i dogodek, da je bil samoglasnik med prvimi tremi črkami prikazan kot i-ti, H_{45} pa dogodek, da sploh ni bil prikazan. Nadalje naj bo U_3

dogodek, da Manja ugane besedo po natanko treh prikazanih črkah. Tedaj velja:

$$\mathbb{P}(H_1) = \mathbb{P}(H_2) = \mathbb{P}(H_3) = \frac{1}{5}\,, \quad \mathbb{P}(H_{45}) = \frac{2}{5}\,.$$

Nadalje je:

$$\begin{aligned}
\mathbb{P}(U_3 \mid H_1) &= 0\text{˙}95 \cdot 0\text{˙}9 \cdot 0\text{˙}8 \cdot 0\text{˙}6 = 0\text{˙}4104\,,\\
\mathbb{P}(U_3 \mid H_2) &= 0\text{˙}95 \cdot 0\text{˙}85 \cdot 0\text{˙}8 \cdot 0\text{˙}6 = 0\text{˙}3876\,,\\
\mathbb{P}(U_3 \mid H_3) &= 0\text{˙}95 \cdot 0\text{˙}85 \cdot 0\text{˙}6 \cdot 0\text{˙}6 = 0\text{˙}2907\,,\\
\mathbb{P}(U_3 \mid H_{45}) &= 0\text{˙}95 \cdot 0\text{˙}85 \cdot 0\text{˙}6 \cdot 0\text{˙}7 = 0\text{˙}33915\,.
\end{aligned}$$

Po izreku o polni verjetnosti dobimo:

$$\mathbb{P}(U_3) = \frac{1}{5} \cdot 0\text{˙}4104 + \frac{1}{5} \cdot 0\text{˙}3876 + \frac{1}{5} \cdot 0\text{˙}2907 + \frac{2}{5} \cdot 0\text{˙}33915 = 0\text{˙}3534\,,$$

po Bayesovi formuli pa:

$$\mathbb{P}(H_{45} \mid U_3) = \frac{\frac{2}{5} \cdot 0\text{˙}33915}{0\text{˙}3534} \doteq 0\text{˙}3839\,.$$

Opomba. Namesto H_{45} bi lahko posebej definirali H_4 in H_5. Iz $\mathbb{P}(U_3 \mid H_4) = \mathbb{P}(U_3 \mid H_5) = 0\text{˙}33915$ sledi tudi $\mathbb{P}(U_3 \mid H_4 \cup H_5) = 0\text{˙}33915$. Tak sklep lahko naredimo, ker sta dogodka H_4 in H_5 nezdružljiva – glej 9. nalogo.

18. Da, da, da, ne.

19. Ne. Velja $\mathbb{P}(F \cap G \cap H) = \mathbb{P}(F)\,\mathbb{P}(G)\,\mathbb{P}(H)$, $\mathbb{P}(F \cap G) = \mathbb{P}(F)\,\mathbb{P}(G)$ in $\mathbb{P}(F \cap H) = \mathbb{P}(F)\,\mathbb{P}(H)$, ne pa tudi $\mathbb{P}(G \cap H) = \mathbb{P}(G)\,\mathbb{P}(H)$.

20. a) Ne: $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{3}{8}$, $\mathbb{P}(A \cap B) = \frac{1}{4} \neq \mathbb{P}(A)\,\mathbb{P}(B)$.

b) Iz:

$$\begin{gathered}
\mathbb{P}(A) = \frac{1}{2}\,, \quad \mathbb{P}(B) = 2 \cdot \frac{1}{2} \cdot p(1-p) + \frac{1}{2}\,p^2 = \frac{p(2-p)}{2}\,,\\
\mathbb{P}(A \cap B) = 2 \cdot \frac{1}{2} \cdot p(1-p) = p(1-p)
\end{gathered}$$

dobimo, da sta dogodka A in B neodvisna pri $p = 0$ in $p = 2/3$.

21. Najprej izračunajmo verjetnosti dogodkov A_k in B_l. Dogodek A_k pomeni, da v prvih $k-1$ metih ni bilo enice, da pa je padla v k-tem metu; torej je $P(A_k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$. Dogodek B_l pa pomeni, da v zadnjih prvih $n-l$ metih ni bilo šestice, da pa je padla v l-tem metu; torej je $P(B_l) = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{l-1}$.

Za ugotovitev neodvisnosti moramo izračunati še verjetnosti presekov $A_k \cap B_l$. Ločimo tri primere.

- Za $k < l$ dogodek $A_k \cap B_l$ pomeni, da v prvih $k-1$ metih ni bilo enice, v k-tem metu je padla enica, v l-tem metu je padla šestica in v zadnjih $n-l$ metih ni bilo šestice. Torej je $P(A_k \cap B_l) = \left(\frac{5}{6}\right)^{k-1}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{n-l}$, kar je enako $P(A_k)\,P(B_l)$, torej sta dogodka A_k in B_l neodvisna.
- Za $k > l$ dogodek $A_k \cap B_l$ pomeni, da v prvih $l-1$ metih ni bilo enice, v l-tem metu je padla šestica, v vmesnih $k-l-1$ metih ni bilo niti enice niti šestice, v k-tem metu je padla enica, v zadnjih $n-k$ metih pa ni bilo šestice. Torej je $P(A_k \cap B_l) = \left(\frac{5}{6}\right)^{n-k+l-1}\left(\frac{2}{3}\right)^{k-l-1}\left(\frac{1}{6}\right)^2$, kar ni enako $P(A_k)\,P(B_l)$, torej sta dogodka A_k in B_l odvisna.
- Za $k = l$ je dogodek $A_k \cap B_l$ nemogoč, torej je $P(A_k \cap B_l) = 0$, kar spet pomeni, da sta dogodka A_k in B_l odvisna.

22. Pri ženskah je bilo prvo zdravilo uspešno v $200/2000 = 10\%$, drugo pa v $10/200 = 5\%$ primerov.

Pri moških je bilo prvo zdravilo uspešno v $190/200 = 95\%$, drugo pa v $1000/2000 = 50\%$ primerov.

Kaže torej, da je bilo prvo zdravilo uspešnejše od drugega tako pri moških kot tudi pri ženskah. Toda če pogledamo oboje skupaj, je bilo prvo zdravilo uspešno v $390/2200 \doteq 17{\cdot}7\%$, drugo zdravilo pa v $1010/2200 \doteq 45{\cdot}9\%$ primerov. Ko skupini združimo, kaže, da je uspešnejše drugo zdravilo.

Kako naj si to razložimo? Zapišimo vse skupaj v jeziku verjetnostnega računa. Za verjetnostni prostor vzamemo množico preizkušancev, pri čemer so vsi enako verjetni. Označimo:

$$\begin{aligned} M &:= \text{množica moških} \\ Z &:= \text{množica žensk} \\ S_1 &:= \text{množica tistih, ki so prejemeli prvo zdravilo} \\ S_2 &:= \text{množica tistih, ki so prejemeli prvo zdravilo} \\ U &:= \text{množica tistih, pri katerih je bilo zdravljenje uspešno} \end{aligned}$$

Tedaj velja:

$$\begin{gathered} \mathbb{P}(U \mid S_1 \cap Z) > \mathbb{P}(U \mid S_2 \cap Z) \\ \mathbb{P}(U \mid S_1 \cap M) > \mathbb{P}(U \mid S_2 \cap M) \\ \mathbb{P}(U \mid S_1) < \mathbb{P}(U \mid S_2)\,. \end{gathered}$$

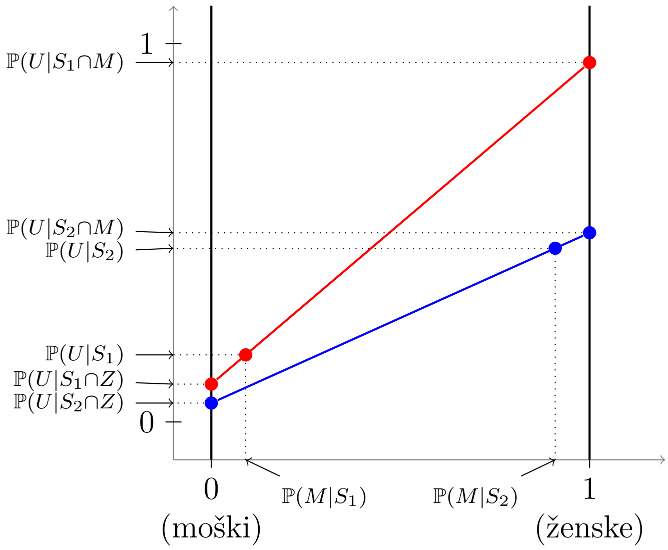
Oglejmo si zdaj, kako se $\mathbb{P}(U \mid S_i)$ izraža z $\mathbb{P}(U \mid S_i \cap Z)$ in $\mathbb{P}(U \mid S_i \cap M)$. Velja:

$$\begin{aligned} \mathbb{P}(U \mid S_1) &= \frac{\mathbb{P}(U \cap S_1)}{\mathbb{P}(S_1)} = \\ &= \frac{\mathbb{P}(U \cap S_1 \cap Z)}{\mathbb{P}(S_1)} + \frac{\mathbb{P}(U \cap S_1 \cap M)}{\mathbb{P}(S_1)} = \\ &= \frac{\mathbb{P}(S_1 \cap Z)\,\mathbb{P}(U \mid S_1 \cap Z)}{\mathbb{P}(S_1)} + \frac{\mathbb{P}(S_1 \cap M)\,\mathbb{P}(U \mid S_1 \cap M)}{\mathbb{P}(S_1)} = \\ &= \mathbb{P}(Z \mid S_1)\,\mathbb{P}(U \mid S_1 \cap Z) + \mathbb{P}(M \mid S_1)\,\mathbb{P}(U \mid S_1 \cap M)\,. \end{aligned}$$

in podobno:

$$\mathbb{P}(U \mid S_2) = \mathbb{P}(Z \mid S_2)\,\mathbb{P}(U \mid S_2 \cap Z) + \mathbb{P}(M \mid S_2)\,\mathbb{P}(U \mid S_2 \cap M)\,.$$

Za $i = 1, 2$ sta $\mathbb{P}(Z \mid S_i)$ in $\mathbb{P}(M \mid S_i)$ uteži z vsoto 1, zato lahko verjetnosti oz. deleže takole prikažemo:



Do paradoksa ne bi prišlo, če bi bili obe vmesni bunkici navpično poravnani, tj. če bi veljalo $\mathbb{P}(M \mid S_1) = \mathbb{P}(M \mid S_2)$ ali ekvivalentno $\mathbb{P}(Z \mid S_1) = \mathbb{P}(Z \mid S_2)$. Ker sta dogodka S_1 in S_2 nasprotna, je to ekvivalentno zahtevi, da sta dogodka Z in S_1 neodvisna, oziroma zahtevi, da sta spol preizkušanca in zdravilo, ki ga prejme, neodvisna. Z drugimi besedami, do paradoksa ne bi prišlo, če bi pri raziskavi pazili na delež žensk, ki so jim dali posamezno zdravilo – natančneje, če bi imeli skupini preizkušancev, ki so prejemali posamezno zdravilo, enako razporeditev spolov.

Posploševanje (statistično sklepanje) z vzorca na populacijo se obnese, če je vzorec *reprezentativen* glede na populacijo, torej če so lastnosti (v našem primeru spol), ki vplivajo na eksperimentalne spremenljivke (v našem primeru uspešnost zdravljenja), vsaj približno enako zastopane kot v populaciji. Do paradoksa torej ne bi prišlo, če bi bili obe skupini reprezentativni glede na spol.

V končni fazi bi lahko izrekli sklep, da je učinkovitejše *prvo* zdravilo, če bi bila različna zastopanost spolov edina šibka točka te raziskave (tj. če bi bile vse ostale lastnosti, ki pomembno vplivajo na delovanje zdravila, reprezentativno zastopane, to pa je dostikrat težko doseči, ker jih ne poznamo).

23. Naj bodo A, B, C, D in E dogodki, da posamezna stikala prepuščajo tok: A in B za stikali na levi, C za stikalo v sredini, D in E pa za stikali na desni. Naj bo še V dogodek, da vezje prepušča tok. Tedaj najprej po izreku o polni verjetnosti velja:

$$\mathbb{P}(V) = \mathbb{P}(C)\,\mathbb{P}(V \mid C) + \mathbb{P}(C^{\mathrm{c}})\,\mathbb{P}(V \mid C^{\mathrm{c}}) = \frac{1}{3}\,\mathbb{P}(V \mid C) + \frac{2}{3}\,\mathbb{P}(V \mid C^{\mathrm{c}})\,.$$

Nadalje je:

$$\mathbb{P}(V \mid C) = \mathbb{P}\big((A \cup B) \cap (D \cup E) \bigm| C\big)\,.$$

Dogodek C pa je neodvisen od katerega koli dogodka iz $\sigma(A, B, D, E)$, torej tudi od dogodka $(A \cup B) \cap (D \cup E)$. Sledi:

$$\mathbb{P}(V \mid C) = \mathbb{P}\big((A \cup B) \cap (D \cup E)\big)\,.$$

Nadalje sta tudi dogodka $A \cup B$ in $D \cup E$ neodvisna, torej je:

$$\begin{aligned}\mathbb{P}(V \mid C) &= \mathbb{P}(A \cup B)\,\mathbb{P}(D \cup E) = \\ &= \big(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\big)\big(\mathbb{P}(D) + \mathbb{P}(E) - \mathbb{P}(D \cap E)\big) = \\ &= \left(\frac{1}{3} + \frac{1}{3} - \frac{1}{9}\right)^2 = \\ &= \frac{25}{81}\,.\end{aligned}$$

Pri drugem členu pa je:

$$\mathbb{P}(V \mid C^{\mathrm{c}}) = \mathbb{P}\big((A \cap D) \cup (B \cup E) \bigm| C^{\mathrm{c}}\big)\,.$$

Ker je dogodek C neodvisen od katerega koli dogodka iz $\sigma(A, B, D, E)$, enako velja tudi za dogodek C^{c}: le-ta je torej neodvisen od dogodka $(A \cap D) \cup (B \cap E)$. Sledi:

$$\begin{aligned}\mathbb{P}(V \mid C^{\mathrm{c}}) &= \mathbb{P}\big((A \cap D) \cup (B \cap E)\big) = \\ &= \mathbb{P}(A \cap D) + \mathbb{P}(B \cap E) - \mathbb{P}(A \cap B \cap D \cap E) = \\ &= \frac{17}{81}\,.\end{aligned}$$

Poberemo skupaj in dobimo $\mathbb{P}(V) = \dfrac{1}{3} \cdot \dfrac{25}{81} + \dfrac{2}{3} \cdot \dfrac{17}{81} = \dfrac{59}{243} \doteq 0{\cdot}243$.

24. a) Označimo z J, F in T dogodke, da je Janez, Francelj oz. Tone zadel. Naj bo še $Z = J \cup F \cup T$ dogodek, da je sploh kdo zadel. Iskana pogojna verjetnost je torej enaka:

$$\mathbb{P}(J \mid Z) = \frac{\mathbb{P}(J \cap Z)}{\mathbb{P}(Z)} = \frac{\mathbb{P}(J)}{1 - \mathbb{P}\big(J^{\mathrm{c}} \cap F^{\mathrm{c}} \cap T^{\mathrm{c}}\big)} = \frac{0{\cdot}1}{1 - 0{\cdot}9 \cdot 0{\cdot}8 \cdot 0{\cdot}7} \doteq 0{\cdot}202\,.$$

b) Naj bo Z_1 dogodek, da je zajca zadel natanko en lovec. Zapišemo lahko:

$$Z_1 = (J \cap F^{\mathrm{c}} \cap T^{\mathrm{c}}) \cup (J^{\mathrm{c}} \cap F \cap T^{\mathrm{c}}) \cup (J^{\mathrm{c}} \cap F^{\mathrm{c}} \cap T)\,,$$

torej je:

$$\mathbb{P}(Z_1) = 0\text{'}1 \cdot 0\text{'}8 \cdot 0\text{'}7 + 0\text{'}9 \cdot 0\text{'}2 \cdot 0\text{'}7 + 0\text{'}9 \cdot 0\text{'}8 \cdot 0\text{'}3 = 0\text{'}398\,.$$

Sledi:

$$\mathbb{P}(J \mid Z_1) = \frac{\mathbb{P}(J \cap Z)}{\mathbb{P}(Z_1)} = \frac{\mathbb{P}(J \cap F^c \cap T^c)}{\mathbb{P}(Z_1)} = \frac{0\text{'}1 \cdot 0\text{'}8 \cdot 0\text{'}7}{0\text{'}398} \doteq 0\text{'}141\,.$$

25. Označimo z A, B, C in D dogodke, da zadene Andraž, Bojan, Cilka oz. Darja. Nadalje označimo z M_1 dogodek, da je v tarči natanko ena modra, z R_1 pa dogodek, da je v tarči natanko ena rdeča puščica. Iskana pogojna verjetnost je torej enaka:

$$\mathbb{P}(B \cap D \mid M_1 \cap R_1) = \frac{\mathbb{P}(B \cap D \cap M_1 \cap R_1)}{\mathbb{P}(M_1 \cap R_1)} = \frac{\mathbb{P}\big(A^c \cap B \cap C^c \cap D\big)}{\mathbb{P}(M_1 \cap R_1)}\,.$$

Po izreku o neodvisnosti izpeljanih dogodkov so dogodki A^c, B, C^c in D neodvisni, prav tako tudi dogodka $M_1 = (A \cap B^c) \cup (A^c \cap B)$ in $R_1 = (C \cap D^c) \cup (C^c \cap D)$. Sledi:

$$\begin{aligned}
&\mathbb{P}(B \cap D \mid M_1 \cap R_1) = \\
&\quad = \frac{\mathbb{P}(A^c)\,\mathbb{P}(B)\,\mathbb{P}(C^c)\,\mathbb{P}(D)}{\Big(\mathbb{P}(A)\,\mathbb{P}(B^c) + \mathbb{P}(A^c)\,\mathbb{P}(B)\Big)\Big(\mathbb{P}(C)\,\mathbb{P}(D^c) + \mathbb{P}(C^c)\,\mathbb{P}(D)\Big)} = \\
&\quad = \frac{0\text{'}4 \cdot 0\text{'}7 \cdot 0\text{'}5 \cdot 0\text{'}9}{(0\text{'}6 \cdot 0\text{'}3 + 0\text{'}4 \cdot 0\text{'}7) \cdot (0\text{'}5 \cdot 0\text{'}1 + 0\text{'}5 \cdot 0\text{'}9)} \doteq \\
&\quad \doteq 0\text{'}548\,.
\end{aligned}$$

26. Naj bodo H_0, H_1, H_2 in H_3 dogodki, da se je študent naučil učil 0, 1, 2 oz. 3 vprašanja, ki jih je dobil na izpitu. Velja:

$$\mathbb{P}(H_i) = \frac{\binom{30}{i}\binom{20}{3-i}}{\binom{50}{3}}\,.$$

Nadalje naj bo A dogodek, da študent naredi izpit.

- Če se ni naučil nobenega vprašanja, mora, če želi narediti izpit, uganiti vsaj dve. Zato je $\mathbb{P}(A \mid H_0) = 3 \cdot 0\text{'}1^2 \cdot 0\text{'}9 + 0\text{'}1^3 = 0\text{'}028$.
- Če se je študent naučil natanko eno vprašanje, mora, če želi narediti izpit, bodisi uganiti obe vprašanji, ki se ju ni naučil, bodisi ne sme pozabiti vprašanja, ki se ga je naučil, obenem pa mora uganiti še eno vprašanje, ki se ga ni učil. Zato je $\mathbb{P}(A \mid H_1) = 0\text{'}1^2 + 2 \cdot 0\text{'}1 \cdot 0\text{'}9 \cdot 0\text{'}7 = 0\text{'}136$.
- Če se je študent naučil natanko dve vprašanji, naredi izpit natanko tedaj, ko bodisi ne pozabi nobenega vprašanja, ki se ga je učil, bodisi pozabi eno vprašanje, ki se ga je naučil, obenem pa ugane vprašanje, ki se ga ni naučil. Zato je $\mathbb{P}(A \mid H_2) = 0\text{'}7^2 + 2 \cdot 0\text{'}7 \cdot 0\text{'}3 \cdot 0\text{'}1 = 0\text{'}532$.

- Če se je naučil vsa tri vprašanja, sme, če želi narediti izpit, pozabiti največ enega. Zato je $\mathbb{P}(A \mid H_3) = 3 \cdot 0{\cdot}7^2 \cdot 0{\cdot}3 + 0{\cdot}7^3 = 0{\cdot}784$.

Sledi:

$$\mathbb{P}(A) = \frac{\binom{20}{3}}{\binom{50}{3}} \cdot 0{\cdot}028 + \frac{\binom{30}{1}\binom{20}{2}}{\binom{50}{3}} \cdot 0{\cdot}136 + \frac{\binom{30}{2}\binom{20}{1}}{\binom{50}{3}} \cdot 0{\cdot}532 + \frac{\binom{30}{3}}{\binom{50}{3}} \cdot 0{\cdot}784 \doteq 0{\cdot}440\,.$$

27. Označimo z J, L in S dogodke, da Janez, Lojz oz. Štefan Mihi ponudi šmarnico. Nadalje naj bodo H_0, H_1, H_2 in H_3 dogodki glede na to, koliko vinogradnikov Mihi ponudi šmarnico, B pa dogodek, da Miho boli glava. Tedaj vemo, da je:

$$\mathbb{P}(B \mid H_0) = 0{\cdot}1\,, \quad \mathbb{P}(B \mid H_1) = 0{\cdot}4\,, \quad \mathbb{P}(B \mid H_2) = 0{\cdot}7\,, \quad \mathbb{P}(B \mid H_3) = 1\,,$$

obenem pa je:

$$\begin{aligned}
\mathbb{P}(H_0) &= 0{\cdot}4 \cdot 0{\cdot}6 \cdot 0{\cdot}9 = 0{\cdot}216\,,\\
\mathbb{P}(H_1) &= 0{\cdot}6 \cdot 0{\cdot}6 \cdot 0{\cdot}9 + 0{\cdot}4 \cdot 0{\cdot}4 \cdot 0{\cdot}9 + 0{\cdot}4 \cdot 0{\cdot}6 \cdot 0{\cdot}1 = 0{\cdot}492\,,\\
\mathbb{P}(H_2) &= 0{\cdot}6 \cdot 0{\cdot}4 \cdot 0{\cdot}9 + 0{\cdot}6 \cdot 0{\cdot}6 \cdot 0{\cdot}1 + 0{\cdot}4 \cdot 0{\cdot}4 \cdot 0{\cdot}1 = 0{\cdot}268\,,\\
\mathbb{P}(H_3) &= 0{\cdot}6 \cdot 0{\cdot}4 \cdot 0{\cdot}1 = 0{\cdot}024\,.
\end{aligned}$$

Torej je:

$$\mathbb{P}(B) = 0{\cdot}216 \cdot 0{\cdot}1 + 0{\cdot}492 \cdot 0{\cdot}4 + 0{\cdot}268 \cdot 0{\cdot}7 + 0{\cdot}024 \cdot 1 = 0{\cdot}43\,.$$

Sledi:

$$\mathbb{P}(J \cap L \mid B) = \frac{\mathbb{P}(J \cap L \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(J \cap L \cap S^{\mathrm{c}} \cap B) + \mathbb{P}(J \cap L \cap S \cap B)}{\mathbb{P}(B)}\,.$$

Seveda je $\mathbb{P}(B \mid J \cap L \cap S) = \mathbb{P}(B \mid H_3) = 1$. Toda velja tudi $\mathbb{P}\big(B \mid J \cap L \cap S^{\mathrm{c}}\big) = 0{\cdot}7$, ne le $\mathbb{P}(B \mid H_2) = 0{\cdot}7$. Sledi:

$$\mathbb{P}(J \cap L \mid B) = \frac{0{\cdot}6 \cdot 0{\cdot}4 \cdot 0{\cdot}9 \cdot 0{\cdot}7 + 0{\cdot}6 \cdot 0{\cdot}4 \cdot 0{\cdot}1 \cdot 1}{0{\cdot}43} \doteq 0{\cdot}407\,.$$

28. a) Označimo s K_Z dogodek, da je Zdravko okužen, s T_Z pa dogodek, da je Zdravkov test pozitiven. Tedaj je $\mathbb{P}(K_Z) = p$, $\mathbb{P}(T_Z \mid K_Z) = a$ in $\mathbb{P}(T_Z \mid K_Z^{\mathrm{c}}) = 1 - b$. Po izreku o popolni verjetnosti je:

$$\mathbb{P}(T_Z) = \mathbb{P}(T_Z \mid K_Z)\,\mathbb{P}(K_Z) + \mathbb{P}(T_Z \mid K_Z^{\mathrm{c}})\,\mathbb{P}(K_Z^{\mathrm{c}}) = ap + (1-b)(1-p)\,.$$

b) *Prvi način.* Pri vsakem delavcu je izvid pozitiven z verjetnostjo $t := ap + (1 - b)(1-p)$ in negativen z verjetnostjo $1-t = a(1-b)+bp$. Nadalje naj bo T_k dogodek, da je pozitivnih natanko k delavcev. Zaradi neodvisnosti je:

$$\mathbb{P}(T_k) = \binom{n}{k} t^k (1-t)^{n-k}\,.$$

Za iskano pogojno verjetnost $\mathbb{P}(K_Z \mid T_k)$ je treba izračunati še:

$$\begin{aligned}\mathbb{P}(K_Z \cap T_k) &= \mathbb{P}(K_Z \cap T_Z \cap T_k) + \mathbb{P}(K_Z \cap T_Z^c \cap T_k) \\ &= \mathbb{P}(K_Z)\,\mathbb{P}(T_Z \mid K_Z)\,\mathbb{P}(T_k \mid K_Z \cap T_Z) \\ &\qquad + \mathbb{P}(K_Z)\,\mathbb{P}(T_Z^c \mid K_Z)\,\mathbb{P}(T_k \mid K_Z \cap T_Z^c)\,.\end{aligned}$$

Označimo s T_l' dogodek, da je, če izvzamemo Zdravka, izvid pozitiven pri natanko l delavcih; posebej definirajmo še T_{-1}' kot nemogoč dogodek. Če je Zdravkov izvid pozitiven, se dogodek T_k ujema z dogodkom T_{k-1}' in iz neodvisnosti dobimo $\mathbb{P}(T_k \mid K_Z \cap T_Z) = \mathbb{P}(T_{k-1}' \mid K_Z \cap T_Z) = \mathbb{P}(T_{k-1}') = \binom{n-1}{k-1} t^{k-1}(1-t)^{n-k}$. Če pa je Zdravkov izvid negativen, se dogodek T_k ujema z dogodkom T_k', zato je $\mathbb{P}(T_k \mid K_Z \cap T_Z^c) = \mathbb{P}(T_k' \mid K_Z \cap T_Z^c) = \mathbb{P}(T_k') = \binom{n-1}{k} t^k (1-t)^{n-k-1}$. Sledi:

$$\mathbb{P}(K_Z \cap T_k) = pa\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k} + p(1-a)\binom{n-1}{k} t^k(1-t)^{n-k-1}\,,$$

iskana verjetnost pa je enaka:

$$\begin{aligned}\mathbb{P}(K_Z \mid T_k) &= \frac{\mathbb{P}(K_Z \cap T_k)}{\mathbb{P}(T_k)} \\ &= \frac{pa\binom{n-1}{k-1} t^{k-1}(1-t)^{n-k} + p(1-a)\binom{n-1}{k} t^k(1-t)^{n-k-1}}{\binom{n}{k} t^k (1-t)^{n-k}} \\ &= p\left(\frac{k}{n}\,\frac{a}{t} + \frac{n-k}{n}\,\frac{1-a}{1-t}\right).\end{aligned}$$

Drugi način. Uporabimo naslednjo pogojno različico izreka o polni verjetnosti:

$$\mathbb{P}(K_Z \mid T_k) = \mathbb{P}(K_Z \mid T_Z)\,\mathbb{P}(T_Z \mid T_k) + \mathbb{P}(K_Z \mid T_Z^c)\,\mathbb{P}(T_Z^c \mid T_k)\,. \qquad (*)$$

Ta enakost sicer v splošnem ne velja, zato jo je treba za dani primer posebej utemeljiti. Še prej pa poračunajmo do konca. Najprej iz simetrije dobimo, da je $\mathbb{P}(T_Z \mid T_k) = k/n$, torej $\mathbb{P}(T_Z^c \mid T_k) = (n-k)/n$. Nato pa po Bayesovi formuli izračunamo:

$$\begin{aligned}\mathbb{P}(K_Z \mid T_Z) &= \frac{\mathbb{P}(T_Z \mid K_Z)\,\mathbb{P}(K_Z)}{\mathbb{P}(T_Z)} = \frac{ap}{t}\,, \\ \mathbb{P}(K_Z \mid T_Z^c) &= \frac{\mathbb{P}(T_Z^c \mid K_Z)\,\mathbb{P}(K_Z)}{\mathbb{P}(T_Z^c)} = \frac{(1-a)p}{1-t}\,.\end{aligned}$$

Sledi:

$$\mathbb{P}(K_Z \mid T_k) = \frac{k}{n}\,\frac{ap}{t} + \frac{n-k}{n}\,\frac{(1-a)p}{1-t}\,,$$

kar je isto kot prej.

A utemeljiti moramo še enakost $(*)$. Izrek o polni verjetnosti nam da:

$$\mathbb{P}(K_Z \cap T_k) = \mathbb{P}(K_Z \mid T_Z \cap T_k)\,\mathbb{P}(T_Z \cap T_k) + \mathbb{P}(K_Z \mid T_Z^c \cap T_k)\,\mathbb{P}(T_Z^c \cap T_k)\,.$$

Delimo s $\mathbb{P}(T_k)$ in dobimo:

$$\mathbb{P}(K_Z \mid T_k) = \mathbb{P}(K_Z \mid T_Z \cap T_k)\,\mathbb{P}(T_Z \mid T_k) + \mathbb{P}(K_Z \mid T_Z^c \cap T_k)\,\mathbb{P}(T_Z^c \mid T_k)\,,$$

kar še ni $(*)$: utemeljiti je treba še, da je $\mathbb{P}(K_Z \mid T_Z \cap T_k) = \mathbb{P}(K_Z \mid T_Z)$ in $\mathbb{P}(K_Z \mid T_Z^c \cap T_k) = \mathbb{P}(K_Z \mid T_Z)$. Za to pa podobno kot pri prvem načinu opazimo, da je $T_Z \cap T_k = T_Z \cap T'_{k-1}$ in $T_Z^c \cap T_k = T_Z^c \cap T'_k$. A dogodka T_Z in T'_{k-1} sta neodvisna, prav tako tudi T_Z^c in T'_k. Prav tako sta neodvisna dogodka $K_Z \cap T_Z$ in T'_{k-1} ter dogodka $K_Z \cap T_Z^c$ in T'_k. Sledi:

$$\begin{aligned}\mathbb{P}(K_Z \mid T_Z \cap T_k) &= \mathbb{P}(K_Z \mid T_Z \cap T'_{k-1}) = \frac{\mathbb{P}(K_Z \cap T_Z \cap T'_{k-1})}{\mathbb{P}(T_Z \cap T'_{k-1})} = \\ &= \frac{\mathbb{P}(K_Z \cap T_Z)\,\mathbb{P}(T'_{k-1})}{\mathbb{P}(T_Z)\,\mathbb{P}(T'_{k-1})} = \mathbb{P}(K_Z \mid T_Z)\end{aligned}$$

in podobno:

$$\begin{aligned}\mathbb{P}(K_Z \mid T_Z^c \cap T_k) &= \mathbb{P}(K_Z \mid T_Z^c \cap T'_k) = \frac{\mathbb{P}(K_Z \cap T_Z^c \cap T'_k)}{\mathbb{P}(T_Z^c \cap T'_k)} = \\ &= \frac{\mathbb{P}(K_Z \cap T_Z^c)\,\mathbb{P}(T'_k)}{\mathbb{P}(T_Z^c)\,\mathbb{P}(T'_k)} = \mathbb{P}(K_Z \mid T_Z^c)\,.\end{aligned}$$

Enakost $(*)$ je s tem utemeljena, s tem pa tudi legitimnost celotnega izračuna.

29. a) *Prvi način.* Oglejmo si nasprotni dogodek, tj. da Albert nikoli ne bo izpuščen. Le-ta je presek padajočega zaporedja dogodkov, da bo Albert v ječi prebil več kot n noči. Verjetnosti teh dogodkov so enake:

$$\frac{1}{2} \cdot \frac{2}{3} \cdot \cdots \cdot \frac{n}{n+1} = \frac{1}{n+1}\,,$$

torej imajo limito nič, se pravi, da je nič tudi verjetnost dogodka, da Albert nikoli ne bo izpuščen.

Drugi način. Označimo z J_n dogodek, da je Albert v ječi prespal natanko n noči. Velja:

$$\mathbb{P}(J_n) = \frac{1}{2} \cdot \frac{2}{3} \cdot \cdots \cdot \frac{n-1}{n} \cdot \frac{1}{n+1} = \frac{1}{n(n+1)}\,.$$

Dogodek, da Albert nekoč pride iz ječe, je unija dogodkov $J_1, J_2, J_3, \ldots$, ki so nezdružljivi. Njegova verjetnost je enaka:

$$\sum_{n=1}^{\infty} \mathbb{P}(J_n) = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{m\to\infty} \sum_{n=1}^{m} \left(\frac{1}{n} - \frac{1}{n+1}\right) = \lim_{m\to\infty} \left(1 - \frac{1}{m+1}\right) = 1\,.$$

b) Označimo z N_5 dogodek, da je Albert v dani deželi prespal natanko petkrat. Ta dogodek je možen, če je v ječi prespal trikrat, štirikrat ali petkrat. Za $n = 3, 4, 5$ velja:

$$\mathbb{P}(N_5 \mid J_n) = \frac{1}{n+1}\,.$$

Po Bayesovi formuli dobimo:

$$\mathbb{P}(J_3 \mid N_5) = \frac{\frac{1}{3 \cdot 4} \cdot \frac{1}{4}}{\frac{1}{3 \cdot 4} \cdot \frac{1}{4} + \frac{1}{4 \cdot 5} \cdot \frac{1}{5} + \frac{1}{5 \cdot 6} \cdot \frac{1}{6}} = \frac{75}{131} \doteq 0{\cdot}573\,.$$

30. Za $n = 1, 2, 3, \ldots$ označimo z V_n dogodek, da pride do n-tega vlečenja, z V pa dogodek, da nikoli ne nehamo vleči. Očitno je $V_1 \supseteq V_2 \supseteq V_3 \supseteq \cdots$ in:

$$V = \bigcap_{n=1}^{\infty} V_n\,,$$

torej je:

$$\mathbb{P}(V) = \lim_{n\to\infty} \mathbb{P}(V_n)\,.$$

Izračunamo in ocenimo:

$$\begin{aligned}
\mathbb{P}(V_n) &= \frac{b}{a+b}\,\frac{b+d}{a+b+d}\cdots\frac{b+(n-2)d}{a+b+(n-2)d} = \\
&= \prod_{k=0}^{n-2}\left(1 - \frac{a}{a+b+kd}\right) \le \\
&\le \prod_{k=0}^{n-2}\exp\left(-\frac{a}{a+b+kd}\right) = \\
&= \exp\left(-\sum_{k=0}^{n-2}\frac{a}{a+b+kd}\right).
\end{aligned}$$

Ker vrsta $\sum_{k=0}^{\infty}\frac{a}{a+b+kd}$ divergira, je $\lim_{n\to\infty}\exp\bigl(-\sum_{k=0}^{n-2}\frac{a}{a+b+kd}\bigr) = 0$, od tod pa sledi, da je $\mathbb{P}(V) = \lim_{n\to\infty}\mathbb{P}(V_n) = 0$.

Opomba. Za izpeljavo, da je $\mathbb{P}(V) = 0$, v resnici sploh ne potrebujemo števne, temveč le končno aditivnost verjetnosti. Že po slednji namreč lahko iz dejstva, da je $V \subseteq V_n$, sklepamo, da je $\mathbb{P}(V) \le \mathbb{P}(V_n)$, potem pa iz $\lim_{n\to\infty}\mathbb{P}(V_n) = 0$ sledi $\mathbb{P}(V) = 0$. Števno aditivnost pa bi zares potrebovali, če bi bilo $\lim_{n\to\infty}\mathbb{P}(V_n) > 0$.

31. Označimo z F_k dogodek, da Ferdinand Mirando prvič pokliče k-ti dan po zabavi. Tedaj so dogodki $F_1, F_2, \ldots$ nezdružljivi, njihova unija, ki jo označimo z F, pa je dogodek, da Ferdinand Mirando sploh pokliče. Verjetnost tega dogodka je enaka:

$$\mathbb{P}(F) = \sum_{k=1}^{\infty} 3^{-k} = \frac{1}{2}\,.$$

Označimo še z M dogodek, da Miranda spozna novega fanta, preden jo Ferdinand pokliče (če je ne pokliče, je torej to dogodek, da Miranda sploh spozna novega fanta). Tedaj velja:

$$\mathbb{P}(M^{\mathrm{c}} \mid F_k) = \left(\frac{9}{10}\right)^{k-1},$$

torej je:

$$\mathbb{P}(M^{\mathrm{c}} \mid F) = \frac{\mathbb{P}(M^{\mathrm{c}} \cap F)}{\mathbb{P}(F)} = \frac{1}{\mathbb{P}(F)} \sum_{k=1}^{\infty} \mathbb{P}(F_k)\,\mathbb{P}(M^{\mathrm{c}} \mid F_k) = \frac{20}{21}\,.$$

iskana pogojna verjetnost pa je enaka:

$$\mathbb{P}(M \mid F) = 1 - \mathbb{P}(M^{\mathrm{c}} \mid F) = \frac{1}{21}\,.$$

4. Slučajne spremenljivke

1. $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$

2. Glede na meta kovancev in kocke ločimo 24 enako verjetnih izidov, ki jih lahko predstavimo z naslednjo tabelo:

	1	2	3	4	5	6
GG	·	·	·	·	·	·
GC	·	·	·	·	·	·
CG	·	·	·	·	·	·
CC	·	·	·	·	·	·

Te izide pa lahko predstavimo tudi tako, da stolpec predstavlja skupni dobitek v evrih. Dobimo:

	1	2	3	4	5	6	7	8	9	10
GG	·	·	·	·	·	·				
GC			·	·	·	·	·	·		
CG			·	·	·	·	·	·		
CC					·	·	·	·	·	·

Tako dobimo, da je:

$$S \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \frac{1}{24} & \frac{1}{24} & \frac{1}{8} & \frac{1}{8} & \frac{1}{6} & \frac{1}{6} & \frac{1}{8} & \frac{1}{8} & \frac{1}{24} & \frac{1}{24} \end{pmatrix}.$$

3. Prema sorazmernost pomeni, da je:

$$\mathbb{P}(X = k) = ck\,; \quad k = 1, 2, \ldots, 10$$

za neko konstanto c. Ker X zavzame vrednosti le na omenjeni končni množici, velja tudi $\sum_{k=1}^{10} \mathbb{P}(X = k) = 1$. Od tod dobimo $c = 1/55$ in $\mathbb{P}(X > 3) = 49/55$.

4. Velja:

$$\mathbb{P}(X = k) = \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k}.$$

Približno velja:

k	$\mathbb{P}(X = k)$
0	0˙162
1	0˙323
2	0˙291
3	0˙155
4	0˙0543
5	0˙0130
6	0˙00217
7	$2˙48 \cdot 10^{-4}$
8	$1˙86 \cdot 10^{-5}$
9	$8˙27 \cdot 10^{-7}$
10	$1˙65 \cdot 10^{-8}$

5. $1 - \left(\frac{2}{3}\right)^6 - 6 \cdot \frac{1}{3}\left(\frac{2}{3}\right)^5 - \binom{6}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4 \doteq 0˙320.$

6. Posamezna kroglica se po določenem številu zamenjav znajde v svoji začetni posodi natanko tedaj, ko je bila premeščena sodo mnogokrat. Na vsakem koraku pa je verjetnost, da bo izbrana kroglica premeščena (ne glede na to, v kateri posodi trenutno je), enaka $1/n$ in ti dogodki so med seboj neodvisni. Iskana verjetnost je torej enaka:

$$p_k := \sum_{\substack{m \le k \\ m \text{ sod}}} \binom{k}{m} \left(\frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{k-m}.$$

Če definiramo *indikator* izjave $\mathcal{A}$ kot:

$$\mathbf{1}(\mathcal{A}) := \begin{cases} 1 & ; \mathcal{A} \text{ je resnična} \\ 0 & ; \mathcal{A} \text{ ni resnična,} \end{cases}$$

lahko iskano verjetnost zapišemo kot:

$$p_k = \sum_{m=0}^{k} \mathbf{1}(m \text{ sod}) \binom{k}{m} \left(\frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{k-m}.$$

Indikator pa se da zapisati na naslednji način:

$$\mathbf{1}(m \text{ sod}) = \frac{1 + (-1)^m}{2}$$

Sledi:

$$p_k = \frac{1}{2} \sum_{m=0}^{k} \binom{k}{m} \left(\frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{k-m} + \frac{1}{2} \sum_{m=0}^{k} \binom{k}{m} \left(-\frac{1}{n}\right)^m \left(1 - \frac{1}{n}\right)^{k-m}$$

in iz binomske formule končno dobimo zaključen izraz:

$$p_k = \frac{1}{2}\left[1 + \left(1 - \frac{2}{n}\right)^k\right].$$

7. Poisson: 0˙315, Laplace: 0˙289.
Točen rezultat: 0˙323.

8. Tabela rezultatov:

	točno	Poisson	Laplace
$\mathbb{P}(X = 20)$	0˙11456	0˙08884	0˙11516
$\mathbb{P}(X = 25)$	0˙04046	0˙04459	0˙04063

9. Točen rezultat: 0˙12574.
Poissonov obrazec, uporabljen neposredno (kar je neustrezno): 0˙01268.
Poissonov obrazec, uporabljen na slučajni spremenljivki, ki pove, kolikokrat se voda *ne* potegne: 0˙12511.
Laplaceova lokalna formula: 0˙12679.

10. Iz besedila naloge izhaja, da je $np = 10$ (natančneje, to je *ocena* pričakovane vrednosti na podlagi povprečne vrednosti). Če z X označimo število obiskov na določen dan, je torej:

$$\mathbb{P}(X = k) \approx \frac{10^k e^{-10}}{k!}.$$

Izračunamo približke verjetnosti $\mathbb{P}(X = k)$ za prvih nekaj k od 16 naprej:

k	$\mathbb{P}(X = k)$
16	0˙0217
17	0˙0127
18	0˙0070
19	0˙0037
20	0˙0019
21	0˙0009
22	0˙0004
23	0˙0002
24	0˙0001

Seštejemo, zaokrožimo in dobimo, da je zahtevana verjetnost enaka približno 5%.

11. Označimo z S število zasedenih sob.

a) Laplaceova aproksimacija za $\mathbb{P}(1420 \le S \le 1450)$: $0{.}74988$.
Laplaceova aproksimacija za $\mathbb{P}(1419 \le S \le 1451)$: $0{.}78028$.
Laplaceova aproksimacija za $\mathbb{P}(1419{.}5 < S < 1450{.}5)$: $0{.}76543$.
Točen rezultat: $0{.}76286$.

b) Laplaceova aproksimacija za $\mathbb{P}(-\infty < S < 1420)$: $0{.}04779$.
Laplaceova aproksimacija za $\mathbb{P}(-\infty < S \le 1419)$: $0{.}04006$.
Laplaceova aproksimacija za $\mathbb{P}(-\infty < S < 1419{.}5)$: $0{.}04379$.
Točen rezultat: $0{.}04567$.

c) Označimo z x minimalno število prenovljenih sob. Po Laplaceovi integralski formuli je x približno najmanjše celo število, za katerega velja:

$$\frac{1}{2} - \Phi\left(\frac{x + \frac{1}{2} - 1600 \cdot 0{.}9}{\sqrt{1600 \cdot 0{.}9 \cdot 0{.}1}}\right) \le 0{.}05$$

oziroma:

$$\Phi\left(\frac{x - 1439{.}5}{12}\right) \ge 0{.}45\,.$$

Torej približno velja $x = \lceil y \rceil$, kjer je y rešitev enačbe:

$$\Phi\left(\frac{y - 1439{.}5}{12}\right) = 0{.}45$$

in iz $y \doteq 1459{.}24$ dobimo oceno $x = 1460$.

Dejansko je verjetnost, da bo zasedenih več kot 1460 sob, enaka $0{.}0417$, da jih zasedenih več kot 1459, pa $0{.}05003$. Torej je ocena $x = 1460$ pravilna.

d) Označimo z n število sprejetih rezervacij, z S pa število sob, ki jih bodo zahtevali gostje. Tedaj je $S \sim \mathrm{Bin}(n, 0{.}9)$. Veljati mora $\mathbb{P}(S > 1600) \le 0{.}05$. Tukaj je verjetnost neželenega dogodka $\{S > 1600\}$ naraščajoča funkcija števila n.

Po Laplaceovi integralski formuli je:

$$\mathbb{P}(S > 1600) \approx \frac{1}{2} - \Phi\left(\frac{1600{.}5 - 0{.}9\,n}{\sqrt{n \cdot 0{.}1 \cdot 0{.}9}}\right)\,.$$

Torej bo število sprejetih rezervacij ustrezalo približno tedaj, ko bo

$$\frac{16005 - 9n}{3\sqrt{n}} \ge \Phi^{-1}(0{.}45)$$

oziroma:

$$9n + 3\,\Phi^{-1}(0{.}45)\,\sqrt{n} - 16005 \le 0\,.$$

Označimo $y := \sqrt{n}$ in $q := \Phi^{-1}(0{.}45) \doteq 1{.}644854$. Dobimo kvadratno neenačbo:

$$9y^2 + 3qy - 16005 \le 0\,.$$

Za splošen y ima ta neenačba rešitev $y_1 \le y \le y_2$, kjer je:

$$y_1 := \frac{-q - \sqrt{q^2 + 64020}}{6} \qquad \text{in} \qquad y_2 := \frac{-q + \sqrt{q^2 + 64020}}{6}\,.$$

Toda ker je $y = \sqrt{n} \ge 0$, je dejanska rešitev $0 \le y \le y_2$ oziroma $n \le y_2^2 \doteq 1755{\cdot}37$, kar je res za $n \le 1755$.

Tudi v resnici je pri 1755 naročilih verjetnost, da ne bodo mogli sprejeti vseh gostov, enaka $0{\cdot}0454$, pri 1756 naročilih pa je enaka že $0{\cdot}0530$.

12. Če imamo tabelo normalne porazdelitve, lahko verjetnost dokaj hitro ocenimo po Laplaceovi integralski formuli, ki nam da:

$$\frac{1}{2} - \Phi\left(-\frac{0{\cdot}5}{2{\cdot}91}\right) \doteq \frac{1}{2} + \Phi(0{\cdot}2931) \doteq 0{\cdot}6152\,.$$

Uporaba Laplaceove lokalne formule terja malo več dela. Dobimo:

$$1 - \frac{1}{\sqrt{5{\cdot}82\,\pi}}\left(e^{-9/5.82} + e^{-4/5.82} + e^{-1/5.82}\right) \doteq 0{\cdot}6356\,.$$

Vendar pa je glede na številke primernejša Poissonova aproksimacija, ki nam da:

$$1 - \left(1 + 3 + \frac{3^2}{2}\right)e^{-3} = 1 - \frac{17}{2}\,e^{-3} \doteq 0{\cdot}5768\,.$$

Ta terja več dela kot Laplaceova integralska formula, a manj kot Laplaceova lokalna formula. Največ dela seveda terja točen rezultat:

$$1 - 0{\cdot}97^{100} - 100 \cdot 0{\cdot}03 \cdot 0{\cdot}97^{99} - 4950 \cdot 0{\cdot}03^2 \cdot 0{\cdot}97^{98} \doteq 0{\cdot}5802\,.$$

Če nam je hitrost dosti pomembnejša od natančnosti, bi torej uporabili Laplaceovo integralsko formulo. To je smiselno tudi zato, ker najbrž tudi deleža okuženih ne poznamo zelo natančno. Uporaba Laplaceove lokalne formule ni smiselna, saj terja več dela, natančna pa ni nič bolj. Poissonova aproksimacija je po pričakovanjih dosti natančnejša od obeh Laplaceovih in je prav tako smiselna, saj tudi ne terja toliko dela. A če lahko rezultat brez večjega truda izračunamo eksaktno, je to vsekakor najbolj smiselno.

13. Označimo z n število izdelkov v pošiljki, z S pa število prvovrstnih. Tedaj je $S \sim \mathrm{Bin}(n, 0{\cdot}6)$. Veljati mora:

$$\mathbb{P}(S \ge 0{\cdot}59\,n) \ge 0{\cdot}99\,.$$

Po Laplaceovi integralski formuli je:

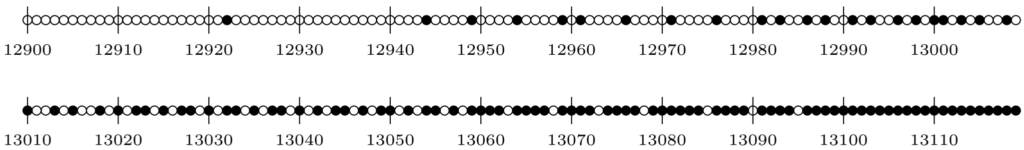
$$\mathbb{P}(S \ge 0{\cdot}59\,n) \approx \frac{1}{2} - \Phi\left(\frac{0{\cdot}59\,n - 0{\cdot}6\,n}{\sqrt{n \cdot 0{\cdot}6 \cdot 0{\cdot}4}}\right) = \frac{1}{2} + \Phi\left(\sqrt{\frac{n}{2400}}\right).$$

Torej bo število naročenih izdelkov ustrezalo približno tedaj, ko bo $\sqrt{n/2400} \ge \Phi^{-1}(0{\cdot}49)$ oziroma $n \ge 2400\big(\Phi^{-1}(0{\cdot}49)\big)^2 \doteq 12988{\cdot}55$, torej $n \ge 12989$.

V resnici je najmanjše možno število naročenih izdelkov, ki ustrezajo zahtevi, že 12922. Ne ustreza pa *vsako* število izdelkov, ki je večje ali enako 12922: prvo število, od katerega naprej vsako število naročenih ustreza, je šele 13096. Nekaj točnih verjetnosti, kjer z n označimo število izdelkov, z S pa število prvovrstnih:

$$\begin{aligned}
n = 12921 :&\quad \mathbb{P}(S \ge 7624) \doteq 0{\cdot}9897146436 \\
n = 12922 :&\quad \mathbb{P}(S \ge 7624) \doteq 0{\cdot}9900021378 \\
n = 12923 :&\quad \mathbb{P}(S \ge 7625) \doteq 0{\cdot}9898072270 \\[1ex]
n = 12094 :&\quad \mathbb{P}(S \ge 7726) \doteq 0{\cdot}9906203073 \\
n = 13095 :&\quad \mathbb{P}(S \ge 7727) \doteq 0{\cdot}9899715692 \\
n = 13096 :&\quad \mathbb{P}(S \ge 7727) \doteq 0{\cdot}9902509306
\end{aligned}$$

Na naslednji sliki so prikazana števila naročenih izdelkov, ki ustrezajo (polni krogci) in števila, ki ne ustrezajo (prazni krogci).



14. Označimo z n število naročenih izdelkov, z S pa število prvovrstnih. Tedaj je $S \sim \mathrm{Bin}(n, 0{\cdot}1)$. Veljati mora $\mathbb{P}(S \ge 100) \ge 0{\cdot}95$. Tukaj je verjetnost želenega dogodka naraščajoča funkcija števila n.

Po Laplaceovi integralski formuli je:

$$\mathbb{P}(S \ge 100) \approx \frac{1}{2} - \Phi\left(\frac{99{\cdot}5 - 0{\cdot}1\,n}{\sqrt{n \cdot 0{\cdot}1 \cdot 0{\cdot}9}}\right) = \frac{1}{2} + \Phi\left(\frac{n - 1000}{3\sqrt{n}}\right).$$

Torej bo število naročenih izdelkov ustrezalo približno tedaj, ko bo

$$\frac{n - 995}{3\sqrt{n}} \ge \Phi^{-1}(0{\cdot}45)$$

oziroma:

$$n - 3\,\Phi^{-1}(0{\cdot}45)\,\sqrt{n} - 995 \ge 0\,.$$

Če označimo $q = \Phi^{-1}(0{\cdot}45) \doteq 1{\cdot}644854$, je to natanko tedaj, ko je:

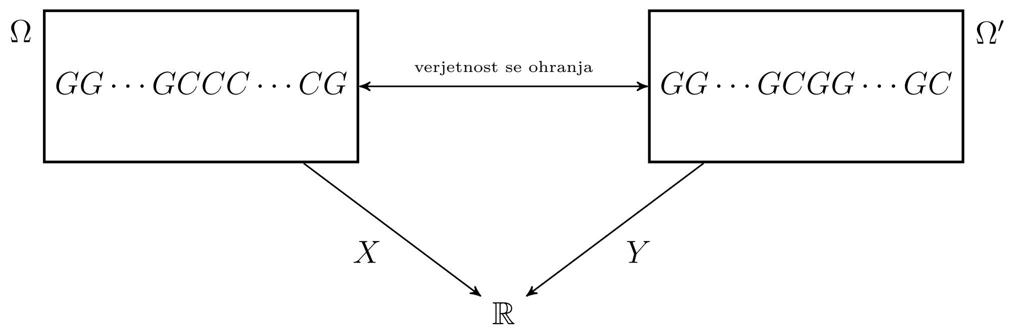
$$n \ge \left(\frac{3q + \sqrt{9q^2 + 3980}}{2}\right)^2 \doteq 1163{\cdot}32 \quad \text{oziroma} \quad n \ge 1164\,.$$

V resnici je dovolj naročiti 1161 izdelkov: verjetnost, da je prvovrstnih vsaj 100, pride pri 1160 izdelkih $0{\cdot}9493$, pri 1161 izdelkih pa $0{\cdot}9502$.

15. $\mathbb{P}(X = k) = \dfrac{1}{6}\left(\dfrac{5}{6}\right)^{k-1}$; $k \in \mathbb{N}$.

16. $\mathbb{P}(X = k) = \dbinom{k-1}{9}\left(\dfrac{1}{6}\right)^{10}\left(\dfrac{5}{6}\right)^{k-10}$; $k = 10, 11, 12, \ldots$.

17. $\mathbb{P}(X = k) = (k-1)\,2^{-k}$ za $k = 2, 3, \ldots$. Porazdelitev je negativna binomska $\mathrm{NB}(2, 1/2)$.
Verjetnostna razlaga: če takoj za prvo cifro vse cifre zamenjamo z grbi, grbe pa s ciframi, se problem iz naloge prevede na čakanje, dokler ne padeta dve cifri. Natančneje, če Y označuje število metov, dokler ne padeta dve cifri, sta X in Y enako porazdeljeni – pišemo $X \stackrel{\mathrm{d}}{=} Y$. Slika:



Če kovanec ni pošten, zgornja verjetnostna razlaga ne zdrži in tudi formula za porazdelitev je znatno bolj zapletena: če je p verjetnost, da pri posameznem metu pade grb, velja:

$$\mathbb{P}(X = k) = \sum_{i=0}^{k-2} p^{i+1}(1-p)^{k-i-1} = \frac{p(1-p)}{1-2p}\Big[(1-p)^{k-1} - p^{k-1}\Big].$$

18. Označimo število izvlečenih kart z X. Ta slučajna spremenljivka lahko zavzame vrednosti $2, 3, 4, 5$ ali 6. Dogodki, da je X enaka $2, 3, 4$ in 5, ustrezajo naslednjim zaporedjem izvlečenih barv:

$$\begin{array}{ll}
X = 2: & \text{R}, \text{Č}, (2 \times \text{R in } 2 \times \text{Č}) \\
X = 3: & \text{R}, \text{R}, \text{Č}, (1 \times \text{R in } 2 \times \text{Č}) \\
 & \text{Č}, \text{R}, \text{Č}, (2 \times \text{R in } 1 \times \text{Č}) \\
X = 4: & \text{R}, \text{R}, \text{R}, \text{Č}, 2 \times \text{Č} \\
 & \text{Č}, \text{R}, \text{R}, \text{Č}, (1 \times \text{R in } 1 \times \text{Č}) \\
 & \text{Č}, \text{Č}, \text{R}, \text{Č}, 2 \times \text{R} \\
X = 5: & \text{Č}, \text{R}, \text{R}, \text{R}, \text{Č}, \text{Č} \\
 & \text{Č}, \text{Č}, \text{R}, \text{R}, \text{Č}, \text{R} \\
X = 6: & \text{Č}, \text{Č}, \text{R}, \text{R}, \text{R}, \text{Č} \\
 & \text{Č}, \text{Č}, \text{Č}, \text{R}, \text{R}, \text{R}
\end{array}$$

Verjetnosti teh dogodkov izračunamo tako, da verjetnostni prostor razbijemo na $\binom{6}{3} = 20$ enako verjetnih izidov, ki predstavljajo razporeditve obeh barv v kupu. Ko preštejemo izide za vse dogodke, dobimo iskano porazdelitev:

$$X \sim \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ \frac{6}{20} & \frac{6}{20} & \frac{4}{20} & \frac{2}{20} & \frac{2}{20} \end{pmatrix} = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 0.3 & 0.3 & 0.2 & 0.1 & 0.1 \end{pmatrix} .$$

19. $\mathbb{P}(X = k) = 2^{-(k-1)}$ za $k = 2, 3, \ldots$, $X - 1 \sim \text{Geom}(1/2)$.
Verjetnostna razlaga: če prvič pade cifra, pri nadaljnjih metih cifre zamenjamo z grbi, grbe pa s ciframi. Tako se problem prevede na čakanje na cifro (od drugega meta dalje).

Če kovanec ni pošten, zgornja razlaga spet ne zdrži; če je p verjetnost, da pri posameznem metu pade grb, velja:

$$\mathbb{P}(X = k) = p(1-p)\left[p^{k-2} + (1-p)^{k-2}\right] .$$

20. Označimo število potrebnih metov z X. Ta slučajna spremenljivka lahko zavzame vrednosti $m, m+1, \ldots, 2m-1$. Naj bo k v množici teh vrednosti. Dogodek $\{X = k\}$ se lahko zgodi na dva disjunktna načina:

- V prvih $k-1$ metih pade natanko $m-1$ grbov in v k-tem metu pade grb (opazimo, da tedaj v prvih $k-1$ metih pade manj kot m številk).
- V prvih $k-1$ metih pade natanko $m-1$ številk in v k-tem metu pade številka (opazimo, da tedaj v prvih $k-1$ metih pade manj kot m grbov).

Sledi:

$$\mathbb{P}(X = k) = \binom{k-1}{m-1}\left[p^m(1-p)^{k-m} + (1-p)^m p^{k-m}\right]; \quad k = m, m+1, \ldots, 2m-1 .$$

Opomba: od tod dobimo netrivialno vsoto:

$$\sum_{k=m}^{2m-1} \binom{k-1}{m-1} \left[p^m (1-p)^{k-m} + (1-p)^m p^{k-m}\right] = 1\,,$$

ki se pri $p = 1/2$ prevede na:

$$\sum_{k=m}^{2m-1} 2^{1-k} \binom{k-1}{m-1} = 1\,.$$

Od tod pa sledi, da za slučajno spremenljivko $Y \sim \mathrm{NB}(m, 1/2)$ velja $\mathbb{P}(Y < 2m) = \mathbb{P}(Y \ge 2m) = 1/2$.

21. V prvem primeru pride:

$$X \sim \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{3}{32} & \frac{5}{64} & \frac{21}{64} \end{pmatrix}.$$

Če umaknemo omejitev števila metov, pa lahko X zavzame vrednosti $2, 3, 4, \ldots$ Porazdelitev lahko izračunamo na vsaj dva načina.

Prvi način. Za $k = 2, 3, 4, 5, 6$ so verjetnosti takšne kot v prejšnjem primeru. Sicer pa se za $k \ge 4$ dogodek $\{X = k\}$ zgodi, če v zadnjih treh metih pade GCC, prej pa ni dveh zaporednih cifer. Torej je:

$$\mathbb{P}(X = k) = \frac{n_{k-3}}{2^k}\,,$$

kjer je n_r število zaporedij cifer in grbov dolžine r, pri katerih ne padeta dve zaporedni cifri. Velja $n_1 = 2$ in $n_2 = 3$, za $n \ge 6$ pa lahko pri ustreznih zaporedjih ločimo naslednji dve možnosti:

- v zadnjem metu pade grb, prej pa ni dveh zaporednih cifer;
- v zadnjem metu pade cifra, v predzadnjem metu grb, prej pa ni dveh zaporednih cifer.

Tako dobimo $n_r = n_{r-1} + n_{r-2}$, kar pomeni, da gre za (zamaknjeno) Fibonaccijevo zaporedje: če je $F_1, F_2, F_3, F_4, F_5, F_6 \ldots = 1, 1, 2, 3, 5, 8, \ldots$ standardno Fibonaccijevo zaporedje, je $n_r = F_{r+2}$. Torej je $\mathbb{P}(X = k) = F_{k-1}/2^k$ (za $k = 2$ in $k = 3$ preverimo posebej).

Drugi način. V 13. nalogi iz 3. razdelka smo izračunali:

$$\mathbb{P}(X > k) = \frac{F_{k+2}}{2^k}\,; \quad k = 1, 2, 3, \ldots\,.$$

Za $k = 2, 3, 4, \ldots$ sledi:

$$\mathbb{P}(X = k) = \mathbb{P}(X > k-1) - \mathbb{P}(X > k-2) = \frac{2\,F_{k+1} - F_{k+2}}{2^k} = \frac{F_{k+1} - F_k}{2^n} = \\ = \frac{F_{k-1}}{2^k}\,,$$

kar je isto kot prej.

Opomba. Iz dobljenih verjetnosti smo dobili še zvezo:

$$\sum_{k=2}^{\infty} \frac{F_{k-1}}{2^k} = 1 \quad \text{oziroma} \quad \sum_{n=1}^{\infty} \frac{F_n}{2^n} = 2\,.$$

22. Velja:

$$\mathbb{P}(X = k) = \frac{\binom{4}{k}\binom{12}{7-k}}{\binom{16}{7}} = \frac{\binom{7}{k}\binom{9}{4-k}}{\binom{16}{4}}$$

oziroma:

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ \frac{18}{260} & \frac{84}{260} & \frac{108}{260} & \frac{45}{260} & \frac{5}{260} \end{pmatrix} \doteq \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0'0692 & 0'3231 & 0'4154 & 0'1731 & 0'0192 \end{pmatrix}.$$

23. *Prvi način.* Za izide vzamemo vse možne kombinacije Janezkovih in Maričkinih izbir škatel: vsak izid torej sestavljata dve podmnožici množice škatel, ena moči 3 in ena moči 4. Vseh izidov je $\binom{12}{4}\binom{12}{3}$.

Označimo z X število praznih škatel. Ta slučajna spremenljivka lahko zavzame vrednosti od 5 do 8. Za izračun porazdelitve predstavimo izide malo drugače – kot razbitja množice škatel na štiri podmnožice – glede na to, ali je Janezek v posamezno škatlo vrgel kroglico ali ne in glede na to, ali je to storila Marička ali ne.

Označimo z J množico škatel, v katere vrže kroglico Janezek, z M pa množico škatel, v katere vrže kroglico Marička. Velja:

$$\sharp(J) = 3\,, \quad \sharp(M) = 4\,, \quad \sharp\big((J \cup K)^{\mathrm{c}}\big) = X\,,$$

od koder izračunamo še:

$$\sharp(J \cup M) = 12 - X\,, \quad \sharp(J \cap M) = X - 5\,, \quad \sharp(J \setminus M) = 8 - X\,, \quad \sharp(M \setminus J) = 9 - X\,.$$

Izide, pri katerih je $X = k$, lahko dobimo tako, da najprej izberemo k praznih škatel, nato med preostalimi $12 - k$ škatlami izberemo $8 - k$ škatel iz $J \setminus M$, nato med preostalimi 4 škatlami izberemo $9 - k$ iz $M \setminus J$, preostalih $k - 5$ škatel pa sestavlja množico $J \cap M$: to so škatle, v katerih sta dve kroglici. Velja torej:

$$\mathbb{P}(X = k) = \frac{\binom{12}{k}\binom{12-k}{8-k}\binom{4}{9-k}}{\binom{12}{4}\binom{12}{3}}\,.$$

Če vse binomske simbole izrazimo s faktorielami, po ureditvi dobimo:

$$\mathbb{P}(X = k) = \frac{\binom{12}{k}\binom{12-k}{8-k}\binom{4}{9-k}}{\binom{12}{4}\binom{12}{3}} = \frac{4!\,8!\,3!\,9!}{12!\,k!\,(8-k)!\,(9-k)!\,(k-5)!}\,.$$

Drugi način. Fiksiramo škatle z Janezkovimi kroglicami. Za Maričkine imamo $\binom{12}{4}$ možnih enako verjetnih izbir. Če tako kot pri prejšnjem načinu definiramo slučajni množici J in M, izbire ustrezajo vsem možnim množicam M. Glede na to, da je $\sharp(J \cap M) = X - 5$ in $\sharp(M \setminus J) = 9 - X$, izbire, pri katerih je $X = k$, dobimo tako, da izberemo $k - 5$ elementov iz J in $9 - k$ elementov iz J^c. Sledi:

$$\mathbb{P}(X = k) = \frac{\binom{3}{k-5}\binom{9}{9-k}}{\binom{12}{4}}.$$

To pa lahko zapišemo tudi v obliki:

$$\mathbb{P}(X = k) = \frac{\binom{3}{8-k}\binom{9}{k}}{\binom{12}{8}},$$

iz katere se vidi, da je $X \sim \mathrm{Hip}(8, 9, 12)$. Podobno bi, če bi fiksirali škatle z Maričkinimi kroglicami, dobili, da je $X \sim \mathrm{Hip}(9, 8, 12)$. Če bi dobljene binomske simbole izrazili s faktorielami, bi po ureditvi dobili isto kot pri prvem načinu.

Tretji način. Slučajna spremenljivka X predstavlja število škatel, v katere Marička *ne vrže* kroglice, med škatlami, v katere Janezek *ne vrže* kroglice. Sledi, da je $X \sim \mathrm{Hip}(8, 9, 12) = \mathrm{Hip}(9, 8, 12)$.

Četrti način. Označimo z Y število škatel, v katerih sta dve kroglici. Velja $Y \sim \mathrm{Hip}(4, 3, 12) = \mathrm{Hip}(3, 4, 12)$ oziroma:

$$\mathbb{P}(Y = l) = \frac{\binom{3}{l}\binom{9}{4-l}}{\binom{12}{4}} = \frac{\binom{4}{l}\binom{8}{3-l}}{\binom{12}{3}}.$$

Če spet prikličemo slučajni množici J in M iz prvega načina, velja $Y = \sharp(J \cap M)$. Pri prvem načinu pa smo izpeljali tudi, da je $\sharp(J \cap M) = X - 5$, torej je $X = Y + 5$. Sledi:

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k - 5) = \frac{\binom{3}{k-5}\binom{9}{9-k}}{\binom{12}{4}} = \frac{\binom{4}{k-5}\binom{8}{8-k}}{\binom{12}{3}},$$

kar je isto kot pri drugem načinu.

Numerično je $X \sim \begin{pmatrix} 5 & 6 & 7 & 8 \\ 0{.}2545 & 0{.}5091 & 0{.}2182 & 0{.}0182 \end{pmatrix}$.

24. Označimo z N število vprašanj, ki se jih učenec ni učil. Za $j = 0, 1, 2$ velja:

$$\mathbb{P}(N = j) = \frac{\binom{5}{j}\binom{5}{2-j}}{\binom{10}{2}}, \qquad \mathbb{P}(U = i \mid N = j) = \binom{j}{i} \cdot 0{.}2^i \cdot 0{.}8^{j-i}$$

oziroma:

$$
\begin{aligned}
&\mathbb{P}(N=0)=\frac{10}{45}\,,\\
&\mathbb{P}(U=0 \mid N=0)=1\,,\quad \mathbb{P}(U=1 \mid N=0)=0\,,\quad \mathbb{P}(U=2 \mid N=0)=0\,,\\
&\mathbb{P}(N=1)=\frac{25}{45}\,,\\
&\mathbb{P}(U=0 \mid N=1)=0\text{˙}8\,,\quad \mathbb{P}(U=1 \mid N=1)=0\text{˙}2\,,\quad \mathbb{P}(U=2 \mid N=1)=0\,,\\
&\mathbb{P}(N=2)=\frac{10}{45}\,,\\
&\mathbb{P}(U=0 \mid N=2)=0\text{˙}64\,,\quad \mathbb{P}(U=1 \mid N=2)=0\text{˙}32\,,\quad \mathbb{P}(U=2 \mid N=2)=0\text{˙}04\,.
\end{aligned}
$$

Torej je:

$$
\begin{aligned}
\mathbb{P}(U=0)&=\frac{10}{45}\cdot 1+\frac{25}{45}\cdot 0\text{˙}8+\frac{10}{45}\cdot 0\text{˙}64 \doteq 0\text{˙}8089\,,\\
\mathbb{P}(U=1)&=\frac{10}{45}\cdot 0+\frac{25}{45}\cdot 0\text{˙}2+\frac{10}{45}\cdot 0\text{˙}32 \doteq 0\text{˙}1822\,,\\
\mathbb{P}(U=2)&=\frac{10}{45}\cdot 0+\frac{25}{45}\cdot 0+\frac{10}{45}\cdot 0\text{˙}04 \doteq 0\text{˙}0089\,,
\end{aligned}
$$

oziroma približno:

$$
U \sim \begin{pmatrix} 0 & 1 & 2 \\ 0\text{˙}8089 & 0\text{˙}1822 & 0\text{˙}0089 \end{pmatrix}.
$$

Opomba: seveda bi bilo dovolj izračunati $\mathbb{P}(U=1)$ in $\mathbb{P}(U=2)$ in nato upoštevati, da je vsota verjetnosti enaka 1.

25. Izvlečemo lahko od 0 do n belih kroglic. Najprej izračunajmo verjetnost dogodka, da v *prvih* k vlečenjih izvlečemo belo, v vseh ostalih vlečenjih pa rdečo kroglico:

$$
p_{n,k} := \frac{b(b+1)\cdots(b+k-1)\cdot r(r+1)\cdots(r+n-k-1)}{(b+r)(b+r+1)\cdots(b+r+n-1)}\,.
$$

Ta verjetnost pa se ohrani, tudi če zamenjamo predpisani vrstni red barv – da le ostane k belih in $n-k$ rdečih. Verjetnost lahko izrazimo s faktorielami:

$$
p_{n,k} = \frac{(b+k-1)!\,(r+n-k-1)!\,(b+r-1)!}{(b-1)!\,(r-1)!\,(b+r+n-1)!}\,.
$$

To je torej verjetnost, da je pri prvih n vlečenjih poljubnih *predpisanih* k kroglic belih. Če z X_n označimo število belih kroglic pri prvih n vlečenjih, torej za $k = 0, 1, 2, \ldots, n$ velja:

$$
\mathbb{P}(X_n = k) = \binom{n}{k} p_{n,k} = \frac{n!\,(b+k-1)!\,(r+n-k-1)!\,(b+r-1)!}{k!\,(n-k)!\,(b-1)!\,(r-1)!\,(b+r+n-1)!}\,.
$$

Za $b = r = 1$ dobimo enakomerno porazdelitev na množici $\{0, 1, 2, \ldots, n\}$.

26. Za izide lahko vzamemo razporeditve r rdečih in $n-r$ zelenih kroglic, pri čemer so kroglice neoznačene. Vseh možnih razporeditev je $\binom{n}{r}$. Označimo z X želeno število vlečenj. Ta slučajna spremenljivka lahko zavzame vrednosti $s, s+1, \ldots, n-r+s$. Če je k dopustna vrednost, so izidi, pri katerih je $X = k$, tiste razporeditve, pri katerih je med prvimi $k-1$ kroglicami natanko $s-1$ rdečih, k-ta kroglica je rdeča, med zadnjimi $n-k$ kroglicami pa je natanko $r-s$ rdečih. Sledi:

$$\mathbb{P}(X = k) = \frac{\binom{k-1}{s-1}\binom{n-k}{r-s}}{\binom{n}{r}}\,.$$

Opomba. Podobno kot pri opombi k hipergeometrijski porazdelitvi lahko tudi tu opazujemo, kaj se dogaja, ko gre n proti neskončno, r/n pa proti fiksnemu številu p. Če je $r_1, r_2, \ldots$ zaporedje z $\lim_{n\to\infty} r_n/n = p$, velja:

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\binom{k-1}{s-1}\binom{n-k}{r_n-s}}{\binom{n}{r_n}} \\
&= \lim_{n\to\infty} \binom{k-1}{s-1} \frac{(n-k)!\, r_n!\, (n-r_n)!}{(r_n-s)!\, (n-r_n-k+s)!\, n!} = \\
&= \lim_{n\to\infty} \binom{k-1}{s-1} \times \\
&\quad \times \frac{r_n(r_n-1)\cdots(r_n-s+1)\cdot(n-r_n)(n-r_n-1)\cdots(n-r_n-k+s+1)}{n(n-1)\cdots(n-k+1)} = \\
&= \lim_{n\to\infty} \binom{k-1}{s-1} p^s (1-p)^{k-s}\,.
\end{aligned}$$

Ustrezna negativna hipergeometrijska porazdelitev, ki izhaja iz izvirnega poskusa, torej glede na točkaste verjetnosti konvergira proti negativni binomski porazdelitvi $\mathrm{NB}(s, p)$, ki izhaja iz limitnega Bernoullijevega zaporedja poskusov.

27. Označimo število vlečenj z X. Ta slučajna spremenljivka zavzame vrednosti na naravnih številih in za vsako naravno število n velja:

$$\mathbb{P}(X = n) = \frac{b}{a+b}\,\frac{b+k}{a+b+k}\,\frac{b+2k}{a+b+2k}\cdots\frac{b+(n-2)k}{a+b+(n-2)k}\,\frac{a}{a+b+(n-1)k}\,.$$

28. Najprej preštejemo, na koliko načinov lahko $2n$ oseb razporedimo po parih. Predstavljajmo si, da to storimo tako, da jih razporedimo v vrsto, nakar poberemo po dva in dva. Če vse ločimo, lahko to storimo na $(2n)!$ načinov. A pri marsikaterih načinih dobimo isto razporeditev v pare: vseeno je, v katerem vrstnem redu so pari, poleg tega pa je v posameznem paru vseeno, kdo pride prej. Tako dobimo, da je razporeditev

$$\frac{(2n)!}{n!\cdot 2^n}\,.$$

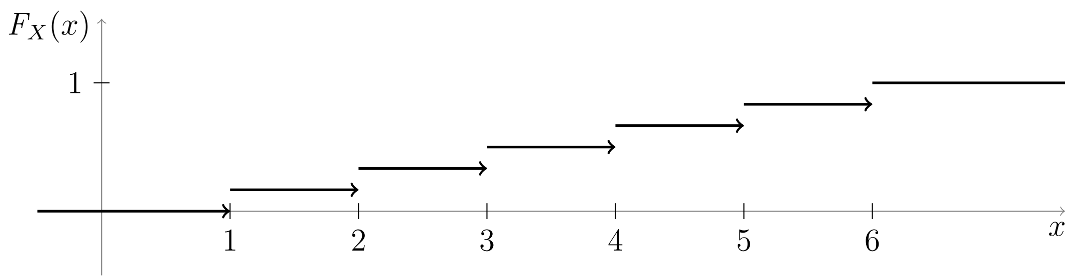
Če število parov, v katerih sta osebi nasprotnega spola, označimo z X, so možne vrednosti te slučajne spremenljivke vsa števila $k \in \{0, 1, \ldots, n\}$, pri katerih je $n - k$ sodo število. Dogodek $\{X = k\}$ se zgodi, če je natanko k moških in k žensk v parih, v ostalih parih pa sta osebi istega spola. Prešteti moramo, koliko razmestitev ustreza temu opisu. Izberemo k moških in k žensk, kar lahko naredimo na $\binom{n}{k}^2$ načinov. Izbrane osebe lahko razmestimo v mešane pare na $k!$ načinov. Ostale osebe moramo razmestiti v pare tako, da med njimi ne bo mešanih parov, torej razvrstimo posebej moške in posebej ženske. To lahko naredimo na

$$\frac{1}{2^{n-k}} \left(\frac{(n-k)!}{\left(\frac{n-k}{2}\right)!} \right)^2$$

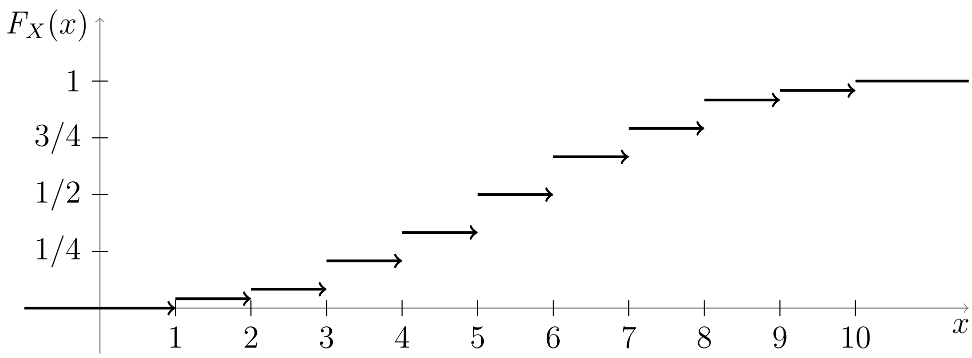
načinov. Sledi:

$$\mathbb{P}(X = k) = \binom{n}{k}^2 \cdot k! \cdot \frac{\frac{1}{2^{n-k}} \left(\frac{(n-k)!}{\left(\frac{n-k}{2}\right)!} \right)^2}{\frac{(2n)!}{n! \cdot 2^n}} = \frac{2^k \cdot (n!)^3}{k! \cdot (2n)! \cdot \left[\left(\frac{n-k}{2}\right)!\right]^2} \,.$$

29. Kumulativna porazdelitvena funkcija slučajne spremenljivke iz 1. naloge:



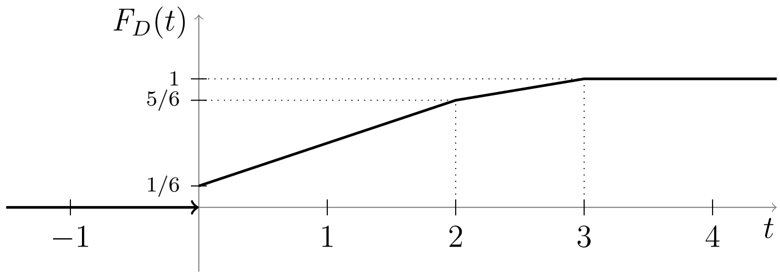
Kumulativna porazdelitvena funkcija slučajne spremenljivke iz 2. naloge:



30. Lahko privzamemo, da je prostor izidov kar kvadrat $\Omega = [-3, 3]$, verjetnost pa je seveda dolžina, deljena z dolžino celotnega intervala, torej s 6. Ločimo več primerov:

- Za $t < 0$ je očitno $\{D \le t\} = \emptyset$, torej je $F_D(t) = \mathbb{P}(D \le t) = 0$.
- Za $t = 0$ je $\{D \le t\} = \{D = 0\} = [0, 1]$, kar ima dolžino 1, torej je $F_D(0) = 1/6$.
- Za $0 \le t \le 2$ je $\{D \le t\} = [-t, 1 + t]$, kar ima dolžino $1 + 2t$, torej je $F_D(t) = (1 + 2t)/6$.
- Za $2 \le t \le 3$ je $\{D \le t\} = [-t, 3]$, kar ima dolžino $3 + t$, torej je $F_D(t) = (3 + t)/6$.
- Za $t \ge 3$ je dogodek $\{D \le t\}$ gotov, torej je $F_D(t) = 1$.

Graf:

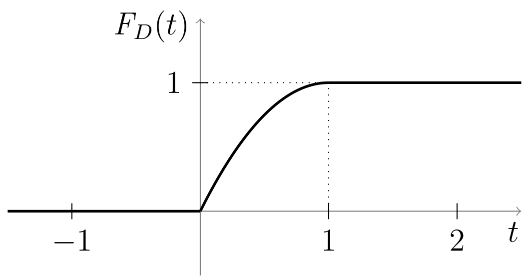


31. Podobno kot pri prejšnji nalogi lahko privzamemo, da je prostor izidov kar $\Omega = [-1, 1]^2$, verjetnost pa je seveda dolžina, deljena s ploščino tega kvadrata, torej s 4.

Označimo dano oddaljenost z D. Ločimo naslednje možnosti:

- Za $t < 0$ je očitno $\{D \le t\} = \emptyset$, torej je $F_D(t) = \mathbb{P}(D \le t) = 0$.
- Za $0 \le t \le 1$ je $\{D \le t\} = [-1, 1]^2 \setminus [-1+t, 1-t]^2$, kar ima ploščino $4-4(1-t)^2$, torej je $F_D(t) = 1 - (1 - t)^2 = 2t - t^2$.
- Za $t \ge 1$ je dogodek $\{D \le t\}$ gotov, torej je $F_D(t) = 1$.

Graf:



32. Iz:

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = \int_0^9 \frac{c}{\sqrt{x}}\,\mathrm{d}x = 2c\sqrt{x}\Big|_0^9 = 6c$$

dobimo $c = 1/6$. Nadalje je:

$$\begin{aligned}\mathbb{P}(1 < X < 4) &= \int_1^4 f_X(x)\,\mathrm{d}x = \frac{1}{6}\int_1^4 \frac{\mathrm{d}x}{\sqrt{x}} = \frac{\sqrt{x}}{3}\Big|_1^4 = \frac{1}{3}\,,\\ \mathbb{P}(X > 1) &= \int_1^\infty f_X(x)\,\mathrm{d}x = \frac{1}{6}\int_1^9 \frac{\mathrm{d}x}{\sqrt{x}} = \frac{\sqrt{x}}{3}\Big|_1^9 = \frac{2}{3}\,.\end{aligned}$$

33. Naj bo D najprej slučajna spremenljivka iz 30. naloge. Ker je $\sum_{t\in\mathbb{R}} \mathbb{P}(D = t) = 1/6 < 1$, D ni porazdeljena diskretno. Ker je $\mathbb{P}(D = 0) = 1/6 > 0$, D ni porazdeljena zvezno.

Naj bo zdaj D slučajna spremenljivka iz 31. naloge. Za poljubna $a, b \in \mathbb{R}$, za katera je $a \le b$, je $\mathbb{P}(a < X \le b) = F_X(b) - F_X(a)$. Ker pa je F_X zvezna, za vse $x \in \mathbb{R}$ velja:

$$\begin{aligned}\mathbb{P}(X = x) &= \lim_{n\to\infty} \mathbb{P}\left(x - \frac{1}{n} < X \le x\right) = F_X(x) - \lim_{n\to\infty} F_X\left(x - \frac{1}{n}\right) = \\ &= F_X(x) - F_X(x) = 0\,.\end{aligned}$$

Posledično je tudi:

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \le X \le b) = F_X(b) - F_X(a)\,.$$

Funkcija F_X pa je tudi odsekoma zvezno odvedljiva z odvodom:

$$F_X'(x) = \begin{cases} 0 & ;\ x < 0 \\ 2(1-x) & ;\ 0 < x \le 1 \\ 0 & ;\ x \ge 1\,. \end{cases}$$

V točki 0 odvod sicer ne obstaja, še vseeno pa je F_X integral svojega odvoda, se pravi, da za poljubna $a, b \in \mathbb{R}$ velja:

$$F_X(b) - F_X(a) = \int_a^b F_X'(x)\,\mathrm{d}x$$

(integral se ne spremeni, če integrand spremenimo v končno mnogo točkah, zato tudi v končno mnogo točkah ni treba definirati). Sledi, da je X porazdeljena zvezno z gostoto:

$$f(x) = \begin{cases} 2(1-x) & ;\ 0 < x < 1 \\ 0 & ;\ \text{sicer.} \end{cases}$$

34. Kumulativna porazdelitvena funkcija:

$$F_X(x) = \begin{cases} 0 & ;\ x \le 0 \\ \dfrac{x}{10} & ;\ 0 \le x \le 10 \\ 1 & ;\ x \ge 10 \end{cases}$$

je zvezna in odsekoma zvezno odvedljiva, torej gre res za zvezno porazdelitev. Gostota:

$$f_X(x) = \begin{cases} \dfrac{1}{10} & ; 0 < x < 2\pi \\ 0 & ; \text{sicer.} \end{cases}$$

35. $F_M(t) = \begin{cases} 0 & ; t \le 0 \\ \dfrac{t^2}{100} & ; 0 \le t \le 10 \\ 1 & ; t \ge 10 \end{cases}, \qquad f_M(t) = \begin{cases} \dfrac{t}{50} & ; 0 < t < 10 \\ 0 & ; \text{sicer} \end{cases}$

Verjetnost, da je Maks čakal od 5 do 6 minut, lahko izračunamo bodisi kot:

$$F_M(6) - F_M(5) = \frac{6^2 - 5^2}{100} = \frac{11}{100}$$

bodisi kot:

$$\int_5^6 \frac{t}{50}\,\mathrm{d}t = \frac{11}{100}\,.$$

36. $F_Z(z) = \begin{cases} 0 & ; z \le 0 \\ \dfrac{2z^2}{5} & ; 0 \le z \le 1 \\ \dfrac{1+z^2}{5} & ; 1 \le z \le 2 \\ 1 & ; z \ge 2 \end{cases}, \quad f_Z(z) = \begin{cases} \dfrac{4z}{5} & ; 0 < z < 1 \\ \dfrac{2z}{5} & ; 1 < z < 2 \\ 0 & ; \text{sicer.} \end{cases}$

37. Velja $c = \lambda$ in:

$$F_X(x) = \begin{cases} 0 & ; x \ge 0 \\ 1 - e^{-\lambda x} & ; x \ge 0 \end{cases},$$
$$\mathbb{P}(1 < X < 2) = e^{-\lambda} - e^{-2\lambda}\,.$$

38. Označimo iskano dolžino z X. Ta slučajna spremenljivka lahko zavzame vrednosti $1, 2, \ldots, m$.

Opazimo, da je $\{X \ge k\}$ dogodek, da prvih k števil tvori strogo naraščajoče zaporedje. Le-teh pa je toliko kot podmnožic množice naravnih števil od 1 do m moči k, torej $\binom{m}{k}$. Sledi:

$$\mathbb{P}(X \ge k) = \frac{1}{m^k}\binom{m}{k}\,.$$

Ob dogovoru, da je $\binom{n}{r} = 0$, brž ko je $r > n$, to velja za vse $k = 1, 2, 3, \ldots$ Med drugim dobimo $\mathbb{P}(X \ge 1) = 1$ in $\mathbb{P}(X \ge k) = 0$, brž ko je $k > m$, saj nobeno strogo naraščajoče podzaporedje ni daljše od m. Odštejemo in dobimo, da za vse

$k = 1, 2, 3, \ldots$ velja:

$$\begin{aligned} \mathbb{P}(X = k) &= \mathbb{P}(X \geq k) - \mathbb{P}(X \geq k+1) = \\ &= \frac{1}{m^k}\binom{m}{k} - \frac{1}{m^{k+1}}\binom{m}{k+1} = \\ &= \frac{1}{m^k}\binom{m}{k} - \frac{1}{m^{k+1}}\,\frac{m-k}{k+1}\binom{m}{k} = \\ &= \frac{1}{m^{k+1}}\left(m - \frac{m-k}{k+1}\right)\binom{m}{k} = \\ &= \frac{1}{m^{k+1}}\,\frac{k(m+1)}{k+1)}\binom{m}{k} = \\ &= \frac{k}{m^{k+1}}\binom{m+1}{k+1}. \end{aligned}$$

Iz zadnje oblike izraza pa se vidi tudi alternativna rešitev naloge: če izbrana števila tvorijo zaporedje $Y_1, Y_2, \ldots$, je:

$$\{X = k\} = \{Y_1 < Y_2 < \cdots < Y_k \geq Y_{k+1}\}\,.$$

Označimo z B množico zaporedij $(y_1, y_2, \ldots, y_{k+1})$ števil iz množice $\{1, 2, \ldots, m\}$, za katera je $y_1 < y_2 < \cdots < y_k \geq y_{k+1}$. Pokažimo, da ima B natanko $k\binom{m+1}{k+1}$ elementov. Toliko elementov pa ima zagotovo množica A, ki jo sestavljajo vse celoštevilske $(m+2)$-terice $(x_1, x_2, \ldots, x_{k+1}, i)$, za katere je $1 \leq x_1 < x_2 < \cdots < x_{k+1} \leq m$ in $i = 1, 2, \ldots, k$. Konstruirajmo bijekcijo iz A v B: vzemimo $(x_1, x_2, \ldots, x_{k+1}, i) \in A$.

- Če je $x_{k+1} = m$, naj bo $y_j = x_j$ za vse $j = 1, 2, \ldots, k$ in $y_{k+1} := x_i$.
- Če je $x_{k+1} < m$, naj bo $y_j = x_j$ za $j = 1, 2, \ldots, i-1$, $y_j = x_{j+1}$ za $j = i, i+1, \ldots, m$ in spet $y_{k+1} := x_i$.

Ni težko preveriti, da je ta predpis res bijekcija. Spet sledi $\mathbb{P}(X = k) = \frac{k}{m^{k+1}}\binom{m+1}{k+1}$.

39. $Y \sim \begin{pmatrix} 0 & 1 & 4 \\ 0{\cdot}1 & 0{\cdot}7 & 0{\cdot}2 \end{pmatrix}$

40. $Y \sim \begin{pmatrix} -1 & 0 & 1 \\ \dfrac{(1-p)^2}{(2-p)(2-2p+p^2)} & \dfrac{1-p}{2-p} & \dfrac{1}{(2-p)(2-2p+p^2)} \end{pmatrix}$

41. Iskana prva decimalka se izraža v obliki $D = \lfloor 10^U \rfloor$ in lahko s pozitivno verjetnostjo zavzame vrednosti $1, 2, \ldots 9$. Za d iz te množice velja:

$$\begin{aligned} \mathbb{P}(D = d) &= \mathbb{P}(d \leq 10^U < d+1) = \mathbb{P}(\log_{10} d \leq U \leq \log_{10}(d+1)) = \\ &= \log_{10}(d+1) - \log_{10} d = \log_{10}\left(1 + \frac{1}{d}\right). \end{aligned}$$

Numerična porazdelitvena shema te porazdelitve:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 0{\cdot}3010 & 0{\cdot}1761 & 0{\cdot}1249 & 0{\cdot}0969 & 0{\cdot}0792 & 0{\cdot}0669 & 0{\cdot}0580 & 0{\cdot}0512 & 0{\cdot}0458 \end{pmatrix}$$

Porazdelitev lahko posplošimo na poljubno osnovo številskega sistema in ji pravimo *Benfordova*[67] *porazdelitev.* Trditev, da ima D Benfordovo porazdelitev z osnovo b, torej pomeni, da za $d = 1, 2, \ldots, b-1$ velja:

$$\mathbb{P}(D = d) = \log_b(d+1) - \log_b d = \log_b\left(1 + \frac{1}{d}\right).$$

42. Nalogo rešimo prek kumulativne porazdelitvene funkcije. Očitno je $Y \geq 0$. Za $y \geq 0$ velja:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\big(-\sqrt{y} - a \leq X \leq \sqrt{y} - a\big).$$

To moramo izračunati na podlagi dejstva, da je X porazdeljena eksponentno. Ločimo več primerov. Za $a \geq 0$ velja:

$$F_Y(y) = \begin{cases} 0 & ; \ y \leq a^2 \\ 1 - e^{\lambda(a-\sqrt{y})} & ; \ y \geq a^2 \end{cases}.$$

Ker je ta funkcija zvezna in odsekoma zvezno odvedljiva, je Y porazdeljena zvezno z gostoto:

$$f_Y(y) = \begin{cases} 0 & ; \ y < a^2 \\ \dfrac{\lambda}{2\sqrt{y}}\, e^{\lambda(a-\sqrt{y})} & ; \ y > a^2 \end{cases}.$$

Naj bo zdaj $a < 0$. V tem primeru pa dobimo:

$$F_Y(y) = \begin{cases} 0 & ; \ y \leq 0 \\ e^{\lambda(a+\sqrt{y})} - e^{\lambda(a-\sqrt{y})} & ; \ 0 \leq y \leq a^2 \\ 1 - e^{\lambda(a-\sqrt{y})} & ; \ y \geq a^2 \end{cases},$$

kar je spet zvezna in odsekoma zvezno odvedljiva funkcija, ki nam da gostoto:

$$f_Y(y) = \begin{cases} 0 & ; \ y < 0 \\ \dfrac{\lambda}{2\sqrt{y}}\left(e^{\lambda(a+\sqrt{y})} + e^{\lambda(a-\sqrt{y})}\right) & ; \ 0 < y < a^2 \\ \dfrac{\lambda}{2\sqrt{y}}\, e^{\lambda(a-\sqrt{y})} & ; \ y > a^2 \end{cases},$$

43. Naj bo $U \sim \mathrm{Unif}(0, 1)$. Če postavimo:

$$X := \begin{cases} -1 & ; \ U < 0{\cdot}4 \\ 0 & ; \ 0{\cdot}4 \leq U < 0{\cdot}5 \\ 1 & ; \ 0{\cdot}5 \leq U < 0{\cdot}8 \\ 2 & ; \ U \geq 0{\cdot}8 \end{cases},$$

je $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0{\cdot}4 & 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}2 \end{pmatrix}$.

[67] Frank Albert Benford, Jr. (1883–1948), ameriški elektroinženir in fizik

Za generiranje eksponentne porazdelitve nastavimo $Y = h(U)$, kjer je $h\colon\ (0,1) \to (0,\infty)$ zvezna, strogo naraščajoča in surjektivna (torej bijektivna). Pomagamo si s kumulativno porazdelitveno funkcijo: ker mora biti Y porazdeljena eksponentno $\mathrm{Exp}(\lambda)$, mora za vsak $y > 0$ veljati:

$$\mathbb{P}(Y \le y) = 1 - e^{-\lambda y}\,.$$

Velja $\{Y \le y\} = \bigl\{h(U) \le y\bigr\} = \bigl\{U \le h^{-1}(y)\bigr\}$, torej je tudi $\mathbb{P}(Y \le y) = h^{-1}(y)$. Ko primerjamo obe vrednosti in enačbo obrnemo, dobimo $h(u) = -\frac{\ln(1-u)}{\lambda}$. Ta funkcija je res zvezna in strogo naraščajoča, interval $(0,1)$ pa preslika na $(0,\infty)$. Sklep: če postavimo:

$$Y := -\frac{\ln(1-U)}{\lambda}\,,$$

bo $Y \sim \mathrm{Exp}(\lambda)$.

Opomba. Konstrukciji slučajnih spremenljivk X in Y nista edini izbiri: lahko bi npr. definirali tudi $Y = -(\ln U)/\lambda$.

44. Kvantil za verjetnost α je enak $\dfrac{\alpha}{1-\alpha}$. Torej je $m = 1$, iz $x_{1/4} = 1/3$ in $x_{3/4} = 3$ pa dobimo še $\mathrm{IQR} = 8/3$.

45. Tabelirajmo verjetnosti levih repov:

	$x<1$	$x=1$	$1<x<2$	$x=2$	$2<x<3$	$x=3$	$3<x<4$	$x=4$	$4<x<5$	$x=5$	$5<x<6$	$x=6$	$x>6$
$\mathbb{P}(X<x)$	0	0	0˙05	0˙05	0˙2	0˙2	0˙35	0˙35	0˙5	0˙5	0˙65	0˙65	1
$\mathbb{P}(X\le x)$	0	0˙05	0˙05	0˙2	0˙2	0˙35	0˙35	0˙5	0˙5	0˙65	0˙65	1	1

Torej je:

$$\begin{aligned}
&\mathbb{P}(X < x) \le 0\dot{}3 \iff x \le 3\,, \qquad &&\mathbb{P}(X < x) \le 0\dot{}5 \iff x \le 5\,,\\
&\mathbb{P}(X \le x) \ge 0\dot{}3 \iff x \ge 3\,, \qquad &&\mathbb{P}(X \le x) \ge 0\dot{}5 \iff x \ge 4\,.
\end{aligned}$$

Tretji decil je torej natančno določen – $x_{0.3} = 3$, mediana pa je lahko koli iz intervala $[4,5]$.

46. a) Dovolj je obravnavati le implikaciji na levi, saj sta jima implikaciji na desni ekvivalentni. Brž ko je $x < q$, je $F(x) = \mathbb{P}(X \le x) \le \mathbb{P}(X < q) \le p$, torej implikacija velja. Toda okrepiti je ne moremo na nobeni strani: če je $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2$, je $q = 1$ kvantil slučajne spremenljivke X za verjetnost $p = 3/4$ (tretji kvartil). Toda če je $x = q = 1$, je $F(x) = 1 > p$, zato implikaciji na levi strani ne moremo dodati enačaja. Če pa postavimo $p = 1/2$, je $q = 2/3$ ena od median, toda za $x = 1/3 < q$ velja $F(x) = 1/2 = p$, zato implikaciji na desni strani enačaja ne moremo odvzeti.

Oglejmo si še spodnjo implikacijo. Brž ko je $x \le q$, je $F(x) = \mathbb{P}(X \le x) \ge \mathbb{P}(X \le q) \ge p$, torej implikacija velja, in sicer tudi, če ji na levi strani dodamo enačaj. Spet pa ji enačaja na desni strani ne moremo odvzeti: $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = 1/2$ in $p = 1/2$, je $q = 1/3$ ena od median, a za $x = 2/3 > q$ velja $F(x) = 1/2 = p$.

b) Naj bo $u < v$ ter $x = Q(u)$ in $y = Q(v)$. Tedaj je:

$$\mathbb{P}(X < x) \le u < v \le \mathbb{P}(X \le y)\,,$$

od koder sledi $x \le y$.

c) Naj bo $x \in \mathbb{R}$ in $\varepsilon > 0$. Tedaj velja:

$$\begin{aligned}
&\mathbb{P}(Q(U) \le x - \varepsilon) \le \mathbb{P}(Q(U) < x) \le \mathbb{P}(U \le F(x)) = F(x)\,,\\
&\mathbb{P}(Q(U) \ge x + \varepsilon) \le \mathbb{P}(Q(U) > x) \le \mathbb{P}(U \ge F(x)) = 1 - F(x)\,,\\
&\mathbb{P}(Q(U) < x + \varepsilon) \ge F(x)\,.
\end{aligned}$$

Ko pošljemo ε proti nič, dobimo $\mathbb{P}(Q(U) < x) \le F(x) \le \mathbb{P}(Q(U) \le x)$. Zdaj pa spet vzemimo $\varepsilon > 0$ in pišimo:

$$F(x) \le \mathbb{P}(Q(U) \le x) \le \mathbb{P}(Q(U) < x + \varepsilon) \le F(x + \varepsilon)\,.$$

Ko pošljemo ε proti nič in upoštevamo desno zveznost kumulativne porazdelitvene funkcije F, dobimo, da mora biti $\mathbb{P}(Q(U) \le x) = F(x)$, torej je $Q(U)$ res porazdeljena enako kot X.

47. 0˙93319.

48. Če je $a \ne 0$, funkcija $h(x) = ax + b$ bijektivno preslika celotno realno os (ki je po definiciji odprta) nase, njena inverzna funkcija $h^{-1}(y) = \frac{y-b}{a}$ pa ima zvezen odvod $(h^{-1})'(y) = \frac{1}{a}$. Po krajšem računu dobimo:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a\mu-b)^2}{2a^2\sigma^2}}\,,$$

torej je $Y \sim \mathrm{N}(a\mu + b, |a|\sigma)$. Jasno je, da to velja tudi za primer, ko je $a = 0$.

49. $1/2 - \Phi(9/5) \doteq 0\dot{}03593$.

50. Iz $f_Y(y) = \frac{1}{k} f_X\left(\frac{y}{k}\right)$ dobimo $Y \sim \mathrm{Gama}(a, \lambda/k)$.

51. Funkcija $h(x) = e^x$ realno os bijektivno preslika na odprt poltrak $(0, \infty)$, njena inverzna funkcija $h^{-1}(y) = \ln y$ pa ima zvezen odvod $(h^{-1})'(y) = \frac{1}{y}$. Za $y > 0$ je torej:

$$f_Y(y) = \frac{f_X(\ln y)}{|y|} = \frac{1}{y\sqrt{2\pi}}\, e^{-(\ln y)^2/2}\,,$$

medtem ko za $y \le 0$ velja $f_Y(y) = 0$.

52. a) Slučajna spremenljivka R je skoncentrirana na odprtem poltraku $(0, \infty)$ in privzamemo lahko, da zavzame vrednosti izključno tam. Funkcija $h(x) = x^2$ ta poltrak bijektivno preslika samega nase, njena inverzna funkcija $h^{-1}(y) = \sqrt{y}$ pa ima zvezen odvod $h^{-1}(y) = \frac{1}{2\sqrt{y}}$. Če torej označimo $S := R^2$, za $s > 0$ velja:

$$f_S(s) = \frac{1}{2\sqrt{s}} f_R\big(\sqrt{s}\,\big) = \frac{a}{2}\, e^{-s/2}\,,$$

medtem ko lahko za $s \le 0$ postavimo $f_S(s) = 0$. Slučajna spremenljivka S je torej porazdeljena eksponentno $\mathrm{Exp}(a/2)$.

b) Najprej poiščimo 95. centil slučajne spremenljivke S, ki ga označimo z $s_{0.95}$. Velja:

$$\mathbb{P}(S \le s_{0.95}) = 1 - e^{-a s_{0.95}/2} = 0{\dot{}}95\,,$$

od koder sledi:

$$s_{0.95} = -\frac{2}{a} \ln 0{\dot{}}05 \doteq \frac{5{\dot{}}99}{a}\,.$$

95. centil slučajne spremenljivke R pa je enak:

$$r_{0.95} = \sqrt{s_{0.95}} \doteq \frac{2{\dot{}}45}{\sqrt{a}}\,.$$

53. Slučajna spremenljivka X je skoncentrirana na odprtem poltraku $(0, \infty)$ in smemo privzeti, da zavzame vrednosti izključno tam. Naj bo $h(x) := e^{ax}$. Za $a > 0$ funkcija h poltrak $(0, \infty)$ bijektivno preslika na odprt poltrak $(1, \infty)$, za $a < 0$ pa ga bijektivno preslika na odprt interval $(0, 1)$. V obeh primerih ima inverzna funkcija $h^{-1}(y) = \frac{1}{a} \ln y$ zvezen odvod $(h^{-1})'(y) = \frac{1}{ay}$. Za vsak y v sliki poltraka $(0, \infty)$ glede na funkcijo h torej velja:

$$f_Y(y) = \frac{1}{|ay|} f_X\left(\frac{\ln y}{a}\right) = \frac{\lambda}{a}\, y^{-\lambda/a-1}\,,$$

za ostale y pa lahko postavimo $f_Y(y) = 0$.

Za $a > 0$ je torej:

$$f_Y(y) = \begin{cases} \frac{\lambda}{a}\, y^{-\lambda/a-1} & ;\ y > 1 \\ 0 & ;\ \text{sicer,} \end{cases}$$

medtem ko za $a < 0$ dobimo:

$$f_Y(y) = \begin{cases} -\frac{\lambda}{a}\, y^{-\lambda/a-1} & ;\ 0 < y < 1 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Za $a = 0$ je kar $Y = 1$.

Opomba: za $a = -\lambda$ je Y porazdeljena enakomerno na intervalu od 0 do 1.

54. Ne le, da smemo privzeti, da je vselej $X \ne 1$, privzeti smemo, da X zavzame vrednosti izključno v odprti množici $A := (0, 1) \cup (1, \infty)$. Funkcija h množico A

bijektivno preslika na odprto množico $B := (-\infty, -1) \cup (0, \infty)$, njena inverzna funkcija:

$$h^{-1}(y) = \frac{y}{y+1}$$

pa je povsod na B zvezno odvedljiva in velja:

$$\left(h^{-1}\right)'(y) = \frac{1}{(y+1)^2}\,.$$

Torej je Y porazdeljena zvezno z gostoto f_Y, ki je za $y \in B$ enaka:

$$f_Y(y) = f_X\left(\frac{y}{y+1}\right)\frac{1}{(y+1)^2} = \frac{1}{(2y+1)^2}\,,$$

za $y \notin B$ pa je $f_Y(y) = 0$.

55. a) Privzeti smemo, da X zavzame vrednosti izključno na odprtem poltraku $(0, \infty)$. Funkcija $h(x) = ax - b/x$ ta poltrak bijektivno preslika na celo realno os, inverzna funkcija:

$$h^{-1}(y) = \frac{y + \sqrt{y^2 + 4ab}}{2a}\,,$$

pa ima zvezen odvod:

$$\left(h^{-1}\right)'(y) = \frac{y + \sqrt{y^2 + 4ab}}{2a\sqrt{y^2 + 4ab}}\,.$$

Opazimo še, da za $x > 0$ velja:

$$f_X(x) = c\,e^{-(h(t))^2 - 2ab}\,.$$

Sledi, da je Y porazdeljena zvezno z gostoto:

$$f_Y(y) = \frac{c\left(y + \sqrt{y^2 + 4ab}\right)}{2a\sqrt{y^2 + 4ab}}\,e^{-y^2 - 2ab}\,.$$

b) Velja:

$$f_W(w) = \frac{c}{2a}\,e^{-w^2 - 2ab}\,,$$

torej je $W \sim \mathrm{N}\left(0, 1/\sqrt{2}\right)$.

c) Gostota normalne porazdelitve $\mathrm{N}\left(0, 1/\sqrt{2}\right)$ je:

$$f_W(w) = \frac{1}{\sqrt{\pi}}\,e^{-w^2}\,.$$

Izenačimo z izrazom iz prejšnje točke in dobimo

$$\frac{c}{2a}\,e^{-2ab} = \frac{1}{\sqrt{\pi}}$$

oziroma

$$c = \frac{2a\,e^{2ab}}{\sqrt{\pi}}\,.$$

Sledi

$$\int_0^\infty e^{-a^2x^2-b^2/x^2}\,\mathrm{d}x = \frac{\sqrt{\pi}\,e^{-2ab}}{2a}\,.$$

56. Funkcija $h(x) = x^2$ je definirana na celotni množici $A := \mathbb{R}$ (ki je odprta po definiciji) in velja $h'(x) = 2x$. Očitno je $\mathbb{P}\big(h'(X) = 0\big) = \mathbb{P}(X = 0) = 0$. Sledi, da je Y porazdeljena zvezno, za izračun gostote f_Y pa opazimo naslednje:

- Za $y \le 0$ enačba $x^2 = y$ nima rešitve na $x \ne 0$, zato je $f_Y(y) = 0$.
- Za $y > 0$ ima enačba $x^2 = y$ dve rešitvi na $x \ne 0$, in sicer $x = \pm\sqrt{y}$. Torej je v tem primeru:

$$f_Y(y) = \frac{f_X\big(\sqrt{y}\big)}{\big|2\sqrt{y}\big|} + \frac{f_X\big(\sqrt{y}\big)}{\big|-2\sqrt{y}\big|} = \frac{1}{\sqrt{2\pi y}}\,e^{-y/2}\,.$$

Podobno lahko tudi za $z \le 0$ postavimo $f_Z(z) = 0$, za $z > 0$ pa je:

$$\begin{aligned} f_Z(z) &= \frac{f_X\big(1+\sqrt{z}\big)}{\big|2\sqrt{z}\big|} + \frac{f_X\big(1-\sqrt{z}\big)}{\big|-2\sqrt{z}\big|} = \\ &= \frac{1}{2\sqrt{2\pi z}}\Big(e^{-(1+\sqrt{z})^2/2} + e^{-(1-\sqrt{z})^2/2}\Big) = \\ &= \frac{1}{\sqrt{2\pi z}}\,e^{(1+z)/2}\,\mathrm{ch}\sqrt{z}\,. \end{aligned}$$

Opomba: slučajna spremenljivka Y ima porazdelitev *hi hvadrat z eno prostostno stopnjo*, ki jo označimo s $\chi^2(1)$ (glej 5. razdelek).

5. Slučajni vektorji

1. Navzkrižno in robni porazdelitvi lahko predstavimo s tabelo:

	$M=0$	$M=1$	$M=2$	
$R=0$	$\frac{10}{120}$	$\frac{20}{120}$	$\frac{5}{120}$	$\frac{35}{120}$
$R=1$	$\frac{30}{120}$	$\frac{30}{120}$	$\frac{3}{120}$	$\frac{63}{120}$
$R=2$	$\frac{15}{120}$	$\frac{6}{120}$	0	$\frac{21}{120}$
$R=3$	$\frac{1}{120}$	0	0	$\frac{1}{120}$
	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$	1

Slučajni spremenljivki R in M sta odvisni. Velja:

$$R+M \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{10}{120} & \frac{50}{120} & \frac{50}{120} & \frac{10}{120} \end{pmatrix}.$$

Navzkrižno porazdelitev lahko dobimo iz formule:

$$\mathbb{P}(R=r, M=m) = \frac{\binom{3}{r}\binom{2}{m}\binom{5}{3-r-m}}{\binom{10}{3}},$$

vse ostale pa iz dejstva, da gre za hipergeometrijsko porazdelitev: $R \sim \mathrm{Hip}(3,3,10)$, $M \sim \mathrm{Hip}(3,2,10)$, $R+M \sim \mathrm{Hip}(3,5,10)$.

2. $R' + M' \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{245}{1800} & \frac{686}{1800} & \frac{623}{1800} & \frac{217}{1800} & \frac{28}{1800} & \frac{1}{1800} \end{pmatrix}.$

Nauk: za porazdelitev funkcije dveh slučajnih spremenljivk ni dovolj poznati le porazdelitve posameznih slučajnih spremenljivk: v splošnem je potrebno poznati navzkrižno porazdelitev.

3. a) Najprej preverimo, da je vsota deklariranih verjetnosti res ena. Ker je to res, je dovolj preveriti, da so vse deklarirane verjetnosti nenegativne. Po ureditvi dobimo, da to velja, če je $a \ge 0$ in $2a^2 + a - 5 \le 0$, torej če je $0 \le a \le \frac{-1+\sqrt{41}}{4} \doteq 1{\cdot}351$.

b) Velja:

$$X \sim \begin{pmatrix} 1 & 2 \\ \frac{5}{6} - \frac{1}{6}a & \frac{1}{6} + \frac{1}{6}a \end{pmatrix}, \qquad Y \sim \begin{pmatrix} 1 & 2 \\ \frac{5}{6} - \frac{1}{3}a^2 & \frac{1}{6} + \frac{1}{3}a^2 \end{pmatrix}$$

Slučajni spremenljivki sta torej enako porazdeljeni, če je $\frac{1}{6}a = \frac{1}{3}a^2$, torej če je $a = 0$ ali $a = \frac{1}{2}$.

c) Slučajni spremenljivki X in Y sta skoraj gotovo enaki, če sta verjetnosti v poljih, kjer nista enaki, enaki nič. To je res za $a = 0$.

d) Slučajni spremenljivki sta neodvisni, če je izpolnjen naslednji sistem enačb:

$$\begin{aligned}\frac{5}{6}-\frac{1}{6}\,a-\frac{1}{3}\,a^2 &= \left(\frac{5}{6}-\frac{1}{6}\,a\right)\left(\frac{5}{6}-\frac{1}{3}\,a^2\right)\\ \frac{1}{3}\,a^2 &= \left(\frac{5}{6}-\frac{1}{6}\,a\right)\left(\frac{1}{6}+\frac{1}{3}\,a^2\right)\\ \frac{1}{6}\,a &= \left(\frac{5}{6}-\frac{1}{3}\,a^2\right)\left(\frac{1}{6}+\frac{1}{6}\,a\right)\\ \frac{1}{6} &= \left(\frac{1}{6}+\frac{1}{6}a\right)\left(\frac{1}{6}+\frac{1}{3}a^2\right).\end{aligned}$$

Po ureditvi dobimo, da je to natanko tedaj, ko je $2a^3+2a^2+a-5=0$, kar je ekvivalentno $(a-1)(2a^2+4a+5)=0$, to pa je res natanko za $a=1$.

4. Porazdelitev slučajnega vektorja (X,Y) z robnima porazdelitvama:

	$Y=0$	$Y=1$	$Y=2$	
$X=-1$	0˙05	0˙1	0˙1	0˙25
$X=0$	0˙1	0˙2	0˙2	0˙5
$X=1$	0˙05	0˙1	0˙1	0˙25
	0˙2	0˙4	0˙4	1

Velja še $Y-X\sim\begin{pmatrix}-1 & 0 & 1 & 2 & 3\\ 0˙05 & 0˙2 & 0˙35 & 0˙3 & 0˙1\end{pmatrix}$

5. Najprej se splača iti računat robni porazdelitvi. Za $k=0,1,2,\ldots$ izračunamo:

$$\begin{aligned}\mathbb{P}(X=k) &= c\sum_{l=0}^{\infty}\frac{(k+l)!}{k!\,l!}a^k b^l =\\ &= ca^k\sum_{l=0}^{\infty}\frac{(k+1)(k+2)\cdots(k+l)}{l!}b^l =\\ &= ca^k\sum_{l=0}^{\infty}\binom{-k-1}{l}(-b)^l =\\ &= \frac{ca^k}{(1-b)^{k+1}}\,,\end{aligned}$$

sicer pa je seveda $\mathbb{P}(X=k)=0$. Sledi $X+1\sim\mathrm{Geom}\left(\frac{1-a-b}{1-b}\right)$ in $c=1-a-b$. Iz simetrije dobimo še $Y+1\sim\mathrm{Geom}\left(\frac{1-a-b}{1-a}\right)$.

Opomba. Mehanizem, ki pripelje do te porazdelitve, je naslednji: vrtimo kovanec, pri katerem grb pade z verjetnostjo a, cifra z verjetnostjo b, možno pa je tudi, da kovanec obstane na robu. Če je X število grbov, Y pa število cifer, dokler kovanec prvič ne obstane na robu, imata ti dve slučajni spremenljivki natančno isto skupno porazdelitev kot v nalogi.

6. Velja:

$$\begin{aligned} S &\sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0{\cdot}7^3 & 3 \cdot 0{\cdot}7^2 \cdot 0{\cdot}3 & 3 \cdot 0{\cdot}7 \cdot 0{\cdot}3^2 & 0{\cdot}3^3 \end{pmatrix} = \\ &= \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0{\cdot}343 & 0{\cdot}441 & 0{\cdot}189 & 0{\cdot}027 \end{pmatrix} = \\ &= \mathrm{Bin}(3, 0{\cdot}3)\,. \end{aligned}$$

Če so $X_1, \ldots X_n$ neodvisne in imajo vse Bernoullijevo porazdelitev $\mathrm{Ber}(p)$, je njihova vsota porazdeljena binomsko $\mathrm{Bin}(n, p)$. Vzemimo namreč zaporedje n neodvisnih slučajnih poskusov, pri katerih vsak uspe z verjetnostjo p. Naj bo:

$$X_i = \begin{cases} 1 & ; \text{če } i\text{-ti poskus uspe} \\ 0 & ; \text{sicer.} \end{cases}$$

Tedaj je $S = X_1 + X_2 + \cdots + X_n$ ravno število uspelih poskusov, za to pa vemo, da je porazdeljeno binomsko $\mathrm{Bin}(n, p)$.

7. *Prvi način.* Za $n \in \mathbb{N}_0$ velja:

$$\begin{aligned} \mathbb{P}(U = n) &= \sum_{k=0}^{n} \mathbb{P}(S = k, T = n - k) = \\ &= \sum_{k=0}^{n} \mathbb{P}(S = k)\, \mathbb{P}(T = n - k) = \\ &= \sum_{k=0}^{n} \frac{\lambda^k \mu^{n-k} e^{-\lambda-\mu}}{k!\,(n-k)!} = \\ &= \frac{e^{-\lambda-\mu}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k} = \\ &= \frac{(\lambda+\mu)^n\, e^{-\lambda-\mu}}{n!}\,, \end{aligned}$$

torej je $U \sim \mathrm{Pois}(\lambda + \mu)$.

Drugi način. Naj bosta $S_1, S_2, S_3, \ldots$ in $T_1, T_2, T_3, \ldots$ zaporedji slučajnih spremenljivk, pri čemer naj bosta S_k in T_k neodvisni za vsak k ter še $S_k \sim \mathrm{Bin}(m_k, p_k)$ in $T_k \sim \mathrm{Bin}(n_k, p_k)$. Tedaj je $S_k + T_k \sim \mathrm{Bin}(m_k + n_k, p_k)$.

Če gre m_k in n_k proti neskončno ter $m_k p_k \to \lambda$ in $n_k p_k \to \mu$ (taka zaporedja m_k, n_k in p_k se dajo konstruirati za poljubna λ in μ), se po Poissonovem obrazcu porazdelitev slučajnih spremenljivk S_k bliža Poissonovi porazdelitvi $\mathrm{Pois}(\lambda)$, porazdelitev slučajnih spremenljivk T_k pa Poissonovi porazdelitvi $\mathrm{Pois}(\mu)$. Še več: ko gre k proti neskončno, se porazdelitev slučajnega vektorja (S_k, T_k) bliža porazdelitvi slučajnega vektorja (S, T). Zato se morajo tudi porazdelitve vsot $S_k + T_k$ bližati porazdelitvi vsote $S + T$. Ker pa se porazdelitve teh vsot, ki so binomske, spet po Poissonovem obrazcu bližajo tudi Poissonovi porazdelitvi $\mathrm{Pois}(\lambda+\mu)$, od tod zaključimo, da mora biti $S + T \sim \mathrm{Pois}(\lambda + \mu)$.

8. Označimo z X_1 število prošenj, ki jih je ministrstvo dobilo prvi dan, z X_2 pa število prošenj, ki jih je dobilo drugi dan. Iskano verjetnost lahko dobimo kot:

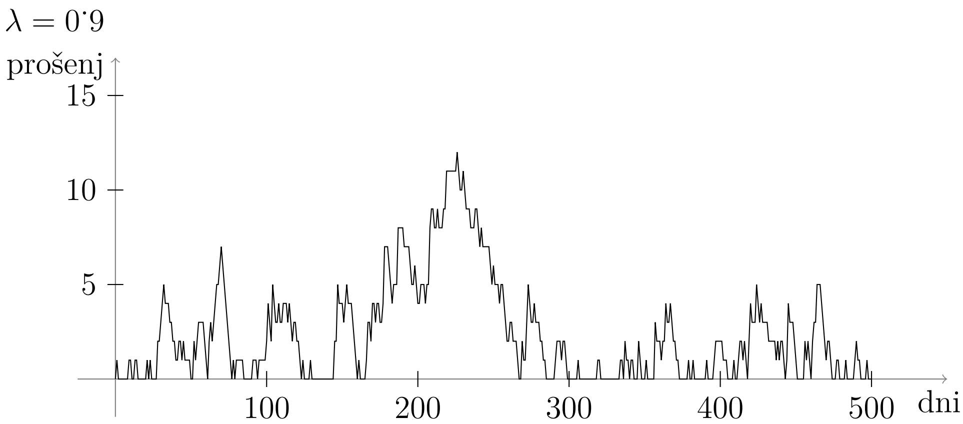
$$1 - \mathbb{P}\big(X_1 \in \{0,1\}, X_2 \in \{0,1,2\}\big) - \mathbb{P}\big(X_2 = 2, X_3 \in \{0,1\}\big) - \mathbb{P}(X_1 = 3, X_2 = 0)$$

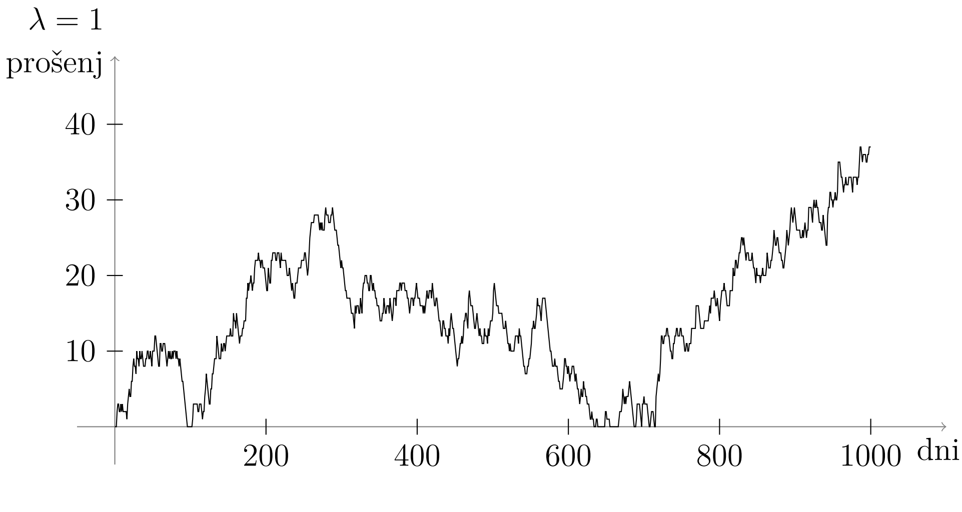
ali tudi kot:

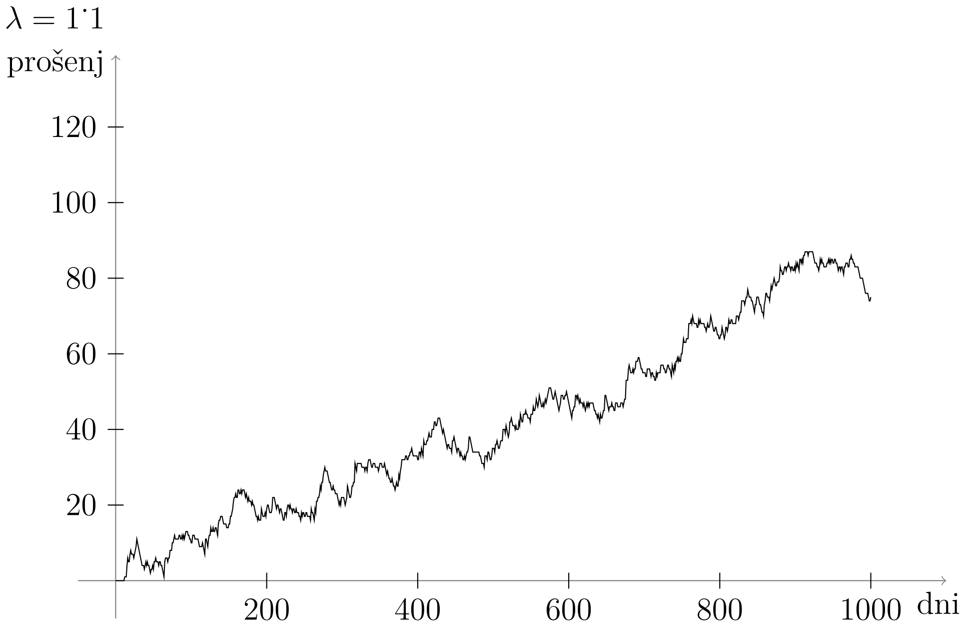
$$1 - \mathbb{P}\big(X_1 + X_2 \in \{0,1,2,3\}\big) + \mathbb{P}(X_1 = 0, X_2 = 3)\,;$$

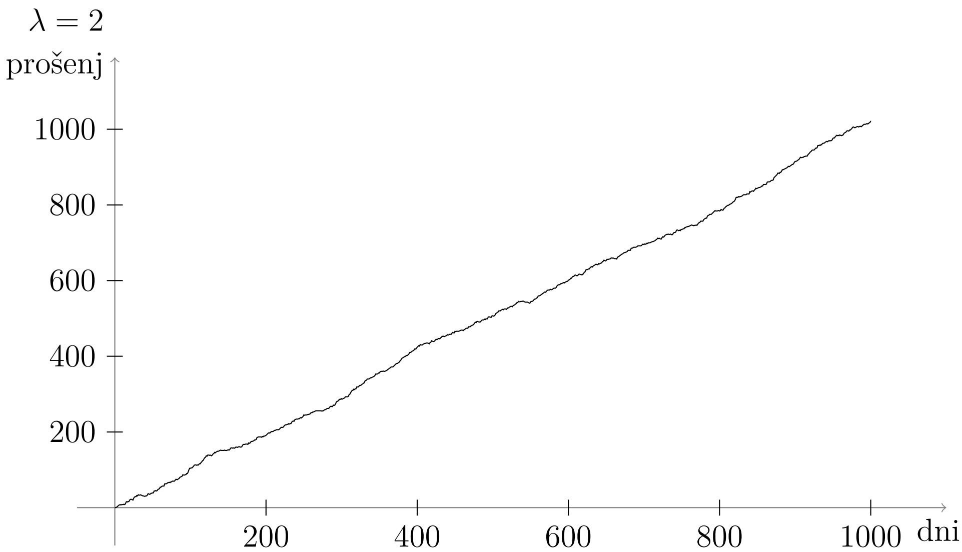
pri slednji možnosti upoštevamo, da je $X_1 + X_2 \sim \mathrm{Pois}(2\lambda)$. Končni rezultat je $1 - e^{-2\lambda}\left(1 + 2\lambda + 2\lambda^2 + \dfrac{7\lambda^3}{6}\right)$.

Simulacije nadaljnjega obnašanja števila prošenj, ki ostanejo na kupu, za različne λ:









9. $$\left(\frac{29}{30}\right)^{30} \cdot \left[1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2}\right)\right] + \\ + \binom{30}{1} \cdot \frac{1}{30} \cdot \left(\frac{29}{30}\right)^{29} \cdot \left[1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6}\right)\right] + \\ + \binom{30}{2} \cdot \left(\frac{1}{30}\right)^{2} \cdot \left(\frac{29}{30}\right)^{28} \cdot \left[1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right)\right] \doteq$$

$\doteq 0{\cdot}036831$.

Točen rezultat (brez zanemarjanja): $0{\cdot}036867$.

10. a) Dani slučajni vektor lahko zavzame vrednosti na vseh $(r+1)$-tericah $(j_0, j_1, \ldots, j_r)$ nenegativnih celih števil z $j_0 + j_1 + \cdots + j_r = n - r$; recimo jim dopustne. Dopustne $(r+1)$-terice pa so v bijektivni korespondenci z razporeditvami r rdečih in $n-r$ črnih neoznačenih kart v kup. Te razporeditve so vse enako verjetne in jih je $\binom{n}{r}$. Torej za vsako dopustno $(r+1)$-terico $(j_0, j_1, \ldots, j_r)$ velja

$$\mathbb{P}(\xi_0 = j_0, \xi_1 = j_1, \ldots, \xi_r = j_r) = \frac{1}{\binom{n}{r}} = \frac{r!\,(n-r)!}{n!}\,.$$

b) Posamezna slučajna spremenljivka ξ_k lahko zavzame vrednosti $0, 1, \ldots, n-r$. Vse možne razporeditve rdečih in črnih kart, pri katerih je $\xi_k = j$, lahko dobimo tako, da vzamemo razporeditve $r-1$ rdečih in $n-r-j$ črnih kart, nakar pod k-to rdečo karto vrinemo j črnih in še eno rdečo karto; pri $k = 0$ te karte dodamo na vrh, pri $k = r+1$ pa na dno dodamo najprej rdečo, nato pa še k črnih kart. Vseh razporeditev, pri katerih je $\xi_k = j$, je torej $\binom{n-j-1}{r-1}$. Sledi

$$\mathbb{P}(\xi_k = j) = \frac{\binom{n-j-1}{r-1}}{\binom{n}{r}} = r\,\frac{(n-r)!}{n!}\,\frac{(n-j-1)!}{(n-j-r)!}\,.$$

Opomba: slučajne spremenljivke ξ_k imajo negativno hipergeometrijsko porazdelitev, pomaknjeno za ena v levo (glej 26. nalogo iz 4. razdelka).

c) Slučajne spremenljivke so vselej odvisne, saj je npr. $\mathbb{P}(\xi_0 = n-r, \xi_1 = n-r) = 0$, medtem ko je $\mathbb{P}(\xi_0 = n-r)\,\mathbb{P}(\xi_1 = n-r) = 1/\binom{n}{r}^2 > 0$.

11. Slučajni vektor $(X_1, X_2, \ldots, X_n)$ lahko zavzame vrednosti le na n-tericah $(k_1, k_2, \ldots, k_n)$ nenegativnih celih števil, za katere je $k_1 + 2k_2 + 3k_3 + \cdots + nk_n = n$. Ker vemo, da je permutacij $n!$, moramo le še prešteti število permutacij z dano ciklično strukturo. Če označimo cikle in mesta v njih, imamo $n!$ možnih razporeditev. Toda nekatere izmed njih sovpadejo: vsak cikel dolžine k lahko predstavimo na k načinov, poleg tega pa lahko permutiramo cikle iste dolžine. Za nenegativne cele $k_1, \ldots, k_n$ s $k_1 + 2k_2 + 3k_3 + \cdots + nk_n = n$ torej velja:

$$\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n) = \frac{1}{1^{k_1} 2^{k_2} \cdots n^{k_n}\, k_1! k_2! \cdots k_n!}\,,$$

sicer pa je iskana verjetnost enaka nič.

Porazdelitev slučajne spremenljivke X_1 se sicer izraža z vsoto – za $k = 0, 1, 2, \ldots, n$ velja:

$$\mathbb{P}(X_1 = k) = \sum_{\substack{k_2, k_3, \ldots, k_n \in \mathbb{Z} \\ k_2, k_3, \ldots, k_n \ge 0 \\ k_1 + 2k_2 + 3k_3 + \cdots + nk_n = n}} \frac{1}{1^{k_1} 2^{k_2} \cdots n^{k_n}\, k_1! k_2! \cdots k_n!}\,,$$

toda možno jo je zapisati preprosteje. Slučajna spremenljivka X_1 namreč predstavlja število negibnih točk naše permutacije. Verjetnost dogodka, da ima permutacija *predpisanih* k točk negibnih, je $(n-k)!/n!$. Nadalje iz 18. naloge iz 2. razdelka razberemo, da je pogojno na ta dogodek verjetnost, da preostalih $n-k$ točk ne bo fiksnih, enaka:

$$\frac{1}{(n-k)!}\left\langle\frac{(n-k)!}{e}\right\rangle.$$

Končno lahko na $\binom{n}{k}$ načinov izberemo negibne točke naše permutacije. Ko vse skupaj sestavimo, dobimo:

$$\mathbb{P}(X_1=k)=\binom{n}{k}\frac{(n-k)!}{n!}\,\frac{1}{(n-k)!}\left\langle\frac{(n-k)!}{e}\right\rangle=\frac{1}{k!\,(n-k)!}\left\langle\frac{(n-k)!}{e}\right\rangle.$$

12. a) Za $j=0,1,2,\ldots$ izračunajmo:

$$\mathbb{P}(\alpha_i\ge j)=\mathbb{P}\big(X\text{ je deljiv s }p_i^j\big)=\sum_{m=1}^{\infty}\frac{1}{\zeta(s)\big(mp_i^j\big)^s}=p_i^{-js}.$$

Sledi:

$$\mathbb{P}(\alpha_i=j)=p_i^{-sj}-p_i^{-s(j+1)}.$$

Torej je slučajna spremenljivka $1+\alpha_i$ porazdeljena geometrijsko $\mathrm{Geom}\big(1-p_i^{-s}\big)$.

b) Izračunajmo najprej:

$$\begin{aligned}\mathbb{P}\big(\alpha_1\ge j_1,\ldots,\alpha_n\ge j_n\big)&=\mathbb{P}\big(X\text{ je deljiv s }p_1^{j_1},p_2^{j_2},\ldots,p_n^{j_n}\big)=\\&=\mathbb{P}\big(X\text{ je deljiv s }p_1^{j_1}\cdots p_n^{j_n}\big)=\\&=\sum_{m=1}^{\infty}\frac{1}{\zeta(s)\big(mp_1^{j_1}\cdots p_n^{j_n}\big)^s}=\\&=\frac{1}{\big(p_1^{j_1}\cdots p_n^{j_n}\big)^s}=\\&=\prod_{i=1}^{n}\mathbb{P}(\alpha_i\ge j_i)\,,\end{aligned}$$

Sledi, da so slučajne spremenljivke $\alpha_1,\alpha_2,\ldots$ neodvisne. Torej je:

$$\mathbb{P}\big(\alpha_1=j_1,\ldots,\alpha_n=j_n\big)=\prod_{i=1}^{n}\big(p_i^{-sj_i}-p_i^{-s(j_i+1)}\big).$$

Opomba. Izražava večrazsežnih točkastih verjetnosti z večrazsežnimi kumulativnimi nasploh ni tako zelo preprosta – velja:

$$\begin{aligned}&\mathbb{P}(X_1=k_1,X_2=k_2,\ldots,X_n=k_n)\\&=\sum_{d_1,d_2,\ldots,d_n\in\{0,1\}}(-1)^{d_1+d_2+\cdots+d_n}\,\mathbb{P}(X_1\ge k_1+d_1,X_2\ge k_2+d_2,\ldots,X_n\ge k_n+d_n)\\&=\sum_{d_1,d_2,\ldots,d_n\in\{0,1\}}(-1)^{d_1+d_2+\cdots+d_n}\,\mathbb{P}(X_1\le k_1-d_1,X_2\le k_2-d_2,\ldots,X_n\le k_n-d_n)\,.\end{aligned}$$

To lahko izpeljemo iz ustrezne identitete za indikatorje:

$$\begin{aligned}\mathbf{1}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n) &= \prod_{i=1}^{n} \bigl(\mathbf{1}(X_i \geq k_1) - \mathbf{1}(X_i \geq k_1 + 1)\bigr) = \\ &= \prod_{i=1}^{n} \bigl(\mathbf{1}(X_i \leq k_1) - \mathbf{1}(X_i \leq k_1 - 1)\bigr)\,.\end{aligned}$$

13. a) Podobno kot v 30. nalogi iz 4. razdelka najprej izračunamo verjetnost, da je prvih s izvlečenih kroglic rdečih, preostalih $n-s$ pa je belih. Verjetnost tega dogodka je enaka:

$$p_s := \frac{r(r+d)\cdots\bigl(r+(s-1)d\bigr)b(b+d)\cdots\bigl(b+(n-s-1)d\bigr)}{(b+r)(b+r+d)\cdots\bigl(b+r+(n-1)d\bigr)}\,.$$

Nato pa opazimo, da je to tudi verjetnost, da je poljubnih *predpisanih* s izvlečenih kroglic rdečih, preostale pa so bele. Sledi:

$$\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n) = p_{k_1+k_2+\cdots+k_n}\,.$$

Iskana verjetnost je torej odvisna zgolj od vsote $k_1 + k_2 + \cdots + k_n$.

b) Iskana verjetnost je enaka $\mathbb{P}(X_n = 1)$, kar se sicer lahko izrazi kot:

$$\sum_{k_1,\ldots,x_{k-1}\in\{0,1\}^{n-1}} \binom{n-1}{k_1+k_2+\cdots+k_{n-1}} p_{k_1+k_2+\cdots+k_{n-1}+1} = \sum_{s=0}^{n-1}\binom{n-1}{s}p_{s+1}\,,$$

vendar pa se da izraziti dosti preprosteje. Ker se namreč verjetnost dogodka $\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n)$ ohrani, če premešamo vrednosti $k_1, k_2, \ldots, k_n$ (pravimo, da so slučajne spremenljivke $X_1, X_2, \ldots, X_n$ *izmenljive*), so vse robne porazdelitve enake. Torej je:

$$\mathbb{P}(n\text{-ta izvlečena kroglica je rdeča}) = \mathbb{P}(X_n = 1) = \mathbb{P}(X_1 = 1) = \frac{r}{b+r}\,.$$

14. Dogodek $\{X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n\}$ se lahko zgodi le, če sta izpolnjena naslednja dva pogoja:

- $a_1 + a_2 + \cdots + a_n = n$;
- obstaja tak m, da je $a_1, a_2, \ldots, a_m > 0$ in $a_{m+1}, a_{m+2}, \ldots, a_m = 0$.

Tako je tudi $a_1 + a_2 + \cdots + a_m = n$. Naborom $(a_1, \ldots, a_n)$, ki izpolnjujejo ta dva pogoja, bomo rekli *dopustni.*

Če je $(a_1, \ldots, a_n)$ dopustna n-terica s pripadajočim številom m, na dogodku $\{X_1 = a_1, \ldots, X_n = a_n\}$ velja, da smo $(a_1 - 1)$-krat izvlekli kroglico z oznako 1, $(a_2 - 1)$-krat kroglico z oznako 2, …, $(a_m - 1)$-krat kroglico z oznako m, pri preostalih $m - 1$ vlečenjih pa smo izvlekli črno kroglico. Izračunajmo najprej verjetnost, da

$(a_1 - 1)$-krat izvlečemo kroglico z oznako 1, nato črno kroglico, potem $(a_2 - 1)$-krat kroglico z oznako 2 in tako naprej. Verjetnost za to je:

$$\begin{aligned}
&\frac{1}{2} \cdots \frac{(a_1-1)}{a_1} \cdot \frac{1}{a_1+1} \cdot \frac{1}{(a_1+2)} \cdots \frac{(a_2-1)}{(a_1+a_2)} \cdots \\
&\qquad \cdots \frac{1}{a_1+\cdots+a_{m-1}+1} \cdot \frac{1}{a_1+\cdots+a_{m-1}+2} \cdots \frac{(a_m-1)}{n} = \\
&= \frac{(a_1-1)!(a_2-1)!\cdots(a_m-1)!}{n!} =: \\
&=: p(a_1, a_2, \ldots, a_n)\,.
\end{aligned}$$

Zdaj pa opazimo, da se produkt ne spremeni, če zamenjamo vrstni red vlečenja. Torej lahko zapišemo:

$$\mathbb{P}(X_1 = a_1, \ldots, X_n = a_n) := N(a_1, \ldots, a_n)\, p(a_1, \ldots, a_n)\,,$$

kjer je $N(a_1, \ldots, a_n)$ število dopustnih vrstnih redov vlečenj $n-1$ kroglic, ki so dopustni za (dopustno) n-terico $(a_1, a_2, \ldots, a_n)$. Dopustni pa so takšni, ki za $k = 1, 2, \ldots, m$ zajemajo natanko $(a_k - 1)$ kroglic z oznako k, poleg tega pa še $m-1$ črnih kroglic. Nadalje pa mora biti pred vsako kroglico z oznako k vsaj $k-1$ črnih kroglic.

Razmišljamo takole: najprej lahko poljubno razporedimo $a_1 - 1$ kroglic z oznako 1 med vseh $n-1$ vlečenj. Pri prvem vlečenju, kjer nismo izvlekli kroglice z oznako 1, smo morali izvleči črno kroglico. Nato razporedimo $a_2 - 1$ kroglic z oznako 2 med $n - a_1 - 1$ preostalih vlečenj. Spet mora biti pri prvem vlečenju še nerazporejene oznake izvlečena črna kroglica. Tako nadaljujemo in dobimo, da je možnih razporeditev:

$$\begin{aligned}
N(a_1, a_2, \ldots, a_n) &= \binom{n-1}{a_1-1} \cdots \binom{n-a_1-1}{a_2-1} \cdots \binom{n-a_1-\cdots-a_{m-1}-1}{a_m-1} = \\
&= \frac{(n-1)!}{\prod_{i=1}^m (a_i-1)!} \cdot \frac{1}{(n-a_1)(n-a_1-a_2)\cdots(n-a_1-\cdots-a_{m-1})}\,.
\end{aligned}$$

Sledi, da za vsako dopustno n-terico $(a_1, \ldots, a_n)$ velja:

$$\begin{aligned}
\mathbb{P}(X_1 = a_1, \ldots, X_n = a_n) &= \frac{1}{n(n-a_1)(n-a_1-a_2)\cdots(n-a_1-\cdots-a_{m-1})} = \\
&= \frac{1}{a_m(a_m+a_{m-1})\cdots(a_m+a_{m-1}+\cdots+a_1)}\,.
\end{aligned}$$

15. a) Za $k_1, k_2, \ldots, k_m \in \mathbb{N}$ je $\{Y_1 = k_1, Y_2 = k_2, \ldots, Y_m = k_m\}$ dogodek, da so k_1-ti, $(k_1 + k_2)$-ti, ..., $(k_1 + k_2 + \cdots + k_m)$-ti poskus uspeli, ostali poskusi med prvimi $k_1 + k_2 + \cdots + k_m$ poskusi pa niso uspeli. Sledi:

$$\mathbb{P}(Y_1 = k_1, Y_2 = k_2, \ldots, Y_m = k_m) = p^m (1-p)^{k_1+k_2+\cdots+k_m-m}\,.$$

b) Očitno je $Y_1 \sim \mathrm{Geom}(p)$. Nadalje, ker je izraz, ki predstavlja točkaste verjetnosti iz prejšnje točke, simetričen v spremenljivkah $k_1, k_2, \ldots, k_m$, so slučajne spremenljivke $Y_1, Y_2, \ldots, Y_m$ izmenljive (glej 13. nalogo), torej so tudi enako porazdeljene.

Ne torej le Y_1, vse slučajne spremenljivke $Y_1, Y_2, \ldots$ so porazdeljene geometrijsko $\mathrm{Geom}(p)$: za vsak $i \in \mathbb{N}$ in vsak $k \in \mathbb{N}$ velja $\mathbb{P}(Y_i = k) = p(1-p)^{k-1}$. Sledi:

$$\mathbb{P}(Y_1 = k_1, Y_2 = k_2, \ldots, Y_m = k_m) = \mathbb{P}(Y_1 = k_1)\,\mathbb{P}(Y_2 = k_2) \cdots \mathbb{P}(Y_m = k_m)\,,$$

torej so slučajne spremenljivke $Y_1, Y_2, \ldots, Y_m$ res neodvisne.

c) V nalogi so konstruirane neodvisne slučajne spremenljivke $Y_1, Y_2, \ldots, Y_m$, porazdeljene geometrijsko $\mathrm{Geom}(p)$. Iz konstrukcije teh slučajnih spremenljivk dobimo, da je $Y_1 + Y_2 + \cdots + Y_m \sim \mathrm{NB}(m, p)$. Toda porazdelitev vsote $Y_1 + Y_2 + \cdots + Y_m$ je natančno določena s skupno porazdelitvijo seštevancev $Y_1, Y_2, \ldots, Y_m$, slednja pa je ob predpostavki neodvisnosti natančno določena z robnimi porazdelitvami. To pa pomeni, da mora za *poljubne* slučajne spremenljivke $Y_1, Y_2, \ldots, Y_m$, ki so neodvisne in porazdeljene geometrijsko $\mathrm{Geom}(p)$, veljati, da je njihova vsota porazdeljena negativno binomsko $\mathrm{NB}(m, p)$.

d) Naj bodo $Y_1, Y_2, \ldots, Y_{m+n}$ neodvisne in porazdeljene geometrijsko $\mathrm{Geom}(p)$. Če definiramo $S := Y_1 + Y_2 + \cdots + Y_m$ in $T := Y_{m+1} + Y_{m+1} + \cdots + Y_{m+n}$, sta S in T neodvisni, iz prejšnje točke pa dobimo, da je $S \sim \mathrm{NB}(m, p)$, $T \sim \mathrm{NB}(n, p)$ in $S + T \sim \mathrm{NB}(m + n, p)$.

Zdaj pa argumentirajmo enako kot v prejšnji točki: tu smo konstruirali neodvisni slučajni spremenljivki $S \sim \mathrm{NB}(m, p)$ in $T \sim \mathrm{NB}(n, p)$, za kateri velja $S + T \sim \mathrm{NB}(m+n, p)$. Toda porazdelitev vsote $S+T$ je natančno določena s skupno porazdelitvijo seštevancev S in T, slednja pa je ob predpostavki neodvisnosti natančno določena z robnima porazdelitvama. To pa pomeni, da mora za *poljubni* neodvisni slučajni spremenljivki $S \sim \mathrm{NB}(m, p)$ in $T \sim \mathrm{NB}(n, p)$ veljati $S+T \sim \mathrm{NB}(m+n, p)$.

16. a) in b) Za $x > 0$ velja:

$$f_X(x) = \int_0^{x/2} c\,e^{-x}\,\mathrm{d}y = \frac{cx}{2}\,e^{-x}$$

in za $y > 0$ velja:

$$f_Y(y) = \int_{2y}^{\infty} c\,e^{-x}\,\mathrm{d}x = c\,e^{-2y}\,.$$

Ko integriramo še ti dve funkciji, dobimo $c = 2$, torej:

$$f_X(x) = \begin{cases} x\,e^{-x} & ; x > 0 \\ 0 & ; \text{sicer}, \end{cases} \qquad \text{in} \qquad f_Y(y) = \begin{cases} 2\,e^{-2y} & ; y > 0 \\ 0 & ; \text{sicer}. \end{cases}$$

Še drugače, velja $X \sim \mathrm{Gama}(2, 1)$ in $Y \sim \mathrm{Exp}(2) = \mathrm{Gama}(1, 2)$.

c) Če je $x > 2y > 0$, je tudi $3x > y$, torej je $\mathbb{P}(3X > Y) = 1$. Drugo verjetnost pa lahko dobimo kot:

$$\mathbb{P}(X > 3Y) = 2 \iint_{\substack{x>3y \\ x>2y>0}} c\,e^{-x}\,\mathrm{d}x\,\mathrm{d}y = 2\int_0^{\infty}\int_{3y}^{\infty} e^{-x}\,\mathrm{d}x\,\mathrm{d}y = 2\int_0^{\infty} e^{-3y}\,\mathrm{d}y = \frac{2}{3}$$

ali kot

$$\mathbb{P}(X > 3Y) = 2\int_0^\infty \int_0^{x/3} e^{-x}\,\mathrm{d}y\,\mathrm{d}x = \frac{2}{3}\int_0^\infty x\,e^{-x}\,\mathrm{d}x = \frac{2}{3}\,.$$

Lahko tudi opazimo, da je gostota $f_{X,Y}$ konstantna na daljicah $\{(x, y)\ ;\ 0 < y < x/2\}$ za vse $x > 0$. Iskani dogodek pa ustreza poddaljicam $\{(x, y)\ ;\ 0 < y < x/3\}$ in vsaka od teh daljic je dolga natanko dve tretjini celotne. To lahko povemo tudi v jeziku nadaljevalne verjetnosti: za vsak $x > 0$ je slučajna spremenljivka Y pogojno na $X = x$ porazdeljena zvezno enakomerno na intervalu od 0 do $x/2$ (presenetljivo se da to "skoraj eksaktno" definirati, čeprav je $\mathbb{P}(X = x) = 0$). Pišemo:

$$Y \mid X = x \sim \mathrm{Unif}\left(0, \frac{x}{2}\right) \qquad \text{ali kar} \qquad Y \mid X \sim \mathrm{Unif}\left(0, \frac{X}{2}\right).$$

Iz pogojne porazdelitve pa sledi

$$\mathbb{P}(X > 3Y \mid X = x) = \mathbb{P}\left(Y < \frac{X}{3}\ \middle|\ X = x\right) = \frac{2}{3}\,,$$

kar se lahko spet napiše tudi kar kot $\mathbb{P}(X > 3Y \mid X) = 2/3$. Ker je pogojna verjetnost enaka za vsak x, je toliko enaka tudi brezpogojna verjetnost: slednja se računa podobno kot pri izreku o popolni verjetnosti.

d) Če bi bili slučajni spremenljivki X in Y neodvisni, bi imel slučajni vektor (X, Y) gostoto:

$$g(x, y) := f_X(x)\,f_Y(y) = \begin{cases} 2x\,e^{-x-2y} & ;\ x, y > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Ta funkcija pa ni enaka dani gostoti $f_{X,Y}$. Natančneje, poljubni gostoti se morata ujemati povsod razen na množici z mero nič. Taki funkciji, ki sta na določeni odprti množici zvezni, pa se morata ujemati prav povsod na njej. Funkciji $f_{X,Y}$ in g sta zvezni tako na odprti množici $\{(x, y)\ ;\ x > 2y > 0\}$ kot tudi na odprti množici $\{(x, y)\ ;\ 2y > x > 0\}$, a se na nobeni od teh množic ne ujemata. Slučajni spremenljivki X in Y sta torej odvisni.

Da sta slučajni spremenljivki X in Y odvisni, pa lahko preverimo tudi neposredno, saj je $\mathbb{P}(X < 2, Y > 1) = 0$, medtem ko je $\mathbb{P}(X < 2) = 1 - 3\,e^{-2} > 0$ in $\mathbb{P}(Y > 1) = e^{-2} > 0$.

e) Velja:

$$F_Z(z) = \mathbb{P}(Z \le z) = 2\iint_{\substack{x+y\le z\\ x>2y>0}} e^{-x}\,\mathrm{d}x\,\mathrm{d}y = 2\iint_{\substack{y>0\\ 2y<x\le z-y}} e^{-x}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Če želimo, da je $2y \le z - y$, mora biti $y \le z/3$. Poleg tega pa mora biti tudi $y > 0$, kar pomeni, da mora biti tudi $z > 0$. Za $z > 0$ torej velja:

$$\begin{aligned} F_Z(z) = \mathbb{P}(Z \le z) &= 2\int_0^{z/3}\int_{2y}^{z-y} e^{-x}\,\mathrm{d}x\,\mathrm{d}y = 2\int_0^{z/3}\left(e^{-2y} - e^{-(z-y)}\right)\mathrm{d}y = \\ &= 1 - 3\,e^{-2z/3} + 2\,e^{-z}\,, \end{aligned}$$

medtem ko za $z \le 0$ velja $F_Z(z) = 0$. Za $z = 0$ velja tudi zgornja formula. Torej je:

$$F_Z(z) = \begin{cases} 0 & ; \, z \le 0 \\ 1 - 3\,e^{-2z/3} + 2\,e^{-z} & ; \, z \ge 0 \end{cases} .$$

Od tod sledi, da je slučajna spremenljivka Z porazdeljena zvezno z gostoto:

$$f_Z(z) = \begin{cases} 0 & ; \, z \le 0 \\ 2\big(e^{-2z/3} - e^{-z}\big) & ; \, z \ge 0 \end{cases} .$$

17. Iz formule za volumen krogle dobimo, da je slučajni vektor (X, Y, Z) porazdeljen zvezno z gostoto:

$$f_{X,Y,Z}(x, y, z) = \begin{cases} \frac{3}{4\pi} & ; \, x^2 + y^2 + z^2 < 1 \\ 0 & ; \, \text{sicer.} \end{cases}$$

Najprej se lotimo dvorazsežnih robnih gostot. Za $x^2 + y^2 < 1$ dobimo:

$$f_{X,Y}(x, y) = \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} \frac{3}{4\pi}\, dx = \frac{3}{2\pi}\sqrt{1 - x^2 - y^2}\,.$$

Ker je skupna gostota $f_{X,Y,Z}$ simetrična v x, y in z, so slučajne spremenljivke X, Y in Z izmenljive. Sledi:

$$f_{X,Y}(u, v) = f_{X,Z}(u, v) = f_{Y,Z}(u, v) = \begin{cases} \frac{3}{2\pi}\sqrt{1 - u^2 - v^2} & ; \, u^2 + v^2 < 1 \\ 0 & ; \, \text{sicer.} \end{cases}$$

Enorazsežne robne gostote lahko zdaj dobimo bodisi neposredno iz trirazsežne gostote bodisi iz dvorazsežnih gostot. Naj bo $|x| < 1$. Iz trirazsežne gostote dobimo:

$$f_X(x) = \iint_{y^2+z^2<1-x^2} \frac{3}{4\pi}\, dy\, dz\,.$$

Če konstanto nesemo iz integrala, dobimo natančno ploščino kroga s polmerom $\sqrt{1 - y^2 - z^2}$. Sledi, da za $|x| < 1$ velja:

$$f_X(x) = \frac{3(1 - x^2)}{4}\,.$$

Iz dvorazsežne gostote $f_{X,Y}$ pa dobimo:

$$f_X(x) = \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{3}{2\pi}\sqrt{1 - x^2 - y^2}\, dy\,.$$

Če spet konstanto nesemo iz integrala, dobimo natančno ploščino polkroga s polmerom $\sqrt{1 - x^2}$ in pride isto kot prej. Spet upoštevamo izmenljivost in dobimo:

$$f_X(t) = f_Y(t) = f_Z(t) = \begin{cases} \frac{3}{4}(1 - t^2) & ; \, -1 < t < 1 \\ 0 & ; \, \text{sicer.} \end{cases}$$

Opomba. Preizkusimo lahko, da se dobljene robne gostote res zintegrirajo v 1. Pri enorazsežnih dobimo:

$$\frac{3}{4}\int_{-1}^{1}(1-x^2)\,\mathrm{d}x = \frac{3}{2}\int_{0}^{1}(1-x^2)\,\mathrm{d}x = \frac{3}{2}\left(1-\frac{1}{3}\right) = 1\,,$$

pri dvorazsežnih pa z uvedbo polarnih koordinat dobimo:

$$\begin{aligned}\iint_{x^2+y^2<1}\frac{3}{2\pi}\sqrt{1-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y &= \frac{3}{2\pi}\iint_{\substack{0<r<1\\0<\varphi<2\pi}}\sqrt{1-r^2}\,\mathrm{d}r\,\mathrm{d}\varphi = \\ &= \frac{3}{2\pi}\int_0^1\sqrt{1-r^2}\,r\,\mathrm{d}r\int_0^{2\pi}\mathrm{d}\varphi = \\ &= 3\int_0^1 r\sqrt{1-r^2}\,\mathrm{d}r\,.\end{aligned}$$

Nadalje s substitucijo $s = \sqrt{1-r^2}$ dobimo:

$$\iint_{x^2+y^2<1}\frac{3}{2\pi}\sqrt{1-x^2-y^2}\,\mathrm{d}x\,\mathrm{d}y = \frac{3}{2}\int_0^1\sqrt{s}\,\mathrm{d}s = 1\,.$$

18. a) Slučajni vektor (X,Y) ima dvorazsežno gostoto:

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\,e^{-(x^2+y^2)/2}\,.$$

Postavimo lahko $h(x,y) = \left(\sqrt{x^2+y^2},\ \arg(x,y)\right)$, kar je bijektivna preslikava iz odprte množice $A := \mathbb{R}^2\setminus\left((-\infty,0]\times\{0\}\right)$ na odprto množico $B := (0,\infty)\times(-\pi,\pi)$. Ker ima poltrak $(-\infty,0]\times\{0\}$ mero nič, lahko brez škode za splošnost privzamemo, da slučajni vektor (X,Y) zavzame vrednosti izključno v A. Za $(r,\theta)\in B$ velja:

$$h^{-1}(r,\theta) = \left(r\cos\theta, r\sin\theta\right) \qquad\text{in}\qquad \det J(h^{-1})(r,\theta) = \begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix} = r$$

in vsi parcialni odvodi so seveda zvezni. Sledi:

$$f_{R,\Theta}(r,\theta) = \begin{cases}\dfrac{r}{2\pi}\,e^{-r^2/2} & ;\ r>0,\ -\pi<\theta<\pi\\[2mm] 0 & ;\ \text{sicer.}\end{cases}$$

b) in c) Slučajni spremenljivki R in Θ sta neodvisni, pri čemer je slučajna spremenljivka Θ porazdeljena enakomerno na intervalu $(-\pi,\pi)$, slučajna spremenljivka R pa je porazdeljena zvezno z gostoto:

$$f_R(r) = \begin{cases} r\,e^{-r^2/2} & ;\ r>0\\ 0 & ;\ \text{sicer.}\end{cases}$$

Opomba. Porazdelitvi slučajne spremenljivke R pravimo *Rayleighova*[68] *porazdelitev* (glej tudi 52. nalogo iz 4. razdelka).

[68] John William Strutt, 3rd Baron Rayleigh (1842–1919), angleški fizik

Opomba. Tako je možno generirati psevdonaključna števila za normalno porazdelitev. Če sta namreč U in V neodvisni slučajni spremenljivki, porazdeljeni enakomerno na intervalu $(0,1)$, izračunamo:

$$R = \sqrt{-2\ln U}\,, \qquad \Theta = \pi(2V-1)\,,$$
$$X = R\cos\Theta\,, \qquad Y = R\sin\Theta\,.$$

Da se preveriti, da imata v tem primeru R in Θ ustrezni porazdelitvi. Tako sta potem tudi X in Y neodvisni in porazdeljeni standardno normalno.

19. Velja $(U,V) = h(X,Y)$, kjer je:

$$h(x,y) := \bigl(x+y,\ \sigma^2 y - \tau^2 x\bigr)\,.$$

Če je vsaj eden od parametrov σ in τ različen od nič, je preslikava $h\colon\ \mathbb{R}^2 \to \mathbb{R}^2$ bijektivna z inverzom:

$$h^{-1}(u,v) = \frac{1}{\sigma^2+\tau^2}\bigl(\sigma^2 u - v,\ \tau^2 u + v\bigr)\,,$$

ki je seveda zvezno parcialno odvedljiv z $\det J(h^{-1})(u,v) = \frac{1}{\sigma^2+\tau^2}$. Če privzamemo, da je $\mu = \nu = 0$, torej velja:

$$\begin{aligned}
&f_{U,V}(u,v) = \\
&= \frac{1}{2\pi\sigma\tau(\sigma^2+\tau^2)}\exp\left(-\frac{(\sigma^2 u - v)^2}{2\sigma^2(\sigma^2+\tau^2)^2} - \frac{(\tau^2 u + v)^2}{2\tau^2(\sigma^2+\tau^2)^2}\right) = \\
&= \frac{1}{2\pi\sigma\tau(\sigma^2+\tau^2)}\exp\left(-\frac{u^2}{2(\sigma^2+\tau^2)} - \frac{v^2}{2\sigma^2\tau^2(\sigma^2+\tau^2)}\right) = \\
&= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}}\exp\left(-\frac{u^2}{2(\sigma^2+\tau^2)}\right)\cdot\frac{1}{\sigma\tau\sqrt{2\pi(\sigma^2+\tau^2)}}\exp\left(-\frac{v^2}{2\sigma^2\tau^2(\sigma^2+\tau^2)}\right).
\end{aligned}$$

Torej sta U in V neodvisni z $U \sim \mathrm{N}\bigl(0, \sqrt{\sigma^2+\tau^2}\bigr)$ in $V \sim \mathrm{N}\bigl(0, \sigma\tau\sqrt{\sigma^2+\tau^2}\bigr)$. To seveda velja tudi za primer, ko je $\sigma = \tau = 0$.

Za splošna μ in ν pa označimo:

$$X' := X - \mu\,, \quad Y' := Y - \nu\,, \quad U' := X' + Y'\,, \quad V' := Y' - \tau^2 X\,.$$

Očitno sta tudi $X' \sim \mathrm{N}(0,\sigma)$ in $Y' \sim \mathrm{N}(0,\tau)$ neodvisni. Po prejšnjem sta U' in V' neodvisni z $U' \sim \mathrm{N}\bigl(0, \sqrt{\sigma^2+\tau^2}\bigr)$ in $V' \sim \mathrm{N}\bigl(0, \sigma\tau\sqrt{\sigma^2+\tau^2}\bigr)$. Sledi, da sta U in V neodvisni z:

$$U = U' + \mu + \nu \sim \mathrm{N}\bigl(\mu+\nu,\ \sqrt{\sigma^2+\tau^2}\bigr)\,,$$
$$V = V' + \sigma^2\nu - \tau^2\mu \sim \mathrm{N}\bigl(\sigma^2\nu - \tau^2\mu,\ \sigma\tau\sqrt{\sigma^2+\tau^2}\bigr)\,.$$

20. Če z A označimo odhod avtobusa, z S pa prihod študenta, Dogodek, da študent ujame avtobus, se ujema z dogodkom, da je $S \le A$ oziroma $S - A \le 0$. Iz prejšnje naloge sledi, da je $S - A \sim \mathrm{N}\bigl(1\text{ min}, (5\text{ min})^2\bigr)$, torej je iskana verjetnost enaka $1/2 - \Phi(1/5) \doteq 0{\cdot}42074$.

21. Slučajni vektor (X, Y) ima dvorazsežno gostoto:

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\, e^{-(x^2+(y-1)^2)/2}\,.$$

Tega preslikamo s preslikavo $h(x,y) = \left(xy, \frac{y}{x}\right)$, ki je definirana na odprti množici $A := \{(x,y)\ ;\ x \neq 0\}$. Velja:

$$\det Jh = \begin{vmatrix} y & x \\ -\frac{y}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{2y}{x}$$

in opazimo, da je $\mathbb{P}\big(\det Jh(X,Y) = 0\big) = \mathbb{P}(Y = 0) = 0$.

Pišimo $U = XY$ in $V = Y/X$. Če je $u > 0$ in $v > 0$, ima sistem enačb $h(x,y) = (u,v)$ dve rešitvi v $(x,y) \in A' := \{(x,y)\ ;\ x \neq 0, y \neq 0\}$:

$$x = \sqrt{\frac{u}{v}}\,, \quad y = \sqrt{uv}$$

in

$$x = -\sqrt{\frac{u}{v}}\,, \quad y = -\sqrt{uv}$$

Če je $u < 0$ in $v < 0$, ima sistem prav tako dve rešitvi v $(x,y) \in A'$:

$$x = \sqrt{\frac{u}{v}}\,, \quad y = -\sqrt{uv}$$

in

$$x = -\sqrt{\frac{u}{v}}\,, \quad y = \sqrt{uv}$$

Sicer pa sistem nima rešitev v $(x,y) \in A'$.

Opazimo še, da vselej velja $\det Jh(x,y) = 2v$. Ko vse skupaj sestavimo, dobimo:

$$\begin{aligned} f_{U,V}(u,v) &= \begin{cases} \dfrac{1}{4\pi|v|}\Big(e^{-(u/v+(\sqrt{uv}-1)^2)/2} + e^{-(u/v+(\sqrt{uv}+1)^2)/2}\Big) & ;\ u,v > 0 \text{ ali } u,v < 0 \\ 0 & ;\ \text{sicer} \end{cases} \\ &= \begin{cases} \dfrac{1}{4\pi|v|}\, e^{-(u/v+uv+1)/2} \operatorname{ch}\sqrt{uv} & ;\ u,v > 0 \text{ ali } u,v < 0 \\ 0 & ;\ \text{sicer.} \end{cases} \end{aligned}$$

22. Tu lahko vzamemo kar $A = \mathbb{R}^2$ (kar je po definiciji odprta množica) in $h(x,y) = x - y$. Enačbo $h(x,y) = z$ lahko rešujemo na x ali na y. Če jo rešujemo na y, je edina rešitev $y = x - z$. Nadalje je $J_y h(x,y) = h_y(x,y) = -1$ in zato $\det J_y h(x,y) = -1$.

Sledi:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x)\, f_Y(x-z)\,\mathrm{d}x = \\
&= \int_{\substack{x>0\\ x-z>0}} \lambda^2\, e^{-\lambda x-\lambda(x-z)}\,\mathrm{d}x = \\
&= \lambda^2\, e^{\lambda z} \int_{\max\{z,0\}}^{\infty} e^{-2\lambda x}\,\mathrm{d}x = \\
&= \frac{\lambda}{2}\, e^{\lambda(z-2\max\{z,0\})} = \\
&= \frac{\lambda}{2}\, e^{-\lambda|z|}\,.
\end{aligned}
$$

Tej porazdelitvi pravimo *Laplaceova*[69] *porazdelitev*.

23. Splača se najprej izračunati porazdelitev slučajne spremenljivke Z. Definirajmo torej $h(x,y) = y/x$. Ta funkcija je definirana na množici $\{(x,y)\ ;\ x \neq 0\}$, a splača se vzeti manjšo množico $A := \{(x,y)\ ;\ x,y > 0\}$, ki je prav tako odprta in tudi zanjo lahko privzamemo, da slučajni vektor (X,Y) zavzame vrednosti izključno v njej.

Enačbo $h(x,y) = z$ bomo reševali na y. Velja $J_y h(x,y) = h_y(x,y) = 1/x$ in zato $\det J_y h(x,y) = 1/x$, torej je $\det J_y h(x,y) \neq 0$ za vse $(x,y) \in A$.

Pri danih x in z torej iščemo rešitve enačbe $y/x = z$, za katere velja $(x,y) \in A$. Za $x,z > 0$ ima ta enačba edino rešitev $y = xz$, sicer pa enačba nima rešitev zahtevane oblike. Za $z > 0$ je torej:

$$
\begin{aligned}
f_Z(z) &= \int_0^{\infty} F_{X,Y}(x,xz)|x|\,\mathrm{d}x = \\
&= \int_0^{\infty} \frac{cx}{(1+x+xz)^3}\,\mathrm{d}x = \\
&= \frac{c}{(1+z)^2} \int_1^{\infty} \frac{t-1}{t^3}\,\mathrm{d}t = \\
&= \frac{c}{2(1+z)^2}\,,
\end{aligned}
$$

medtem ko lahko za $z \le 0$ postavimo $f_Z(z) = 0$.

Z integracijo pravkar dobljene gostote dobimo, da je $c = 2$. Torej je:

$$
f_Z(z) = \begin{cases} (1+z)^{-2} & ;\ z > 0 \\ 0 & ;\ \text{sicer.} \end{cases}
$$

Končno je še:

$$
\mathbb{P}(X < 2Y) = \mathbb{P}\left(Z > \frac{1}{2}\right) = \int_{1/2}^{\infty} \frac{1}{(1+z)^2}\,\mathrm{d}z = \frac{2}{3}\,.
$$

[69] Pierre-Simon de Laplace (1749–1827), francoski matematik, astronom, fizik in politik

24. *Prvi način*: s pomočjo transformacije gostote. Funkcija $h(x, y) = xy$ je definirana na celotni ravnini, a splača se omejiti na množico $A := \{(x, y) \; ; \; x, y > 0\}$, ki je odprta in za katero lahko privzamemo, da slučajni vektor (X, Y) zavzame vrednosti izključno v njej.

Če se odločimo, da enačbo $h(x, y) = z$ rešujemo y, velja $J_y h(x, y) = h_y(x, y) = x$ in zato $\det J_y h(x, y) = x$, torej je $\det J_y h(x, y) \neq 0$ za vse $(x, y) \in A$.

Pri danih x in z torej iščemo rešitve enačbe $y/x = z$, za katere velja $0 < x, y < 1$. Naj bo $0 < x, z/x < 1$, kar je ekvivalentno $0 < z < 1$, $z < x < 1$. Če je to res, ima ta enačba edino rešitev $y = z/x$, sicer pa enačba nima rešitev zahtevane oblike. Za $0 < z < 1$ je torej:

$$f_Z(z) = \int_z^1 \frac{1}{x} \, \mathrm{d}x = -\ln z \, ,$$

za ostale z pa lahko postavimo $f_Z(z) = 0$.

Drugi način: s pomočjo kumulativne porazdelitvene funkcije. Velja:

$$F_Z(z) = \iint_{\substack{0<x,y<1 \\ xy \le z}} \mathrm{d}x \, \mathrm{d}y \, .$$

Za $0 \le z \le 1$ se to prevede na:

$$F_Z(z) = z + \int_z^1 \frac{z}{x} \, \mathrm{d}x = z(1 - \ln z) \, ,$$

za ostale z pa je rezultat očiten – povzamemo v:

$$F_Z(z) = \begin{cases} 0 & ; \; z \le 0 \\ z(1 - \ln z) & ; \; 0 \le z \le 1 \\ 1 & ; \; z \ge 1 \, . \end{cases}$$

Ta funkcija je zvezna in odsekoma zvezno odvedljiva z odvodom f_Z, ki je enak kot v prejšnji točki.

25. *Prvi način*: s pomočjo transformacije gostote. Definirajmo torej $h(x, y) = x^2 + y^2$. Ta funkcija je definirana na celotni ravnini. Če se odločimo, da enačbo $h(x, y) = z$ rešujemo na y, velja $J_y h(x, y) = h_y(x, y) = 2y$ in zato $\det J_y h(x, y) = 2y$ in opazimo, da je $\mathbb{P}\big(\det J_y(X, Y) = 0\big) = \mathbb{P}(Y = 0) = 0$.

Pri danih x in z torej iščemo rešitve enačbe $x^2 + y^2 = z$, za katere velja $y \neq 0$. Za $z - x^2 > 0$ ima ta enačba rešitvi $y = \sqrt{z - x^2}$ in $y = -\sqrt{z - x^2}$, sicer pa enačba nima rešitve. Za $z > 0$ je torej:

$$\begin{aligned} f_Z(z) &= \int_{-\sqrt{z}}^{\sqrt{z}} \left(\frac{f_{X,Y}\big(x, \sqrt{z - x^2}\,\big)}{\big|2\sqrt{z - x^2}\big|} + \frac{f_{X,Y}\big(x, -\sqrt{z - x^2}\,\big)}{\big|-2\sqrt{z - x^2}\big|} \right) \mathrm{d}x = \\ &= \frac{C}{z^{5/2}(1 + z)} \int_{-\sqrt{z}}^{\sqrt{z}} x^2 \sqrt{z - x^2} \, \mathrm{d}x \, , \end{aligned}$$

medtem ko lahko za $z \le 0$ postavimo $f_Z(z) = 0$. Za $z > 0$ z upoštevanjem sodosti in substitucijo $t = x^2/z$ dobimo:

$$f_Z(z) = \frac{C_1}{(1+z)\sqrt{z}}\,,$$

kjer je:

$$C_1 = C \int_0^1 \sqrt{t(1-t)}\,\mathrm{d}t\,.$$

S substitucijo $t = 1/(1+z)$ izračunamo:

$$\int_0^\infty \frac{\mathrm{d}z}{(1+z)\sqrt{z}} = \int_0^1 \frac{\mathrm{d}t}{\sqrt{t(1-t)}} = \mathrm{B}\big(\tfrac{1}{2}, \tfrac{1}{2}\big) = \frac{\Big[\Gamma\big(\frac{1}{2}\big)\Big]^2}{\Gamma(1)} = \pi\,,$$

od koder sledi $C_1 = 1/\pi$, torej:

$$f_Z(z) = \begin{cases} \dfrac{1}{\pi(1+z)\sqrt{z}} & ;\ z > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Sedaj lahko izračunamo tudi konstanto C. Iz:

$$\int_0^1 \sqrt{t(1-t)}\,\mathrm{d}t = \mathrm{B}\big(\tfrac{3}{2}, \tfrac{3}{2}\big) = \frac{\Big[\Gamma\big(\frac{3}{2}\big)\Big]^2}{\Gamma(3)} = \frac{\pi}{8}$$

sledi $C = 8C_1/\pi = 8/\pi^2$.

Drugi način: z uporabo kumulativne porazdelitvene funkcije in polarnih koordinat. Za $z \ge 0$ velja:

$$F_Z(z) = \iint_{x^2+y^2<z} \frac{Cx^2y^2}{(1+x^2+y^2)(x^2+y^2)^{5/2}}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Z uporabo polarnih koordinat dobimo:

$$F_Z(z) = C_1 \int_0^{\sqrt{z}} \frac{\mathrm{d}r}{1+r^2} = C_2 \operatorname{arctg}\sqrt{r}\,,$$

kjer je:

$$C_2 = C \int_0^{2\pi} \cos^2\varphi \sin^2\varphi\,\mathrm{d}\varphi\,.$$

Ker mora biti $\lim_{z\to\infty} F_Z(z) = 1$, velja $C_2 = 2/\pi$. Za $z > 0$ torej velja:

$$f_Z(z) = F_Z'(z) = \frac{1}{\pi(1+z)\sqrt{z}}\,,$$

kar je isto kot prej. Sedaj lahko ponovno izračunamo tudi konstanto C. Iz:

$$\int_0^{2\pi} \cos^2\varphi \sin^2\varphi\,\mathrm{d}\varphi = \frac{1}{4}\int_0^{2\pi} \sin^2(2\varphi)\,\mathrm{d}\varphi = \frac{\pi}{4}$$

sledi $C = 4C_2/\pi = 8/\pi^2$, spet isto kot prej.

26. Ne le, da je $X \neq 0$, privzamemo lahko, da slučajni vektor (X, Y, Z) zavzame vrednosti izključno v odprti množici $A := \left\{(x, y, z) \in \mathbb{R}^3 \; ; \; x, y > 0 \, , \; 0 < z < \sqrt{xy}\,\right\}$. Definirajmo:

$$h(x, y, z) = \frac{xy - z^2}{x}$$

in enačbo $h(x, y, z) = s$ rešujmo na z. Natančneje, pri danih $x, y > 0$ poglejmo, pri katerih $s \in \mathbb{R}$ ima enačba rešitev, za katero je $0 < z < \sqrt{xy}$.

- Če je $s > y$, enačba nima rešitve.
- Če je $s = y$, ima enačba sicer dvojno rešitev $z = 0$, a ta ne izpolnjuje pogoja.
- Če je $s < y$, ima enačba rešitvi $z = \pm\sqrt{x(y - s)}$, toda:
  - Le pri $z = \sqrt{x(y - s)}$ je $z > 0$.
  - Le pri $s > 0$ je $z < \sqrt{xy}$.

Sklep: če je $0 < s < y$, ima obravnavana enačba v okviru danega pogoja natanko eno rešitev $z = \sqrt{x(y - s)}$, sicer pa enačba v okviru tega pogoja nima rešitve. Nadalje velja:

$$h_z(x, y, z) = -\frac{2z}{x}\,,$$

kar je za vse $(x, y, z) \in A$ različno od nič. Sledi, da je S porazdeljena zvezno in da za $s > 0$ velja:

$$\begin{aligned}
f_S(s) &= \iint_{\substack{x>0\\y>s}} f_{X,Y,Z}\Big(x, y, \sqrt{x(y-s)}\,\Big) \left|\frac{x}{2\sqrt{x(y-s)}}\right| \mathrm{d}x\, \mathrm{d}y = \\
&= \iint_{\substack{x>0\\y>s}} \frac{1}{4\pi\sqrt{s(y-s)}}\, e^{-(x+y)/2}\, \mathrm{d}x\, \mathrm{d}y = \\
&= \frac{1}{4\pi\sqrt{s}} \int_0^\infty e^{-x/2}\, \mathrm{d}x \int_s^\infty \frac{1}{\sqrt{y-s}}\, e^{-y/2}\, \mathrm{d}y = \\
&= \frac{1}{2\pi\sqrt{s}} \int_s^\infty \frac{1}{\sqrt{y-s}}\, e^{-y/2}\, \mathrm{d}y = \\
&= \frac{1}{\pi\sqrt{s}}\, e^{-s/2} \int_0^\infty e^{-t^2/2}\, \mathrm{d}t = \\
&= \frac{1}{\sqrt{2\pi s}}\, e^{-s/2}\,,
\end{aligned}$$

medtem ko lahko za $s \le 0$ postavimo $f_S(s) = 0$. Dobljeno gostoto prepoznamo kot $S \sim \mathrm{Gama}\left(\frac{1}{2}, \frac{1}{2}\right) = \chi^2(1)$.

27. *Prvi način.* Najprej bomo izračunali kumulativno porazdelitveno funkcijo slučajne spremenljivke T_n. Dogodek $T_n \le t$ se zgodi, če se je do časa T_n zgodilo že vsaj n danih pojavov, torej za $t \ge 0$ velja:

$$F_{T_n}(t) = \mathbb{P}(T_n \le t) = 1 - e^{-\lambda t}\left(1 + \lambda t + \frac{\lambda^2 t^2}{2!} + \cdots + \frac{\lambda^{n-1} t^{n-1}}{(n-1)!}\right).$$

Po odvajanju dobimo:

$$f_{T_n}(t) = \begin{cases} \dfrac{\lambda^n}{(n-1)!}\, t^{n-1}\, e^{-\lambda t} & ; \ t \geq 0 \\ 0 & ; \ \text{sicer,} \end{cases}$$

torej je $T_n \sim \mathrm{Gama}(n, \lambda)$. Za $n = 1$ velja $\mathrm{Gama}(1, \lambda) = \mathrm{Exp}(\lambda)$.

Drugi način. Označimo z N_t število pojavov, ki se je zgodilo do vključno časa t. Za $0 \leq t < t + h$ velja:

$$F_{T_n}(t+h) - F_{T_n}(t) = \mathbb{P}(N_{t+h} \geq n) - \mathbb{P}(N_t \geq n)\,.$$

Situacija, ko je $N_t \sim \mathrm{Pois}(\lambda t)$ in $N_{t+h} \sim \mathrm{Pois}(\lambda(t+h))$, nastopi tudi v primeru, ko je razlika $N_{t+h} - N_t$ neodvisna od N_t in porazdeljena po Poissonu $\mathrm{Pois}(\lambda h)$; to sicer velja tudi pri Poissonovem procesu, a pri izračunu zgornje verjetnosti lahko to privzamemo brez škode za splošnost. Tedaj velja:

$$\begin{aligned} F_{T_n}(t+h) - F_{T_n}(t) &= \mathbb{P}(N_t < n, N_{t+h} \geq n) = \\ &= \sum_{i=0}^{n-1} \sum_{j=n}^{\infty} \mathbb{P}(N_t = i, N_{t+h} = j) = \\ &= \sum_{i=0}^{n-1} \sum_{j=n}^{\infty} \mathbb{P}(N_t = i, N_{t+h} - N_t = j - i) = \\ &= \sum_{i=0}^{n-1} \sum_{j=n}^{\infty} \frac{\lambda^j t^i h^{j-i} e^{-\lambda(t+h)}}{i!\,(j-i)!}\,. \end{aligned}$$

Opazimo, da je potenca pri h vselej najmanj 1. Če delimo s h in naredimo limito, ko gre h proti nič, na levi strani dobimo natanko odvod, to je gostoto porazdelitve. Na desni strani pa ostanejo le členi s h^1 (utemeljitev konvergence bomo opustili), to pa je le člen z $i = n - 1$ in $n = k$. Torej velja:

$$f_{T_n}(t) = \frac{\lambda^n t^{n-1}\, e^{-\lambda t}}{(n-1)!}\,,$$

kar je isto kot prej.

28. Za $a, b \in \mathbb{N}$ se da porazdelitev izpeljati z verjetnostnim premislekom. Vzemimo Poissonov tok pojavov z intenziteto λ ter naj bo S čas, ob katerem se zgodi a-ti, T pa čas, ki mine od a-tega do $(a + b)$-tega pojava. Tedaj iz prejšnje naloge dobimo, da je $S \sim \mathrm{Gama}(a, \lambda)$. Nadalje je $S + T$ čas, ob katerem se zgodi $(a + b)$-ti pojav, zato je $S + T \sim \mathrm{Gama}(a + b, \lambda)$.

Tudi za Poissonov tok velja *krepka časovno homogena lastnost Markova*: pogojno na zgodovino, tj. dogajanje do časa S, se nadaljevanje Poissonovega toka obnaša enako kot Poissonov tok od začetka. Zato ima ob zgornji definiciji slučajna spremenljivka

T porazdelitev Gama(b, λ) in je neodvisna od S. S tem je navzkrižna porazdelitev slučajnih spremenljivk S in T natančno določena.

Ker je porazdelitev vsote natančno določena z navzkrižno porazdelitvijo, lahko sklepamo, da, brž ko sta slučajni spremenljivki $S \sim$ Gama(a, λ) in $T \sim$ Gama(b, λ) neodvisni in je $a, b \in \mathbb{N}$, velja $U = S + T \sim$ Gama$(a + b, \lambda)$.

Za splošni primer pa bomo porazdelitev izpeljali računsko. Za $u > 0$ velja:

$$f_U(u) = \int_{-\infty}^{\infty} f_S(s)\, f_T(u - s)\,\mathrm{d}s = \frac{\lambda^{a+b}\, e^{-\lambda u}}{\Gamma(a)\,\Gamma(b)} \int_0^u s^{a-1}(u - s)^{b-1}\,\mathrm{d}s\,.$$

S substitucijo $t = s/u$ dobimo:

$$f_U(u) = C\, u^{a+b-1}\, e^{-\lambda u}\,,$$

kjer je:

$$C = \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)} \int_0^1 t^{a-1}(1 - t)^{b-1}\,\mathrm{d}t = \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)}\,\mathrm{B}(a, b)\,.$$

Od tod dobimo, da je res $U \sim$ Gama$(a + b, \lambda)$. Ker od tod sledi, da mora biti $C = \lambda^{a+b}/\,\Gamma(a + b)$, sledi še znana zveza:

$$\mathrm{B}(a, b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a + b)}\,,$$

29. Za $0 \le t \le 1$ označimo z N_t število gostov, ki so prišli na zabavo do časa t. Tedaj je $\{T_k \le t\} = \{N_t \ge k\}$.

Prvi način. Ker je $N_t \sim \mathrm{Bin}(n, t)$, ima T_k kumulativno porazdelitveno funkcijo:

$$F_{T_k}(t) = \sum_{l=k}^{n} \binom{n}{l} t^l (1 - t)^{n-l}\,; \quad 0 \le t \le 1\,.$$

Z odvajanjem dobimo še porazdelitveno gostoto (pazimo na zadnji člen):

$$\begin{aligned}
f_{T_k}(t) &= \sum_{l=k}^{n} \binom{n}{l} \Big[l\, t^{l-1}(1 - t)^{n-l} - (n - l) t^l (1 - t)^{n-l-1} \Big] = \\
&= \sum_{l=k}^{n} \binom{n}{l} l\, t^{l-1}(1 - t)^{n-l} - \sum_{l=k}^{n-1} \binom{n}{l} (n - l) t^l (1 - t)^{n-l-1} = \\
&= \sum_{l=k}^{n} \frac{n!}{(l - 1)!\,(n - l)!}\, t^{l-1}(1 - t)^{n-l} - \sum_{l=k}^{n-1} \frac{n!}{l!\,(n - l - 1)!}\, t^l (1 - t)^{n-l-1} = \\
&= \sum_{l=k}^{n} \frac{n!}{(l - 1)!\,(n - l)!}\, t^{l-1}(1 - t)^{n-l} - \sum_{j=k+1}^{n} \frac{n!}{(j - 1)!\,(n - j)!}\, t^{j-1}(1 - t)^{n-j} = \\
&= \frac{n!}{(k - 1)!\,(n - k)!}\, t^{k-1}(1 - t)^{n-k}\,.
\end{aligned}$$

Dobili smo *porazdelitev beta*: $T_k \sim \mathrm{Beta}(k, n-k+1)$ (glej tudi 31. nalogo).

Drugi način. Gostoto zlahka dobimo z naslednjim fizikalnim, a matematično ne ravno korektnim premislekom: dogodek $\{T_k \in [t, t+\mathrm{d}t]\}$ pomeni, da je do časa t prišlo $k-1$ gostov, en gost je prišel v infinitezimalnem časovnem intervalu od t do $t+\mathrm{d}t$, preostalih $n-k$ gostov pa je prišlo od časa $t+\mathrm{d}t$ naprej. Torej je:

$$\mathbb{P}(T_k \in [t, t+\mathrm{d}t]) = \frac{n!}{(k-1)!\,1!\,(n-k)!}\, t^{k-1}\,\mathrm{d}t\,(1-t)^{n-k},$$

od koder sledi:

$$f_{T_k}(t) = \frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k},$$

kar je isto kot prej. Korektna izpeljava pa je naslednja: za $0 \le t < t+h \le 1$ velja:

$$F_{T_k}(t+h) - F_{T_k}(t) = \mathbb{P}(N_{t+h} \ge k) - \mathbb{P}(N_t \ge k).$$

Ker je $\{N_t \ge k\} \subseteq \{N_{t+h} \ge k\}$, je tudi:

$$\begin{aligned}
F_{T_k}(t+h) - F_{T_k}(t) &= \mathbb{P}(N_{t+h} \ge k) - \mathbb{P}(N_t \ge k) = \\
&= \mathbb{P}(N_t < k, N_{t+h} \ge k) = \\
&= \mathbb{P}(N_t < k, N_1 - N_{t+h} \le n-k) = \\
&= \sum_{i=0}^{k-1} \sum_{j=k}^{n} \mathbb{P}(N_t = i, N_{t+h} - N_t = j-i, N_1 - N_{t+h} = n-j) = \\
&= \sum_{i=0}^{k-1} \sum_{j=k}^{n} \frac{n!}{i!\,(j-i)!\,(n-j)!}\, t^i h^{j-i} (1-t-h)^{n-j}.
\end{aligned}$$

Opazimo, da je eksponent pri h vselej najmanj 1. Če delimo s h in naredimo limito, ko gre h proti nič, na levi strani dobimo natanko *desni* odvod funkcije F_{T_k} v točki t. Na desni strani pa ostanejo le členi s h^1, to pa je le člen z $i = k-1$ in $j = k$. Torej je prej omenjeni desni odvod enak:

$$\frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k},$$

kar je spet isto kot prej. Dokazati moramo le še, da enako velja tudi za *levi* odvod. Za $0 \le t-h < t \le 1$ podobno kot prej izračunamo:

$$F_{T_k}(t) - F_{T_k}(t-h) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} \frac{n!}{i!\,(j-i)!\,(n-j)!}\, (t-h)^i h^{j-i} (1-t)^{n-j}.$$

in limita, ko gre h proti nič, je ista kot prej. To pomeni, da je F_{T_k} na celotnem intervalu $(0,1)$ odvedljiva, in ni se težko prepričati, da je odvod zvezen. Poleg tega je F_{T_k} povsod zvezna. Sledi:

$$f_{T_k}(t) = \frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k}.$$

30. a) Če z f_X in f_Y označimo ustrezni gostoti:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/(2\sigma^2)}\,, \quad f_Y(y) = \frac{\lambda^a}{\Gamma(a)}\, y^{a-1}\, e^{-\lambda y}\,;\ y > 0\,,$$

velja:

$$\begin{aligned} f_T(t) &= \int_0^\infty f_X\big(t\sqrt{y}\big)\, f_Y(y)\sqrt{y}\,\mathrm{d}y = \\ &= \frac{\lambda^a}{\sigma\sqrt{2\pi}\,\Gamma(a)}\int_0^\infty y^{a-1/2}\, e^{-(\lambda+t^2/(2\sigma^2))y}\,\mathrm{d}y = \\ &= \frac{\Gamma\big(a+\frac{1}{2}\big)}{\Gamma(a)}\,\frac{1}{\sqrt{2\pi\lambda\sigma^2}\left(1+\dfrac{t^2}{2\lambda\sigma^2}\right)^{a+1/2}} = \\ &= \frac{1}{\mathrm{B}\big(a,\frac{1}{2}\big)\sqrt{2\lambda\sigma^2}\left(1+\dfrac{t^2}{2\lambda\sigma^2}\right)^{a+1/2}}\,. \end{aligned}$$

b) Spomnimo se, da je:

$$\frac{X_1^2+\cdots+X_n^2}{\sigma^2} \sim \chi^2(n) = \mathrm{Gama}\left(\frac{n}{2}, \frac{1}{2}\right)\,,$$

torej je (glej 50. nalogo iz 4. razdelka):

$$\frac{X_1^2+\cdots+X_n^2}{n} \sim \mathrm{Gama}\left(\frac{n}{2}, \frac{n}{2\sigma^2}\right)\,.$$

Tako lahko primer prevedemo na prejšnjo točko in dobimo, da ima Studentova porazdelitev z n prostostnimi stopnjami gostoto:

$$f_T(t) = \frac{1}{\mathrm{B}\big(\frac{n}{2},\frac{1}{2}\big)\sqrt{n}\left(1+\dfrac{t^2}{n}\right)^{(n+1)/2}}\,,$$

kar je neodvisno od σ. Ko gre n proti neskončno, to konvergira proti gostoti standardne normalne porazdelitve.

31. Najprej opazimo, da Z skoraj gotovo zavzame vrednosti na intervalu $(0,1)$. Za $0 < z < 1$ ima enačba $x/(x+y) = z$ enolično rešitev $y = x(1-z)/z$. Za ta y je:

$$\frac{\partial}{\partial y}\frac{x}{x+y} = -\frac{x}{(x+y)^2} = -\frac{z^2}{x}\,.$$

Po krajšem računu dobimo, da za $0 < z < 1$ velja:

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)}\,\frac{(1-z)^{b-1}}{z^{b+1}}\int_0^\infty x^{a+b-1}e^{-\lambda x/z}\,\mathrm{d}x = \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, z^{a-1}(1-z)^{b-1}\,. \end{aligned}$$

Gostota je ravno integrand v integralski definiciji funkcije beta in tudi dobljeno porazdelitev imenujemo *porazdelitev beta* in pišemo $Z \sim \mathrm{Beta}(a, b)$. Ker je $\int_{-\infty}^{\infty} f_Z(z)\,\mathrm{d}z = \int_0^1 f_Z(z)\,\mathrm{d}z = 1$, se predfaktor izraža tudi s funkcijo beta, tako da velja:

$$f_Z(z) = \begin{cases} \dfrac{1}{\mathrm{B}(a,b)}\, z^{a-1}(1-z)^{b-1} & ;\ 0 < z < 1 \\ 0 & ;\ \text{sicer} \end{cases}$$

in spet sledi dobro znana zveza $\mathrm{B}(a,b) = \dfrac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$.

32. *Prvi način.* Če ustrezni gostoti označimo z f_X in f_Y, za $q > 0$ velja:

$$\begin{aligned} f_Q(q) &= \int_0^\infty f_X(qy)\, f_Y(y)\, y\,\mathrm{d}y = \\ &= \frac{\lambda^a \mu^b}{\Gamma(a)\,\Gamma(b)}\, q^{a-1} \int_0^\infty y^{a+b-1} e^{-(\lambda q+\mu)y}\,\mathrm{d}y = \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, \frac{\lambda^a \mu^b}{(\lambda q+\mu)^{a+b}}\, q^{a-1} = \\ &= \frac{1}{\mathrm{B}(a,b)\left(1+\dfrac{\mu}{\lambda q}\right)^a \left(1+\dfrac{\lambda q}{\mu}\right)^b q}\,. \end{aligned}$$

Torej je:

$$f_Q(q) = \begin{cases} \left[\mathrm{B}(a,b)\left(1+\frac{\mu}{\lambda q}\right)^a\left(1+\frac{\lambda q}{\mu}\right)^b q\right]^{-1} & ;\ q > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

Drugi način: prevedemo na prejšnjo nalogo, najprej za poseben primer, ko je $\lambda = \mu$. Če definiramo $Z := X/(X+Y) \sim \mathrm{Beta}(a,b)$, velja $Q = h(Z)$, kjer je:

$$h(z) = \frac{z}{z-1}\,, \quad h^{-1}(q) = \frac{q}{q+1}\,, \quad (h^{-1})'(q) = -\frac{1}{(q+1)^2}\,,$$

torej za $q > 0$ velja:

$$f_Q(q) = f_Z\left(\frac{q}{q+1}\right)\frac{1}{(q+1)^2} = \frac{q^{a-1}}{\mathrm{B}(a,b)(q+1)^{a+b}}$$

ali tudi:

$$f_Q(q) = \frac{1}{\mathrm{B}(a,b)\left(1+\frac{1}{q}\right)^a (1+q)^b\, q}\,.$$

Za splošni primer definiramo:

$$X' := \lambda X \sim \mathrm{Gama}(a,1)\,, \quad Y' := \mu Y \sim \mathrm{Gama}(b,1) \quad \text{in} \quad Q' := \frac{X'}{Y'} = \frac{\lambda}{\mu}\,Q\,.$$

Tedaj velja $f_Q(q) = \dfrac{\lambda}{\mu}\, f_{Q'}\left(\dfrac{\lambda}{\mu}\, q\right)$, ker je isto kot prej.

33. Naj bosta U in V neodvisni ter $U \sim \chi^2(m)$ in $V \sim \chi^2(n)$. Tedaj je

$$F := \frac{n}{m}\frac{U}{V} \sim \mathrm{F}(m,n) \quad \text{in} \quad B := \frac{U}{U+V} \sim \mathrm{Beta}\left(\frac{m}{2}, \frac{n}{2}\right).$$

Med tema slučajni spremenljivki velja funkcijska zveza:

$$B = \frac{mF}{mF+n} \quad \text{oziroma} \quad F = \frac{n}{m}\frac{B}{1-B}.$$

34. Verjetnost bomo izrazili s standardnim dvorazsežnim normalnim vektorjem: poiskali bomo matriko $\mathbf{A}$, za katero velja $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 8 & 2 \\ 2 & 5 \end{bmatrix} =: \mathbf{\Sigma}$. Tedaj bo $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \stackrel{\mathrm{d}}{=} \mathbf{A}\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$, kjer sta Z_1 in Z_2 neodvisni standardni enorazsežni normalni spremenljivki. To gre lahko na več načinov.

Prvi način: postavimo $\mathbf{A} = \mathbf{\Sigma}^{1/2}$, kar izračunamo s pomočjo lastnih vrednosti. Matrika $\mathbf{\Sigma}$ ima normiran lastni vektor $\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ za lastno vrednost 4 in normiran lastni vektor $\frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ za lastno vrednost 9. Sledi:

$$\mathbf{\Sigma}^{1/2} = \frac{1}{5}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 14 & 2 \\ 2 & 11 \end{bmatrix}.$$

Torej je:

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \mathbb{P}(14Z_1 + 2Z_2 > 0,\ 2Z_1 + 11Z_2 > 0).$$

Zdaj pa uporabimo, da je slučajni kot $\Theta := \arg(Z_1, Z_2)$ porazdeljen enakomerno na intervalu od $-\pi$ do π (glej 18. nalogo). Sledi:

$$\begin{aligned} \mathbb{P}(X_1 > 0, X_2 > 0) &= \mathbb{P}\left(-\operatorname{arctg}\frac{2}{11} < \Theta < \pi - \operatorname{arctg} 7\right) = \\ &= \frac{1}{2\pi}\left(\pi - \operatorname{arctg} 7 + \operatorname{arctg}\frac{2}{11}\right) = \\ &= \frac{1}{2} - \frac{\operatorname{arctg} 7 - \operatorname{arctg}\frac{2}{11}}{2\pi} \doteq \\ &\doteq 0{\cdot}301\,. \end{aligned}$$

Zadevo pa lahko pogledamo tudi še bolj geometrijsko. Pogoj $14Z_1 + 2Z_2 > 0$ pomeni, da je $\theta_1 - \frac{\pi}{2} < \Theta < \theta_1 + \frac{\pi}{2}$, kjer je $\theta_1 = \operatorname{arctg}\frac{1}{7}$ kot, ki pripada vektorju $\boldsymbol{v}_1 = (14, 2)$. Podobno pogoj $2Z_1 + 11Z_2 > 0$ pomeni, da je $\theta_2 - \frac{\pi}{2} < \Theta < \theta_2 + \frac{\pi}{2}$, kjer je $\theta_2 = \operatorname{arctg}\frac{11}{2}$ kot, ki pripada vektorju $\boldsymbol{v}_2 = (2, 11)$. Presek obeh pogojev je pogoj $\theta_2 - \frac{\pi}{2} < \Theta < \theta_1 + \frac{\pi}{2}$ in verjetnost tega dogodka je:

$$\frac{\pi - (\theta_2 - \theta_1)}{2\pi} = \frac{1}{2} - \frac{\theta_2 - \theta_1}{2\pi}.$$

Razlika $\theta_2 - \theta_1$ pa je kot med vektorjema $\boldsymbol{v}_1$ in $\boldsymbol{v}_2$, ki je enak $\arccos \frac{\sqrt{10}}{10}$, torej je iskana verjetnost enaka $\frac{1}{2} - \frac{1}{2\pi} \arccos \frac{\sqrt{10}}{10}$.

Iz identitete $\arccos x = \operatorname{arctg} \frac{\sqrt{1-x^2}}{x}$, ki velja za vse $x \in (0, 1]$, dobimo, da je $\theta_2 - \theta_1 = \operatorname{arctg} 3$. Iz ujemanja s prej izračunano verjetnostjo dobimo trigonometrijsko zvezo:

$$\operatorname{arctg} 3 + \operatorname{arctg} \frac{2}{11} = \operatorname{arctg} 7\,.$$

Ta enakost pa sledi iz znane zveze $\operatorname{arctg} x + \operatorname{arctg} y = \operatorname{arctg} \dfrac{x+y}{1-xy}$, ki velja, brž ko je $xy < 1$.

Drugi način: uporabimo *razcep Choleskega*[70]:

$$\begin{bmatrix} \alpha & \boldsymbol{b}^{\mathrm{T}} \\ \boldsymbol{b} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \sqrt{\alpha} & \mathbf{0} \\ \frac{1}{\sqrt{\alpha}}\boldsymbol{b} & \mathbf{I} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \frac{1}{\alpha}\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \sqrt{\alpha} & \frac{1}{\sqrt{\alpha}}\boldsymbol{b}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

V našem konkretnem primeru je:

$$\begin{bmatrix} 8 & 2 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ \frac{\sqrt{2}}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{9}{2} \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & \frac{\sqrt{2}}{2} \\ 0 & 1 \end{bmatrix} = \mathbf{A}\mathbf{A}^{\mathrm{T}}\,,$$

kjer je:

$$\mathbf{A} = \begin{bmatrix} 2\sqrt{2} & 0 \\ \frac{\sqrt{2}}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{3\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{3\sqrt{2}}{2} \end{bmatrix}.$$

Dobimo:

$$\mathbb{P}(X_1 > 0, X_2 > 0) = \mathbb{P}(Z_1 > 0,\ Z_1 + 3Z_2 > 0)\,.$$

Kot med vektorjema $(1, 0)$ in $(1, 3)$ se seveda ujema s kotom med vektorjema $(14, 2)$ in $(2, 11)$ iz prvega načina.

Tretji način: uporabimo razcep $\mathbf{LDL}^{\mathrm{T}}$, kjer je korak:

$$\begin{bmatrix} \alpha & \boldsymbol{b}^{\mathrm{T}} \\ \boldsymbol{b} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \frac{1}{\alpha}\boldsymbol{b} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \alpha & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \frac{1}{\alpha}\boldsymbol{b}\boldsymbol{b}^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\alpha}\boldsymbol{b}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

V našem konkretnem primeru je:

$$\begin{bmatrix} 8 & 2 \\ 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & \frac{9}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{4} \\ 0 & 1 \end{bmatrix} = \mathbf{A}\mathbf{A}^{\mathrm{T}}\,,$$

kjer je:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \frac{3\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 2\sqrt{2} & 0 \\ \frac{\sqrt{2}}{2} & \frac{3\sqrt{2}}{2} \end{bmatrix},$$

kar je isto kot pri drugem načinu.

[70] André-Louis Cholesky (1875–1918), francoski oficir in matematik

35. *Prvi način.* Slučajni vektor:

$$\begin{bmatrix} X+Y \\ X+Z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

je porazdeljen dvorazsežno normalno s pričakovano vrednostjo nič in kovariančno matriko:

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} .$$

Poiščimo matriko $\mathbf{A}$, za katero velja $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}$. Ena izmed možnosti je spet, da matriko korenimo. Matrika $\mathbf{\Sigma}$ ima normiran lastni vektor $\frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ za lastno vrednost 1 in normiran lastni vektor $\frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ za lastno vrednost 3. Sledi:

$$\mathbf{\Sigma}^{1/2} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} \sqrt{3}+1 & \sqrt{3}-1 \\ \sqrt{3}-1 & \sqrt{3}+1 \end{bmatrix} .$$

Če je torej $\begin{bmatrix} U \\ V \end{bmatrix}$ standardni dvorazsežni normalni vektor, je $\begin{bmatrix} X+Z \\ Y+Z \end{bmatrix}$ porazdeljen enako kot:

$$\mathbf{\Sigma}^{1/2}\begin{bmatrix} U \\ V \end{bmatrix} = \frac{1}{2}\begin{bmatrix} (\sqrt{3}+1)U + (\sqrt{3}-1)V \\ (\sqrt{3}-1)U + (\sqrt{3}+1)V \end{bmatrix} .$$

Podobno kot pri prejšnji nalogi iz kota med vektorjema $(\sqrt{3}+1, \sqrt{3}-1)$ in $(\sqrt{3}-1, \sqrt{3}+1)$, ki je enak $\pi/3$, dobimo, da je iskana verjetnost enaka $1/3$.

Drugi način[71]: uporabimo radialno simetrijo porazdelitve slučajnega vektorja $\boldsymbol{W} := (X, Y, Z)$. Le-ta ima namreč radialno simetrično gostoto:

$$f_{X,Y,Z}(x, y, z) = (2\pi)^{-3/2} e^{-(x^2+y^2+z^2)/2} ,$$

Dogodek $A := \{X + Y \geq 0, X + Z \geq 0\}$ je možno zapisati tudi kot:

$$A = \big\{\langle \boldsymbol{W}, \boldsymbol{a}\rangle \geq 0,\ \langle \boldsymbol{W}, \boldsymbol{b}\rangle \geq 0\big\} ,$$

kjer je $\boldsymbol{a} = (1, 1, 0)$ in $\boldsymbol{b} = (1, 0, 1)$. Naj bo Θ orientiran kot, ki ga projekcija vektorja $\boldsymbol{W}$ na ravnino, določeno z vektorjema $\boldsymbol{a}$ in $\boldsymbol{b}$, oklepa z vektorjem $\boldsymbol{a}$, gledano proti vektorju $\boldsymbol{b}$. Kot naj se meri od $-\pi$ do π. Če je γ kot med vektorjema $\boldsymbol{a}$ in $\boldsymbol{b}$, dogodek A ustreza dogodku $\big\{\gamma - \frac{\pi}{2} < \Theta < \frac{\pi}{2}\big\}$.

Ker je porazdelitev slučajnega vektorja $\boldsymbol{W}$ invariantna za vse ortogonalne transformacije, je invariantna tudi za rotacije vzdolž ravnine, ki jo določata vektorja $\boldsymbol{a}$ in

[71] Idejo dal Teo Kukuljan

b. Od tod sledi, da je slučajna spremenljivka Θ porazdeljena enakomerno od $-\pi$ do π. Torej podobno kot pri prejšnji nalogi velja:

$$\mathbb{P}(A) = \frac{\pi - \gamma}{2\pi} \,.$$

Kot γ se seveda ujema s kotom med vektorjema iz prvega načina.

36. Velja:

$$\boldsymbol{X}_1 + \boldsymbol{X}_2 = \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \,,$$

kar je porazdeljeno n-razsežno normalno s pričakovano vrednostjo $\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$ in kovariančno matriko $\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$.

V enorazsežnem primeru torej velja: če sta $X_1 \sim \mathrm{N}(\mu_1, \sigma_1)$ in $X_2 \sim \mathrm{N}(\mu_2, \sigma_2)$ neodvisni, je $X_1 + X_2 \sim \mathrm{N}\big(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\,\big)$.

37. Slučajni vektor:

$$\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 + \mathbf{C}\boldsymbol{X}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$$

je porazdeljen $(m+n)$-razsežno normalno s pričakovano vrednostjo:

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 + \mathbf{C}\boldsymbol{\mu}_1 \end{bmatrix}$$

in kovariančno matriko:

$$\begin{aligned} &\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} = \\ &\qquad = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{11}\mathbf{C}^{\mathrm{T}} \\ \boldsymbol{\Sigma}_{21} + \mathbf{C}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{22} + \mathbf{C}\boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{21}\mathbf{C}^{\mathrm{T}} + \mathbf{C}\boldsymbol{\Sigma}_{11}\mathbf{C}^{\mathrm{T}} \end{bmatrix} . \end{aligned}$$

Torej sta slučajna vektorja $\boldsymbol{X}_1$ in $\boldsymbol{X}_2 + \mathbf{C}\boldsymbol{X}_1$ neodvisna natanko tedaj, ko je $\mathbf{C} = -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$.

6. Pričakovana vrednost in sorodne karakteristike

1. $\mathbb{E}(X) = (-1)\cdot 0\dot{}1 + 0\cdot 0\dot{}3 + 2\cdot 0\dot{}1 + 3\cdot 0\dot{}1 + 4\cdot 0\dot{}1 = 2$,
$\mathbb{E}(X^2) = (-1)^2\cdot 0\dot{}1 + 0^2\cdot 0\dot{}3 + 2^2\cdot 0\dot{}1 + 3^2\cdot 0\dot{}1 + 4^2\cdot 0\dot{}1 = 7\dot{}8$.

2. Število piskov X je porazdeljeno diskretno po shemi:

$$\begin{pmatrix} 1 & 2 & 4 & 8 & \cdots \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \cdots \end{pmatrix}.$$

Pričakovana vrednost:

$$\mathbb{E}(X) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

ne obstaja. Pač pa obstaja pričakovana vrednost:

$$\mathbb{E}\big(\sqrt{X}\,\big) = \frac{1}{2} + \frac{1}{2\sqrt{2}} + \frac{1}{4} + \frac{1}{4\sqrt{2}} + \cdots = \frac{1}{2-\sqrt{2}} = 1 + \frac{\sqrt{2}}{2} \doteq 1\dot{}707\,.$$

Primerljivo z X pa bi bilo število:

$$\Big(\mathbb{E}\big(\sqrt{X}\,\big)\Big)^2 = \frac{3}{2} + \sqrt{2} \doteq 2\dot{}914\,.$$

3. Lahko sicer izračunamo neposredno, a če računamo na roko, je lažje, če za slučajno spremenljivko:

$$Y := \frac{X - 5000}{10} \sim \begin{pmatrix} -10 & 0 & 20 \\ 0\dot{}1 & 0\dot{}3 & 0\dot{}6 \end{pmatrix}.$$

izračunamo $\mathbb{E}(Y) = 11$. Sledi:

$$\mathbb{E}(X) = 10\,\mathbb{E}(Y) + 5000 = 5110\,.$$

4. $3\dot{}55$.

5. Označimo število metov z X. Ta slučajna spremenljivka lahko zavzame vrednosti $m, m+1, m+2, \ldots 2m-1$. Če je k dopustna vrednost, se lahko dogodek $\{X = k\}$ zgodi na dva nezdružljiva načina: ali v k-tem metu prvič dobimo natanko m grbov ali pa dobimo prvič natanko m cifer. Sledi:

$$\mathbb{P}(X = k) = 2\binom{k-1}{m-1}\left(\frac{1}{2}\right)^m\left(\frac{1}{2}\right)^{k-m} = 2^{-(k-1)}\binom{k-1}{m-1}.$$

Po definiciji izračunamo:

$$\mathbb{E}(X) = \sum_{k=m}^{2m-1} n\,2^{-(k-1)}\binom{k-1}{m-1} = m\sum_{k=m}^{2m-1} 2^{-(k-1)}\binom{k}{m} = m\sum_{k=m+1}^{2m} 2^{-(k-2)}\binom{k-1}{m}.$$

Zdaj pa v identiteti $\sum_{k=m}^{2m-1}\mathbb{P}(X = k) = 1$ parameter m zamenjamo z $m+1$. Dobimo:

$$\sum_{k=m+1}^{2m+1} 2^{-(k-1)}\binom{k-1}{m} = 1\,.$$

Sledi:

$$\mathbb{E}(X) = 2m\left[1 - 2^{-2m}\binom{2m}{m}\right].$$

6. $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $\quad \mathbb{E}(e^{-tX}) = \dfrac{\lambda}{\lambda + t}$ obstaja za $t > -\lambda$.

7. Pričakovana vrednost bi bila enaka:

$$\int_0^\infty \frac{x}{(1+x)^2}\,\mathrm{d}x = \lim_{b\to\infty}\left(\ln(1+b) - \frac{b}{1+b}\right)\Big|_0^\infty,$$

kar ne obstaja.

8. $\mathrm{var}(X) = 3\dot{}8$, $\quad \sigma(X) \doteq 1\dot{}949$.

9. Spet definiramo $Y = (X - 5000)/10$ in izračunamo $\mathrm{var}(Y) = 129$. Sledi $\mathrm{var}(X) = 12900$ in $\sigma(X) \doteq 113\dot{}6$.

10. Naj bo $S \sim \mathrm{Pois}(\lambda)$, tj.:

$$\mathbb{P}(S = k) = \frac{\lambda^k}{k!}\,e^{-\lambda}\,; \qquad k = 0, 1, 2, \ldots$$

Velja:

$$\mathbb{E}(S) = \sum_{k=0}^\infty k\,\frac{\lambda^k}{k!}\,e^{-\lambda} = e^{-\lambda}\sum_{k=1}^\infty \frac{\lambda^k}{(k-1)!} = \lambda\,e^{-\lambda}\sum_{j=0}^\infty \frac{\lambda^j}{j!} = \lambda\,.$$

Nadalje je:

$$\mathbb{E}\big[S(S-1)\big] = \sum_{k=0}^\infty k(k-1)\,\frac{\lambda^k}{k!}\,e^{-\lambda} = e^{-\lambda}\sum_{k=2}^\infty \frac{\lambda^k}{(k-2)!} = \lambda^2\,e^{-\lambda}\sum_{j=0}^\infty \frac{\lambda^j}{j!} = \lambda^2\,.$$

Sledi $\mathbb{E}(S^2) = \lambda^2 + \lambda$ in $\mathrm{var}(S) = \lambda$.

11. Naj bo $S \sim \mathrm{Gama}(a, \lambda)$. Velja:

$$\mathbb{E}(S^p) = \frac{\lambda^a}{\Gamma(a)}\int_0^\infty s^p\,s^{a-1}\,e^{-\lambda s}\,\mathrm{d}s = \frac{\lambda^a}{\Gamma(a)}\int_0^\infty s^{a+p-1}\,e^{-\lambda s}\,\mathrm{d}s\,.$$

S substitucijo $t = \lambda s$ dobimo:

$$\mathbb{E}(S^p) = \frac{1}{\lambda^p\,\Gamma(a)}\int_0^\infty t^{a+p-1}\,e^{-\lambda t}\,\mathrm{d}t = \frac{1}{\lambda^p}\,\frac{\Gamma(a+p)}{\Gamma(a)}\,.$$

Če je p naravno število, torej velja:

$$\mathbb{E}(S^p) = \frac{a(a+1)(a+2)\cdots(a+p-1)}{\lambda^p}\,.$$

Torej je $\mathbb{E}(S^2) = (a + a^2)/\lambda^2$, od koder sledi $\mathrm{var}(S) = a/\lambda^2$.

12. Če je n lih, je $\mathbb{E}(Z^n) = 0$, saj ima slučajna spremenljivka Z enako porazdelitev kot $-Z$ in posledično Z^n enako porazdelitev kot $-Z^n$.

Če pa je n sod, nastavimo:

$$\mathbb{E}(Z^n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^n \, e^{-z^2/2} \, \mathrm{d}z \, .$$

Z integracijo per partes ($u = z^{n-1}$, $\mathrm{d}u = (n-1) z^{n-2} \, \mathrm{d}z$, $\mathrm{d}v = z \, e^{-z^2/2} \, \mathrm{d}z$, $v = -e^{-z^2/2}$) dobimo:

$$\mathbb{E}(Z^n) = -\frac{1}{\sqrt{2\pi}} \, z^{n-1} \, e^{-z^2/2} \Big|_{-\infty}^{\infty} + (n-1) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{n-2} \, e^{-z^2/2} \, \mathrm{d}z = (n-1) \, \mathbb{E}(Z^{n-2}) \, .$$

Ker je $\mathbb{E}(Z^0) = 1$, od tod sledi:

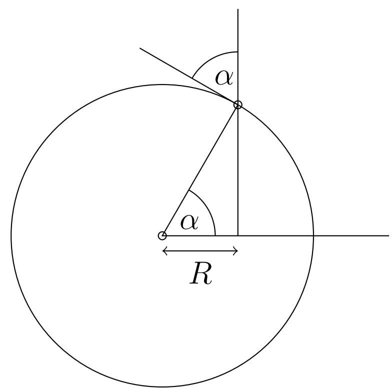
$$\mathbb{E}(Z^n) = 1 \cdot 3 \cdot \cdots \cdot (n-1) = (n-1)!! \, .$$

13. $\mathbb{E}(1 + 4X^2) = 2, \quad \mathrm{var}(X) = 1/4.$

14. Velja:

$$\mathbb{E}\big(e^{Y-X}\big) = \iint_{x > 2y > 0} 2 \, e^{y - 2x} \, \mathrm{d}x \, \mathrm{d}y = 2 \int_0^{\infty} e^y \int_{2y}^{\infty} e^{-2x} \, \mathrm{d}x \, \mathrm{d}y = \int_0^{\infty} e^{-3y} \, \mathrm{d}y = \frac{1}{3} \, .$$

15. Zaradi enakomerne porazdelitve smeri lahko vesoljsko smer, iz katere pade meteorit, fiksiramo. Ravne trajektorije, ki sekajo planet, ustrezajo točkam na krogu. Za enoto lahko brez škode za splošnost vzamemo kar polmer planeta. Če trajektorija ustreza točki (X, Y) iz enotskega kroga, meteorit pade pod kotom $\arccos R$, kjer je $R = \sqrt{X^2 + Y^2}$. Slika:



Tako je pričakovani kot enak:

$$\bar{\alpha} := \mathbb{E}\Big[\arccos\sqrt{X^2+Y^2}\Big]\,,$$

kjer je točka (X, Y) porazdeljena enakomerno po enotskem krogu. Sledi:

$$\bar{\alpha} = \frac{1}{\pi}\iint_{x^2+y^2\le 1} \arccos\sqrt{x^2+y^2}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Prevedba na polarne koordinate nam da:

$$\bar{\alpha} = \frac{1}{\pi}\int_0^{2\pi}\int_0^1 r\arccos r\,\mathrm{d}r\,\mathrm{d}\varphi = 2\int_0^1 r\arccos r\,\mathrm{d}r\,\mathrm{d}\varphi\,.$$

S substitucijo $r = \cos t$ dobimo:

$$\bar{\alpha} = 2\int_0^{\pi/2} t\cos t\sin t\,\mathrm{d}t = \int_0^{\pi/2} t\sin(2t)\,\mathrm{d}t = -\left.\frac{t\cos(2t)}{2}\right|_0^{\pi/2} + \frac{1}{2}\int_0^{\pi/2}\cos(2t)\,\mathrm{d}t = \\ = \frac{\pi}{4}\,.$$

Opomba. Porazdelitev smeri, iz katere pade meteorit, je enakomerna *le za cel planet*, ni pa enakomerna za posamezno točko, na katero pade meteorit. Če si namreč zamislimo neko majhno površino okrog izbrane točke, vsaki smeri ustreza prizma žarkov, ki ponazarjajo trajektorije meteoritov, ki padejo iz izbrane smeri na to površino. Toda smeri, ki je pravokotna na površino, ustreza prizma z večjim presekom kot pa smeri, ki je na površino poševna. Zato bi, če bi gledali porazdelitev smeri za posamezno točko, bolj poševne smeri (tj. tiste, ki ustrezajo manjšim kotom) imele manjšo gostoto verjetnosti.

Opomba. Pričakovana vrednost, ki smo jo dobili, se ujema s pričakovano vrednostjo kota, ki bi bil porazdeljen enakomerno na intervalu od 0 do $\pi/2$. Vendar pa kot, pod katerim pade meteorit, ni porazdeljen enakomerno, saj za $0 \le t \le \pi/2$ velja:

$$F_\alpha(t) = \mathbb{P}(\alpha \le t) = \mathbb{P}\big(\arccos\sqrt{X^2+Y^2} \le t\big) = \mathbb{P}\big(\sqrt{X^2+Y^2} \ge \cos t\big) = 1 - \cos^2 t\,,$$

od koder sledi $f_\alpha(t) = 2\sin t\cos t$.

16. $c = \dfrac{2}{\pi}$, $\quad\mathbb{E}(X) = \mathbb{E}(Y) = 0$, $\quad\mathbb{E}(X^2) = \mathbb{E}(Y^2) = \dfrac{1}{2}$, $\quad\mathbb{E}(X^2+Y^2) = 1$.

17. Zaradi simetrije so slučajne spremenljivke:

$$\frac{X^2}{X^2+Y^2+Z^2}\,,\quad \frac{Y^2}{X^2+Y^2+Z^2}\quad\text{in}\quad \frac{Z^2}{X^2+Y^2+Z^2} = S^2$$

enako porazdeljene. Ker je njihova vsota enaka 1, mora biti $\mathbb{E}(S^2) = 1/3$.

18. Velja $X = Y_1 + Y_2 + Y_3 + Y_4 + Y_5$, kjer je:

$$\begin{aligned} A_1 &= \{\text{prva ploščica z oznako } \boxed{1} \text{ je bila odkrita}\}\,, & Y_1 &:= \mathbf{1}\,A_1\,,\\ A_2 &= \{\text{druga ploščica z oznako } \boxed{1} \text{ je bila odkrita}\}\,, & Y_2 &:= \mathbf{1}\,A_2\,,\\ A_3 &= \{\text{tretja ploščica z oznako } \boxed{1} \text{ je bila odkrita}\}\,, & Y_3 &:= \mathbf{1}\,A_3\,,\\ A_4 &= \{\text{ploščica z oznako } \boxed{2} \text{ je bila odkrita}\}\,, & Y_4 &:= 2\cdot\mathbf{1}\,A_4\,,\\ A_5 &= \{\text{ploščica } \boxed{3} \text{ je bila odkrita}\}\,. & Y_5 &:= 3\cdot\mathbf{1}\,A_5\,, \end{aligned}$$

Za $i = 1, 2, 3$ je $\mathbb{E}(Y_i) = \mathbb{E}\big[\mathbf{1}_{A_i}\big] = \mathbb{P}(A_i) = 1/2$; slednja verjetnost velja zaradi simetrije. Nadalje je $\mathbb{E}(Y_4) = 2\,\mathbb{P}(A_4) = 1$ in $\mathbb{E}(Y_5) = 3\,\mathbb{P}(A_5) = 3/2$. Sledi $\mathbb{E}(X) = \sum_{i=1}^{5} \mathbb{E}(Y_i) = 4$.

Za izračun variance pišimo:

$$\mathbb{E}(X^2) = \sum_{i=1}^{5} \sum_{j=1}^{4} \mathbb{E}(Y_i Y_j)\,.$$

Najprej opazimo, da je $\mathbf{1}_{A_i}^2 = \mathbf{1}\,A_i$, torej $\mathbb{E}\big[\mathbf{1}_{A_i}^2\big] = 1/2$. Nadalje je $\mathbb{E}\big[\mathbf{1}_{A_i}\,\mathbf{1}_{A_j}\big] = \mathbb{P}(A_i \cap A_j)$.

Če sta i in j različna indeksa, si lahko predstavljamo, da permutacijo ploščic generiramo tako, da najprej razporedimo i-to, j-to in tisto z oznako S, nakar med te tri drugo za drugo vrinemo še preostale ploščice. Ne glede na to, kako razporedimo prve tri, imamo za preostale enako mnogo možnosti vrivanj, zato lahko vrinjene ploščice odmislimo. Presek $A_i \cap A_j$ je dogodek, da je ploščica z oznako S med aktualnimi tremi na zadnjem mestu, torej je $\mathbb{E}\big[\mathbf{1}_{A_i}\,\mathbf{1}_{A_j}\big] = \mathbb{P}(A_i \cap A_j) = 1/3$.

Ko pomnožimo z vrednostmi oznak na ploščicah, dobimo naslednjo tabelo pričakovanih vrednosti produktov:

$\mathbb{E}(Y_iY_j)$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
$i=1$	$1/2$	$1/3$	$1/3$	$2/3$	1
$i=2$	$1/3$	$1/2$	$1/3$	$2/3$	1
$i=3$	$1/3$	$1/3$	$1/2$	$2/3$	1
$i=4$	$2/3$	$2/3$	$2/3$	2	2
$i=5$	1	1	1	2	$9/2$

Seštejemo in dobimo $\mathbb{E}(X^2) = 24$. Sledi $\mathrm{var}(X) = 8$.

19. Iskano število negibnih točk lahko zapišemo kot vsoto:

$$S = X_1 + X_2 + \cdots + X_n\,,$$

kjer je X_i indikator dogodka, da se i-ti element preslika sam vase. Ker je verjetnost tega dogodka enaka $\frac{1}{n}$, je:

$$\mathbb{E}(S) = \sum_{i=1}^{n} \mathbb{E}(X_i) = n \cdot \frac{1}{n} = 1\,.$$

Nadalje je:

$$\begin{aligned}\mathbb{E}(S^2) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}(X_iX_j) = \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{P}(i\text{-ti in } j\text{-ti element se oba preslikata sama vase})\,.\end{aligned}$$

Za različna indeksa i in j je verjetnost dogodka, da se i-ti in j-ti element oba preslikata sama vase, enaka $\frac{1}{n(n-1)}$. Sledi:

$$\mathbb{E}(S^2) = n\cdot\frac{1}{n} + n(n-1)\cdot\frac{1}{n(n-1)} = 2\,,$$

torej je $\mathrm{var}(S) = 1$.

20. Pišimo:

$$Y = \sum_{i=1}^{r}\mathbf{1}_{A_i}\,,$$

kjer je $A_i = \{\text{med števili } X_1,\ldots,X_n \text{ se pojavi } i\}$. Velja:

$$\mathbb{E}\big[\mathbf{1}_{A_i}\big] = \mathbb{P}(A_i) = 1 - \mathbb{P}(A_i^{\mathrm{c}}) = 1 - \left(\frac{r-1}{r}\right)^n.$$

Sledi:

$$\mathbb{E}(Y) = r\left[1 - \left(\frac{r-1}{r}\right)^n\right].$$

Izračunajmo še varianco. Za $r = 1$ je seveda $\mathrm{var}(Y) = 0$. Naj bo zdaj $r > 1$. Za $i \neq j$ izračunajmo:

$$\begin{aligned}\mathbb{E}\big[\mathbf{1}_{A_i}\mathbf{1}_{A_j}\big] &= \mathbb{P}(A_i\cap A_j) = \\ &= 1 - \mathbb{P}(A_i^{\mathrm{c}}\cup A_j^{\mathrm{c}}) = \\ &= 1 - \mathbb{P}(A_i^{\mathrm{c}}) - \mathbb{P}(A_j^{\mathrm{c}}) + \mathbb{P}(A_i^{\mathrm{c}}\cap A_j^{\mathrm{c}}) = \\ &= 1 - 2\left(\frac{r-1}{r}\right)^n + \left(\frac{r-2}{r}\right)^n.\end{aligned}$$

Sledi:

$$\begin{aligned}\mathbb{E}(Y^2) &= r\left[1 - \left(\frac{r-1}{r}\right)^n\right] + r(r-1)\left[1 - 2\left(\frac{r-1}{r}\right)^n + \left(\frac{r-2}{r}\right)^n\right] = \\ &= r^2 - r(2r-1)\left(\frac{r-1}{r}\right)^n + r(r-1)\left(\frac{r-2}{r}\right)^n,\end{aligned}$$

torej:

$$\begin{aligned}\operatorname{var}(Y) = r^2 - r(2r-1)\left(\frac{r-1}{r}\right)^n + r(r-1)\left(\frac{r-2}{r}\right)^n - \\ - r^2 + 2r^2\left(\frac{r-1}{r}\right)^n - r^2\left(\frac{r-1}{r}\right)^{2n} = \\ = r\left(\frac{r-1}{r}\right)^n + r(r-1)\left(\frac{r-2}{r}\right)^n - r^2\left(\frac{r-1}{r}\right)^{2n}.\end{aligned}$$

21. *Prvi način.* Naj bo $N \sim \operatorname{Hip}(s; r, n)$. Za $s \geq 1$ velja:

$$\mathbb{E}(N) = \sum_{k=1}^{s} k\,\frac{\binom{s}{k}\binom{n-s}{r-k}}{\binom{n}{r}} = \frac{rs}{n}\sum_{k=1}^{s}\frac{\binom{s-1}{k-1}\binom{n-s}{r-k}}{\binom{n-1}{r-1}}.$$

Zdaj pa opazimo, da je vsota na desni ravno vsota verjetnosti za hipergeometrijsko porazdelitev $\operatorname{Hip}(s-1; r, n-1)$, torej je enaka 1. Sledi:

$$\mathbb{E}(N) = \frac{rs}{n}.$$

Rezultat je pravilen tudi za $s = 0$, saj je v tem primeru slučajna spremenljivka N identično enaka 0. Nadalje za $s \geq 2$ velja:

$$\mathbb{E}\big[N(N-1)\big] = \sum_{k=2}^{s} k(k-1)\,\frac{\binom{s}{k}\binom{n-s}{r-k}}{\binom{n}{r}} = \frac{r(r-1)s(s-1)}{n(n-1)}\sum_{k=2}^{s}\frac{\binom{s-2}{k-2}\binom{n-s}{r-k}}{\binom{n-2}{r-2}}.$$

Spet opazimo, da je vsota na desni ravno vsota verjetnosti za hipergeometrijsko porazdelitev $\operatorname{Hip}(s-2; r, n-2)$, torej je enaka 1. Sledi:

$$\mathbb{E}\big[N(N-1)\big] = \frac{r(r-1)s(s-1)}{n(n-1)}.$$

Rezultat je pravilen tudi za $s = 0$ in $s = 1$, saj je v tem primeru slučajna spremenljivka N lahko enaka kvečjemu 0 ali 1, torej je $N(N-1) = 0$. Sledi:

$$\operatorname{var}(N) = \mathbb{E}\big[N(N-1)\big] + \mathbb{E}(N) - \big(\mathbb{E}(N)\big)^2 = \frac{rs(n-r)(n-s)}{n^2(n-1)}.$$

Drugi način. Spomnimo se, da ima hipergeometrijsko porazdelitev $\operatorname{Hip}(s; r, n)$ slučajna spremenljivka N, ki šteje število rdečih kroglic med s kroglicami, ki jih na slepo in brez vračanja izvlečemo iz posode, v kateri je n kroglic, od tega r rdečih. Pišemo lahko $N = X_1 + X_2 + \cdots + X_s$, kjer je X_i indikator dogodka, da je i-ta izvlečena kroglica rdeča. Torej je:

$$\mathbb{E}(N) = \sum_{i=1}^{s}\mathbb{E}(X_i) = \sum_{i=1}^{s}\mathbb{P}(i\text{-ta izvlečena kroglica je rdeča}).$$

Verjetnost, da je posamezna izvlečena kroglica rdeča, je enaka r/n, zato je:

$$\mathbb{E}(N) = \frac{rs}{n}\,.$$

Nadalje je:

$$\mathbb{E}(N^2) = \sum_{i=1}^{s}\sum_{j=1}^{s}\mathbb{E}(X_iX_j) = \sum_{i=1}^{s}\sum_{j=1}^{s}\mathbb{P}(i\text{-ta in } j\text{-ta kroglica sta obe rdeči})\,.$$

Za $i \neq j$ je verjetnost dogodka, da sta i-ta in j-ta kroglica obe rdeči, enaka $\frac{r(r-1)}{n(n-1)}$, zato je:

$$\mathbb{E}(N^2) = s\cdot\frac{r}{n} + s(s-1)\cdot\frac{r(r-1)}{n(n-1)} = \frac{rs(rs-r-s+n)}{n(n-1)}$$

in končno:

$$\operatorname{var}(N) = \mathbb{E}(N^2) - \big(\mathbb{E}(N)\big)^2 = \frac{rs(n-r)(n-s)}{n^2(n-1)}\,.$$

Opomba. V 4. razdelku smo videli, da, če pošljemo n proti neskončno in r/n proti nekemu fiksnemu številu p, točkaste verjetnosti pri hipergeometrijski porazdelitvi $\operatorname{Hip}(s,r,n)$ konvergirajo proti točkastim verjetnostim pri binomski porazdelitvi $\operatorname{Bin}(s,p)$. To velja tudi za pričakovano vrednost in varianco: brž ko je namreč $r_1, r_2, \ldots$ zaporedje z $\lim_{n\to\infty} r_n/n = p$, velja:

$$\lim_{n\to\infty}\frac{r_n s}{n} = sp \qquad \text{in} \qquad \lim_{n\to\infty}\frac{r_n s(n-r_n)(n-s)}{n^2(n-1)} = sp(1-p)\,.$$

22. Za $i = 1,\ldots,8$ definirajmo slučajne spremenljivke X_i, kjer naj bo X_i enaka 1, če i-ti igralec stavi zlatnik, sicer pa 0. Tedaj je očitno:

$$\mathbb{E}(X_i) = p_0 := \mathbb{P}(i\text{-ti igralec stavi zlatnik}) = \frac{4}{52}\cdot\frac{48}{51}\cdot\frac{47}{50} \doteq 0\text{˙}06805\,.$$

Ker je $S = X_1 + \cdots + X_8$, je tudi $\mathbb{E}(S) = 8p_0 \doteq 0\text{˙}5444$.

Izračunajmo še varianco: $\operatorname{var}(S) = \mathbb{E}(S^2) - \big(\mathbb{E}(S)\big)^2$. Velja:

$$\begin{aligned}
\mathbb{E}(S^2) &= \sum_{i=1}^{8}\sum_{j=1}^{8}\mathbb{E}(X_iX_j) = \\
&= \sum_{i=1}^{8}\sum_{j=1}^{8}\mathbb{P}\big(i\text{-ti in } j\text{-ti igralec oba stavita zlatnik}\big) = \\
&= 8(p_0 + 2p_1 + 2p_2 + 2p_3 + p_4)
\end{aligned}$$

kjer je p_k verjetnost, da stavita zlatnik posamezna igralca, ki sta oddaljena za k (tj. med njima je $k-1$ sedežev). Verjetnost p_0 smo že izračunali, velja pa še:

$$\begin{aligned}
p_1 &= 0 \\
p_2 &= \frac{4}{52}\cdot\frac{3}{51}\cdot\frac{48}{50}\cdot\frac{47}{49}\cdot\frac{46}{48} \doteq 0\text{˙}00399\,, \\
p_3 = p_4 &= \frac{4}{52}\cdot\frac{3}{51}\cdot\frac{48}{50}\cdot\frac{47}{49}\cdot\frac{46}{48}\cdot\frac{45}{47} \doteq 0\text{˙}00382\,.
\end{aligned}$$

Torej je:

$$\mathrm{var}(S) = 8(p_0 + 2p_2 + 3p_3) - (8p_0)^2 \doteq 0{\cdot}4037\,.$$

23. Problem lahko ekvivalentno zastavimo tako, da v vrsto razporedimo r rdečih in $n-r$ zelenih kroglic; vse možne razporeditve so enako verjetne. Želena količina, ki nas zanima, je število zelenih kroglic pred prvo rdečo, povečano za 1. To označimo z X.

Za kroglice lahko privzamemo, da so označene. Za $i = 1, 2, \ldots, n-r$ naj bo X_i indikator dogodka, da je i-ta zelena kroglica pred vsemi rdečimi. Tedaj je $X = 1 + \sum_{i=1}^{n-r} X_i$.

Permutacije kroglic lahko generiramo tako, da najprej razporedimo vse rdeče kroglice, nato pa mednje drugo za drugo vrivamo zelene kroglice. Če izberemo eno izmed zelenih kroglic, lahko privzamemo, da njo vrinemo prvo. Za to imamo $r+1$ možnosti. Ne glede na to, kako jo vrinemo, imamo nato za preostale še enako mnogo, tj. $(r+2)(r+3)\cdots n$ možnosti. Če so torej vse permutacije enako verjetne, je izbrana zelena kroglica na posameznem položaju med rdečimi z verjetnostjo $1/(r+1)$. Z drugimi besedami, za vse i velja $\mathbb{E}(X_i) = 1/(r+1)$. Sledi:

$$\mathbb{E}(X) = 1 + \sum_{i=1}^{n-r} \mathbb{E}(X_i) = 1 + \frac{n-r}{r+1} = \frac{n+1}{r+1}\,.$$

Prvi korak k izračunu variance je izračun $\mathbb{E}(X_i X_j)$, kjer je $i \neq j$. Produkt $X_i X_j$ je indikator dogodka, da sta i-ta in j-ta zelena kroglica obe pred vsemi rdečimi. Spet si lahko predstavljamo, da permutacijo kroglic genetriramo tako, da najprej razporedimo vse rdeče kroglice, nato pa mednje drugo za drugo vrivamo zelene kroglice, pri čemer najprej vrinemo i-to, nato j-to in šele nato ostale zelene kroglice. Spet imamo ne glede na to, kako vrinemo izbrani zeleni kroglici, za preostale še enako mnogo, tj. $(r+3)(r+4)\cdots n$ možnosti. Če so torej vse permutacije enako verjetne, sta izbrani zeleni kroglici pred vsemi rdečimi z verjetnostjo $\frac{2}{(r+1)(r+2)}$. To je torej $\mathbb{E}(X_i X_j)$. Sledi:

$$\begin{aligned}
\mathbb{E}\big[(X-1)^2\big] &= \sum_{i=1}^{n-r} \mathbb{E}(X_i) + \sum_{\substack{1 \le i,j \le n-r \\ i \neq j}} \mathbb{E}(X_i X_j) = \\
&= \frac{n-r}{r+1} + \frac{2(n-r)(n-r-1)}{(r+1)(r+2)} = \\
&= \frac{(n-r)(2n-r)}{(r+1)(r+2)}
\end{aligned}$$

in končno:

$$\begin{aligned}
\mathrm{var}(X) &= \mathrm{var}(X-1) = \\
&= \mathbb{E}\big[(X-1)^2\big] - \big[\mathbb{E}(X-1)\big]^2 = \\
&= \frac{(n-r)(2n-r)}{(r+1)(r+2)} - \frac{(n-r)^2}{(r+1)^2} = \\
&= \frac{r(n+1)(n-r)}{(r+1)^2(r+2)}\,.
\end{aligned}$$

24. Označimo cene v posameznih trgovinah z X_1, X_2 in X_3, poleg tega pa naj bo še Z_i indikator dogodka, da iskani čevlji v i-ti trgovini stanejo največ 50 evrov. Tedaj je:

$$\begin{aligned} C &= C\,\mathbf{1}(X_1 \le 50) + C\,\mathbf{1}(X_1 > 50, X_2 < 50) + C\,\mathbf{1}(X_1 > 50, X_2 > 50) = \\ &= X_1\,\mathbf{1}(X_1 \le 50) + X_2\,\mathbf{1}(X_1 > 50)\,\mathbf{1}(X_2 \le 50) + X_3\,\mathbf{1}(X_1 > 50)\,\mathbf{1}(X_2 > 50)\,. \end{aligned}$$

Sledi:

$$\begin{aligned} \mathbb{E}(C) &= \mathbb{E}\big[X_1\,\mathbf{1}(X_1 \le 50)\big] + \mathbb{E}\big[\mathbf{1}(X_1 > 50)\big]\,\mathbb{E}\big[X_2\,\mathbf{1}(X_2 \le 50)\big] + \\ &\qquad + \mathbb{E}\big[\mathbf{1}(X_1 > 50)\big]\,\mathbb{E}\big[\mathbf{1}(X_2 > 50)\big]\,\mathbb{E}(X_3) = \\ &= \mathbb{E}\big[X_1\,\mathbf{1}(X_1 \le 50)\big] + \mathbb{P}(X_1 > 50)\,\mathbb{E}\big[X_2\,\mathbf{1}(X_2 \le 50)\big] + \\ &\qquad + \mathbb{P}(X_1 > 50)\,\mathbb{P}(X_2 > 50)\,\mathbb{E}(X_3) = \\ &= \frac{36+45}{3} + \frac{1}{3}\int_{40}^{50} x\,\frac{60-x}{200}\,\mathrm{d}x + \frac{1}{3}\int_{50}^{60}\frac{60-x}{200}\,\mathrm{d}x \cdot 54 = \\ &= 42\tfrac{11}{18} \doteq 42{\cdot}61\,. \end{aligned}$$

25. Velja $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$ in $\mathbf{1}_{A\cap B} = \mathbf{1}_A\,\mathbf{1}_B$. Načelo vključitev in izključitev predstavlja formulo za verjetnost unije dogodkov, toda glede na pravkar izpeljani zvezi nam pri uniji bolj prav pride zapis:

$$A_1 \cup A_2 \cup \cdots \cup A_n = (A_1^c \cap A_2^c \cap \cdots \cap A_n^c)^c\,.$$

Sledi:

$$\begin{aligned} \mathbf{1}_{A_1\cup A_2\cup\cdots\cup A_n} &= 1 - (1-\mathbf{1}_{A_1})(1-\mathbf{1}_{A_2})\cdots(1-\mathbf{1}_{A_n}) = \\ &= 1 - 1 + \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n} - \\ &\qquad - \mathbf{1}_{A_1\cap A_2} - \mathbf{1}_{A_1\cap A_3} - \cdots - \mathbf{1}_{A_{n-1}\cap A_n} + \\ &\qquad + \mathbf{1}_{A_1\cap A_2\cap A_3} + \mathbf{1}_{A_1\cap A_2\cap A_4} + \cdots + \mathbf{1}_{A_{n-2}\cap A_{n-1}\cap A_n} - \\ &\qquad\qquad \cdots \\ &\qquad + (-1)^{n+1}\,\mathbf{1}_{A_1\cap A_2\cap\cdots\cap A_n} \end{aligned}$$

Načelo vključitev in izključitev od tod takoj dobimo z uporabo pričakovane vrednosti in upoštevanjem njene linearnosti.

26. Fiksirajmo N in opazimo, da je:

$$\mathbf{1}(N \ge n) = \begin{cases} 1 & ;\ n = 1, 2, \ldots, N \\ 0 & ;\ \text{sicer}, \end{cases}$$

$$\mathbf{1}(N > n) = \begin{cases} 1 & ;\ n = 0, 1, \ldots, N-1 \\ 0 & ;\ \text{sicer}, \end{cases}$$

Torej je:

$$N = \sum_{n=1}^{\infty} \mathbf{1}(N \ge n) = \sum_{n=0}^{\infty} \mathbf{1}(N > n)\,,$$

kar vstavimo v pričakovano vrednost, upoštevamo aditivnost in dobimo.

27. Označimo število vlečenj z N. Dogodek $\{N > n\}$ pomeni, da smo v prvih n vlečenjih izvlekli same bele kroglice. Verjetnost tega dogodka je enaka:

$$\mathbb{P}(N > n) = \frac{b}{b+2} \cdot \frac{b+1}{b+3} \cdots \frac{b+n-1}{b+n+1} = \frac{b(b+1)}{(b+n)(b+n+1)}$$

(glej tudi 30. nalogo iz 3. razdelka). V skladu s prejšnjo nalogo je torej:

$$\begin{aligned}\mathbb{E}(N) &= \sum_{n=0}^{\infty} P(N > n) = \\ &= b(b+1) \sum_{n=0}^{\infty} \left(\frac{1}{b+n} - \frac{1}{b+n+1} \right) = \\ &= b+1\,.\end{aligned}$$

Opomba. Iskano pričakovano vrednost bi lahko izračunali tudi neposredno: $\{X = n\}$ je namreč dogodek, da smo v prvih $n-1$ vlečenjih izvlekli same bele kroglice, v n-tem pa rdečo. Sledi:

$$\mathbb{P}(X = n) = \frac{2b(b+1)}{(b+n-1)(b+n)(b+n+1)}$$

in

$$\mathbb{E}(X) = 2b(b+1) \sum_{n=1}^{\infty} \frac{n}{(b+n-1)(b+n)(b+n+1)}\,.$$

Tudi slednja vsota se da izračunati z razbitjem na parcialne ulomke, le da je le-to nekoliko zapletenejše.

28. *Prvi način.* Naj bo $X \sim \mathrm{Geom}(p)$, tj.:

$$\mathbb{P}(X = k) = p(1-p)^{k-1}]\,; \quad k = 1, 2, 3, \ldots$$

Torej je:

$$\mathbb{E}(X) = p \sum_{k=1}^{\infty} k q^{k-1}\,,$$

kjer je $q = 1-p$. Zdaj pa to zvezo odmislimo in si predstavljajmo, da je p konstanta, q pa spremenljivka. Opazimo, da so členi v dobljeni vrsti odvodi preprostejših izrazov – velja:

$$\sum_{k=1}^{\infty} k q^{k-1} = \sum_{k=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}q}(q^k) = \frac{\mathrm{d}}{\mathrm{d}q} \sum_{k=1}^{\infty} q^k = \frac{\mathrm{d}}{\mathrm{d}q} \frac{q}{1-q} = \frac{1}{(1-q)^2}\,.$$

To velja za vse $q \in (-1, 1)$. Zdaj pa spet vstavimo $q = 1-p$ in dobimo:

$$\mathbb{E}(X) = \frac{p}{(1-q)^2} = \frac{1}{p}\,.$$

Podobno nastavimo:

$$\mathbb{E}\big[X(X-1)\big] = p\sum_{k=1}^{\infty} k(k-1)q^{k-1} = pq\sum_{k=1}^{\infty} k(k-1)q^{k-2}$$

in vsoto spet gledamo kot funkcijo spremenljivke q, ki je drugi odvod neke funkcije. Dobimo:

$$\sum_{k=1}^{\infty} k(k-1)q^{k-2} = \sum_{k=1}^{\infty} \frac{\mathrm{d}^2}{\mathrm{d}q^2}(q^k) = \frac{\mathrm{d}^2}{\mathrm{d}q^2}\sum_{k=1}^{\infty} q^k = \frac{\mathrm{d}^2}{\mathrm{d}q^2}\frac{q}{1-q} = \frac{2}{(1-q)^3}\,.$$

Torej je:

$$\mathbb{E}\big[X(X-1)\big] = \frac{2pq}{(1-q)^3} = \frac{2(1-p)}{p^2}$$

in končno:

$$\operatorname{var}(X) = \mathbb{E}\big[X(X-1)\big] + \mathbb{E}(X) - \big(\mathbb{E}(X)\big)^2 = \frac{1-p}{p^2}\,.$$

Drugi način. Uporabimo rezultat iz prejšnje naloge, pri čemer se še spomnimo, da lahko geometrijsko porazdelitev $\operatorname{Geom}(p)$ dobimo kot porazdelitev števila poskusov do vključno prvega uspelega v Bernoullijevem zaporedju poskusov, kjer vsak poskus uspe z verjetnostjo p. Če je torej $X \sim \operatorname{Geom}(p)$, za $k = 0, 1, 2, \ldots$ velja $\mathbb{P}(X > k) = (1-p)^k$. Sledi:

$$\mathbb{E}(X) = \sum_{k=0}^{\infty}(1-p)^k = \frac{1}{p}\,.$$

Nadalje z uporabo formule $X = \sum_{k=0}^{\infty}\{X > k\}$ dobimo:

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\mathbb{P}(X > i, X > j) = \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(1-p)^{\max\{i,j\}} = \\
&= \sum_{k=0}^{\infty}(1-p)^k + 2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}(1-p)^j = \\
&= \frac{1}{p} + \frac{2}{p}\sum_{i=0}^{\infty}(1-p)^{i+1} = \\
&= \frac{2-p}{p^2}\,.
\end{aligned}$$

Sledi:

$$\operatorname{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{1-p}{p^2}\,.$$

Opomba. Pri obeh načinih smo v vmesnem koraku izračunali:

$$\mathbb{E}\big[X(X-1)\big] = pq\,\frac{\mathrm{d}^2}{\mathrm{d}q^2}\sum_{k=1}^{\infty} q^k = 2\sum_{i=0}^{\infty}\sum_{j=i+1}^{\infty}\mathbb{P}(X > i, X > j)\,.$$

Tej količini pravimo *drugi faktorski moment*. Faktorski momenti $\mathbb{E}\big[X(X-1)\cdots(X-r+1)\big]$ igrajo marsikje v verjetnosti pomembno vlogo. Z njimi se bomo še srečali pri rodovnih funkcijah v 8. razdelku.

Opomba. Kot je dobro znano, geometrijsko porazdelitev $\mathrm{Geom}(p)$ dobimo kot porazdelitev števila metov, kjer kovanec mečemo, dokler ne pade grb, meti pa so neodvisni. Ta poskus lahko dobimo kot limito poskusov iz 23. naloge, kjer iz posode, v kateri je n kroglic in od tega r rdečih, vlečemo kroglice, dokler ne dobimo rdeče, pri čemer gre n proti neskončno in r/n proti p (glej tudi opombo pri 26. nalogi iz 4. razdelka). V kakšnem smislu dobimo limito, bi sicer morali še precizirati, a tu tega ne bomo storili. Opazimo pa, da pričakovano vrednost in varianca ustrezne negativne hipergeometrijske porazdelitve, ki izhaja iz vlečenja kroglic, konvergirata proti pričakovani vrednosti in varianci geometrijske porazdelitve $\mathrm{Geom}(p)$, ki izhaja iz limitnega Bernoullijevega zaporedja poskusov: brž je $r_1, r_2, \ldots$ zaporedje z $\lim_{n\to\infty} r_n/n = p$, velja tudi:

$$\lim_{n\to\infty}\frac{n+1}{r_n+1}=\frac{1}{p} \qquad \text{in} \qquad \lim_{n\to\infty}\frac{r_n(n+1)(n-r_n)}{(r_n+1)^2(r_n+2)}=\frac{1-p}{p^2}\,.$$

29. Velja $\mathbb{E}(X)=0{\cdot}4$, $\mathbb{E}(Y)=2$ in $\mathbb{E}(XY)=0{\cdot}8$, torej sta X in Y res nekorelirani. Sta pa odvisni, kar lahko razberemo iz poljubne navzkrižne verjetnosti. To sledi tudi iz dejstva, da je pri $X=1$ skoraj gotovo $Y=2$, medtem ko pri $X=0$ skoraj gotovo velja $Y\neq 2$.

30. a) $0\le t\le 3/2$.
b) $t=0$, $t=1$.
c) $t=1$.

31. $\mathrm{var}(3X-Y)=64$, $\quad\mathbb{E}\big((X-2Y)^2\big)=95$.

32. Naj bo $S\sim\mathrm{Bin}(n,p)$.

Prvi način. Za $n\ge 1$ računamo:

$$\begin{aligned}\mathbb{E}(S)&=\sum_{k=1}^{\infty}k\binom{n}{k}p^k(1-p)^{n-k}=\\&=n\sum_{k=1}^{\infty}\binom{n-1}{k-1}p^k(1-p)^{n-k}=\\&=np\sum_{j=0}^{\infty}\binom{n-1}{j}p^j(1-p)^{n-1-j}\,.\end{aligned}$$

Opazimo, da je vsota na desni natanko vsota verjetnosti za binomsko porazdelitev $\mathrm{Bin}(n-1,p)$, torej je enaka 1. Sledi:

$$\mathbb{E}(S)=np\,.$$

Rezultat je pravilen tudi za $n = 0$, saj je v tem primeru slučajna spremenljivka S identično enaka 0. Nadalje za $n \geq 2$ velja:

$$\begin{aligned}\mathbb{E}\big[S(S-1)\big] &= \sum_{k=2}^{\infty} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} = \\ &= n(n-1)\sum_{k=2}^{\infty}\binom{n-2}{k-2} p^k (1-p)^{n-k} = \\ &= n(n-1)p^2 \sum_{j=0}^{\infty}\binom{n-2}{j} p^j (1-p)^{n-2-j} \,.\end{aligned}$$

Opazimo, da je vsota na desni natanko vsota verjetnosti za binomsko porazdelitev $\mathrm{Bin}(n-2, p)$, torej je enaka 1. Sledi:

$$\mathbb{E}\big[S(S-1)\big] = n(n-1)p^2 \,.$$

Rezultat je pravilen tudi za $n = 0$ in $n = 1$, saj je v tem primeru slučajna spremenljivka S lahko enaka kvečjemu 0 ali 1, torej je $S(S-1) = 0$. Sledi:

$$\mathrm{var}(S) = \mathbb{E}\big[S(S-1)\big] + \mathbb{E}(S) - \big(\mathbb{E}(S)\big)^2 = np(1-p) \,.$$

Drugi način. Spomnimo se, da je S porazdelitev števila uspelih poskusov v Bernoullijevem zaporedju n poskusov, pri katerih vsak uspe z verjetnostjo p. Pišemo lahko $S = X_1 + X_2 + \cdots + X_n$, kjer je X_i indikator dogodka, da i-ti poskus uspe. Velja $\mathbb{E}(X_i) = p$ in $\mathrm{var}(X_i) = p(1-p)$, torej je:

$$\mathbb{E}(S) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = np$$

in zaradi neodvisnosti tudi:

$$\mathrm{var}(S) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n) = np(1-p) \,.$$

33. Naj bo $S \sim \mathrm{NB}(n, p)$.

Prvi način. Računamo:

$$\begin{aligned}\mathbb{E}(S) &= \sum_{k=n}^{\infty} k\binom{k-1}{n-1} p^n (1-p)^{k-n} = \\ &= n\sum_{k=n}^{\infty}\binom{k}{n} p^n (1-p)^{k-n} = \\ &= \frac{n}{p}\sum_{j=n+1}^{\infty}\binom{j-1}{n} p^{n+1}(1-p)^{j-n-1} \,.\end{aligned}$$

Opazimo, da je vsota na desni natanko vsota verjetnosti za negativno binomsko $\mathrm{NB}(n+1, p)$, torej je enaka 1. Sledi:

$$\mathbb{E}(S) = \frac{n}{p} \,.$$

Nadalje velja:

$$\begin{aligned}\mathbb{E}\big[S(S+1)\big] &= \sum_{k=n}^{\infty} k(k+1)\binom{k-1}{n-1}p^n(1-p)^{k-n} = \\ &= n(n+1)\sum_{k=n}^{\infty}\binom{k+1}{n+1}p^n(1-p)^{k-n} = \\ &= \frac{n(n+1)}{p^2}\sum_{j=n+2}^{\infty}\binom{j-1}{n+1}p^{n+2}(1-p)^{j-n-2}\,.\end{aligned}$$

Opazimo, da je vsota na desni natanko vsota verjetnosti za binomsko porazdelitev $\mathrm{Bin}(n+2,p)$, torej je enaka 1. Sledi:

$$\mathbb{E}\big[S(S+1)\big] = \frac{n(n+1)}{p^2}\,.$$

Sledi:

$$\mathrm{var}(S) = \mathbb{E}\big[S(S+1)\big] - \mathbb{E}(S) - \big(\mathbb{E}(S)\big)^2 = \frac{n(1-p)}{p^2}\,.$$

Drugi način. Spomnimo se, da je S porazdelitev vsote $X_1 + X_2 + \cdots + X_n$, kjer so $X_1, X_2, \ldots, X_n$ neodvisne in porazdeljene geometrijsko $\mathrm{Geom}(p)$ (glej 15. nalogo iz 5. razdelka). V 28. nalogi smo izračunali, da je:

$$\mathbb{E}(X_i) = \frac{1}{p} \quad \text{in} \quad \mathrm{var}(X_i) = \frac{1-p}{p^2}\,.$$

Iz aditivnosti pričakovane vrednosti in aditivnosti variance za neodvisne slučajne spremenljivke dobimo želena rezultata:

$$\mathbb{E}(S) = \frac{n}{p} \quad \text{in} \quad \mathrm{var}(S) = \frac{n(1-p)}{p^2}\,.$$

34. a) Vsota $X + Y$ prav tako lahko zavzame vrednosti $0, 1, 2, \ldots$ in za vsak n iz te množice je:

$$\begin{aligned}\mathbb{P}(X+Y=n) &= \sum_{k=0}^{n}\mathbb{P}(X=k, Y=n-k) = \\ &= q^n(1-4q)\sum_{k=0}^{n}\binom{2k}{k}\binom{2n-2k}{n-k} = \\ &= 4^n q^n(1-4q)\,.\end{aligned}$$

Dobili smo, da je $X + Y + 1 \sim \mathrm{Geom}(1-4q)$.

b) Iz prejšnje točke sledi, da je:

$$\mathrm{var}(X+Y) = \mathrm{var}(X+Y+1) = \frac{4q}{(1-4q)^2}\,.$$

Toda ker sta X in Y neodvisni in enako porazdeljeni, je $\operatorname{var}(X+Y) = 2\operatorname{var}(X)$. Sledi:

$$\operatorname{var}(X) = \frac{\operatorname{var}(X+Y)}{2} = \frac{2q}{(1-4q)^2}\,.$$

35. Ker je $\mathbb{E}(X+Y) = 0$ ter sta X in Y neodvisni, velja:

$$\mathbb{E}\big((X+Y)^2\big) = \operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) = 2\sigma^2$$

in:

$$\begin{aligned}\mathbb{E}\big((X+Y)^4\big) &= \mathbb{E}(X^4) + 4\,\mathbb{E}(X^3)\,\mathbb{E}(X) + 6\,\mathbb{E}(X^2)\,\mathbb{E}(Y^2) + 4\,\mathbb{E}(X)\,\mathbb{E}(Y^3) + \mathbb{E}(Y^4) = \\ &= 2\kappa^4 + 6\sigma^4\,.\end{aligned}$$

Sledi:

$$\operatorname{var}\big((X+Y)^2\big) = 2(\kappa^4 + \sigma^4)\,.$$

Splošneje, če so $X_1, X_2, \ldots, X_n$ neodvisne slučajne spremenljivke z $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^2) = \sigma^2$ in $\mathbb{E}(X_i^4) = \kappa^4$ ter $S = X_1 + X_2 + \cdots + X_n$, velja $\mathbb{E}(S) = 0$ in zato:

$$\mathbb{E}(S^2) = \operatorname{var}(S) = \sum_{i=1}^{n} \operatorname{var}(X_i) = n\sigma^2\,.$$

Nadalje je:

$$\mathbb{E}(S^4) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} \mathbb{E}(X_i X_j X_k X_l)\,.$$

Med n^4 četvericami (i, j, k, l) je:

- n takih, kjer so vsi indeksi enaki; v tem primeru je $\mathbb{E}(X_i X_j X_k X_l) = \kappa^4$;
- $4n(n-1)$ takih, kjer so trije indeksi enaki, eden pa različen od prej omenjenih; v tem primeru je $\mathbb{E}(X_i X_j X_k X_l) = 0$;
- $3n(n-1)$ takih, kjer sta po dva in dva indeksa enaka; v tem primeru je $\mathbb{E}(X_i X_j X_k X_l) = \sigma^4$;
- $6n(n-1)(n-2)$ takih, kjer dva indeksa enaka, preostala dva pa sta različna tako med seboj kot tudi od prej omenjenih dveh; v tem primeru je $\mathbb{E}(X_i X_j X_k X_l) = 0$;
- $n(n-1)(n-2)(n-3)$ takih, kjer so vsi indeksi različni; v tem primeru je $\mathbb{E}(X_i X_j X_k X_l) = 0$.

Sledi $\mathbb{E}(S^4) = n\kappa^4 + 3n(n-1)\sigma^4$ in končno:

$$\operatorname{var}(S^4) = n\kappa^4 + (2n^2 - 3n)\sigma^4\,.$$

36. $\operatorname{cov}(X, Y) = -93$, $\quad\operatorname{corr}(X, Y) \doteq -0{\cdot}3952$.

37. Za poljubna $a, b > -2$ lahko izračunamo:

$$\mathbb{E}\big[X^a Y^b\big] = \int_0^1 \int_0^1 3x^a y^b (x^2 y + xy^2)\,\mathrm{d}x\,\mathrm{d}y = 3\left(\frac{1}{(a+3)(b+2)} + \frac{1}{(a+2)(b+3)}\right).$$

Od tod dobimo:

$$\mathbb{E}(X) = \mathbb{E}(Y) = \frac{17}{24}, \qquad \mathbb{E}(X^2) = \mathbb{E}(Y^2) = \frac{11}{20},$$
$$\mathrm{var}(X) = \mathrm{var}(Y) = \frac{139}{2880}, \qquad \mathbb{E}(XY) = \frac{1}{2}$$

in končno:

$$\mathrm{cov}(X,Y) = -1/576 \doteq -0{\cdot}001736, \qquad \mathrm{corr}(X,Y) = -5/139 \doteq -0{\cdot}03597.$$

38. *Prvi način.* Pišimo:

$$R = \sum_{k=1}^{n} R_k \quad \text{in} \quad Z = \sum_{l=1}^{n} Z_l,$$

kjer je R_k indikator dogodka, da je k-ta izvlečena kroglica rdeča, Z_l pa je indikator dogodka, da je l-ta izvlečena kroglica zelena. Iz $\mathbb{E}(R_k) = \frac{r}{N}$ in $\mathbb{E}(Z_l) = \frac{z}{N}$ dobimo, da je $\mathbb{E}(R) = \frac{nr}{N}$ in $\mathbb{E}(Z) = \frac{nz}{N}$; slednje sicer sledi tudi iz dejstva, da je $R \sim \mathrm{Hip}(r, n, N)$ in $Z \sim \mathrm{Hip}(z, n, N)$. Nadalje je:

$$\mathbb{E}(RZ) = \sum_{k=1}^{n} \sum_{l=1}^{n} \mathbb{E}(R_k Z_l).$$

Če je $k = l$, je $\mathbb{E}(R_k Z_l) = 0$, za $k \neq l$ pa je $\mathbb{E}(R_k Z_l) = \frac{rz}{N(N-1)}$. Sledi:

$$\mathbb{E}(RZ) = \frac{n(n-1)rz}{N(N-1)} \quad \text{in} \quad \mathrm{cov}(R,Z) = \mathbb{E}(RZ) - \mathbb{E}(R)\,\mathbb{E}(Z) = -\frac{rzn(N-n)}{N^2(N-1)}.$$

Drugi način. Pišimo:

$$R = \sum_{k=1}^{r} R'_k \quad \text{in} \quad Z = \sum_{l=1}^{z} Z'_l,$$

kjer je R'_k indikator dogodka, da je k-ta rdeča kroglica izvlečena, Z'_l pa je indikator dogodka, da je l-ta zelena kroglica izvlečena. Iz $\mathbb{E}(R'_k) = \mathbb{E}(Z'_l) = \frac{n}{N}$ dobimo, da je $\mathbb{E}(R) = \frac{nr}{N}$ in $\mathbb{E}(Z) = \frac{nz}{N}$, kar je isto kot pri prvem načinu. Nadalje je:

$$\mathbb{E}(RZ) = \sum_{k=1}^{r} \sum_{l=1}^{z} \mathbb{E}\big(R'_k Z'_l\big)$$

in za poljubna k, l je $\mathbb{E}(R'_k Z'_l) = \frac{n(n-1)}{N(N-1)}$. Sledi:

$$\mathbb{E}(RZ) = \frac{n(n-1)rz}{N(N-1)},$$

kar je seveda prav tako isto kot prej.

39. Za $i = 1, 2, \ldots, n$ naj bo X_i indikator dogodka, da i-ta ženska stavi biser, Y_i pa naj bo indikator dogodka, da i-ti moški stavi cekin. Očitno je:

$$\mathbb{E}(X_i) = \mathbb{E}(Y_i) = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 = \frac{25}{216}\,,$$
$$\mathrm{var}(X_i) = \mathrm{var}(Y_i) = \frac{25}{216} - \left(\frac{25}{216}\right)^2 = \frac{4775}{46656}\,.$$

Če je med i-to in j-to žensko natanko en moški, je X_iX_j indikator dogodka, da obe vržeta šestico in je hkrati ne vrže nobeden od določenih treh moških. Torej je:

$$\mathbb{E}(X_iX_j) = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = \frac{125}{7776}\,,$$
$$\mathrm{cov}(X_i, X_j) = \frac{125}{7776} - \left(\frac{25}{216}\right)^2 = \frac{125}{46656}\,.$$

Če pa je med ženskama več kot en moški, sta X_i in X_j neodvisni, saj je slučajna spremenljivka X_i natančno določena zgolj z izidi kock pri i-ti ženski in njenih sosedih. Zato je $\mathrm{cov}(X_i, X_j) = 0$. Sledi:

$$\mathrm{var}(X) = \mathrm{var}(Y) = \sum_{i=1}^n \sum_{j=1}^n \mathrm{cov}(X_i, X_j) = n\,\frac{4775}{46656} + 2n\,\frac{125}{46656} = \frac{5025\,n}{46656}\,.$$

Nadalje, če i-ta ženska in j-ti moški sedita skupaj, je X_iY_j indikator dogodka, da dana ženska vrže šestico, dani moški enojko, drugi moški, ki sedi poleg dane ženske, ne vrže šestice, druga ženska, ki sedi poleg danega moškega, pa ne vrže enojke. Sledi:

$$\mathbb{E}(X_iY_j) = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = \frac{25}{1296}\,,$$
$$\mathrm{cov}(X_i, Y_j) = \frac{25}{1296} - \left(\frac{25}{216}\right)^2 = \frac{275}{46656}\,.$$

Če pa i-ta ženska in j-ti moški ne sedita skupaj, sta slučajni spremenljivki X_i in Y_j neodvisni, zato je $\mathrm{cov}(X_i, Y_j) = 0$. Sledi:

$$\mathrm{cov}(X, Y) = \sum_{i=1}^n \sum_{j=1}^n \mathrm{cov}(X_i, Y_j) = 2n\,\frac{275}{46656} = \frac{275\,n}{23328}\,,$$
$$\mathrm{corr}(X, Y) = \frac{\frac{275\,n}{23328}}{\frac{5025\,n}{46656}} = \frac{22}{201} \doteq 0{\cdot}109\,.$$

Opomba. Izračun verjetnosti $\mathbb{P}(X = r)$, $\mathbb{P}(Y = s)$ in $\mathbb{P}(X = r, Y = s)$, s pomočjo katerih bi lahko bolj neposredno izračunali korelacijski koeficient, še malo ni preprost. Dogodek $\{X = 1\}$ se npr. lahko zgodi, če je prva ženska vrgla šestico, te pa ni vrgel nihče drug (ne ženska ne moški), lahko pa se zgodi tudi, če vse ženske vržejo šestico in jo vržeta tudi dva moška, med katerima je ena ženska, preostala dva moška pa je ne vržeta. Vmes je še veliko drugih možnosti.

40. Preslikava iz množice $\{0, 1, 2, \ldots, 2^n - 1\}$ v $\{0, 1\}^n$, ki število preslika v binarni zapis ali Grayevo kodo (z ustreznimi vodilnimi ničlami), je injektivna. Ker slika med množicama iste končne moči, je tudi bijektivna. Torej so vse možne n-terice ničel in enic enako verjetne. To pa se zgodi tudi, če vzamemo nabor n neodvisnih števil, porazdeljenih enakomerno na $\{0, 1\}$. Ker so robne porazdelitve in neodvisnost določene s skupno porazdelitvijo, sledi, da so tudi števke iz binarnega zapisa ali Grayeve kode neodvisne in porazdeljene enakomerno na $\{0, 1\}$.

b) Za $i = 0, 1, \ldots, n-1$ naj bo X_i števka na i-tem mestu v binarnem zapisu, Y_i pa naj bo števka v Grayevi kodi števila N; pri tem mesta gledamo z desne proti levi, tako da je $N = \sum_{i=0}^{n-1} 2^i X_i$. Velja $X = \sum_{i=0}^{n-1} X_i$ in $Y = \sum_{i=0}^{n-1} Y_i$. Iz prejšnje točke sledi, da sta slučajni spremenljivki X in Y porazdeljeni binomsko $\mathrm{Bin}(n, 1/2)$, torej je $\mathrm{var}(X) = \mathrm{var}(Y) = n/4$.

Velja $\mathrm{cov}(X, Y) = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \mathrm{cov}(X_i, Y_j)$. Za kovariance ločimo naslednje možnosti:

- Če je $i = j < n - 1$, je $X_i Y_j = 1$ natanko tedaj, ko je števka na i-tem mestu v binarnem zapisu enaka `1`, števka na $(i+1)$-tem mestu pa `0`. Torej je $\mathbb{E}(X_i Y_j) = 1/4$. Ker je $\mathbb{E}(X_i) = \mathbb{E}(Y_j) = 1/2$, je $\mathrm{cov}(X_i, X_j) = 0$.
- Če je $i = j = n - 1$, je $X_i Y_j = 1$ natanko tedaj, ko je števka na i-tem mestu v binarnem zapisu enaka `1`. Torej je $\mathbb{E}(X_i Y_j) = 1/2$ in zato $\mathrm{cov}(X_i, X_j) = 1/4$.
- Če je $i = j + 1$, je $X_i Y_j = 1$ natanko tedaj, ko je števka na i-tem mestu v binarnem zapisu enaka `1`, števka na $(i+1)$-tem mestu pa `0`. Torej je spet $\mathbb{E}(X_i Y_j) = 1/4$ in zato $\mathrm{cov}(X_i, X_j) = 0$.
- V vseh ostalih primerih sta X_i in Y_j neodvisni, ker je X_i natančno določena s števko na i-tem mestu, Y_j pa s števkama na j-tem in $(j+1)$-tem mestu, množici $\{i\}$ in $\{j, j+1\}$ pa sta disjunktni. Spet sledi $\mathrm{cov}(X_i, X_j) = 0$.

Le za $i = j = n - 1$ je torej $\mathrm{cov}(X_i, X_j) = 1/4$, sicer pa je $\mathrm{cov}(X_i, X_j) = 0$. Sledi $\mathrm{cov}(X, Y) = 1/4$ in končno $\mathrm{corr}(X, Y) = 1/n$.

41. $\mathrm{cov}(U, V) = a - a^2$, U in V sta nekorelirani pri $a = 0$ in $a = 1$.

42. a) $\mathbf{K}(\mathbf{A}\boldsymbol{X}) = \mathbf{A}\,\mathbf{K}(\boldsymbol{X})\,\mathbf{A}^{\mathrm{T}}$.

b) $\mathbf{K}(\mathbf{A}\boldsymbol{X}, \mathbf{B}\boldsymbol{Y}) = \mathbf{A}\,\mathbf{K}(\boldsymbol{X}, \boldsymbol{Y})\,\mathbf{B}^{\mathrm{T}}$.

c) $\mathrm{cov}\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle, \langle \boldsymbol{Y}, \boldsymbol{v}\rangle\big) = \langle \mathbf{K}(\boldsymbol{X}, \boldsymbol{Y})\boldsymbol{v}, \boldsymbol{u}\rangle$.

d) Iz definicije sledi, da je vsaka kovariančna matrika, ki se nanaša na en slučajni vektor, simetrična. Iz prejšnje točke pa dobimo, da za vsak slučajni vektor $\boldsymbol{X}$ s kovariančno matriko $\boldsymbol{\Sigma}$ in vsak determinističen vektor $\boldsymbol{u}$ velja $\langle \boldsymbol{\Sigma}\boldsymbol{u}, \boldsymbol{u}\rangle = \mathrm{var}\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle\big) \ge 0$. Torej mora biti vsaka kovariančna matrika, ki se nanaša na en slučajni vektor, pozitivno definitna.

Velja pa tudi obratno: za vsako simetrično pozitivno definitno matriko $\boldsymbol{\Sigma}$ obstaja slučajni vektor $\boldsymbol{X}$, za katerega je $\mathbf{K}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. Najprej to opazimo za primer, ko je $\mathbf{\Sigma} = \mathbf{I}_n$: za ta primer lahko vzamemo kar neodvisne slučajne spremenljivke $Z_1, \ldots, Z_n$ z varianco 1 in jih zložimo v slučajni vektor $\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$; velja $\mathbf{K}(\mathbf{Z}) = \mathbf{I}_n$.

Za splošno simetrično pozitivno definitno matriko $\mathbf{\Sigma}$ pa uporabimo *razcep Choleskega*[72]: obstaja taka matrika $\mathbf{A}$, da je $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$. Po prejšnjem obstaja slučajni vektor $\boldsymbol{Z}$ z $\mathbf{K}(\boldsymbol{Z}) = \mathbf{I}_n$, kjer je n število stolpcev matrike $\mathbf{A}$. Po točki a) je tedaj $\mathbf{K}(\mathbf{A}\boldsymbol{Z}) = \mathbf{A}\mathbf{I}_n\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}$, kot zahtevano.

43. V prejšnji nalogi smo izpeljali, da velja $\mathrm{var}\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle\big) = \langle \mathbf{\Sigma}\boldsymbol{u}, \boldsymbol{u}\rangle$. Ta izraz je minimalen oz. maksimalen, če je $\boldsymbol{u}$ lastni vektor matrike $\mathbf{\Sigma}$ za minimalno oz. maksimalno lastno vrednost. Iz karakterističnega polinoma:

$$\det(\mathbf{\Sigma} - \lambda\mathbf{I}) = -\lambda^3 + 12\lambda^2 - 21\lambda + 10 = -(\lambda - 1)^2(\lambda - 10)$$

dobimo dvakratno lastno vrednost $\lambda_{1,2} = 1$ in še lastno vrednost $\lambda_3 = 10$. Najmanjša možna varianca je 1, dosežena pa je, če je $\boldsymbol{u}$ lastni vektor matrike $\mathbf{\Sigma}$ za to lastno vrednost. To pa so vsi (enotski) vektorji, pravokotni na vektor $(2, 1, 2)$. Največja možna varianca pa je 10 in dosežena je za $\boldsymbol{u} = \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)$ in za $\boldsymbol{u} = \left(-\frac{2}{3}, -\frac{1}{3}, -\frac{2}{3}\right)$.

44. Naj bo slučajni vektor $\boldsymbol{X}$ porazdeljen večrazsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}$ in kovariančno matriko $\mathbf{\Sigma}$ (obojim iz definicije večrazsežne normalne porazdelitve). Dokazati moramo, da je $\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ in $\mathbf{K}(\boldsymbol{X}) = \mathbf{\Sigma}$. Najprej to dokažimo za primer, ko je $\boldsymbol{X}$ porazdeljen *standardno* normalno. Če je torej $\boldsymbol{X} = (X_1, \ldots, X_n)$ porazdeljen standardno n-razsežno normalno, so njegove komponente neodvisne in porazdeljene standardno enorazsežno normalno, torej je $\mathbb{E}(X_i) = 0$, $\mathrm{var}(X_i) = 1$ in $\mathrm{cov}(X_i, X_j) = 0$ za poljubna $i \neq j$. To pomeni, da je $\boldsymbol{E}(\boldsymbol{X}) = \mathbf{0}$ in $\mathbf{K}(\boldsymbol{X}) = \mathbf{I}_n$, tako kot mora biti.

V splošnem pa je X enako porazdeljen kot $\mathbf{A}\boldsymbol{Z}_n + \boldsymbol{\mu}$, kjer je $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}$ (npr. $\mathbf{A} = \mathbf{\Sigma}^{1/2}$). Sledi:

$$\boldsymbol{E}(\boldsymbol{X}) = \mathbf{A}\,\boldsymbol{E}(\boldsymbol{Z}_n) + \boldsymbol{\mu} = \boldsymbol{\mu}\,, \qquad \mathbf{K}(\boldsymbol{X}) = \mathbf{A}\,\mathbf{K}(\boldsymbol{X})\mathbf{A}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \mathbf{\Sigma}\,,$$

kot mora biti.

[72] André-Louis Cholesky (1875–1918), francoski oficir in matematik

7. Pogojne porazdelitve in pogojevanje na slučajne spremenljivke

1. Naj bo B dogodek, da je po Pepčkovi zamenjavi na obeh prvih kovancih grb. Če se zgodi B, je očitno $X > 0$. Iz:

$$\mathbb{P}\big(\{X = 1\} \cap B\big) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{12}, \quad \mathbb{P}\big(\{X = 2\} \cap B\big) = \frac{1}{4}\left(\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2}\right) = \frac{1}{6}$$

najprej dobimo $\mathbb{P}(B) = 1/4$, nato pa še pogojno porazdelitev, ki je določena s shemo $\begin{pmatrix} 1 & 2 \\ 1/3 & 2/3 \end{pmatrix}$.

2. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{27}{65} & \frac{18}{65} & \frac{12}{65} & \frac{8}{65} \end{pmatrix}$.

3. Pogojno na $Y = 2$ je $X \sim \begin{pmatrix} 1 & 2 \\ 1/4 & 3/4 \end{pmatrix}$.

Pogojno na $X = 0$ je $Y \sim \begin{pmatrix} 0 & 1 \\ 1/3 & 2/3 \end{pmatrix}$.

Pogojno na $Y \ge X$ je $Y \sim \begin{pmatrix} 0 & 1 & 2 \\ 1/7 & 2/7 & 4/7 \end{pmatrix}$.

4. To je zvezna porazdelitev z gostoto:

$$f_{X \mid X \ge a}(x) = \begin{cases} \lambda\, e^{-\lambda(x-a)} & ; \ x > a \\ 0 & ; \ \text{sicer.} \end{cases}$$

To pomeni, da je slučajna spremenljivka $X - a$ pogojno glede na dogodek $\{X \ge a\}$ spet porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Pravimo, da je eksponentna porazdelitev *pozabljiva* (beseda dobi smisel, če si zalogo vrednosti slučajne spremenljivke predstavljamo kot čas).

5. To je porazdelitev z gostoto:

$$f_{X \mid X < Y}(x) = \begin{cases} 2(1 - x) & ; \ 0 < x < 1 \\ 0 & ; \ \text{sicer.} \end{cases}$$

6. Najprej bomo izračunali pogojno kumulativno porazdelitveno funkcijo. Označimo s T_1 dogodek, da je Tone vprašal, ali je $X < 2/3$, s T_2 dogodek, da je Tone vprašal, ali je $X > 1/3$, z D pa dogodek, da je bil odgovor na vprašanje pritrdilen. Pišimo:

$$\begin{aligned} F_{X \mid D}(x) &= \frac{\mathbb{P}(\{X \le x\} \cap D)}{\mathbb{P}(D)}, \\ \mathbb{P}(\{X \le x\} \cap D) &= \mathbb{P}(T_1)\,\mathbb{P}\big(\{X \le x\} \cap D \bigm| T_1\big) + \\ &\quad + \mathbb{P}(T_2)\,\mathbb{P}\big(\{X \le x\} \cap D \bigm| T_2\big). \end{aligned}$$

Sedaj ločimo tri primere. Za $0 \le x \le 1/3$ velja:

$$\begin{aligned}\mathbb{P}\big(\{X \le x\} \cap D \bigm| T_1\big) &= \mathbb{P}(X \le x) = x\,,\\ \mathbb{P}\big(\{X \le x\} \cap D \bigm| T_2\big) &= 0\,,\end{aligned}$$

za $1/3 \le x \le 2/3$ velja:

$$\begin{aligned}\mathbb{P}\big(\{X \le x\} \cap D \bigm| T_1\big) &= \mathbb{P}(X \le x) = x\,,\\ \mathbb{P}\big(\{X \le x\} \cap D \bigm| T_2\big) &= \mathbb{P}\big(\tfrac{1}{3} < X \le x\big) = x - \tfrac{1}{3}\,,\end{aligned}$$

za $2/3 < x \le 1$ pa velja:

$$\begin{aligned}\mathbb{P}\big(\{X \le x\} \cap D \bigm| T_1\big) &= \tfrac{2}{3}\,,\\ \mathbb{P}\big(\{X \le x\} \cap D \bigm| T_2\big) &= \mathbb{P}\big(\tfrac{1}{3} < X \le x\big) = x - \tfrac{1}{3}\,.\end{aligned}$$

Če v vse skupaj vstavimo $x = 1$ in seštejemo, najprej dobimo $\mathbb{P}(D) = 2/3$. Od tod dobimo pogojno kumulativno porazdelitveno funkcijo:

$$F_{X|D}(x) = \begin{cases} 0 & ;\ x \le 0 \\ \frac{3}{4}\,x & ;\ 0 \le x \le \frac{1}{3} \\ \frac{3}{2}\,x - \frac{1}{4} & ;\ \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{3}{4}\,x + \frac{1}{4} & ;\ \frac{2}{3} \le x \le 1 \\ 1 & ;\ x \ge 1 \end{cases}$$

in nazadnje še pogojno gostoto:

$$f_{X|D}(x) = \begin{cases} 0 & ;\ x < 0 \\ \frac{3}{4} & ;\ 0 < x < \frac{1}{3} \\ \frac{3}{2} & ;\ \frac{1}{3} < x < \frac{2}{3} \\ \frac{3}{4} & ;\ \frac{2}{3} < x < 1 \\ 0 & ;\ x > 1 \end{cases}\,.$$

7. Za X in B iz 1. naloge velja $\mathbb{E}(X \mid D) = 5/3$ in $\operatorname{var}(X \mid D) = 2/9$. Za X in D iz 6. naloge pa velja $\mathbb{E}(X \mid D) = 1/2$ in $\operatorname{var}(X \mid D) = 7/108$.

8. Če je Z indikator dogodka, da je $2X > Y$, velja:

$$\mathbb{E}(XY \mid 2X > Y) = \frac{\mathbb{E}(XYZ)}{\mathbb{P}(2X > Y)}\,.$$

Po izreku o polni verjetnosti je:

$$\begin{aligned}\mathbb{P}(2X > Y) &= \mathbb{P}(Y = 0)\,\mathbb{P}(2X > Y \mid Y = 0) + \mathbb{P}(Y = 1)\,\mathbb{P}(2X > Y \mid Y = 1) + \\ &\qquad + \mathbb{P}(Y = 2)\,\mathbb{P}(2X > Y \mid Y = 2) = \\ &= \mathbb{P}(Y = 0)\,\mathbb{P}(2X > 0) + \mathbb{P}(Y = 1)\,\mathbb{P}(2X > 1) + \\ &\qquad + \mathbb{P}(Y = 2)\,\mathbb{P}(2X > 2) = \\ &= \frac{1}{2}\,.\end{aligned}$$

Nadalje je $XYZ = f(X)\,g(Y)$, kjer je:

$$f(x) = \begin{cases} x & ; x > 1/2 \\ 0 & ; \text{sicer,} \end{cases} \qquad g(y) = \begin{cases} 1 & ; y = 1 \\ 0 & ; \text{sicer.} \end{cases}$$

Sledi:

$$\mathbb{E}(XY \mid 2X > Y) = 2\,\mathbb{E}\big[f(X)\big]\,\mathbb{E}\big[g(Y)\big] = \frac{2}{3}\int_{1/2}^{1} x\,\mathrm{d}x = \frac{1}{4}\,.$$

9. Naj bo B dogodek, da je bila premeščena kroglica bela, R pa dogodek, da je bila rdeča. Nadalje naj bo X število belih izvlečenih kroglic. Tedaj velja:

$$\mathbb{E}(X \mid B) = \frac{12}{10}\,, \quad \mathbb{E}(X \mid R) = \frac{9}{10}\,.$$

Sledi:

$$\mathbb{E}(X) = \mathbb{P}(B)\,\mathbb{E}(X \mid B) + \mathbb{P}(R)\,\mathbb{E}(X \mid R) = \frac{6}{10}\cdot\frac{12}{10} + \frac{4}{10}\cdot\frac{9}{10} = 1{\cdot}08\,.$$

10. Označimo z N število vseh metov in definirajmo naslednje tri domneve:

$$H_1 = \{\text{v prvem metu pade grb}\}\,.$$
$$H_2 = \{\text{v prvem metu pade cifra, v drugem pa grb}\}\,.$$
$$H_3 = \{\text{v prvih dveh metih pade cifra}\}\,.$$

Če se zgodi H_3, je očitno $N = 2$ in torej tudi $\mathbb{E}(N \mid H_3) = 2$. Če pa se zgodita H_1 ali H_2, je nadaljnje dogajanje spet zaporedje neodvisnih metov kovanca, zato je $\mathbb{E}(N \mid H_1) = 1 + \mathbb{E}(N)$ in $\mathbb{E}(N \mid H_2) = 2 + \mathbb{E}(N)$. Sledi:

$$\mathbb{E}(N) = \tfrac{1}{2}\big(1 + \mathbb{E}(N)\big) + \tfrac{1}{4}\big(2 + \mathbb{E}(N)\big) + \tfrac{1}{4}\cdot 2\,,$$

od koder sledi $\mathbb{E}(N) = 6$.

11. Če z Y označimo število vseh metov, z A pa dogodek, da ne pade nobena enojka, je $\mathbb{P}(A \mid Y) = \big(\tfrac{4}{5}\big)^{Y-1}$.

12. a) $$\mathbb{P}(A \mid Y) = \frac{\frac{1}{4}\cdot\frac{\binom{3}{Y}\binom{12}{4-Y}}{\binom{15}{4}}}{\frac{\binom{4}{Y}\binom{12}{4-Y}}{\binom{16}{4}}} = \frac{\frac{1}{4}\cdot\frac{\binom{4}{Y}\binom{11}{3-Y}}{\binom{15}{3}}}{\frac{\binom{4}{Y}\binom{12}{4-Y}}{\binom{16}{4}}} = \frac{4-Y}{12}.$$

b) Ker je $Y \sim \mathrm{Hip}(4, 4, 16)$, je $\mathbb{E}\big(\mathbb{P}(A \mid Y)\big) = \big(4 - \mathbb{E}(Y)\big)/12 = 1/4$. Opazimo, da je to enako ravno $\mathbb{P}(A)$.

c) Če Y zavzame vrednosti $b_1, b_2, b_3, \ldots$, iz formule za pričakovano vrednost funkcije slučajne spremenljivke in izreka o polni verjetnosti dobimo:

$$\mathbb{E}\big(\mathbb{P}(A \mid Y)\big) = \mathbb{P}(Y = b_1)\,\mathbb{P}(A \mid Y = b_1) + \mathbb{P}(Y = b_2)\,\mathbb{P}(A \mid Y = b_2) + \cdots =$$
$$= \mathbb{P}(A)\,.$$

13. Tako kot v 11. nalogi označimo z Y število vseh metov, z A pa dogodek, da ne pade nobena enojka. Izračunali smo že, da je $\mathbb{P}(A \mid Y) = \left(\frac{4}{5}\right)^{Y-1}$. Torej je:

$$\mathbb{P}(A) = \mathbb{E}\left[\left(\frac{4}{5}\right)^{Y-1}\right] = \sum_{k=1}^{\infty} \left(\frac{4}{5}\right)^{Y-1} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{k-1} = \frac{1}{2}\,.$$

14. a) Porazdelitev je hipergeometrijska $\mathrm{Hip}(2, Y-1, 9)$ oziroma:

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{72}(10-Y)(9-Y) & \frac{1}{36}(Y-1)(10-Y) & \frac{1}{72}(Y-1)(Y-2) \end{pmatrix}.$$

b) $\mathbb{E}(X \mid Y) = \frac{2}{9}(Y-1)$, $\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = \frac{2}{9}\big(\mathbb{E}(Y) - 1\big) = 1$. Iz simetrije lahko sklepamo, da je tudi $\mathbb{E}(X) = 1$: če z Z označimo število zelenih kroglic *za* rdečo, je zaradi simetrije $\mathbb{E}(Z) = \mathbb{E}(X)$, velja pa tudi $X + Z = 2$.

c) Pišimo $\mathbb{E}(X \mid Y) = g(Y)$, kjer je $g(y) = \mathbb{E}(X \mid Y = y)$. Če Y zavzame vrednosti $b_1, b_2, b_3, \ldots$, iz izreka o polni pričakovani vrednosti dobimo:

$$\begin{aligned} \mathbb{E}\big(\mathbb{E}(X \mid Y)\big) &= \mathbb{E}\big[g(Y)\big] = \sum_k \mathbb{P}(Y = b_k)\, g(b_k) = \sum_k \mathbb{P}(Y = b_k)\, \mathbb{E}(X \mid Y = b_k) = \\ &= \mathbb{E}(X)\,. \end{aligned}$$

15. V 7. nalogi iz 5. razdelka smo že izračunali, da je $S \sim \mathrm{Pois}(\lambda + \mu)$. Zdaj pa izračunajmo:

$$\begin{aligned} \mathbb{P}(X = k \mid S = n) &= \frac{\mathbb{P}(X = k, S = n)}{\mathbb{P}(S = n)} = \\ &= \frac{\mathbb{P}(X = k, Y = n-k)}{\mathbb{P}(S = n)} = \\ &= \frac{\mathbb{P}(X = k)\,\mathbb{P}(Y = n-k)}{\mathbb{P}(S = n)} = \\ &= \frac{n!}{k!\,(n-k)!}\,\frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n} = \\ &= \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k \left(\frac{\mu}{\lambda+\mu}\right)^{n-k}. \end{aligned}$$

Sledi, da je iskana pogojna porazdelitev binomska $\mathrm{Bin}\left(S, \dfrac{\lambda}{\lambda+\mu}\right)$.

16. Za izračun pogojne porazdelitve slučajnega vektorja $(X_1, X_2, \ldots, X_n)$ je treba za vse $k_1, k_2, \ldots, k_n, k \in \mathbb{N}_0$ določiti pogojne verjetnosti:

$$\begin{aligned} &\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n \mid S = k) = \\ &\qquad = \frac{\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n, S = k)}{\mathbb{P}(S = k)}\,. \end{aligned}$$

Najprej se spomnimo, da je $S \sim \mathrm{Pois}(\lambda)$, kjer je $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Nadalje, ker je $S = X_1 + X_2 + \cdots + X_n$, bo zgornja pogojna verjetnost različna od nič samo, če bo $k = k_1 + k_2 + \cdots + k_n$. Za ta primer pa bo:

$$\begin{aligned}
\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n \mid S = k) &= \\
&= \frac{\mathbb{P}(X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n)}{\mathbb{P}(S = k)} = \\
&= \frac{\lambda_1^{k_1} e^{-\lambda_1}}{k_1!} \frac{\lambda_2^{k_2} e^{-\lambda_2}}{k_2!} \cdots \frac{\lambda_n^{k_n} e^{-\lambda_n}}{k_n!} \cdot \frac{k!}{\lambda^k e^{-\lambda}} = \\
&= \frac{k!}{k_1!\, k_2! \cdots k_n!} \left(\frac{\lambda_1}{\lambda}\right)^{k_1} \left(\frac{\lambda_2}{\lambda}\right)^{k_2} \cdots \left(\frac{\lambda_n}{\lambda}\right)^{k_n} .
\end{aligned}$$

Dobili smo *multinomsko porazdelitev.*

Opomba. Prejšnjo nalogo lahko štejemo kot poseben primer te naloge, saj pri $n = 2$ in $j = 0, 1, 2, \ldots, k$ dobimo:

$$\begin{aligned}
\mathbb{P}(X_1 = j \mid S = k) = \mathbb{P}(X_1 = j, X_2 = k - j \mid S = k) &= \\
&= \frac{k!}{j!\,(k-j)!} \left(\frac{\lambda_1}{\lambda}\right)^{j} \left(\frac{\lambda_2}{\lambda}\right)^{k-j} = \\
&= \binom{k}{j} \left(\frac{\lambda_1}{\lambda}\right)^{j} \left(1 - \frac{\lambda_1}{\lambda}\right)^{k-j} .
\end{aligned}$$

17. Velja:

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\big[\mathbb{E}(X \mid Y)\big] = \mathbb{E}(Y) = 2 \cdot \frac{1}{2} = 1\,, \\
\mathbb{E}(X^2 \mid Y) &= \mathrm{var}(X \mid Y) + \big(\mathbb{E}(X \mid Y)\big)^2 = Y^2 + Y + 1\,, \\
\mathbb{E}(X^2) &= \mathbb{E}\big[\mathbb{E}(X^2 \mid Y)\big] = \mathbb{E}(Y^2) + \mathbb{E}(Y) + 1 = \mathrm{var}(Y) + \big(\mathbb{E}(Y)\big)^2 + \mathbb{E}(Y) + 1 = \\
&= 2 \cdot \frac{1}{2} \cdot \frac{1}{2} + 3 = \frac{7}{2}\,, \\
\mathrm{var}(X) &= \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{5}{2}\,.
\end{aligned}$$

18. Najprej iz formul za pogojni verjetnosti dobimo formulo za pogojno pričakovano vrednost:

$$\mathbb{E}(X_{n+1} \mid X_n) = X_n \, \frac{n + 1 - X_n}{n+1} + (X_n + 1)\, \frac{X_n}{n+1} = \frac{X_n(n+2)}{n+1}\,.$$

Sledi:

$$\mathbb{E}(X_{n+1}) = \frac{n+2}{n+1}\, \mathbb{E}(X_n)\,.$$

Ker je $X_1 = 1$ in zato tudi $\mathbb{E}(X_1) = 1$, od tod induktivno sledi $\mathbb{E}(X_n) = \frac{n+1}{2}$. Podobno je:

$$\begin{aligned}\mathbb{E}(Y_{n+1} \mid X_n) &= X_n(X_n+1)\,\frac{n+1-X_n}{n+1} + (X_n+1)(X_n+2)\,\frac{X_n}{n+1} = \\ &= \frac{(X_n^2+X_n)(n+3)}{n+1} = \\ &= \frac{n+3}{n+1}\,Y_n\,,\end{aligned}$$

od koder dobimo:

$$\mathbb{E}(Y_{n+1}) = \frac{n+3}{n+1}\,\mathbb{E}(Y_n)\,.$$

Ker je $Y_1 = 2$ in zato tudi $\mathbb{E}(Y_1) = 2$, od tod induktivno sledi $\mathbb{E}(Y_n) = \frac{(n+1)(n+2)}{3}$. Sledi:

$$\operatorname{var}(X_n) = \mathbb{E}(X_n^2) - \big(\mathbb{E}(X_n)\big)^2 = \mathbb{E}(Y_n) - \mathbb{E}(X_n) - \big(\mathbb{E}(X_n)\big)^2 = \frac{n^2-1}{12}\,.$$

19. Velja:

$$\begin{aligned}\mathbb{P}(X = k) &= a(1-a)^{k-1} && ;\ k = 1,2,3,\ldots \\ \mathbb{P}(Y = l \mid X = k) &= \binom{l-1}{k-1} b^k (1-b)^{k-l} && ;\ l = k, k+1, k+2, \ldots\end{aligned}$$

Za $l = 1,2,3,\ldots$ sledi:

$$\begin{aligned}\mathbb{P}(Y = l) &= \sum_{k=1}^{l} \mathbb{P}(X = k)\,\mathbb{P}(Y = l \mid X = k) = \\ &= \sum_{k=1}^{l} a(1-a)^{k-1}\binom{l-1}{k-1} b^k(1-b)^{k-l} = \\ &= ab\sum_{j=0}^{l-1}\binom{l-1}{j}(1-a)^j b^j (1-b)^{l-1-j} = \\ &= ab(1-ab)^{l-1}\,.\end{aligned}$$

Torej je $Y \sim \operatorname{Geom}(ab)$.

Ta rezultat lahko lepo pojasnimo z naslednjo situacijo, ki lahko služi tudi kot izpeljava rešitve: oglejmo si dve neodvisni zaporedji neodvisnih poskusov, recimo jima serija A in serija B. Vsak poskus iz serije A uspe z verjetnostjo a, vsak poskus iz serije B pa z verjetnostjo b. Seriji naj se izvajata vzporedno, torej se hkrati izvedeta n-ti poskus iz serije A in n-ti poskus iz serije B. Temu recimo *združeni poskus*.

Naj bo Y število združenih poskusov, potrebnih, da hkrati uspeta poskusa iz obeh serij. Očitno je $Y \sim \operatorname{Geom}(ab)$.

Združene poskuse razdelimo na skupine glede na uspele poskuse iz serije B: prvo skupino naj sestavljajo združeni poskusi do vključno prvega uspelega poskusa v seriji B, drugo skupino tisti od nevključno prvega do vključno drugega uspelega in tako naprej: i-to skupino naj sestavljajo združeni poskusi od nevključno $(i-1)$-tega do vključno i-tega uspelega poskusa iz serije B.

V vsaki skupini bomo gledali število združenih poskusov, gledali pa bomo tudi, ali je v zadnjem združenem poskusu v posamezni skupini, ko je po definiciji uspel poskus iz serije B, uspel tudi poskus iz serije A. Taki skupini bomo rekli kar *uspela.* Tedaj so števila združenih poskusov po skupinah in tudi uspelost posameznih skupin med seboj neodvisni: to se da izpeljati iz krepke lastnosti Markova.

Označimo z X število skupin do vključno prve uspele. Iz zgoraj povedanega sledi, da je $X \sim \mathrm{Geom}(a)$. Nadalje z Y_i označimo število združenih poskusov v i-ti skupini. Iz zgoraj povedanega sledi, da so slučajne spremenljivke $Y_1, Y_2, \ldots$ med seboj neodvisne in imajo geometrijsko porazdelitev $\mathrm{Geom}(b)$.

Velja pa tudi $Y = Y_1 + Y_2 + \cdots + Y_X$. To pomeni, da se pogojna porazdelitev slučajne spremenljivke Y glede na $X = k$ ujema s pogojno porazdelitvijo vsote $Y_1 + Y_2 + \cdots + Y_k$ glede na ta dogodek. A ker je zaporedje $Y_1, Y_2, \ldots$ neodvisno od X, se slednja pogojna porazdelitev ujema z brezpogojno porazdelitvijo prej omenjene vsote, iz prej povedanega pa sledi, da je le-ta negativna binomska $\mathrm{NB}(k, b)$.

Dobili smo torej isto situacijo kot v besedilu naloge. Ker je (brezpogojna) porazdelitev slučajne spremenljivke Y natančno določena s porazdelitvijo slučajne spremenljivke X in pogojne porazdelitve slučajne spremenljivke Y glede na X, rezultat naloge sledi.

20. Tako kot v 11. nalogi označimo z Y število vseh metov, z X pa označimo število vseh enojk. Pogojno na Y je $X \sim \mathrm{Bin}(Y-1, 1/5)$. Ob dogovoru, da je $\binom{m}{k} = 0$, brž ko je $m \notin \{0, 1, \ldots, m\}$, torej velja:

$$\mathbb{P}(X = n \mid Y) = \binom{Y-1}{n} \left(\frac{1}{5}\right)^n \left(\frac{4}{5}\right)^{Y-1-n} .$$

Za $n = 0, 1, 2, \ldots$ torej velja:

$$\begin{aligned}\mathbb{P}(X = n) &= \mathbb{E}\left[\binom{Y-1}{n} \left(\frac{1}{5}\right)^n \left(\frac{4}{5}\right)^{Y-1-n}\right] = \\ &= \frac{1}{6} \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} \binom{k-1}{n} \left(\frac{1}{5}\right)^n \left(\frac{4}{5}\right)^{k-1-n} = \\ &= \frac{1}{6} \left(\frac{1}{4}\right)^n \sum_{k=n+1}^{\infty} \binom{k-1}{n} \left(\frac{2}{3}\right)^{k-1} .\end{aligned}$$

Zdaj pa se spomnimo na negativno binomsko porazdelitev: če je $Z \sim \mathrm{NB}(n+1, p)$, velja:

$$\mathbb{P}(Z = k) = \binom{k-1}{n} p^{n+1} (1-p)^{k-n-1} ; \quad k = n+1, n+2, \ldots$$

Sledi:

$$\sum_{k=n+1}^{\infty} \binom{k-1}{n} (1-p)^{k-1} = \frac{(1-p)^n}{p^{n+1}} \,.$$

Torej je:

$$\sum_{k=n+1}^{\infty} \binom{k-1}{n} \left(\frac{2}{3}\right)^{k-1} = 3 \cdot 2^n$$

in končno:

$$\mathbb{P}(X = n) = \frac{1}{2^{n+1}} \,.$$

Dobili smo geometrijsko porazdelitev Geom(1/2), pomaknjeno za ena v levo.

21. Pogojno na N je $S \sim \mathrm{Bin}(N, p)$, tj.:

$$\mathbb{P}(S = k \mid N) = \binom{N}{k} p^k (1-p)^{N-k} \,.$$

Sledi:

$$\begin{aligned}
\mathbb{P}(S = k) &= \mathbb{E}\left[\binom{N}{k} p^k (1-p)^{N-k}\right] = \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n \, e^{-\lambda}}{n!} \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \frac{\lambda^n p^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} = \\
&= \frac{(p\lambda)^k \, e^{-p\lambda}}{k!} \,.
\end{aligned}$$

Torej je $S \sim \mathrm{Pois}(p\lambda)$. Podobno je tudi $T \sim \mathrm{Pois}((1-p)\lambda)$. Nadalje velja še:

$$\begin{aligned}
\mathbb{P}(S = k, T = l) = \mathbb{P}(S = k, N = k + l) &= \mathbb{P}(N = k + l)\, \mathbb{P}(S = k \mid N = k + l) = \\
&= \frac{\lambda^{k+l} \, e^{-\lambda}}{(k+l)!} \binom{k+l}{k} p^k (1-p)^l = \frac{\lambda^{k+l} \, p^k (1-p)^l \, e^{-\lambda}}{k! \, l!} = \\
&= \mathbb{P}(S = k)\, \mathbb{P}(T = l) \,,
\end{aligned}$$

torej sta S in T res neodvisni.

Neodvisnost slučajnih spremenljivk S in T pa lahko izpeljemo tudi s sklicevanjem na 15. nalogo: iz le-te namreč sledi, da obstajajo slučajne spremenljivke S, T in N, za katere velja, da je $S \sim \mathrm{Pois}(p\lambda)$ in $N \sim \mathrm{Pois}(\lambda)$, da sta S in T neodvisni, da je $N = S + T$ in da pogojno na N velja $S \sim \mathrm{Bin}(N, p)$. Ker je porazdelitev slučajnega vektorja (S, T, N) natančno določena s porazdelitvijo slučajne spremenljivke N, pogojno porazdelitvijo slučajne spremenljivke S glede na N in zvezo $N = S + T$, od tod sledi, da S in T morata biti neodvisni, brž ko izpolnjujeta pogoje naloge.

Če je $N = n$ kar konstanta, sta S in T odvisni, brž ko je $n \geq 1$ in $0 < p < 1$: v tem primeru namreč S in T zavezameta vsaj dve vrednosti, a če $S = k$, je nujno $T = n - k$.

22. Pogojno glede na N ima slučajna spremenljivka S porazdelitev $\mathrm{Gama}(N, \lambda)$, torej za $s > 0$ velja:

$$f_{S|N}(s) = \frac{\lambda^N s^{N-1}}{(N-1)!}\, e^{-\lambda s}\,.$$

Sledi:

$$f_S(s) = \mathbb{E}\left[\frac{\lambda^N s^{N-1}}{(N-1)!}\, e^{-\lambda s}\right] = p\, e^{-\lambda s} \sum_{n=1}^{\infty} \frac{\lambda^n s^{n-1}(1-p)^{n-1}}{(n-1)!} = \lambda p\, e^{-p\lambda s}\,.$$

Torej je $S \sim \mathrm{Exp}(p\lambda)$.

23. $\mathbb{E}(X \mid Y = 2) = 7/4, \quad \mathbb{E}(X^2 \mid Y = 2) = 13/4,$
$\mathbb{E}(X^2 Y^2 \mid Y = 2) = \mathbb{E}(X^2 \cdot 2^2 \mid Y = 2) = 4\, \mathbb{E}(X^2 \mid Y = 2) = 13.$

24. Tako kot v 20. nalogi označimo z Y število vseh metov, z X pa označimo število vseh enojk. Pogojno na Y je $X \sim \mathrm{Bin}(Y-1, 1/5)$, torej je tudi $\mathbb{E}(X \mid Y) = \frac{Y-1}{5}$. Pričakovani delež enojk med vsemi meti je:

$$\mathbb{E}\left(\frac{X}{Y}\right) = \mathbb{E}\left(\frac{Y-1}{5Y}\right) = \frac{1}{5}\left[1 - \mathbb{E}\left(\frac{1}{Y}\right)\right] = \frac{1}{5} - \frac{1}{30}\sum_{k=1}^{\infty} \frac{1}{k}\left(\frac{5}{6}\right)^{k-1} = \frac{1}{5} - \frac{\ln 6}{25} \doteq$$
$$\doteq 0{\cdot}128\,.$$

25. Pišimo $Z = Z - aY + aY$ in uporabimo, da je slučajna spremenljivka $Z - aY$ neodvisna od para $(X, Y - aX)$. Ker se par (X, Y) deterministično izraža z $(X, Y - aX)$, je $Z - aY$ neodvisna tudi od para (X, Y). Zato je:

$$\mathbb{E}(Z - aY \mid X, Y) = \mathbb{E}(Z - aY) = 0\,.$$

Slučajna spremenljivka Y pa je funkcija para (X, Y), zato je:

$$\mathbb{E}(Y \mid X, Y) = Y\,.$$

Sledi:

$$\mathbb{E}(Z \mid X, Y) = \mathbb{E}(Z - aY \mid X, Y) + a\, \mathbb{E}(Y \mid X, Y) = aY\,.$$

Drugo pogojno pričakovano vrednost izračunamo s pomočjo:

$$\mathbb{E}(YZ \mid X, Y) = Y\, \mathbb{E}(Z \mid X, Y) = aY^2\,.$$

Iz neodvisnosti sledi:

$$\begin{aligned}
\mathbb{E}(YZ \mid X) = a\, \mathbb{E}(Y^2 \mid X) &= \\
&= a\, \mathbb{E}\big[(Y - aX)^2 + 2aX(Y - aX) + a^2X^2 \;\big|\; X\big] = \\
&= a\, \mathbb{E}\big[(Y - aX)^2\big] + 2a^2 X\, \mathbb{E}(Y - aX) + a^3 X^2\,.
\end{aligned}$$

Očitno je $\mathbb{E}(Y - aX) = 0$. Nadalje je:

$$(Y - aX)^2 = Y^2 - 2aXY + a^2X^2 = Y^2 - 2aX(Y - aX) - a^2X^2\,.$$

Z uporabo neodvisnosti ter prvih in drugih momentov dobimo $\mathbb{E}\big[(Y - aX)^2\big] = 1 - a^2$, torej je končno:

$$\mathbb{E}(YZ \mid X) = a - a^3 + a^3 X^2\,.$$

26. Ker slučajna spremenljivka N zavzame le vrednosti v naravnih številih, je dovolj za vsako naravno število n poiskati pogojno verjetnost $h_n(S) = \mathbb{P}(N = n \mid S)$. To bo natanko tedaj, ko bo za vsako primerno merljivo funkcijo g veljala zveza:

$$\mathbb{E}\big[h_n(S)\, g(S)\big] = \mathbb{E}\big[Z_n\, g(S)\big]\,,$$

kjer je Z_n indikator dogodka, da je $N = n$. Z drugimi besedami, veljati mora:

$$\mathbb{E}\big[h_n(S)\, g(S)\big] = \mathbb{P}(N = n)\, \mathbb{E}\big[g(S) \mid N = n\big]\,.$$

Ko vstavimo brezpogojno in pogojno gostoto, dobimo:

$$p\lambda \int_0^\infty h_n(s)\, g(s)\, e^{-p\lambda s}\, \mathrm{d}s = p(1-p)^{n-1}\, \frac{\lambda^n}{(n-1)!} \int_0^\infty g(s)\, s^{n-1}\, e^{-\lambda s}\, \mathrm{d}s\,.$$

Ta pogoj bo izpolnjen za regresijsko funkcijo:

$$h_n(s) = \frac{\lambda^{n-1}(1-p)^{n-1} s^{n-1}}{(n-1)!}\, e^{-\lambda(1-p)s}\,.$$

Z drugimi besedami, velja:

$$\mathbb{P}(N = n \mid S) = \frac{\lambda^{n-1}(1-p)^{n-1} S^{n-1}}{(n-1)!}\, e^{-\lambda(1-p)S}\,.$$

Pogojno na S ima torej slučajna spremenljivka $N-1$ Poissonovo porazdelitev $\mathrm{Pois}\big(\lambda(1-p)S\big)$.

27. Dejstvo, da je X pogojno na U porazdeljena binomsko $\mathrm{Bin}(n, U)$, lahko zapišemo takole:

$$\mathbb{E}\big[\mathbf{1}(X = k) \bigm| U\big] = \mathbb{P}(X = k \mid U) = \binom{n}{k} U^k (1-U)^{n-k}\,; \qquad k = 0, 1, 2, \ldots, n\,.$$

Sledi:

$$\mathbb{E}\big[h(U)\, \mathbf{1}(X = k)\big] = \binom{n}{k} \int_0^1 h(u)\, u^k (1-u)^{n-k}\, \mathrm{d}u\,.$$

Za $h \equiv 1$ dobimo:

$$\mathbb{P}(X = k) = \binom{n}{k} \mathrm{B}(k+1, n-k+1) = \frac{1}{n+1}\,.$$

Slučajna spremenljivka X je torej porazdeljena diskretno enakomerno na $\{0, 1, \ldots, n\}$. Nadalje velja:

$$\begin{aligned}\mathbb{E}\big[h(U) \bigm| X = k\big] &= \frac{\mathbb{E}\big[h(U)\, \mathbf{1}(X = k)\big]}{\mathbb{P}(X = k)} = \\ &= \frac{1}{\mathrm{B}(k+1, n-k+1)} \int_0^1 h(u)\, u^k (1-u)^{n-k}\, \mathrm{d}u\,.\end{aligned}$$

Sledi, da ima slučajna spremenljivka U pogojno na X porazdelitev $\mathrm{Beta}(X+1, n-X+1)$.

28. Brezpogojna porazdelitev slučajne spremenljivke X ima gostoto:

$$f_X(x) = \mathbb{E}\big[f_{X|N}(x)\big] = \sum_{n=1}^{\infty} \mathbb{P}(N = n) f_{X|N}(x \mid n)\,.$$

Za $x > 1$ torej velja:

$$f_X(x) = -\frac{1}{\ln(1-q)} \sum_{n=1}^{\infty} \frac{q^n}{x^{n+1}} = -\frac{1}{\ln(1-q)}\,\frac{q}{x(x-q)}\,,$$

medtem ko za $x \le 1$ lahko postavimo $f_X(x) = 0$.

Ker slučajna spremenljivka N zavzame le vrednosti v naravnih številih, je dovolj za vsako naravno število n poiskati pogojno verjetnost $h_n(X) = \mathbb{P}(N = n \mid X)$. To bo natanko tedaj, ko bo za vsako primerno merljivo funkcijo g veljala zveza:

$$\mathbb{E}\big[h_n(X)\,g(X)\big] = \mathbb{E}\big[Z_n\,g(X)\big]\,,$$

kjer je Z_n indikator dogodka, da je $N = n$. Z drugimi besedami, veljati mora:

$$\mathbb{E}\big[h_n(X)\,g(X)\big] = \mathbb{P}(N = n)\,\mathbb{E}\big[g(X) \mid N = n\big]\,.$$

Ko vstavimo brezpogojno in pogojno gostoto, dobimo:

$$\int_1^{\infty} h_n(x)\,g(x)\,\frac{q}{x(x-q)}\,\mathrm{d}x = q^n \int_1^{\infty} g(x)\,\frac{\mathrm{d}x}{x^{n+1}}\,.$$

Ta pogoj bo izpolnjen za regresijsko funkcijo:

$$h_n(x) = \frac{q^{n-1}(x-q)}{x^n}\,.$$

Z drugimi besedami, za $n = 1, 2, 3, \ldots$ velja:

$$\mathbb{P}(N = n \mid X) = \frac{q^{n-1}(X-q)}{X^n} = \left(1 - \frac{q}{X}\right)\left(\frac{q}{X}\right)^{n-1}\,.$$

Pogojno na X ima torej slučajna spremenljivka N geometrijsko porazdelitev $\mathrm{Geom}(1 - q/X)$.

29. Iz $\mathbb{E}(X \mid Y) \sim \begin{pmatrix} 3/4 & 1 & 7/4 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$ dobimo $\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = 7/6$ in $\mathrm{var}\big(\mathbb{E}(X \mid Y)\big) = 13/72$.

Nadalje iz $\mathrm{var}(X \mid Y) \sim \begin{pmatrix} 3/16 & 1 \\ 2/3 & 1/3 \end{pmatrix}$ dobimo $\mathbb{E}\big(\mathrm{var}(X \mid Y)\big) = 11/24$.

Iz $X \sim \begin{pmatrix} 0 & 1 & 2 \\ 1/4 & 1/3 & 5/12 \end{pmatrix}$ pa dobimo $\mathbb{E}(X) = 7/6$ in $\mathrm{var}(X) = 23/36$.

Seveda je $\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = \mathbb{E}(X)$, opazimo pa še, da je $\mathrm{var}(X) = \mathrm{var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\mathrm{var}(X \mid Y)\big)$.

30. Velja:

$$\begin{aligned}\operatorname{var}\big(\mathbb{E}(X \mid Y)\big) &= \mathbb{E}\big[\big(\mathbb{E}(X \mid Y)\big)^2\big] - \big(\mathbb{E}(X)\big)^2\\ \operatorname{var}(X \mid Y) &= \mathbb{E}(X^2 \mid Y) - \big(\mathbb{E}(X \mid Y)\big)^2,\\ \mathbb{E}\big(\operatorname{var}(X \mid Y)\big) &= \mathbb{E}(X^2) - \mathbb{E}\big[\big(\mathbb{E}(X \mid Y)\big)^2\big].\end{aligned}$$

Ko seštejemo, dobimo:

$$\operatorname{var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\operatorname{var}(X \mid Y)\big) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \operatorname{var}(X).$$

31. Iz Poissonove porazdelitve (glej 10. nalogo iz 6. razdelka) dobimo, da je $\mathbb{E}(X^2 \mid Y) = Y^2 + Y$. Iz eksponentne porazdelitve (glej 6. nalogo iz 6. razdelka) zdaj dobimo:

$$\mathbb{E}(X^2) = \mathbb{E}\big[\mathbb{E}(X^2 \mid Y)\big] = \mathbb{E}(Y^2 + Y) = \frac{2}{\lambda^2} + \frac{1}{\lambda}.$$

32. Najprej izračunamo:

$$\mathbb{E}\big[a^X \mid Y\big] = \sum_{k=0}^{\infty} a^k \, \frac{Y^k e^{-Y}}{k!} = e^{(a-1)Y}.$$

To obstaja za vse $a \in \mathbb{R}$. Formalno gledano je potem:

$$\mathbb{E}(a^X) = \mathbb{E}\big[e^{(a-1)Y}\big] = \lambda \int_0^{\infty} e^{(a-1)y}\, e^{-\lambda y}\, dy = \frac{\lambda}{\lambda - a + 1},$$

in sicer za $a < \lambda + 1$. Toda v resnici $\mathbb{E}(a^X)$ za $a \le -\lambda - 1$ ne obstaja, ker ne obstaja $\mathbb{E}(|a^X|)$. Iskana pričakovana vrednost torej obstaja za $|a| < \lambda + 1$.

33. Za $t > 0$ ima slučajna spremenljivka T pogojno gostoto:

$$f_{T \mid N=n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}\, e^{-\lambda t}.$$

Brezpogojno gostoto dobimo s seštetjem:

$$f_T(t) = \sum_{n=1}^{\infty} \mathbb{P}(N = n)\, f_{T \mid N=n}(t) = \sum_{n=1}^{\infty} p(1-p)^{n-1} \frac{\lambda^n t^{n-1}}{(n-1)!}\, e^{-\lambda t} = p\lambda\, e^{-p\lambda t}.$$

Torej je $T \sim \mathrm{Exp}(p\lambda)$.

34. Za $y > 0$ je $f_{X \mid Y}(x \mid y) = \begin{cases} e^{-(x-2y)} & ;\ x > 2y \\ 0 & ;\ \text{sicer.} \end{cases}$

Za $x > 0$ je $f_{Y \mid X}(y \mid x) = \begin{cases} 2/x & ;\ 0 < y < x/2 \\ 0 & ;\ \text{sicer.} \end{cases}$

Pogojno glede na X ima torej Y enakomerno porazdelitev $\mathrm{Unif}(0, X/2)$. Torej ima Y/X pogojno na X enakomerno porazdelitev $\mathrm{Unif}(0, 1/2)$, ne glede na X. To pa pomeni, da sta Y/X in X neodvisni.

Pogojno glede na Y pa ima X eksponentno porazdelitev $\mathrm{Exp}(1)$, pomaknjeno za $2Y$ v desno. Torej ima $X - 2Y$ pogojno glede na Y eksponentno porazdelitev $\mathrm{Exp}(1)$, ne glede na Y. To pa pomeni, da sta $X - Y$ in Y neodvisni.

35. *Prvi način.* Iz pogojne gostote:

$$f_{Y|X}(y \mid x) = \frac{|x|}{\sqrt{2\pi}}\, e^{-x^2y^2}$$

izračunamo dvorazsežno gostoto:

$$f_{X,Y}(x,y) = f_X(x)\, f_{Y|X}(y \mid x) = \frac{|x|}{2\pi}\, e^{-x^2(1+y^2)/2}\,,$$

z integriranjem katere določimo porazdelitev gostoto spremenljivke Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x = \frac{1}{\pi(1+y^2)}\,.$$

Slučajna spremenljivka Y ima torej standardno Cauchyjevo porazdelitev.

Drugi način (v resnici le drugače pisan prvi način). Uporabimo formulo:

$$f_Y(y) = \mathbb{E}\big[f_{Y|X}(y)\big] = \mathbb{E}\,\frac{|X|}{\sqrt{2\pi}}\, e^{-X^2y^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} |x|\, e^{-x^2(1+y^2)/2}\,\mathrm{d}x\,,$$

ki nam seveda da isto kot prej.

Tretji način. Če postavimo $U := XY$, iz formule za transformacijo enorazsežne normalne porazdelitve, uporabljene pogojno na X, dobimo, da je pogojna porazdelitev slučajne spremenljivke U glede na $X = x$ standardna normalna, ne glede na x. To pa pomeni, da je U neodvisna od X. Po formuli za gostoto funkcije dveh slučajnih spremenljivk dobimo:

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x)\, f_U(xy)\,|x|\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\infty}^{\infty} |x|\, e^{-x^2(1+y^2)/2}\,\mathrm{d}x\,,$$

kar je spet isto kot prej.

Končno je še:

$$f_{X|Y}(x \mid y) = \frac{|x|(1+y^2)}{2}\, e^{-x^2(1+y^2)/2}$$

in po transformacijski formuli za $x > 0$:

$$f_{X^2|Y}(t \mid y) = \frac{f_{X|Y}\big(\sqrt{t} \mid y\big) + f_{X|Y}\big(-\sqrt{t} \mid y\big)}{2\sqrt{t}} = \frac{1+y^2}{2}\, e^{-t(1+y^2)/2}\,,$$

medtem ko lahko za $x \le 0$ postavimo $f_{X^2|Y}(t \mid y) = 0$. Pogojno na Y ima torej X^2 eksponentno porazdelitev $\mathrm{Exp}\big(\frac{1+Y^2}{2}\big)$.

36. Dana pogojna porazdelitev ima pogojno gostoto (za $x > 0$):

$$f_{Y|X}(y \mid x) = \begin{cases} \dfrac{1}{(b-a)x} & ;\ ax < y < bx \\ 0 & ;\ \text{sicer.} \end{cases}$$

Iz tega najprej izračunamo dvorazsežno gostoto:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-x}}{b-a} & ; x > 0,\ ax < y < bx \\ 0 & ; \text{sicer} \end{cases} = \\ = \begin{cases} \dfrac{e^{-x}}{b-a} & ; x > 0,\ \dfrac{y}{b} < x < \dfrac{y}{a} \\ 0 & ; \text{sicer}, \end{cases}$$

nakar integriramo in dobimo brezpogojno gostoto: za $y > 0$ velja:

$$f_Y(y) = \frac{1}{b-a}\int_{y/b}^{y/a} e^{-x}\,\mathrm{d}x = \frac{e^{-y/b} - e^{-y/a}}{b-a},$$

torej je:

$$f_Y(y) = \begin{cases} \dfrac{e^{-y/b} - e^{-y/a}}{b-a} & ; y > 0 \\ 0 & ; \text{sicer.} \end{cases}$$

Zdaj pa uporabimo namig in iz pogojne gostote izračunamo pogojno pričakovano upanje:

$$\mathbb{E}\left(\frac{1}{Y}\,\middle|\, X = x\right) = \frac{1}{(b-a)x}\int_a^b \frac{1}{y}\,\mathrm{d}y = \frac{1}{(b-a)x}\ln\frac{b}{a}\,.$$

Brezpogojno matematično upanje je enako:

$$\mathbb{E}\left(\frac{1}{Y}\right) = \mathbb{E}\left[\mathbb{E}\left(\frac{1}{Y}\,\middle|\, X = x\right)\right] = \frac{1}{b-a}\ln\frac{b}{a}\int_0^\infty \frac{1}{x}\,x\,e^{-x}\,\mathrm{d}x = \frac{1}{b-a}\ln\frac{b}{a}\,.$$

Če to primerjamo z $\int_{-\infty}^{\infty}\frac{1}{y}\,f_Y(y)\,\mathrm{d}y$, dobimo znani rezultat:

$$\int_0^\infty \frac{e^{-y/b} - e^{-y/a}}{y}\,\mathrm{d}y = \ln\frac{b}{a}\,.$$

Če a zamenjamo z $1/b$, b pa z $1/a$, dobimo to v še bolj znani obliki:

$$\int_0^\infty \frac{e^{-ay} - e^{-by}}{y}\,\mathrm{d}y = \ln\frac{b}{a}\,.$$

37. Za $0 \le x, y \le 1$ velja $f_{X|Y}(x \mid y) = \dfrac{6(x^2 + xy)}{2 + 3y}$, od koder dobimo:

$$\mathbb{E}(X \mid Y) = \frac{4Y + 3}{6Y + 4} \quad \text{in} \quad \mathbb{E}(XY \mid Y) = Y\,\mathbb{E}(Y \mid X) = \frac{4Y^2 + 3Y}{6Y + 4}\,.$$

38. a) Dana pogojna porazdelitev ima pogojno gostoto (za $x > 0$):

$$f_{Y|X}(y \mid x) = \begin{cases} \dfrac{1}{(b-a)x} & ; ax < y < bx \\ 0 & ; \text{sicer.} \end{cases}$$

Iz tega najprej izračunamo dvorazsežno gostoto:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{e^{-x}}{b-a} & ;\ x>0,\ ax<y<bx \\ 0 & ;\ \text{sicer} \end{cases} =$$

$$= \begin{cases} \dfrac{e^{-x}}{b-a} & ;\ x>0,\ \dfrac{y}{b}<x<\dfrac{y}{a} \\ 0 & ;\ \text{sicer,} \end{cases}$$

nakar integriramo in dobimo brezpogojno gostoto: za $y > 0$ velja:

$$f_Y(y) = \frac{1}{b-a}\int_{y/b}^{y/a} e^{-x}\,\mathrm{d}x = \frac{e^{-y/b}-e^{-y/a}}{b-a}\,,$$

torej je:

$$f_Y(y) = \begin{cases} \dfrac{e^{-y/b}-e^{-y/a}}{b-a} & ;\ y>0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

b) Zahtevano matematično upanje lahko izračunamo na vsaj dva načina.

Prvi način: uporabimo namig in iz pogojne gostote najprej izračunamo pogojno matematično upanje:

$$\mathbb{E}\left(\frac{1}{Y}\;\middle|\; X=x\right) = \frac{1}{(b-a)x}\int_{ax}^{bx}\frac{1}{y}\,\mathrm{d}y = \frac{1}{(b-a)x}\ln\frac{b}{a}$$

oziroma:

$$\mathbb{E}\left(\frac{1}{Y}\;\middle|\; X\right) = \frac{1}{(b-a)X}\ln\frac{b}{a}\,.$$

Sledi:

$$\mathbb{E}\left(\frac{X^t}{Y}\;\middle|\; X\right) = X^t\,\mathbb{E}\left(\frac{1}{Y}\;\middle|\; X\right) = \frac{X^{t-1}}{(b-a)}\ln\frac{b}{a}\,,$$

zahtevano matematično upanje pa je enako:

$$\mathbb{E}\left(\frac{X^t}{Y}\right) = \mathbb{E}\left[\mathbb{E}\left(\frac{X^t}{Y}\;\middle|\; X\right)\right] = \frac{1}{b-a}\ln\frac{b}{a}\,\mathbb{E}\big(X^{t-1}\big) = \frac{1}{b-a}\ln\frac{b}{a}\int_0^\infty x^t e^{-x}\,\mathrm{d}x\,,$$

kar obstaja za $t > -1$ in je enako:

$$\mathbb{E}\left(\frac{X^t}{Y}\right) = \frac{\Gamma(t+1)}{b-a}\ln\frac{b}{a} = \frac{t!}{b-a}\ln\frac{b}{a}$$

(slednje za $t = 0, 1, 2, \ldots$).

Drugi način: neposredno iz dvorazsežne gostote, tako da najprej integriramo po y:

$$\begin{aligned}\mathbb{E}\left(\frac{1}{Y}\right) &= \iint_{ax<y<bx} \frac{x^t}{y} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = \\ &= \frac{1}{b-a}\int_0^\infty x^t e^{-x}\int_{ax}^{bx}\frac{1}{y}\,\mathrm{d}y\,\mathrm{d}x = \\ &= \frac{1}{b-a}\ln\frac{b}{a}\int_0^\infty x^t\,e^{-x}\,\mathrm{d}x = \\ &= \frac{\Gamma(t+1)}{b-a}\ln\frac{b}{a}\,.\end{aligned}$$

Opomba. Oba načina sta v resnici ekvivalentna: kar prvi način naredi z orodji iz pogojne verjetnosti, drugi naredi z orodji iz analize.

Opomba. Za $t=0$ lahko zahtevano matematično upanje dobimo tudi kot $\int_{-\infty}^{\infty}\frac{1}{y}f_Y(y)\,\mathrm{d}y$. Če ta integral primerjamo z izračunano vrednostjo, dobimo znani rezultat:

$$\int_0^\infty \frac{e^{-y/b}-e^{-y/a}}{y}\,\mathrm{d}y = \ln\frac{b}{a}\,.$$

Če a zamenjamo z $1/b$, b pa z $1/a$, dobimo to v še bolj znani obliki:

$$\int_0^\infty \frac{e^{-ay}-e^{-by}}{y}\,\mathrm{d}y = \ln\frac{b}{a}\,.$$

39. Iz gostote normalne porazdelitve dobimo, da je X porazdeljena eksponentno $\mathrm{Exp}(1)$, poleg tega pa še, da je slučajna spremenljivka Y pogojno na X porazdeljena normalno $\mathrm{N}\big(X,\sqrt{X}\big)$. Povedano v formulah je to:

$$f_X(x) = \begin{cases} e^{-x} & ;\ x>0 \\ 0 & ;\ \text{sicer,}\end{cases} \qquad f_{Y\mid X}(y\mid x) = \frac{1}{\sqrt{2\pi x}}\,e^{-\frac{(y-x)^2}{2x}} \quad (x>0)\,.$$

Podobno je tudi Z porazdeljena eksponentno $\mathrm{Exp}(1)$, W pa je pogojno na Z porazdeljena normalno $\mathrm{N}\big(Z,\sqrt{Z}\big)$. Povedano v formulah je to:

$$f_Z(z) = \begin{cases} e^{-z} & ;\ z>0 \\ 0 & ;\ \text{sicer,}\end{cases} \qquad f_{W\mid Z}(w\mid z) = \frac{1}{\sqrt{2\pi z}}\,e^{-\frac{(w-z)^2}{2z}} \quad (z>0)\,.$$

Ker sta X in Y neodvisni, ima slučajna spremenljivka $U := X+Y$ porazdelitev $\mathrm{Gama}(2,1)$. Povedano s formulo je to:

$$f_U(u) = \begin{cases} u\,e^{-u} & ;\ u>0 \\ 0 & ;\ \text{sicer.}\end{cases}$$

Iz neodvisnosti pa dobimo tudi pogojno porazdelitev slučajnega vektorja (Y,W) glede na (X,Z):

$$\begin{aligned} f_{Y,W\mid X,Z}(y,w\mid x,z) &= \frac{f_{X,Y,Z,W}(x,y,z,w)}{f_{X,Z}(x,z)} = \\ &= \frac{f_{X,Y}(x,y)\,f_{Z,W}(z,w)}{f_X(x)\,f_Z(z)} = \\ &= f_{Y\mid X}(y\mid x)\,f(W\mid Z)(w\mid z)\,.\end{aligned}$$

Poleg tega pa je še:

$$f_{Y|X,Z}(y \mid x,z) = \frac{f_{X,Y,Z}(x,y,z)}{f_{X,Z}(x,z)} = \frac{f_{X,Y}(x,y)\,f_Z(z)}{f_X(x)\,f_Z(z)} = \frac{f_{X,Y}(x,y)}{f_X(x)} = f_{Y|X}(y \mid x)$$

in podobno:

$$f_{W|X,Z}(w \mid x,z) = f_{W|Z}(w \mid z)\,.$$

Pogojno na (X, Z) sta torej slučajni spremenljivki Y in W neodvisni in porazdeljeni normalno $\mathrm{N}\big(X, \sqrt{X}\,\big)$ oziroma $\mathrm{N}\big(Z, \sqrt{Z}\,\big)$. To pa pomeni, da je vsota $V = Y + W$ pogojno na (X, Z) porazdeljena normalno $\mathrm{N}\big(X+Z, \sqrt{X+Z}\,\big)$. Povedano s formulo je to:

$$f_{V|X,Z}(v \mid x,z) = \frac{1}{\sqrt{2\pi(x+z)}}\,e^{-\frac{(w-x-z)^2}{2(x+z)}} \qquad (x, z > 0)\,.$$

Opazimo, da je pogojna porazdelitev vsote V glede na (X, Z) odvisna zgolj od vsote $X + Z$. Po pameti bi potem rekli, da je tudi pogojno na U vsota V porazdeljena normalno $\mathrm{N}\big(U, \sqrt{U}\,\big)$. Tudi v resnici je tako, saj iz skupne gostote:

$$f_{X,Z,V}(x,z,v) = f_{X,Z}(x,z)\cdot\frac{1}{\sqrt{2\pi(x+z)}}\,e^{-\frac{(w-x-z)^2}{2(x+z)}}$$

z dvakratno uporabo transformacijske formule dobimo:

$$\begin{aligned} f_{X,U,V}(x,u,v) &= f_{X,Z,V}(x,u-x,v) = \\ &= f_{X,Z}(x,u-x)\cdot\frac{1}{\sqrt{2\pi u}}\,e^{-\frac{(v-u)^2}{2u}} = \\ &= f_{X,U}(x,u)\cdot\frac{1}{\sqrt{2\pi u}}\,e^{-\frac{(v-u)^2}{2u}}\,, \end{aligned}$$

z integracijo po x pa nadalje:

$$f_{U,V}(u,v) = f_U(u)\cdot\frac{1}{\sqrt{2\pi u}}\,e^{-\frac{(v-u)^2}{2u}}\,.$$

Sledi, da je V pogojno na U res porazdeljena normalno $\mathrm{N}\big(U, \sqrt{U}\,\big)$, iskana skupna porazdelitev pa je na dlani:

$$f_{U,V}(u,v) = \begin{cases} \sqrt{\dfrac{u}{2\pi}}\,e^{-u-\frac{(v-u)^2}{2u}} & ;\ u > 0 \\ 0 & ;\ \text{sicer.} \end{cases}$$

40. *Prvi način*: neposredno z deljenjem gostot. Vemo, da je $\boldsymbol{X}_1$ porazdeljen m-razsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}_1$ in kovariančno matriko $\boldsymbol{\Sigma}_{11}$. Če označimo:

$$\mathbf{B} := \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1}$$

ter uvedemo $\boldsymbol{y}_1 = \boldsymbol{x}_1 - \boldsymbol{\mu}_1$ in $\boldsymbol{y}_2 = \boldsymbol{x}_2 - \boldsymbol{\mu}_2$, sledi:

$$
\begin{aligned}
&f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2 \mid \boldsymbol{x}_1) = \\
&\quad = f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{y}_2 + \boldsymbol{\mu}_2 \mid \boldsymbol{y}_1 + \boldsymbol{\mu}_1) = \\
&\quad = \sqrt{\det \frac{\mathbf{B}}{2\pi} \det\bigl(2\pi\boldsymbol{\Sigma}_{11}\bigr)}\, e^{-(\boldsymbol{y}_1^{\mathrm{T}}(\mathbf{B}_{11}-\boldsymbol{\Sigma}_{11}^{-1})\boldsymbol{y}_1+\boldsymbol{y}_1^{\mathrm{T}}\mathbf{B}_{12}\boldsymbol{y}_2+\boldsymbol{y}_2^{\mathrm{T}}\mathbf{B}_{21}\boldsymbol{y}_1+\boldsymbol{y}_2^{\mathrm{T}}\mathbf{B}_{22}\boldsymbol{y}_2)/2} = \\
&\quad = \sqrt{\det \frac{\mathbf{B}}{2\pi} \det\bigl(2\pi\boldsymbol{\Sigma}_{11}\bigr)}\, e^{-\boldsymbol{y}_1^{\mathrm{T}}(\mathbf{B}_{11}-\mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}-\boldsymbol{\Sigma}_{11}^{-1})\boldsymbol{y}_1^{\mathrm{T}}/2} \times \\
&\qquad\quad \times e^{-(\boldsymbol{y}_2+\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\boldsymbol{y}_1)^{\mathrm{T}}\mathbf{B}_{22}(\boldsymbol{y}_2+\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\boldsymbol{y}_1)/2}\,.
\end{aligned}
$$

Pogojno na $\boldsymbol{X}_1$ je torej $\boldsymbol{X}_2$ porazdeljen n-razsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}_2 - \mathbf{B}_{22}^{-1}\mathbf{B}_{21}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)$ in kovariančno matriko $\mathbf{B}_{22}^{-1}$. Iz formule za inverz bločne matrike:

$$
\begin{aligned}
&\begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} = \\
&\quad = \begin{bmatrix} \bigl(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\bigr)^{-1} & -\bigl(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\bigr)^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ -\bigl(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\bigr)^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \bigl(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\bigr)^{-1} \end{bmatrix},
\end{aligned}
$$

dobimo $\mathbf{B}_{22} = \bigl(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\bigr)^{-1}$ in $\mathbf{B}_{22}^{-1}\mathbf{B}_{21} = -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$. Pogojno na $\boldsymbol{X}_1$ je torej $\boldsymbol{X}_2$ porazdeljen n-razsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)$ in kovariančno matriko $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$.

Drugi način. Tako kot pri prvem načinu z deljenjem gostot dobimo pogojno gostoto, izraženo z matriko $\mathbf{B}$. Matrike $\mathbf{B}$ pa ne izračunamo, temveč le razberemo naslednje:

- Pogojna porazdelitev je večrazsežna normalna.
- Pogojna pričakovana vrednost je afina funkcija slučajnega vektorja $\boldsymbol{X}_1$, tj. $\boldsymbol{E}(\boldsymbol{X}_2 \mid \boldsymbol{X}_1) = \boldsymbol{a} + \mathbf{L}\boldsymbol{X}_1$.
- Pogojna kovariančna matrika je deterministična, torej recimo $\mathbf{K}(\boldsymbol{X}_2 \mid \boldsymbol{X}_1) = \boldsymbol{\Sigma}_{2|1}$.

Za izračun determinističnega vektorja $\boldsymbol{a}$ in deterministične matrike $\mathbf{L}$ nastavimo:

$$
\boldsymbol{\mu}_2 = \boldsymbol{E}(\boldsymbol{X}_2) = \boldsymbol{E}\bigl[\boldsymbol{E}(\boldsymbol{X}_2 \mid \boldsymbol{X}_1)\bigr] = \boldsymbol{E}(\boldsymbol{a} + \mathbf{L}\boldsymbol{X}_1) = \boldsymbol{a} + \mathbf{L}\boldsymbol{\mu}_1
$$

in:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{12} = \mathbf{K}(\boldsymbol{X}_1, \boldsymbol{X}_2) &= \\
&= \boldsymbol{E}\bigl[(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)\,\boldsymbol{E}\bigl(\boldsymbol{X}_2^{\mathrm{T}} \bigm| \boldsymbol{X}_1\bigr)\bigr] = \\
&= \boldsymbol{E}\bigl[(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)(\boldsymbol{a} + \mathbf{L}\boldsymbol{X}_1)^{\mathrm{T}}\bigr] = \\
&= \boldsymbol{E}\bigl[(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)\boldsymbol{X}_1^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\bigr] = \\
&= \mathbf{K}(X_1)\mathbf{L}^{\mathrm{T}}\bigr]\,.
\end{aligned}
$$

Od tod dobimo $\mathbf{L} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$ in $\boldsymbol{a} = \boldsymbol{\mu}_2 - \mathbf{L}\boldsymbol{\mu}_1$, torej je:

$$
\boldsymbol{E}(\boldsymbol{X}_2 \mid \boldsymbol{X}_1) = \boldsymbol{\mu}_2 + \mathbf{L}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1) = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)\,.
$$

Za izračun pogojne kovariančne matrike pa se spomnimo na razcep variance, ki velja tudi v večrazsežnem primeru, torej:

$$\begin{aligned}\boldsymbol{\Sigma}_{22} &= \mathbf{K}(\mathbf{X}_{22}) = \\ &= \mathbf{K}\big(\boldsymbol{E}(\boldsymbol{X}_2 \mid \boldsymbol{X}_2)\big) + \boldsymbol{E}\big(\mathbf{K}(\boldsymbol{X}_2 \mid \boldsymbol{X}_2)\big) = \\ &= \mathbf{K}\big(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)\big) + \boldsymbol{\Sigma}_{2|1} = \\ &= \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{2|1}\,.\end{aligned}$$

Torej je $\boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Ko povzamemo vse skupaj, vidimo, da smo dobili isto pogojno porazdelitev kot pri prvem načinu.

Tretji način: skličemo se na rešitev 37. naloge v 5. razdelku, kjer je dokazano, da je slučajni vektor $\boldsymbol{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1$ neodvisen od $\boldsymbol{X}_1$. Njegova pogojna porazdelitev glede na $\boldsymbol{X}_1$ se torej ujema z brezpogojno porazdelitvijo. V rešitvi prej omenjene naloge je dokazano še, da je slučajni vektor:

$$\boldsymbol{X}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{X}_1 = \begin{bmatrix} -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$$

porazdeljen n-razsežno normalno s pričakovano vrednostjo:

$$\begin{bmatrix} -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\mu}_1$$

in kovariančno matriko:

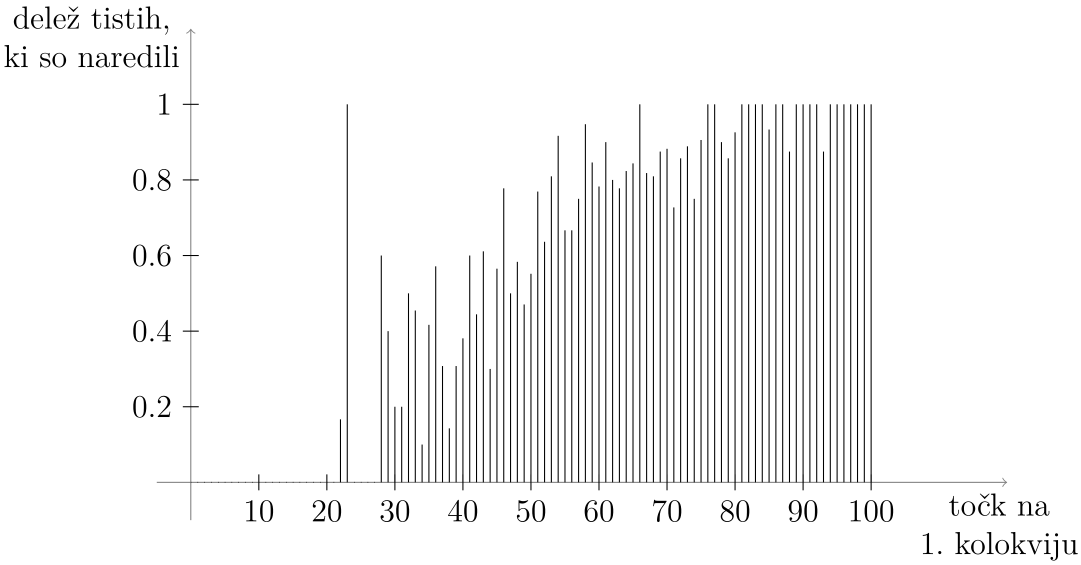
$$\begin{bmatrix} -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{I}_n \end{bmatrix} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\,.$$

Kot že rečeno, to velja tudi pogojno na $\boldsymbol{X}_1$. To pa pomeni, da je pogojno na $\boldsymbol{X}_1$ slučajni vektor $\boldsymbol{X}_2$ porazdeljen n-razsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)$ in kovariančno matriko $\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, kar je isto kot prej.

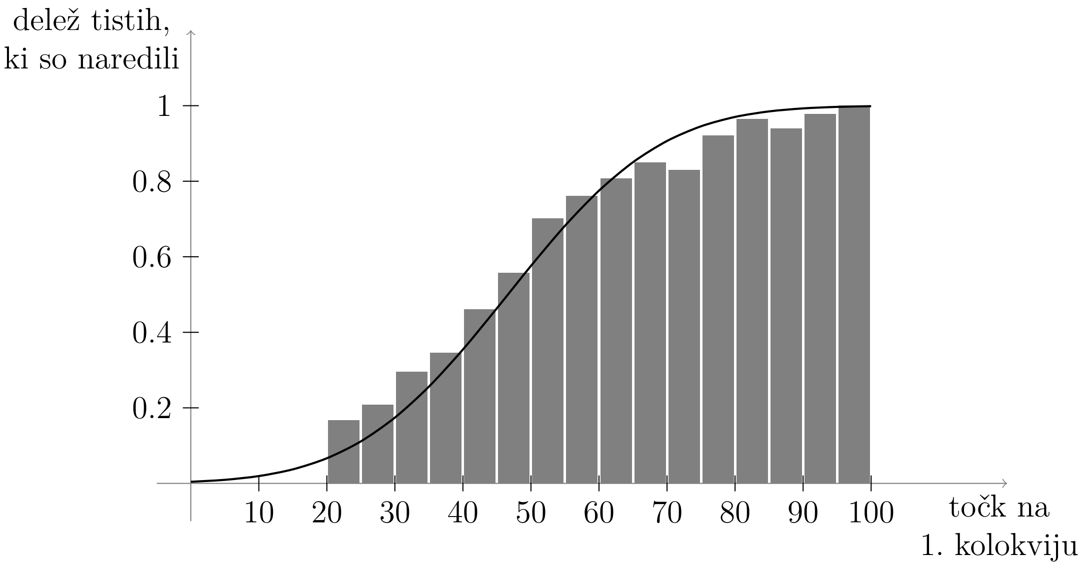
b) $\displaystyle \frac{1}{2} - \Phi\left(\frac{75 - 58{\cdot}0 - 0{\cdot}291 \cdot 25{\cdot}3 \cdot (25 - 59{\cdot}2)/20{\cdot}1}{25{\cdot}3 \cdot \sqrt{1 - 0{\cdot}291^2}}\right) \doteq 0{\cdot}111.$

V resnici je 7 kandidatov na prvem kolokviju zbralo 25 točk in nihče izmed njih ni naredil. Prav tako ni naredil nihče, ki je zbral 26 ali 27 točk, pač pa je naredil en kandidat od 6, ki so zbrali 22 točk, in 3 kandidati od 5, ki so zbrali 28 točk.

Diagram pogojnih deležev:

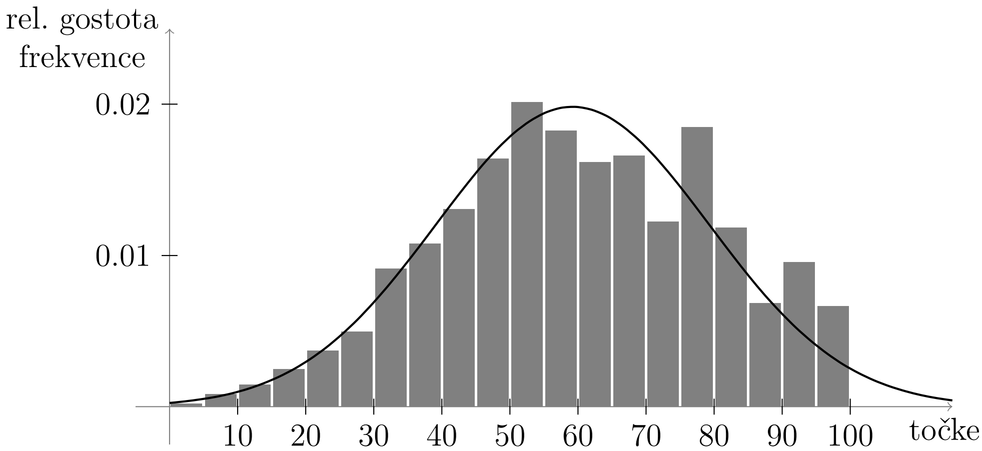


Zaradi velike zaloge vrednosti rezultatov dobimo veliko boljši pregled, če gledamo pogojne verjetnosti glede na 5 točk široke razpone. Krivulja ponazarja ocene za deleže (verjetnosti) na podlagi Gaussovega modela, tj. modela z dvorazsežno normalno porazdelitvijo:

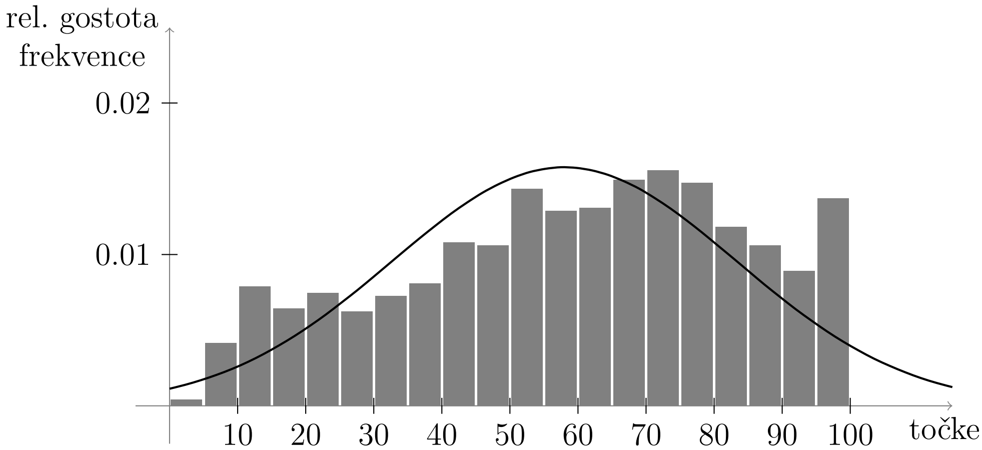


Histogram se kar dobro prilega krivulji glede na to, da se histogram frekvenčnih porazdelitev vsaj za drugi kolokvij ne prilega tako dobro normalni porazdelitvi. Prikazane so *relativne gostote frekvenc*, tj. deleži ustrezajo ploščinam pravokotnikov v histogramu; višine pravokotnikov so tako primerljive z gostoto zvezne porazdelitve, s katero aproksimiramo.

Histogram za prvi kolokvij:



Histogram za drugi kolokvij:



8. Rodovne, momentno-rodovne in karakteristične funkcije

1. $G_X(s) = e^{\lambda(s-1)}$, $\quad G_X^{(r)}(s) = \lambda^r e^{\lambda(s-1)}$, $\quad \mathbb{E}(X) = \operatorname{var}(X) = \lambda$, $\quad f_r = \lambda^r$.

2. $G(s) = \dfrac{ps}{1-(1-p)s} = \dfrac{p}{1-p}\left(\dfrac{1}{1-(1-p)s} - 1\right)$,

$\mathbb{E}(X) = \dfrac{1}{p}$, $\quad f_r = r!\,\dfrac{(1-p)^{r-1}}{p^r}$, $\quad \operatorname{var}(X) = \dfrac{(1-p)}{p^2}$.

3. V formulo iz namiga vstavimo $x = -qs$ in dobimo:

$$G_X(s) = c\sum_{k=1}^\infty \frac{q^k s^k}{k} = -c\ln(1-qs)\,.$$

Zdaj pa upoštevamo, da je $G_X(1) = 1$, in dobimo $c = -1/\ln(1-q)$. Tako je:

$$G_X(s) = \frac{\ln(1-qs)}{\ln(1-q)} = \log_{1-q}(1-qs)\,.$$

Po odvajanju dobimo:

$$G_X^{(r)}(s) = -\frac{(r-1)!}{\ln(1-q)}\,\frac{q^r}{(1-qs)^r}$$

in končno:

$$f_r = -\frac{(r-1)!}{\ln(1-q)}\,\frac{q^r}{(1-q)^r}\,.$$

4. Iz $G_X(s) = 1$ dobimo $a = e^{-2}$. Iz razvoja:

$$G_X(s) = e^{-2}\left[1 + (s+s^2) + \frac{(s+s^2)^2}{2} + \cdots\right] = e^{-2}\left[1 + s + \frac{3s^2}{2} + \cdots\right]$$

dobimo $\mathbb{P}(X=2) = \dfrac{3\,e^{-2}}{2} \doteq 0{\cdot}203$.

5. a) Iz modela dobimo, da za vse $n = 0, 1, 2, \ldots$ in vse $k = 0, 1, \ldots, m$ velja:

$$\mathbb{P}(X_{n+1} = k+1 \mid X_n = k) = \frac{m-k}{m} \quad \text{in} \quad \mathbb{P}(X_{n+1} = k-1 \mid X_n = k) = \frac{k}{m}$$

Sledi:

$$\mathbb{E}\big(s^{X_{n+1}} \bigm| X_n = k\big) = s^{k+1}\cdot\frac{m-k}{m} + s^{k-1}\cdot\frac{k}{m}\,.$$

Obe strani enačbe pomnožimo s $\mathbb{P}(X_n = k)$ in seštejemo po $k = 0, 1, \ldots, m$. Po formuli za popolno pričakovano vrednost je:

$$\begin{aligned} G_{n+1}(s) = \mathbb{E}\bigl(s^{X_{n+1}}\bigr) &= \\ &= \sum_{k=0}^{m} \mathbb{P}(X_n = k)\, \mathbb{E}\bigl(s^{X_{n+1}} \bigm| X_n = k\bigr) = \\ &= \sum_{k=0}^{m} \mathbb{P}(X_n = k) \left(s^{k+1} \cdot \frac{m-k}{m} + s^{k-1} \cdot \frac{k}{m} \right) = \\ &= s \sum_{k=0}^{m} s^k\, \mathbb{P}(X_n = k) + \frac{1-s^2}{m} \sum_{k=0}^{m} k\, s^{k-1}\, \mathbb{P}(X_n = k)\,. \end{aligned}$$

Prvi člen je enak $s\, G_n(s)$. Za drugi člen pa se spomnimo, da je $k\, s^{k-1} = \frac{\mathrm{d}}{\mathrm{d}s} s^k$. Po Abelovem izreku smemo potenčno vrsto znotraj konvergenčnega polmera vselej členoma odvajati. Sledi:

$$G_{n+1}(s) = s\, G_n(s) + \frac{1-s^2}{m}\, G_n'(s)\,.$$

b) Če skupno rodovno funkcijo slučajnih spremenljivk $X_0, X_1, \ldots$ označimo z G, mora le-ta zadoščati diferencialni enačbi prvega reda:

$$G(s) = s\, G(s) + \frac{1-s^2}{m}\, G'(s)\,.$$

Pri tej enačbi lahko ločimo spremenljivki. Vsaj za $0 < s < 1$ je $G(s) > 0$ in lahko preuredimo v:

$$\frac{G'(s)}{G(s)} = \frac{m}{1+s}\,.$$

Po integraciji dobimo:

$$\ln G(s) = m \ln(1+s) + C\,.$$

Iz dejstva, da je $\lim_{s \uparrow 1} G(s) = 1$, dobimo $C = -m \ln 2$. Razrešimo in dobimo:

$$G(s) = \left(\frac{1+s}{2} \right)^m,$$

kar je rodovna funkcija binomske porazdelitve $\mathrm{Bin}\bigl(m, \frac{1}{2}\bigr)$: slučajne spremenljivke $X_0, X_1, X_2, \ldots$ imajo enako porazdelitev natanko tedaj, ko je $X_0 \sim \mathrm{Bin}\bigl(m, \frac{1}{2}\bigr)$.

6. a) Verjetnost, da ima permutacija n elementov natanko j negibnih točk, dobimo tako, da izberemo j negibnih točk, ostalih $n-j$ elementov pa mora biti permutiranih tako, da ni nobene negibne točke. Sledi:

$$\mathbb{P}(X_n = j) = \frac{1}{j!}\, \mathbb{P}(X_{n-j} = 0)\,, \qquad \mathbb{P}(X_{n+1} = j+1) = \frac{1}{(j+1)!}\, \mathbb{P}(X_{n-j} = 0)\,, \quad (*)$$

torej:

$$\frac{\mathbb{P}(X_n = j)}{\mathbb{P}(X_{n+1} = j+1)} = j + 1 \,.$$

b) Prejšnjo enakost prepišimo v obliki:

$$\mathbb{P}(X_n = j) = (j+1)\,\mathbb{P}(X_{n+1} = j+1) \,.$$

Pomnožimo z s^j in seštejemo po $j = 0, 1, \ldots, n$. Dobimo:

$$\begin{aligned} G_n(s) &= \sum_{j=0}^{n} \mathbb{P}(X_n = j)\, s^j \\ &= \sum_{j=0}^{n} (j+1)\,\mathbb{P}(X_{n+1} = j+1)\, s^j \\ &= \sum_{k=1}^{n+1} k\,\mathbb{P}(X_{n+1} = k)\, s^{k-1} \\ &= \sum_{k=0}^{n+1} k\,\mathbb{P}(X_{n+1} = k)\, s^{k-1} \\ &= G'_{n+1}(s) \,. \end{aligned}$$

Glede na to, da je treba rodovne funkcije izraziti s spremenljivko $s - 1$, je ugodno označiti $G_n(s) = H_n(s-1)$. Velja $H_n = H'_{n+1}$.

Permutacija enega elementa ima vedno natanko eno negibno točko, torej je $G_1(s) = s$, se pravi $H_1(t) = 1 + t$. Nadaljnje funkcije dobimo z integracijo. Ob upoštevanju dejstva, da je $H_n(0) = G_n(1) = 1$, izračunamo:

$$H_2(t) = 1 + t + \frac{t^2}{2} \,, \qquad H_3(t) = 1 + t + \frac{t^2}{2} + \frac{t^3}{6} \,,$$

nakar z indukcijo dokažemo, da je:

$$H_n(t) = \sum_{r=0}^{n} \frac{t^r}{r!} \qquad \text{oziroma} \qquad G_n(s) = \sum_{r=0}^{n} \frac{(s-1)^r}{r!} \,.$$

c) V polinomu iz prejšnje točke moramo najti koeficiente pri potenci s^k. Za izraz $\frac{(s-1)^r}{r!}$ je le-ta enak $\frac{(-1)^{r-k}}{r!}\binom{r}{k} = \frac{(-1)^{r-k}}{k!\,(r-k)!}$, če je $r \ge k$, sicer pa je enak nič. Sledi:

$$\mathbb{P}(X_n = k) = \sum_{r=k}^{n} \frac{(-1)^{r-k}}{k!\,(r-k)!} \,.$$

Opomba. To točko bi lahko rešili tudi neposredno, tako da bi podobno kot v 18. nalogi iz 2. razdelka s pomočjo načela vključitev in izključitev izračunali $\mathbb{P}(X_n = 0)$, nato pa uporabili prvo zvezo v formuli $(*)$.

7. Če z Z označimo vsoto, je rodovna funkcija te slučajne spremenljivke enaka:

$$G_Z(s) = \frac{s}{2-s} \cdot \frac{s}{3-2s} = \frac{2s^2}{3-2s} - \frac{s^2}{2-s} =$$
$$= \frac{2s^2}{3}\left[1 + \frac{2s}{3} + \left(\frac{2s}{3}\right)^2 + \cdots\right] -$$
$$- \frac{s^2}{2}\left[1 + \frac{s}{2} + \left(\frac{s}{2}\right)^2 + \cdots\right],$$

torej je:

$$\mathbb{P}(Z = k) = \left(\frac{2}{3}\right)^{k-1} - \frac{1}{2^{k-1}}\,; \quad k = 2, 3, 4, \ldots$$

Nalogo bi seveda lahko rešili tudi neposredno, brez uporabe rodovnih funkcij.

8. Rodovna funkcija slučajne spremenljivke X je:

$$G_X(s) = \frac{9}{16} + \frac{9}{32}\,s + \sum_{k=2}^{\infty} \frac{5}{2^{k+4}}\,s^k = \frac{1}{4} + \frac{1}{8}\,s + \frac{5}{16}\sum_{k=0}^{\infty}\frac{s^k}{2^k} = \frac{1}{4} + \frac{1}{8}\,s + \frac{5}{16\left(1 - \frac{s}{2}\right)} =$$
$$= \frac{9 - s^2}{8(2-s)},$$

rodovna funkcija slučajne spremenljivke Y pa je:

$$G_Y(s) = \frac{8}{9}\sum_{k=1}^{\infty}\frac{s^k}{9^{k-1}} = \frac{8s}{9}\sum_{l=0}^{\infty}\frac{s^l}{9^l} = \frac{8s}{9\left(1 - \frac{s}{9}\right)} = \frac{8s}{9-s}.$$

Rodovna funkcija slučajne spremenljivke $2Y$ pa je enaka:

$$G_{2Y}(s) = G_Y(2s) = \frac{8s^2}{9-s^2}.$$

Rodovna funkcija vsote $Z = X + 2Y$ je torej:

$$G_Z(s) = G_X(s)\,G_{2Y}(s) = \frac{s^2}{2-s} = \frac{s^2}{2\left(1-\frac{s}{2}\right)} = s^2\sum_{k=0}^{\infty}\frac{s^k}{2^{k+1}} = \sum_{l=2}^{\infty}\frac{s^l}{2^{l-1}}.$$

Sledi:

$$\mathbb{P}(Z = l) = 2^{-(l-1)}\,; \quad l = 2, 3, 4, \ldots$$

torej je $Z - 1 \sim \mathrm{Geom}(1/2)$.

9. a) Rodovni funkciji slučajnih spremenljivk X in Z sta:

$$G_X(s) = \frac{4s}{5-s} \quad \text{in} \quad G_Z(s) = \frac{2s}{3-s}.$$

Slučajna spremenljivka Y mora torej imeti rodovno funkcijo:

$$G_Y(s) = \frac{G_Z(s)}{G_X(s)} = \frac{5-s}{6-2s} = \frac{1}{2} + \frac{1}{3-s} = \frac{5}{6} + \sum_{n=1}^{\infty} \frac{s^n}{3^{n+1}}\,,$$

njena porazdelitev pa je podana s predpisom:

$$\mathbb{P}(Y=0) = \frac{5}{6}\,, \qquad \mathbb{P}(Y=n) = \frac{1}{3^{n+1}}\,; \quad n = 1,2,3,\ldots$$

b) Iskane slučajne spremenljivke obstajajo natanko tedaj, ko obstaja porazdelitev z rodovno funkcijo:

$$\begin{aligned} G_Y(s) &= \frac{\frac{bs}{1-(1-b)s}}{\frac{as}{1-(1-a)s}} = \\ &= \frac{b}{a}\,\frac{1-(1-a)s}{1-(1-b)s} = \\ &= \frac{b}{a(1-b)}\left(1 - a + \frac{a-b}{1-(1-b)s}\right) = \\ &= \frac{b}{a} + \frac{b(a-b)}{a}\sum_{n=1}^{\infty}(1-b)^{n-1}s^n\,. \end{aligned}$$

Porazdelitev s tako rodovno funkcijo pa obstaja natanko tedaj, ko je $a \geq b$.

10. Če označimo $S = X + Y$, je:

$$G_S(s) = \sum_{n=0}^{\infty}(1-q)q^n s^n = \frac{1-q}{1-qs}\,.$$

Torej je:

$$G_X(s) = G_Y(s) = \sqrt{1-q}\,(1-qs)^{-1/2} = \sqrt{1-q}\sum_{n=0}^{\infty}\binom{-1/2}{s}(-q)^n s^n\,,$$

kar pomeni, da za $n = 0,1,2,\ldots$ velja:

$$\mathbb{P}(X=n) = \mathbb{P}(Y=n) = \sqrt{1-q}\binom{-1/2}{s}(-q)^n = \sqrt{1-q}\,\frac{(2n-1)!!}{(2n)!!}\,q^n\,,$$

pri čemer se dogovorimo, da je $(-1)!! = 0!! = 1$, saj v tem primeru za vse $m = 0,1,2,\ldots$ velja zveza $(m+2)!! = m!! \cdot (m+2)$.

11. Neposredno ali z uporabo dejstva, da se binomska porazdelitev ujema s porazdelitvijo vsote neodvisnih Bernoullijevih slučajnih spremenljivk, dobimo $G_X(s) = (1-p+ps)^n$. Nadalje iz zveze:

$$\frac{1}{1+X} = \int_0^1 s^X\,\mathrm{d}s$$

dobimo:

$$\mathbb{E}\left(\frac{1}{1+X}\right) = \int_0^1 G_X(s)\,\mathrm{d}s = \frac{1-(1-p)^{n+1}}{(n+1)p}\,.$$

12. Rodovna funkcija:

$$G(s) = \left(\frac{1}{2} + \frac{1}{4}s + \frac{1}{4}s^2\right)^2 = \frac{1}{4} + \frac{1}{4}s + \frac{5}{16}s^2 + \frac{1}{8}s^3 + \frac{1}{16}s^4\,,$$

porazdelitev:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1/4 & 1/4 & 5/16 & 1/8 & 1/16 \end{pmatrix}.$$

13. Rodovno funkcijo števila Nikinih otrok označimo z:

$$G_1(s) = \frac{1}{4} + \frac{1}{4}s + \frac{1}{4}s^2 + \frac{1}{4}s^3\,,$$

rodovno funkcijo števila otrok posameznega Nikinega otroka pa z:

$$G_2(s) = \frac{1}{2} + \frac{1}{4}s + \frac{1}{4}s^2\,,$$

Tedaj je rodovna funkcija števila Nikinih vnukov enaka:

$$G(s) = G_1\big(G_2(s)\big)\,,$$

verjetnost, da Nika ostane brez vnukov, je $G_1(G_2(0)) = 15/32$, pričakovana vrednost pa je:

$$G'(1) = G_1'\big(G_2(1)\big)\, G_2'(1) = \frac{9}{8}\,.$$

14. Označimo število zasedenih miz z X. Pišemo lahko $X = I_1 + I_2 + \ldots + I_N$, kjer je I_n indikator dogodka, da se število n usede za novo mizo. Če je $N = 0$, je seveda $X = 0$. Rodovna funkcija te slučajne spremenljivke je enaka:

$$\begin{aligned} G_X(s) &= \mathbb{E}(s^X) = \\ &= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\, \mathbb{E}\big(s^X \mid N = n\big) = \\ &= \sum_{n=0}^{\infty} p(1-p)^n\, \mathbb{E}\big(s^{I_1+I_2+\cdots+I_n}\big) = \\ &= p \sum_{n=0}^{\infty} \frac{s(s+1)(s+2)\cdots(s+n-1)}{n!}\,(1-p)^n = \\ &= p \sum_{n=0}^{\infty} \binom{-s}{n} (p-1)^n = \\ &= p^{1-s} = \\ &= e^{-(s-1)\ln p}\,. \end{aligned}$$

Sledi $X \sim \mathrm{Pois}(-\ln p)$.

15. *Prvi način.* Če vsoto označimo z Z, velja:

$$G_Z(s) = \frac{b\,\frac{as}{1-(1-a)s}}{1-(1-b)\,\frac{as}{1-(1-a)s}} = \frac{abs}{1-(1-ab)s}\,,$$

torej je $Z \sim \mathrm{Geom}(ab)$.

Drugi način. Gre za isto situacijo kot v 19. nalogi iz 7. razdelka: pogojna porazdelitev vsote $X_1 + X_2 + \ldots + X_N$ glede na $N = n$ se ujema s pogojno porazdelitvijo vsote $X_1 + X_2 + \ldots + X_n$ glede na ta dogodek, ta pa se zaradi neodvisnosti ujema z brezpogojno porazdelitvijo te vsote, ki je negativna binomska $\mathrm{NB}(n, b)$. Rezultat sledi.

16. Naj bo G rodovna funkcija slučajnih spremenljivk X_i. Ker so slučajne spremenljivke X_i porazdeljene geometrijsko, je:

$$G(s) = \frac{\frac{2}{3}s}{1-\frac{1}{3}s} = \frac{2s}{3-s}\,.$$

Rodovna funkcija slučajne spremenljivke N je funkcija:

$$s \mapsto \frac{\frac{3}{4}s}{1-\frac{1}{4}s} = \frac{3s}{4-s}\,.$$

Če s H označimo rodovno funkcijo slučajne spremenljivke $2N$, je torej $H(s) = \dfrac{3s^2}{4-s^2}$. Rodovna funkcija slučajne vsote S je tako funkcija:

$$H(G(s)) = \frac{3\left(\frac{2s}{3-s}\right)^2}{4-\left(\frac{2s}{3-s}\right)^2} = \frac{s^2}{3-2s} = \frac{\frac{1}{3}s^2}{1-\frac{2}{3}s}\,,$$

torej ima Z geometrijsko porazdelitev $\mathrm{Geom}(1/3)$, pomaknjeno za 1 v desno. Z drugimi besedami, $Z - 1 \sim \mathrm{Geom}(1/3)$.

17. Rodovna funkcija slučajnih spremenljivk X_n je enaka:

$$G(s) = \frac{\ln\big(1-(1-p)s\big)}{\ln(p)}$$

(glej 3. nalogo). Rodovna funkcija dane Poissonove porazdelitve pa je enaka $H(s) = e^{a(1-s)\ln p}$ (glej 1. nalogo). Rodovna funkcija dane vsote je torej enaka:

$$G_Z(s) = H(G(s)) = e^{a(\ln p - \ln(1-(1-p)s)} = \left(\frac{p}{1-(1-p)s}\right)^a.$$

Razvijemo v vrsto po Newtonovi formuli in dobimo:

$$G_Z(s) = p^a \sum_{k=0}^{\infty} \binom{-a}{k} (-1)^k (1-p)^k s^k\,.$$

Slučajna spremenljivka Z torej zavzame vrednosti $0, 1, 2, \ldots$ in za k iz te množice velja:

$$\mathbb{P}(Z = k) = \binom{-a}{k}(-1)^k p^a (1-p)^k = \frac{a(a+1)\cdots(a+k-1)p^a(1-p)^k}{k!} .$$

Dobimo torej **Pólyevo**[73] porazdelitev.

18. a) Označimo z G rodovno funkcijo slučajnih spremenljivk $X_1, X_2, \ldots$, s H rodovno funkcijo slučajne spremenljivke N, z G_Z pa rodovno funkcijo slučajne spremenljivke S. Vemo, da je $G_Z(s) = H(G(s))$. S seštetjem ustreznih vrst dobimo

$$H(s) = e^{s-1} \quad \text{in} \quad G_Z(s) = \frac{1}{2-s} .$$

Torej mora veljati

$$e^{G(s)-1} = \frac{1}{2-s} ,$$

od koder dobimo

$$G(s) = 1 - \ln(2-s) = 1 - \ln 2 - \ln\left(1 - \frac{s}{2}\right) = 1 - \ln 2 + \sum_{k=1}^{\infty} \frac{s^k}{k \cdot 2^k} .$$

Torej za $i = 1, 2, 3, \ldots$ velja

$$\mathbb{P}(X_i = 0) = 1 - \ln 2 , \qquad \mathbb{P}(X_i = k) = \frac{1}{k \cdot 2^k} ; \quad k = 1, 2, 3, \ldots$$

b) Zdaj mora veljati

$$e^{\lambda(G(s)-1)} = \frac{1}{2-s} ,$$

od koder dobimo

$$G(s) = 1 - \frac{\ln(2-s)}{\lambda} = 1 - \frac{\ln 2}{\lambda} + \frac{1}{\lambda}\sum_{k=1}^{\infty} \frac{s^k}{k \cdot 2^k} .$$

Očitno je $G(1) = 1$, vsi koeficienti pa bodo nenegativni natanko tedaj, ko bo $\lambda \geq \ln 2$: to je potreben in zadosten pogoj za obstoj zahtevanih slučajnih spremenljivk.

19. Označimo z $G(s) = \frac{1}{6} + \frac{1}{3}s + \frac{1}{2}s^2$ rodovno funkcijo števila otrok. Verjetnost, da Maks v n-ti generaciji ne bo imel potomcev, je enaka $G_n(0)$, kjer je:

$$G_n(s) = G(G(G(\ldots G(s) \ldots))) \qquad (n \text{ znakov } G).$$

Verjetnost, da bo Maksovo potomstvo izumrlo, pa je enaka $p = \lim_{n\to\infty} G_n(0)$. Velja $G(p) = p$, kar ima rešitvi $p = 1/3$ in $p = 1$. Z indukcijo po n dokažemo, da je $G_n(0) < 1/3$ za vse n, torej je $p = 1/3$.

[73] György Pólya (1887–1985), madžarski matematik judovskega rodu

20. a) Naj bo še $\tilde{Z} := Z - 1$ število vseh potomcev (neposrednih ali posrednih) začetnega predstavnika tega procesa, $\tilde{G}$ pa rodovna funkcija te slučajne spremenljivke. Očitno je:

$$G(s) = \mathbb{E}\big[s^Z\big] = \mathbb{E}\big[s^{Z'+1}\big] = s\tilde{G}(s)\,.$$

Slučajna spremenljivka $\tilde{Z}$ pa je porazdeljena enako kot vsota slučajnega števila neodvisnih kopij slučajnih spremenljivk Z, pri čemer je število seštevancev neodvisno od seštevancev samih in porazdeljeno enako kot Z_1. Torej je:

$$\tilde{G}(s) = G_1\big(G(s)\big)\,.$$

Združimo in dobimo:

$$G(s) = s\,G_1\big(G(s)\big)\,.$$

b) Velja $G_1(s) = \frac{1}{2}(1 + s^2)$ in enačba $G_1(s) = s$ ima edino (dvojno) rešitev $s = 1$. To je torej tudi verjetnost, da proces izumre. Vstavimo v zvezo iz prejšnje točke in dobimo:

$$G(s) = \frac{s}{2}\left(1 + \big(G(s)\big)^2\right).$$

Po ureditvi pride:

$$s\big(G(s)\big) - 2\,G(s) + s = 0\,,$$

kar ima rešitvi:

$$\begin{aligned} G(s) &= \frac{1 \pm \sqrt{1 - s^2}}{s} = \\ &= \frac{1}{s}\left(1 \pm \left[1 + \binom{1/2}{1}(-s^2) + \binom{1/2}{2}(-s^2)^2 + \cdots\right]\right). \end{aligned}$$

Pravilen bo tisti predznak, pri katerem se bosta člena z $1/s$ odštela, to pa je minus. Torej bo:

$$G(s) = -\sum_{k=1}^{\infty} \binom{1/2}{k} (-1)^k s^{2k-1}\,.$$

Dobljena rodovna funkcija je skladna z dejstvom, da ima proces vselej liho število predstavnikov. Za $k = 1, 2, 3, \ldots$ izračunamo:

$$\begin{aligned} \binom{1/2}{k} &= (-1)^{k-1}\,\frac{1 \cdot 1 \cdot 3 \cdot \cdots \cdot (2k-3)}{2^k\,k!} = \\ &= (-1)^{k-1}\,\frac{(2k-3)!!}{(2k)!!} = \\ &= \frac{(-1)^{k-1}}{k\,2^{2k-1}}\binom{2k-2}{k-1}\,, \end{aligned}$$

pri čemer se dogovorimo, da je $(-1)!! = 1$: glej 10. nalogo. Zadnja oblika pa velja ob dogovoru, da je $\binom{0}{0} = 1$. Poberemo skupaj in dobimo, da za $k = 1, 2, 3, \ldots$ velja:

$$\mathbb{P}(Z = 2k - 1) = \frac{(2k-3)!!}{(2k)!!}\,.$$

21. a) Odigrane igre tvorijo proces razvejanja, v katerem ima lahko vsaka igra 0 potomcev z verjetnostjo $1-p$ in m potomcev z verjetnostjo p. Ker je pričakovano število potomcev enako $mp < 1$, ta proces izumre z verjetnostjo 1, tako da bo igralec skoraj gotovo odigral samo končno mnogo iger.

b) Označimo z G rodovno funkcijo števila odigranih iger. Na enak način kot v prejšnji nalogi dobimo, da ta funkcija zadošča zvezi:

$$G(s) = (1-p)s + ps\big(G(s)\big)^m .$$

Zvezo odvajamo in dobimo:

$$G'(s) = 1 - p + p\big(G(s)\big)^m + mps\big(G(s)\big)^{m-1}\, G'(s)\,. \tag{$*$}$$

To velja vsaj za $|s| < 1$. Ker je $mp < 1$, smemo za vse te s enačbo obrniti in dobimo:

$$G'(s) = \frac{1 - p + p\big(G(s)\big)^m}{1 - mps\big(G(s)\big)^{m-1}}$$

Pošljemo $s \uparrow 1$, upoštevamo, da je $\lim_{s\uparrow 1} G(s) = 1$, in dobimo, da pričakovana vrednost obstaja in je enaka:

$$\mathbb{E}(Z) = \frac{1}{1-mp}\,.$$

Po odvajanju še zveze $(*)$ dobimo, da za $|s| < 1$ velja:

$$\begin{aligned} G''(s) = 2mp\big(G(s)\big)^{m-1}\, G'(s) + m(m-1)ps\big(G(s)\big)^{m-2}\big(G'(s)\big)^2 + \\ + mps\big(G(s)\big)^{m-1}\, G''(s)\,. \end{aligned}$$

Spet, ker je $mp < 1$, smemo enačbo za vse $|s| < 1$ obrniti in dobimo:

$$G''(s) = \frac{2mp\big(G(s)\big)^{m-1}\, G'(s) + m(m-1)ps\big(G(s)\big)^{m-2}\big(G'(s)\big)^2}{1 - mps\big(G(s)\big)^{m-1}}\,.$$

Spet pošljemo $s \uparrow 1$, upoštevamo, da je $\lim_{s\uparrow 1} G(s) = 1$, in dobimo, da drugi faktorski moment obstaja in je enak:

$$\begin{aligned} \mathbb{E}\big[Z(Z-1)\big] &= \frac{2mp\,\mathbb{E}(Z) + m(m-1)p\big(\mathbb{E}(Z)\big)^2}{1-mp} = \\ &= \frac{2mp(1-mp) + m(m-1)p}{(1-mp)^3} = \\ &= \frac{mp - 2m^2p^2 + m^2p}{(1-mp)^3}\,. \end{aligned}$$

Sledi, da obstaja tudi varianca in je enaka:

$$\begin{aligned} \operatorname{var}(Z) &= \mathbb{E}\big[Z(Z-1)\big] + \mathbb{E}(Z) - \big(\mathbb{E}(Z)\big)^2 = \\ &= \frac{mp - 2m^2p^2 + m^2p}{(1-mp)^3} - \frac{mp}{(1-mp)^2} = \\ &= \frac{m^2p(1-p)}{(1-mp)^3}\,. \end{aligned}$$

22. a) Iz lastnosti Markova sledi, da so slučajne spremenljivke $T_1, T_2 - T_1, T_3 - T_2, \ldots$ neodvisne in porazdeljene tako kot T_1. Sledi $F_r(s) = \bigl(F_1(s)\bigr)^r$.

b) Velja $\mathbb{P}(T_1 = 0 \mid p) = 0$, $\mathbb{P}(T_1 = 1 \mid p) = p$, za $n = 2, 3, 4, \ldots$ pa velja:

$$\mathbb{P}(T_1 = n \mid p) = (1-p)\,\mathbb{P}(T_2 = n-1 \mid p)\,.$$

Sledi:

$$\begin{aligned} F_1(s \mid p) &= \mathbb{P}(T_1 = 1 \mid p)\, s + \sum_{n=2}^{\infty} \mathbb{P}(T_1 = n \mid p)\, s^n = \\ &= ps + (1-p) \sum_{n=2}^{\infty} \mathbb{P}(T_2 = n-1 \mid p)\, s^n = \\ &= ps + (1-p) \sum_{k=1}^{\infty} \mathbb{P}(T_2 = k \mid p)\, s^{k+1} = \\ &= ps + (1-p)s\, F_2(s \mid p) = \\ &= ps + (1-p)s\, \bigl(F_1(s \mid p)\bigr)^2 . \end{aligned}$$

Rešimo kvadratno enačbo, izberemo ustrezni predznak in dobimo:

$$F_1(s \mid p) = \frac{1 - \sqrt{1 - 4p(1-p)s^2}}{2(1-p)s} = \sum_{k=1}^{\infty} (-1)^{k-1} 4^k p^k (1-p)^{k-1} \binom{\frac{1}{2}}{k} s^{2k-1} ,$$

torej za $n = 2k-1$, $k = 1, 2, 3, \ldots$ velja:

$$\mathbb{P}(T_1 = n \mid p) = (-1)^{k-1} 2^{2k-1} p^k (1-p)^{k-1} \binom{\frac{1}{2}}{k} = \frac{p^k (1-p)^{k-1}}{2(2k-1)} \binom{2k}{k} ,$$

medtem ko za vsa ostala števila n velja $\mathbb{P}(T_1 = n \mid p) = 0$. Nadalje je:

$$\mathbb{P}(T_1 < \infty \mid p) = F_1(1 \mid p) = \frac{1 - |1 - 2p|}{2(1-p)} = \min\left\{1,\ \frac{p}{1-p}\right\} .$$

Za čas T_0 pa je po definiciji $\mathbb{P}(T_0 = 0 \mid p) = 0$, medtem ko za $n = 1, 2, 3, \ldots$ velja:

$$\begin{aligned} \mathbb{P}(T_0 = n \mid p) &= (1-p)\,\mathbb{P}(T_1 = n-1 \mid p) + p\,\mathbb{P}(T_{-1} = n-1 \mid p) = \\ &= (1-p)\,\mathbb{P}(T_1 = n-1 \mid p) + p\,\mathbb{P}(T_1 = n-1 \mid 1-p)\,. \end{aligned}$$

Sledi:

$$\begin{aligned} F(s \mid p) &= (1-p)s\, F_1(s \mid p) + ps\, F_1(s \mid 1-p) = \\ &= 1 - \sqrt{1 - 4p(1-p)s^2} = \\ &= \sum_{k=1}^{\infty} (-1)^{k-1} 4^k p^k (1-p)^k \binom{\frac{1}{2}}{k} s^{2k} , \end{aligned}$$

torej za $n = 2k$, $k = 1, 2, 3, \ldots$, velja:

$$\mathbb{P}(T_0 = n \mid p) = (-1)^{k-1} 4^k p^k (1-p)^k \binom{\frac{1}{2}}{k} = \frac{p^k (1-p)^k}{2k-1} \binom{2k}{k},$$

medtem ko za vsa ostala števila n velja $\mathbb{P}(T_0 = n \mid p) = 0$. Nadalje je:

$$\mathbb{P}(T_0 < \infty \mid p) = F(1 \mid p) = 1 - |1 - 2p| = 2 \min\{p, 1-p\}.$$

c) Odvajamo:

$$F'(s \mid p) = \frac{4p(1-p)s}{\sqrt{1 - 4p(1-p)s^2}}, \qquad F_1'(s \mid p) = \frac{1 - \sqrt{1 - 4p(1-p)s^2}}{2(1-p)^2 s^2 \sqrt{1 - 4p(1-p)s^2}}$$

in vstavimo $s = 1$. Dobimo:

$$F'(1 \mid p) = \frac{4p(1-p)}{|1-2p|}, \qquad F_1'(1 \mid p) = \frac{1 - |1-2p|}{2(1-p)^2 |1-2p|},$$

iskani pogojni pričakovani vrednosti pa sta enaki:

$$\begin{aligned}
\mathbb{E}(T_0 \mid T_0 < \infty, p) &= \frac{F'(1 \mid p)}{F'(1 \mid p)} = \frac{4p(1-p)}{|1-2p|\big(1 - |1-2p|\big)}, \\
\mathbb{E}(T_1 \mid T_1 < \infty, p) &= \frac{F_1'(1 \mid p)}{F_1(1 \mid p)} = \frac{1}{(1-p)|1-2p|}.
\end{aligned}$$

Iz $1 - |1-2p| = 2\min\{p, 1-p\}$ in $p(1-p) = \min\{p, 1-p\}\big(1 - \min\{p, 1-p\}\big)$ sledi še druga oblika za T_0:

$$\mathbb{E}(T_0 \mid T_0 < \infty, p) = 1 + \frac{1}{1 - 2\min\{p, 1-p\}}.$$

Za $p = \frac{1}{2}$ pogojni pričakovani vrednosti ne obstajata.

23. $M_X(t) = \begin{cases} \dfrac{e^{bt} - e^{at}}{t(b-a)} & ; t \neq 0 \\ 1 & ; t = 0 \end{cases}.$
Definirana je za vse t.

24. Momentno-rodovna funkcija je:

$$M_X(t) = e^{\lambda(e^t - 1)}$$

in je definirana je za vse t. Momente lahko dobimo na vsaj dva načina.

Prvi način: iz razvoja momentno-rodovne funcije:

$$
\begin{aligned}
M_X(t) &= \exp\left[\lambda\left(t + \frac{t^2}{2} + \frac{t^3}{3} + \frac{t^4}{24} + \cdots\right)\right] = \\
&= 1 + \lambda\left(t + \frac{t^2}{2} + \frac{t^3}{6} + \frac{t^4}{24} + \cdots\right) + \frac{\lambda^2}{2}\left(t + \frac{t^2}{2} + \frac{t^3}{6} + \cdots\right)^2 + \\
&\qquad + \frac{\lambda^3}{6}\left(t + \frac{t^2}{2} + \cdots\right)^3 + \frac{\lambda^4}{24}\Big(t + \cdots\Big)^4 + \cdots = \\
&= 1 + \lambda\left(t + \frac{t^2}{2} + \frac{t^3}{6} + \frac{t^4}{24} + \cdots\right) + \frac{\lambda^2}{2}\left(t^2 + t^3 + \frac{7}{12}t^4 + \cdots\right) + \\
&\qquad + \frac{\lambda^3}{6}\left(t^3 + \frac{3}{2}t^4 + \cdots\right) + \frac{\lambda^4}{24}\Big(t^4 + \cdots\Big) + \cdots = \\
&= 1 + \lambda t + \left(\frac{\lambda}{2} + \frac{\lambda^2}{2}\right)t^2 + \left(\frac{\lambda}{6} + \frac{\lambda^2}{2} + \frac{\lambda^3}{6}\right)t^3 + \\
&\qquad + \left(\frac{\lambda}{24} + \frac{7\lambda^2}{24} + \frac{\lambda^3}{4} + \frac{\lambda^4}{24}\right)t^4 + \cdots
\end{aligned}
$$

Drugi način: iz faktorskih momentov, izračunanih v 1. nalogi. Le-te izrazimo z običajnimi:

$$
\begin{aligned}
f_1 &= \mathbb{E}(X) = z_1 = \lambda\,, \\
f_2 &= \mathbb{E}(X^2 - X) = z_2 - z_1 = \lambda^2\,, \\
f_3 &= \mathbb{E}(X^3 - 3X^2 + 2X) = z_3 - 3z_2 + 2z_1 = \lambda^3\,, \\
f_4 &= \mathbb{E}(X^4 - 6X^3 + 11X^2 - 6X) = z_4 - 6z_3 + 11z_2 - 6z_1 = \lambda^4\,.
\end{aligned}
$$

Iz enega ali drugega načina dobimo, da so prvi štirje momenti enaki:

$$
\begin{aligned}
z_1 &= \lambda\,, \\
z_2 &= \lambda^2 + \lambda\,, \\
z_3 &= \lambda^3 + 3\lambda^2 + \lambda\,, \\
z_4 &= \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda\,.
\end{aligned}
$$

Opomba. V splošnem je $z_r = \sum_{k=1}^r S(r,k)\lambda^k$, kjer je $S(r,k)$ število vseh možnih razbitij r-elementne množice z označenimi elementi na k neoznačenih nepraznih podmnožic. Števila $S(r,k)$ se imenujejo *Stirlingova števila druge vrste.*

25. $z_r = \dfrac{r!}{\lambda^r}$,

$M_X(t) = \dfrac{\lambda}{\lambda - t}$, definirana je za $t < \lambda$
(ustrezna vrsta pa konvergira le za $-\lambda < t < \lambda$).

$m_3 = \dfrac{2}{\lambda^3}$.

26. Za $Z \sim \mathrm{N}(0,1)$ velja:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - t^2/2} \, \mathrm{d}t \,.$$

S substitucijo $u = t - x$ dobimo:

$$M_Z(t) = \frac{e^{u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2} \, \mathrm{d}u = e^{t^2/2} = 1 + \frac{t^2}{2} + \frac{t^4}{2^2 \cdot 2!} + \frac{t^6}{2^3 \cdot 3!} + \cdots \,.$$

Vsi momenti lihih redov so torej enaki 0. Če je n sod, pa je:

$$z_n = \frac{n!}{2^{n/2} \cdot (n/2)!} = 1 \cdot 3 \cdot \cdots \cdot (n-1) = (n-1)!! \,.$$

27. Za $X \sim \mathrm{N}(\mu, \sigma)$ lahko zapišemo $X = \sigma Z + \mu$, kjer je $Z \sim \mathrm{N}(0,1)$. Sledi:

$$M_X(t) = \mathbb{E}\big[e^{t(\sigma X + \mu)}\big] = e^{\mu t} \, M_Z(\sigma t) = e^{\sigma^2 t^2/2 + \mu t} \,.$$

Od tod in iz izreka o enoličnosti sledi, da, če sta $X_1 \sim \mathrm{N}(\mu_1, \sigma_1)$ in $X_2 \sim \mathrm{N}(\mu_2, \sigma_2)$ neodvisni slučajni spremenljivki, velja $X_1 + X_2 \sim \mathrm{N}\big(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\big)$.

28. Če je $X \sim \mathrm{Gama}(a, \lambda)$, velja $M_X(t) = \left(\dfrac{\lambda}{\lambda - t}\right)^a$; momentno-rodovna funkcija je definirana za $t < \lambda$ (v kompleksnem pa za $\operatorname{Re} t < \lambda$).

Od tod in iz izreka o enoličnosti sledi, da, če sta $X_1 \sim \mathrm{Gama}(a_1, \lambda)$ in $X_2 \sim \mathrm{Gama}(a_2, \lambda)$ neodvisni slučajni spremenljivki, velja $X_1 + X_2 \sim \mathrm{Gama}(a_1 + a_2, \lambda)$.

29. $\kappa_1 = \mu, \quad \kappa_2 = \sigma^2, \quad \kappa_r = 0$ za vse $r \ge 3$.

31. $\kappa_2 = m_2 = \sigma^2, \quad \kappa_3 = m_3, \quad \kappa_4 = m_4 - 3m_2^2, \quad A = m_3/\sigma^3, \quad K = m_4/\sigma^4 - 3.$

32. $\kappa_1 = \dfrac{a+b}{2}, \quad \kappa_2 = \dfrac{(b-a)^2}{12}, \quad \kappa_3 = 0, \quad \kappa_4 = -\dfrac{(b-a)^4}{120}, \quad A = 0, \quad K = -\dfrac{6}{5}.$

33. $m_2 = m_3 = \lambda, \quad m_4 = \lambda + 3\lambda^2, \quad A = 1/\sqrt{n}, \quad K = 3 + 1/n.$ **Opomba**. Asimetrija in presežna sploščenost gresta proti nič, ko gre λ proti neskončno. To bi lahko nakazovalo, da je Poissonova porazdelitev čedalje bližje normalni, ko gre λ proti neskončno. Dejansko je to tudi res, če smiselno definiramo, kdaj se ena porazdelitev bliža drugi, kar preciziramo kot konvergenco porazdelitev; ustrezna konvergenca Poissonove porazdelitve proti normalni sledi iz centralnega limitnega izreka.

34. Spomnimo se, da je $\mathbb{E}(S) = 1600 \cdot 0{\cdot}1 = 160$. Neposredno po neenačbi Markova lahko ocenimo $\mathbb{P}(S > 250) = \mathbb{P}(S \ge 251) \le \dfrac{160}{251} \doteq 0{\cdot}638$ (zaokroženo navzgor).

Za oceno po neenačbi Čebiševa potrebujemo $\mathrm{var}(S) = 1600 \cdot 0{\cdot}1 \cdot 0{\cdot}9 = 144$. Ocenimo:

$$\mathbb{P}(S > 250) = \mathbb{P}(S - 160 \ge 91) \le \mathbb{P}(|S - 160| \ge 91) \le \frac{144}{91^2} \doteq 0{\cdot}0174$$

(spet zaokroženo navzgor).

Končno ocenimo še z uporabo momentno-rodovne funkcije. Za $S \sim \mathrm{Bin}(n, p)$ iz pravila za vsote neodvisnih slučajnih spremenljivk dobimo
$M_S(t) = \left(1 - p + p\, e^t\right)^n$. Če je $x < n$, je izraz $e^{-tx}\, M(t)$ najmanjši pri:

$$t = \ln \frac{(1-p)x}{p(n-x)} \,.$$

Za $n = 1600$ in $p = 0{,}1$ dobimo $t \doteq 0{,}5156$ in $\mathbb{P}(S > 250) = \mathbb{P}(S \ge 251) \le 1{,}45 \cdot 10^{-11}$.
Če bi šli po Laplaceovi integralski formuli in upoštevali $\frac{1}{2} - \Phi(x) \approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$, bi dobili naslednjo aproksimacijo, ki ni ocena navzgor:

$$\mathbb{P}(S > 250) \approx \frac{1}{2} - \Phi\left(\frac{90{,}5}{12}\right) \approx \frac{12}{90{,}5\sqrt{2\pi}}\, e^{-90.5^2/288} \doteq 2{,}36 \cdot 10^{-14} \,.$$

Točen rezultat: $9{,}705 \cdot 10^{-13}$.

35. Najprej izpeljimo oceni za splošni primer. Naj bo $x > 0$. Iz simetrije in neenačbe Čebiševa dobimo:

$$\mathbb{P}(S_n > x) = \frac{1}{2}\,\mathbb{P}(|S_n| > x) \le \frac{\mathrm{var}(S_n)}{2x^2} = \frac{n}{x^2} \,,$$

Momentno-rodovna funkcija pa je enaka:

$$M_{S_n}(t) = \frac{1}{1 - t^2}$$

in je definirana za $-1 < t < 1$, izraz $e^{-tx}\, M_{S_n}(t)$ pa je minimalen pri

$$t = \frac{\sqrt{n^2 + x^2} - n}{x} \,.$$

Za $\mathbb{P}(S_5 > 5)$ iz neenačbe Čebiševa dobimo zgornjo mejo $0{,}2$, iz momentno-rodovne funkcije pa (za $t = \sqrt{2} - 1$) $0{,}323$. Ocena iz neenačbe Čebiševa je torej tu ostrejša od ocene iz momentno-rodovnih funkcij. Točen rezultat pa je $0{,}0555$.
Za $\mathbb{P}(S_{20} > 20)$ iz neenačbe Čebiševa dobimo zgornjo mejo $0{,}05$, iz momentno-rodovne funkcije pa (prav tako za $t = \sqrt{2} - 1$) $0{,}0109$. Zdaj je torej ocena iz momentno-rodovnih funkcij ostrejša od ocene iz neenačbe Čebiševa. Točen rezultat pa je $0{,}00114$.

36. $\phi_X(t) = \dfrac{p\, e^{it}}{1 - (1-p)e^{it}}$.

37. $\phi_X(t) = \dfrac{\lambda}{\lambda - it}$.

38. $\phi_X(t) = \dfrac{1}{1 + t^2}$.

39. Če je $X \sim \mathrm{Bin}(n, p)$, je $\phi_X(t) = \left(1 - p + p\, e^{it}\right)^n$.

40. Če je $X \sim \mathrm{NB}(n, p)$, je $\phi_X(t) = \left(\dfrac{p\, e^{it}}{1 - (1-p)e^{it}}\right)^n$.

41. Če je $X \sim \mathrm{Gama}(n, \lambda)$, je $\phi_X(t) = \left(\dfrac{\lambda}{\lambda - it}\right)^n$.

42. $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1/6 & 1/3 & 1/6 & 1/3 \end{pmatrix}$.

43. $f_X(x) = \begin{cases} \dfrac{1 - \cos x}{\pi x^2} & ;\ |x| \neq 0 \\ \dfrac{1}{2\pi} & ;\ x = 0 \end{cases}$

44. Slučajna spremenljivka X ima *Cauchyjevo porazdelitev*, tj.:

$$f_X(x) = \frac{1}{\pi(1 + x^2)}$$

45. Slučajna spremenljivka $\bar{X}$ ima prav tako Cauchyjevo porazdelitev.

9. Limitni izreki

1. Označimo z ℓ_n število vseh zaporednih cifer od n-tega meta nazaj (če torej v n-tem metu pade grb, je $\ell_n = 0$). Tedaj za vsak $\varepsilon > 0$ velja:

$$\mathbb{P}(\ell_n \geq (1+\varepsilon)\log_2 n) \leq n^{-(1+\varepsilon)}\,.$$

Ker vrsta iz izrazov na desni konvergira, se po prvi Borel–Cantellijevi lemi skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\ell_n \geq (1+\varepsilon)\log_2 n$. A če je to res velja tudi $\limsup_{n\to\infty} L_n/\log_2 n \leq 1+\varepsilon$. Ker to za vsak ε skoraj gotovo velja, skoraj gotovo velja tudi:

$$\limsup_{n\to\infty} \frac{L_n}{\log_2 n} \leq 1\,.$$

Za oceno limite v nasprotno smer razdelimo mete na bloke dolžine $\lfloor(1-\varepsilon)\log_2 n\rfloor+1$. Na vsakem od teh blokov je verjetnost, da bodo padle same cifre, enaka $2^{-(\lfloor(1-\varepsilon)\log_2 n\rfloor+1)} \geq n^{-(1-\varepsilon)}/2$. Če je $L_n \leq (1-\varepsilon)\log_2 n$, na vseh blokih velja, da ne padejo same cifre, torej je:

$$\begin{aligned}\mathbb{P}(L_n \leq (1-\varepsilon)\log_2 n) &\leq \left(1 - \frac{n^{-(1-\varepsilon)}}{2}\right)^{\lfloor n/(\lfloor(1-\varepsilon)\log_2 n\rfloor+1)\rfloor} \leq \\ &\leq \exp\left\{-\frac{n^{-(1-\varepsilon)}}{2}\left(\frac{n}{\lfloor(1-\varepsilon)\log_2 n\rfloor+1} - 1\right)\right\}\,.\end{aligned}$$

Za dovolj velike n velja tudi:

$$\mathbb{P}(L_n \leq (1-\varepsilon)\log_2 n) \leq \exp\left(-\frac{n^\varepsilon}{2\log_2 n}\right)\,.$$

Ker vrsta iz izrazov na desni spet konvergira, se skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\{L_n \leq (1-\varepsilon)\log_2 n\}$, a če je to res, je tudi $\liminf_{n\to\infty} L_n/\log_2 n \geq 1-\varepsilon$. Ker to za vsak $\varepsilon > 0$ skoraj gotovo velja, skoraj gotovo velja tudi:

$$\liminf_{n\to\infty} \frac{L_n}{\log_2 n} \geq 1\,.$$

Iz obeh ocen pa že sledi želeni rezultat.

2. Zaporedje konvergira skoraj gotovo in posledično tudi v verjetnosti in šibko.

Konvergenca v verjetnosti sledi iz neenačbe Čebiševa: iz $\mathbb{E}(S_n) = 0$ in $\mathrm{var}(S_n/n) = 1/n$ sledi:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0\,.$$

Za dokaz skoraj gotove konvergence pa se moramo malo bolj potruditi. Spomnimo se na momentno-rodovne funkcije in uporabo neenačbo Markova za eksponentno funkcijo. Iz:

$$\mathbb{E}\big[e^{tS_n}\big] = \left(\frac{e^t + e^{-t}}{2}\right)^n$$

izpeljemo:

$$\mathbb{P}\left(\frac{S_n}{n} \ge \varepsilon\right) = \mathbb{P}(S_n \ge n\varepsilon) \le e^{-nt\varepsilon}\left(\frac{e^t + e^{-t}}{2}\right)^n = \left(e^{-t\varepsilon}\,\frac{e^t + e^{-t}}{2}\right)^n .$$

Za vsak $\varepsilon > 0$ obstaja tak t, da je $e^{-t\varepsilon}\,\frac{e^t+e^{-t}}{2} < 1$. V tem primeru vrsta iz izrazov na desni konvergira, torej se po prvi Borel–Cantellijevi lemi skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\{S_n/n \ge \varepsilon\}$. Zaradi simetrije to velja tudi za dogodke $\{S_n/n \le -\varepsilon\}$, torej velja tudi za dogodke $\{|S_n/n| \ge \varepsilon\}$. Torej skoraj gotovo velja, da se za vsak $m \in \mathbb{N}$ zgodi kvečjemu končno mnogo dogodkov $\{|S_n/n| \ge 1/m\}$. A če je to res, zaporedje S_n/n konvergira proti nič. Sklep: zaporedje slučajnih spremenljivk S_n/n skoraj gotovo konvergira proti nič.

3. Velja $\mathbb{E}(X_n) = 0$ in $\operatorname{var}(X_n) = n/\ln n$, od koder sledi:

$$\mathbb{E}(S_n = 0)\,, \qquad \operatorname{var}(S_n) = \sum_{k=2}^{n} \frac{k}{\ln k} \le \frac{n^2}{\ln n}$$

(slednja ocena velja za $n \ge 3$). Po neenačbi Čebiševa dobimo:

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \ge \varepsilon\right) = \mathbb{P}(|S_n| \ge n\varepsilon) \le \frac{1}{\varepsilon^2 \ln n}\,,$$

kar gre proti nič, ko gre n proti neskončno, zato zaporedje šibko konvergira proti nič.

V nadaljevanju bomo pokazali, da zaporedje S_n/n skoraj gotovo ni Cauchyjevo, ker skoraj gotovo za neskončno mnogo indeksov n velja $\left|\frac{S_{n+1}}{n+1} - \frac{S_n}{n}\right| \ge \frac{1}{2}$. Iz zapisa:

$$\left|\frac{S_{n+1}}{n+1} - \frac{S_n}{n}\right| = \frac{X_{n+1}}{n+1} - \frac{S_n}{n(n+1)}$$

namreč vidimo, da se to zagotovo zgodi, če je $|X_{n+1}| = n+1$ in $|S_n| \le n(n+1)/2$. Po neenačbi Čebiševa ocenimo $\mathbb{P}\big(|S_n| \le \frac{n(n+1)}{2}\big) \le \frac{4}{(n+1)^2 \ln n}$ in iz prve Borel–Cantellijeve leme sledi, da se skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\{|S_n| \le n(n+1)/2\}$. Nadalje je $\mathbb{P}(|X_{n+1}| = n+1) = 1/((n+1)\ln(n+1))$ in ker vrsta iz desnih strani divergira, se po drugi Borel–Cantellijevi lemi skoraj gotovo zgodi neskončno mnogo dogodkov $\{|X_{n+1}| = n+1\}$. Če oboje združimo, dobimo, da se skoraj gotovo zgodi neskončno mnogo dogodkov $\left|\frac{S_{n+1}}{n+1} - \frac{S_n}{n}\right| \ge \frac{1}{2}$, torej je zaporedje S_n/n skoraj gotovo divergentno.

4. Za $k \in \mathbb{Z}$ označimo s π_k verjetnost, da slučajni sprehod pride v točko k; s π_0 označimo verjetnost, da se sprehod še kdaj vrne v izhodišče, in ta verjetnost nas zanima. Iz simetrije in dinamike slučajnega sprehoda (dogajanje razdelimo glede na *prvi* korak) razberemo:

$$\pi_k = \pi_{-k}\,, \tag{1}$$
$$\pi_k = \frac{\pi_{k-1} + \pi_{k+1}}{2}\,; \quad k = 2, 3, 4, \ldots \tag{2}$$
$$\pi_1 = \frac{1 + \pi_2}{2}\,, \tag{3}$$
$$\pi_0 = \frac{\pi_{-1} + \pi_1}{2} = \pi_1\,. \tag{4}$$

Enačba (2) je diferenčna enačba, katere rešitve so linearne kombinacije členov $k^r \lambda^k$, kjer je λ rešitev karakteristične enačbe $\lambda^2 - 2\lambda + 1 = 0$, ki ima dvojno rešitev $\lambda_{1,2} = 1$, torej so za $k \in \mathbb{N}$ rešitve oblike $\pi_k = C_1 + C_2 k$. A le za $C_2 = 0$ je možno, da so vse verjetnosti π_k v intervalu $[0, 1]$, torej so te verjetnosti vse enake. Iz zveze (3) dobimo, da je $\pi_k = 1$ za vse $k \in \mathbb{N}$, iz zvez (1) in (4) pa potem še, da je tudi $\pi_0 = 1$. Sklep: standardni slučajni sprehod se skoraj gotovo vrne v izhodišče (in tudi obišče vsa stanja).

5. Označimo iskano verjetnost s π_k (za π_0 pa se dogovorimo, da je to verjetnost, da se slučajni sprehod *še kdaj* vrne v izhodišče). Iz neodvisnosti in enake porazdeljenosti sledijo naslednje rekurzivne zveze:

$$\pi_k = p\pi_{k-1} + (1-p)\pi_{k+1}\,; \quad |k| > 1 \tag{1}$$
$$\pi_1 = p + (1-p)\pi_2 \tag{2}$$
$$\pi_{-1} = p\pi_{-2} + (1-p) \tag{3}$$
$$\pi_0 = p\pi_{-1} + (1-p)\pi_1\,. \tag{4}$$

Enačba (1) je diferenčna enačba, katere rešitve so linearne kombinacije členov $k^r \lambda^k$, kjer je λ rešitev karakteristične enačbe:

$$(1-p)\lambda^2 - \lambda + p = 0\,, \tag{5}$$

ki ima rešitvi $\lambda_1 = 1$ in $\lambda_2 = p/(1-p)$.

Oglejmo si najprej primer, ko je $p < 1/2$. V tem primeru sta rešitvi različni. Ker se zveza (1) pri izhodišču prekine, rešitev nastavimo v obliki:

$$\pi_k = C_1^+ + C_2^+ \left(\frac{p}{1-p}\right)^k\,; \quad k \ge 1$$
$$\pi_k = C_1^- + C_2^- \left(\frac{p}{1-p}\right)^k\,; \quad k \le -1\,.$$

Najprej za negativne k opazimo, da mora biti $C_2^- = 0$, sicer verjetnosti uidejo izven intervala $[0, 1]$. Torej je $\pi_k = C_1^-$ za vse negativne k. Ko to vstavimo v enačbo (3), dobimo, da je $C_1^- = 1$.

Za pozitivne k pa si pomagamo z dejstvom, da je $\lim_{k\to\infty} \pi_k = 0$, kar je intuitivno jasno in bomo tudi eksaktno dokazali malo kasneje. Iz tega dejstva dobimo, da mora biti $C_1^+ = 0$. Ko vse skupaj vstavimo v enačbo (2) in poračunamo, dobimo, da mora veljati $C_2^+ = 1$. Iz enačbe (4) zdaj dobimo še $\pi_0 = 2p$. Sledi:

$$\pi_k = \begin{cases} \left(\frac{p}{1-p}\right)^k & ; \; k > 1 \\ 2p & ; \; k = 0 \\ 1 & ; \; k < 0 \end{cases} . \tag{$*$}$$

Preostane nam le še dokazati, da je $\pi := \lim_{k\to\infty} \pi_k = 0$. Za ta namen definirajmo dogodke:

$$A_k = \{\text{slučajni sprehod obišče stanje } k\} .$$

Najprej opazimo, da je $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$. Torej je π verjetnost njihovega preseka, to pa je dogodek, da slučajni sprehod obišče vsa pozitivna stanja. Pokazati je torej treba, da ima ta dogodek verjetnost nič.

Po krepkem zakonu velikih števil Kolmogorova z verjetnostjo ena velja:

$$\lim_{n\to\infty} \frac{S_n}{n} = 2p - 1 < 0 .$$

Torej so z verjetnostjo ena vse vrednosti $S_1, S_2, S_3, \ldots$ od nekod naprej negativne ali nič, se pravi, da z verjetnostjo nič velja, da je neskončno mnogo vrednosti $S_1, S_2, S_3, \ldots$ strogo pozitivnih. Dogodek, da slučajni sprehod obišče vsa pozitivna stanja, je način tega dogodka, torej ima tudi sam verjetnost nič. Zveza $(*)$ je tako dokazana.

Za $p > 1/2$ je obnašanje zrcalno simetrično okrog $1/2$ – velja torej:

$$\pi_k = \begin{cases} 1 & ; \; k > 1 \\ 2(1-p) & ; \; k = 0 \\ \left(\frac{p}{1-p}\right)^k & ; \; k < 0 \end{cases} .$$

6. Slučajna spremenljivka T_n se razlikuje od S_n le v primeru, ko je $S_n = 0$. Toda po Laplaceovi lokalni formuli je $\mathbb{P}(S_n = 0) \sim \sqrt{2/n\pi}$, če je n sod, in $\mathbb{P}(S_n = 0) = 0$, če je n lih. Torej je $\lim_{n\to\infty} \mathbb{P}(S_n = 0) = 0$. To pa pomeni, da se mora zaporedje T_n glede konvergence v verjetnosti obnašati enako kot S_n/n. Natančneje, za vsak $\varepsilon > 0$ velja:

$$\mathbb{P}(|T_n| < \varepsilon) = \mathbb{P}\left(\left|\frac{S_n}{n}\right| < \varepsilon, S_n \neq 0\right) \geq \mathbb{P}\left(\left|\frac{S_n}{n}\right| < \varepsilon\right) - \mathbb{P}(S_n = 0)$$

od koder sledi $\lim_{n\to\infty} \mathbb{P}(|T_n| < \varepsilon) = 1$.

Po drugi strani pa smo v 4. nalogi videli, da se standardni slučajni sprehod skoraj gotovo vrne v izhodišče. To pomeni, da se skoraj gotovo tudi *neskončno mnogokrat* vrne v izhodišče, torej je skoraj gotovo neskončno mnogo slučajnih spremenljivk T_n enakih 2. Toda če je $T_n = 2$, je $S_n = 0$ torej $|S_{n+1}| = 1$ in zato $|T_{n+1}| = 1/(n+1)$. Zaporedje, v katerem se neskončno mnogokrat ponovi ta vzorec, pa ne more nikamor konvergirati.

7. Vzemimo slučajne spremenljivke:

$$X_1 = X_2 = \cdots \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

in postavimo še $X := -X_1$. Tedaj so vse slučajne spremenljivke enako porazdeljene, torej zaporedje X_n zagotovo konvergira proti X v porazdelitvi. Po drugi strani pa je:

$$\mathbb{P}(|X_n - X| < 2) = \mathbb{P}(X_n = X) = 0$$

za vse n, zato zaporedje ne more konvergirati v verjetnosti.

Če pa privzamemo, da je $X = c$ konstanta, konvergenca v porazdelitvi pomeni, da je:

$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = 0 \quad \text{za } x < c$$
$$\lim_{n\to\infty} \mathbb{P}(X_n \le x) = 1 \quad \text{za } x > c\,,$$

kar je ekvivalentno trditvi, da za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} \mathbb{P}(X_n \le c - \varepsilon) = \lim_{n\to\infty} \mathbb{P}(X_n > c + \varepsilon) = 0\,,$$

kar je ekvivalentno tudi trditvi, da za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} \mathbb{P}(X_n \le c - \varepsilon) = \lim_{n\to\infty} \mathbb{P}(X_n \ge c + \varepsilon) = 0\,,$$

to pa je ekvivalentno trditvi, da za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} \mathbb{P}\big(|X_n - c| \ge \varepsilon\big) = 0\,,$$

le-ta pa pomeni ravno konvergenco v verjetnosti.

8. Porazdelitve opišimo s kumulativnimi porazdelitvenimi funkcijami F_n, limitna porazdelitev pa naj ima kumulativno porazdelitveno funkcijo F. Slučajne spremenljivke konstruiramo tako kot v 46. nalogi iz 4. razdelka: naj bodo Q_n oz. Q merljive izbire kvantilnih funkcij, ki pripadajo danim kumulativnim porazdelitvenim funkcijam (ena od takih izbir je, če vzamemo spodnje kvantile, tj. $Q(\omega) = \inf\{x\,;\,F(x) \ge \omega\}$ in analogno tudi $Q_n(\omega)$). Nadalje naj bo U porazdeljena enakomerno na intervalu $(0, 1)$. Tedaj imajo slučajne spremenljivke $Q_n(U)$ in $Q(U)$ kumulativne porazdelitvene funkcije F_n in F.

Funkcija Q je naraščajoča, zato ima kvečjemu števno mnogo točk nezveznosti. Dovolj bo torej pokazati, da je $\lim_{n\to\infty} Q_n(u) = Q(u)$, brž ko je Q zvezna v u.

Naj bo u taka točka in $x = Q(u)$. Nadalje naj bo $\varepsilon > 0$. Torej obstajata tak $u^+ > u$, da je $Q(u^+) < x + \varepsilon$, in tak $u^- < u$, da je $Q(u^-) > x - \varepsilon$. Tudi funkcija F ima kvečjemu števno mnogo točk nezveznosti, zato nadalje obstajata taka $x < x^+ < x + \varepsilon$ in $x - \varepsilon < x^- < x$, da je F zvezna v x^+ in x^- ter da je $Q(u^+) < x^+$ in $Q(u^-) > x^-$. Iz implikacij iz točke a) 46. naloge iz 4. razdelka dobimo, da potem velja $F(x^+) \ge u^+ > u$ in $F(x^-) \le u^- < u$. Zaradi zveznosti in šibke konvergence za dovolj velike n velja tudi $F_n(x^+) > u$ in $F_n(x^-) < u$. Spet iz implikacij dobimo $Q_n(u) \le x^+ < x + \varepsilon$ in $Q_n(u) \ge x^- > x - \varepsilon$. Ker to velja za vsak $\varepsilon > 0$, zaporedje $Q_n(u)$ konvergira proti x.

9. Za skoraj gotovo konvergenco je trditev pravilna, saj je:

$$\left\{\lim_{n\to\infty} X_n = X\right\} \cap \left\{\lim_{n\to\infty} Y_n = Y\right\} \subseteq \left\{\lim_{n\to\infty} (X_n + Y_n) = X + Y\right\}.$$

Prav tako je trditev pravilna za konvergenco v verjetnosti, saj je:

$$\left\{\lim_{n\to\infty} |X_n - X| < \frac{\varepsilon}{2}\right\} \cap \left\{\lim_{n\to\infty} |Y_n - Y| < \frac{\varepsilon}{2}\right\} \subseteq \left\{\lim_{n\to\infty} \big|(X_n + Y_n) - (X + Y)\big| < \varepsilon\right\}.$$

Pač pa trditev ni pravilna za konvergenco v porazdelitvi: vzemimo slučajne spremenljivke:

$$X_1 = X_2 = \cdots = Y_1 = Y_2 = \cdots \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

ter še slučajni spremenljivki:

$$X \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad Y = -X$$

Ker so vse omenjene slučajne spremenljivke enako porazdeljene, zaporedje X_n v porazdelitvi zagotovo konvergira proti X, Y_n pa proti Y. Pač pa je:

$$X_n + Y_n \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \text{medtem ko je} \quad X + Y = 0,$$

zato zaporedje $X_n + Y_n$ zagotovo ne konvergira v porazdelitvi proti $X + Y$.

10. Dokazati moramo, da za vsak z, kjer je F_{X+c} zvezna, velja:

$$\lim_{n\to\infty} \mathbb{P}(X_n + Y_n \le z) = \mathbb{P}(X + x \le z),$$

kar je ekvivalentno trditvi, da za vsak x, kjer je F_X zvezna, velja:

$$\lim_{n\to\infty} \mathbb{P}(X_n + Y_n - c \le x) = \mathbb{P}(X \le x). \tag{$*$}$$

Naj bo $\varepsilon > 0$ in naj bo F_X zvezna v $x + \varepsilon$ in $x - \varepsilon$. Iz zveze:

$$\begin{aligned} \mathbb{P}(X_n + Y_n - c \le x) &= \mathbb{P}(X_n + Y_n - c \le x, |Y_n - c| < \varepsilon) + \\ &\qquad + \mathbb{P}(X_n + Y_n - c \le x, |Y_n - c| \ge \varepsilon) \le \\ &\le \mathbb{P}(X_n \le x + \varepsilon) + \mathbb{P}(|Y_n - c| \ge \varepsilon) \end{aligned}$$

in konvergence zaporedij sledi:

$$\limsup_{n\to\infty} \mathbb{P}(X_n + Y_n - c \le x) \le \mathbb{P}(X \le x + \varepsilon).$$

V limiti, ko gre ε proti nič, ob upoštevanju dejstva, da je F_X nezvezna v kvečjemu števno mnogo točkah, dobimo:

$$\limsup_{n\to\infty} \mathbb{P}(X_n + Y_n - c \le x) \le \mathbb{P}(X \le x). \tag{$**$}$$

Naj bo zdaj spet $\varepsilon > 0$ in F_X zvezna v $x - \varepsilon$. Podobno kot prej iz zveze:

$$\begin{aligned}\mathbb{P}(X_n \le x - \varepsilon) &= \mathbb{P}(X_n \le x - \varepsilon, |Y_n - c| < \varepsilon) + \\ &\qquad + \mathbb{P}(X_n \le x - \varepsilon, |Y_n - c| \ge \varepsilon) \le \\ &\le \mathbb{P}(X_n + Y_n - c \le x) + \mathbb{P}(|Y_n - c| \ge \varepsilon)\end{aligned}$$

in konvergence zaporedij sledi:

$$\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n - c \le x) \ge \mathbb{P}(X \le x - \varepsilon)\,.$$

V limiti, ko gre ε proti nič, ob upoštevanju zveznosti funkcije F_X v x in spet dejstva, da je F_X nezvezna v kvečjemu števno mnogo točkah, dobimo:

$$\liminf_{n\to\infty} \mathbb{P}(X_n + Y_n - c \le x) \ge \mathbb{P}(X \le x)\,. \qquad (***)$$

Iz $(**)$ in $(***)$ pa že sledi $(*)$.

11. Pišemo lahko $S = X_1 + X_2 + \cdots + X_{105}$, kjer je X_i število pik pri i-tem metu. Slučajne spremenljivke X_i so porazdeljene enakomerno na množici $\{1, 2, 3, 4, 5, 6\}$, od koder sledi $\mathbb{E}(X_i) = 7/2$ in $\mathrm{var}(X_i) = 35/12$. Torej je $\mathbb{E}(S) = 735/2 = 367{\cdot}5$ in $\mathrm{var}(S) = 1225/4$, torej $\sigma(S) = 17{\cdot}5$. Na prvi pogled se zdi smiselna aproksimacija:

$$\mathbb{P}(S < 350) \approx \Phi\left(\frac{350 - 367{\cdot}5}{17{\cdot}5}\right) + \frac{1}{2} \doteq 0{\cdot}15866\,,$$

vendar je prav tako smiselno aproksimirati tudi:

$$\mathbb{P}(S < 350) = \mathbb{P}(S \le 349) \approx \Phi\left(\frac{349 - 367{\cdot}5}{17{\cdot}5}\right) + \frac{1}{2} \doteq 0{\cdot}14522\,.$$

Kadar slučajna spremenljivka zavzame vrednosti na mreži oblike $a\mathbb{Z} + b$, se v povprečju najbolj splača aproksimirati na sredini, torej:

$$\mathbb{P}(S < 350) = \mathbb{P}(S < 349{\cdot}5) \approx \Phi\left(\frac{349{\cdot}5 - 367{\cdot}5}{17{\cdot}5}\right) + \frac{1}{2} \doteq 0{\cdot}15184\,.$$

Točen rezultat: $0{\cdot}1520449$.

Opomba. Pri 100 metih bi bila iskana verjetnost enaka približno $1/2$. Pet dodatnih metov se torej zelo pozna.

12. Rezultat, dobljen s pomočjo centralnega limitnega izreka:
$\Phi\left(\frac{9{\cdot}5}{\sqrt{120}}\right) + \Phi\left(\frac{29{\cdot}5}{\sqrt{120}}\right) \doteq 0{\cdot}80355$.
Točen rezultat: $0{\cdot}80439$.

13. Naša bilanca se izraža kot vsota 50 neodvisnih slučajnih spremenljivk s porazdelitvijo:

$$\begin{pmatrix} -1 & 0 & 9 \\ 0{\cdot}5 & 0{\cdot}45 & 0{\cdot}05 \end{pmatrix}.$$

Posamezen sumand ima pričakovano vrednost $-0{\cdot}05$, varianco $4{\cdot}5475$ in standardni odklon $2{\cdot}132$; cela bilanca pa ima pričakovano vrednost $-2{\cdot}5$, varianco $227{\cdot}375$ in standardni odklon $15{\cdot}08$. Verjetnost dobička aproksimiramo kot:

$$\frac{1}{2} - \Phi\left(\frac{0{\cdot}5 + 2{\cdot}5}{15{\cdot}08}\right) \doteq 0{\cdot}4211\,.$$

Točen rezultat pa je $0{\cdot}3957429$. Napaka je tokrat večja, kot bi pričakovali iz prejšnjih primerov, to pa zato, ker ima porazdelitev vrednost, ki zelo odstopa.

14. Ker ima X enako porazdelitev kot vsota 100 neodvisnih slučajnih spremenljivk, porazdeljenih po hi kvadrat z eno prostostno stopnjo, so izpolnjeni pogoji centralnega limitnega izreka in lahko porazdelitev aproksimiramo z normalno $\mathrm{N}\big(100, \sqrt{200}\big)$.
$\mathbb{P}(X > 110)$: CLI: $0{\cdot}23975$, točen rezultat: $0{\cdot}23220$.
$\mathbb{P}(90 < X < 110)$: CLI: $0{\cdot}52050$, točen rezultat: $0{\cdot}52099$.

Opomba. Porazdelitev hi kvadrat je poseben primer porazdelitve gama, ta pa je *neskončno deljiva*, kar pomeni, da se da zapisati kot porazdelitev vsote *poljubnega* števila neodvisnih in enako porazdeljenih slučajnih spremenljivk. Natančneje, če so $X_1, X_2, \ldots, X_n$ neodvisne s porazdelitvijo $\mathrm{Gama}(50/n, 1/2)$, ima njihova vsota porazdelitev hi kvadrat s 100 prostostnimi stopnjami. A to ne pomeni, da je porazdelitev hi kvadrat poljubno blizu normalni (oziroma kar normalna), ker se z rastočim n veča tudi razmerje γ_1/σ_1.

15. Iz centralnega limitnega izreka dobimo:

$$\mathbb{P}(S_n > x) \approx \frac{1}{2} - \Phi\left(\frac{x}{\sqrt{2n}}\right)\,.$$

$\mathbb{P}(S_5 > 5)$: CLI: $0{\cdot}0569$, točen rezultat: $0{\cdot}0555$.
$\mathbb{P}(S_{20} > 20)$: CLI: $0{\cdot}00078$, točen rezultat: $0{\cdot}00114$.

Približek za $\mathbb{P}(S_{20} > 20)$ ima veliko relativno napako, ker se že nahajamo v območju *velikih odklonov*.

16. Iz $\mathbb{E}(U_i) = 1/2$, $\mathbb{E}(U_i^2) = 1/3$, $\mathbb{E}(V_i) = 4/3$, $\mathbb{E}(V_i^2) = 2$ in neodvisnosti dobimo $\mathbb{E}(X_i) = \mathbb{E}(U_i)\,\mathbb{E}(V_i) = 2/3$, $\mathbb{E}(X_i^2) = \mathbb{E}(U_i^2)\,\mathbb{E}(V_i^2) = 2/3$ in $\mathrm{var}(X_i) = 2/3 - (2/3)^2 = 2/9$. Sledi $\mathbb{E}(S) = 200/3$ in zaradi neodvisnosti še $\mathrm{var}(S) = 200/9$. Ker je S vsota velikega števila neodvisnih slučajnih spremenljivk, je po centralnem limitnem izreku:

$$\mathbb{P}(S < 60) \approx \Phi\left(\frac{60 - \frac{200}{3}}{\sqrt{\frac{200}{9}}}\right) + \frac{1}{2} = \frac{1}{2} - \Phi\big(\sqrt{2}\big) \doteq 0{\cdot}0787\,.$$

Točen rezultat: $0{\cdot}07642$.

17. Iz:

$$\mathbb{E}(X_i) = 1\,, \quad \mathrm{var}(X_i) = 2p\,, \quad \mathbb{E}(S) = 100\,, \quad \mathrm{var}(S) = 200p$$

in centralnega limitnega izreka dobimo:

$$\mathbb{P}(S < 90) \approx \Phi\left(\frac{89{\dot{}}5 - 100}{\sqrt{200p}}\right) + \frac{1}{2}$$

Torej bo $\Phi\left(\frac{10{\dot{}}5}{\sqrt{200p}}\right) \approx 0{\dot{}}45$ oziroma $\frac{10{\dot{}}5}{\sqrt{200p}} \approx 1{\dot{}}645$ oziroma $p \approx 0{\dot{}}204$.

Točen rezultat: 0˙20414.

18. Iz $\mathbb{E}(X_i) = 0$ in $\mathrm{var}(X_i) = 1/2$ dobimo:

$$\mathbb{P}(|S_n| > 100) = \mathbb{P}(S_n < -100{\dot{}}5) + \mathbb{P}(S_n > 100{\dot{}}5) \approx 1 - 2\,\Phi\left(\frac{100{\dot{}}5}{\sqrt{n/2}}\right).$$

Neenačba:

$$1 - 2\,\Phi\left(\frac{100{\dot{}}5}{\sqrt{n/2}}\right) > 0{\dot{}}05 \qquad \text{oziroma} \qquad \Phi\left(\frac{201}{\sqrt{2n}}\right) < 0{\dot{}}475$$

ima približno rešitev:

$$\frac{201}{\sqrt{2n}} < 1{\dot{}}96 \qquad \text{oziroma} \qquad n > \frac{1}{2}\left(\frac{201}{1{\dot{}}96}\right)^2 \doteq 5258{\dot{}}356\,,$$

kar nam da izbiro $n \geq 5259$.

To je tudi točna meja, saj je verjetnost danega dogodka za $n = 5258$ enaka približno 0˙04998288, za $n = 5259$ pa je enaka približno 0˙05000466.

19. a) 400

b) *Prvi način.* Dogodek, da se bo moral potopiti več kot 450-krat, lahko zapišemo kot dogodek, da bo imel po 450 potopih manj kot 80 biserov. Tako, če binomsko porazdelitev $\mathrm{Bin}(450, 0{\dot{}}2)$ aproksimiramo z normalno $\mathrm{N}\big(90, \sqrt{72}\big)$, dobimo približen rezultat 0˙10796.

Drugi način. Število potopov je porazdeljeno po negativni binomski porazdelitvi $\mathrm{NB}(80, 0{\dot{}}2)$, ki jo, ker jo dobimo iz vsote 80 neodvisnih slučajnih spremenljivk, lahko aproksimiramo z normalno $\mathrm{N}(400, 40)$. Tako dobimo približen rezultat 0˙10338.

Točen rezultat: 0˙10669.

20. a) Velja $\mathbb{E}(X_k) = k$, $\mathrm{var}(X_k) = k^2$ in $\mathbb{E}\big(|X_k - \mathbb{E}(X_k)|^3\big) = k^3\gamma_1^3$, kjer je:

$$\gamma_1^3 = \mathbb{E}\big(|X_1 - \mathbb{E}(X_1)|^3\big) = \int_0^\infty |x-1|^3\, e^{-x}\,\mathrm{d}x = \frac{12}{e} - 2 \doteq 2{\dot{}}415\,.$$

Torej je:

$$\mathbb{E}(S_n) = \frac{n(n+1)}{2}\,, \qquad \operatorname{var}(S_n) = \frac{n(n+1)(2n+1)}{6}\,,$$
$$\sum_{k=1}^{n} \mathbb{E}(|X_k - \mathbb{E}(X_k)|^3) = \frac{n^2(n+1)^2}{4}\,\gamma_1^3\,.$$

Ker razmerje Ljapunova $\gamma_1^3\sqrt{\dfrac{27n(n+1)}{2(2n+1)^3}}$ konvergira proti nič, ko gre n proti neskončno, intervalske verjetnosti enakomerno konvergirajo proti ustreznim verjetnostim za normalno porazdelitev, od tod pa sledi, da standardizirane vsote šibko konvergirajo proti standardni normalni porazdelitvi.

b) Približno velja:

$$\mathbb{P}(S_{100} < 6000) \approx \frac{1}{2} + \Phi\left(\frac{6000 - 5050}{\sqrt{338350}}\right) \doteq 0{\dot{}}948787\,.$$

Iz ocene Ševcove po krajšem računu dobimo, da je napaka, ki smo jo naredili, navzgor omejena z 0˙175. Dejanska napaka, ki smo jo naredili, pa je dosti manjša: točen rezultat pride 0˙942167.

21. Ker sumand X_1 izstopa (**prispeva celo več kot polovico variance**), ga moramo obravnavati posebej. Če z S' označimo vsoto preostalih, tj. $S' = X_2 + X_3 + \cdots + X_{100}$, velja:

$$\begin{aligned}
\mathbb{P}(S < 41) &= \mathbb{P}(X_1 = -1)\,\mathbb{P}(S < 41 \mid X_1 = -1) + \\
&\qquad + \mathbb{P}(X_1 = 1)\,\mathbb{P}(S < 41 \mid X_1 = 1) = \\
&= \mathbb{P}(X_1 = -1)\,\mathbb{P}(S' - 10 < 41 \mid X_1 = -1) + \\
&\qquad + \mathbb{P}(X_1 = 1)\,\mathbb{P}(S' + 10 < 41 \mid X_1 = 1) = \\
&= \frac{1}{3}\,\mathbb{P}(S' < 51) + \frac{2}{3}\,\mathbb{P}(S' < 31)\,.
\end{aligned}$$

Ker slučajna spremenljivka S' zavzame vrednosti na lihih številih, je meji za S' v zgornji formuli smiselno znižati za 1. Iz:

$$\mathbb{E}(X_2) = \frac{1}{3}\,, \quad \operatorname{var}(X_2) = \frac{8}{9}\,; \qquad \mathbb{E}(S') = 33\,, \quad \operatorname{var}(S') = 88$$

dobimo:

$$\mathbb{P}(S < 41) \approx \frac{1}{3}\left[\Phi\left(\frac{50-33}{\sqrt{88}}\right) + \frac{1}{2}\right] + \frac{2}{3}\left[\Phi\left(\frac{30-33}{\sqrt{88}}\right) + \frac{1}{2}\right] \doteq 0{\dot{}}57138\,.$$

Točen rezultat: 0˙5695804.

Opomba. Če bi centralni limitni izrek uporabili neposredno na S, bi dobili približek 0˙609, ki znatno bolj odstopa od prave vrednosti.

22. a) Označimo z X_{kl} število kilometrov, ki jih Bernard prevozi l-ti dan k-tega tedna. Skupno število kilometrov,

$$S_{12} := \sum_{k=1}^{12} \sum_{l=1}^{7} X_{kl}$$

je vsota neodvisnih slučajnih spremenljivk, ki sicer niso vse enako porazdeljene, a recimo najprej, da to zanemarimo. Za $l = 1, 2, \ldots, 6$ velja:

$$\mathbb{E}(X_{kl}) = \frac{20 + 25}{2} = 22\text{˙}5 \,, \qquad \operatorname{var}(X_{kl}) = \frac{5^2}{4} = 6\text{˙}25 \,,$$

za $l = 7$ pa velja:

$$\mathbb{E}(X_{k7}) = \frac{100}{4} = 25 \,, \qquad \operatorname{var}(X_{kl}) = \frac{3 \cdot 100^2}{16} = 1875 \,.$$

Seštejemo in dobimo:

$$\begin{aligned} \mathbb{E}(S_{12}) &= 12 \cdot 6 \cdot 22\text{˙}5 + 12 \cdot 25 = 1620 + 300 = 1920 \,, \\ \operatorname{var}(S_{12}) &= 12 \cdot 6 \cdot 6\text{˙}25 + 12 \cdot 1875 = 450 + 22500 = 22950 \,, \end{aligned}$$

kar nam da približek:

$$\begin{aligned} P(S_{12} > 2000) = P(S_{12} > 2002\text{˙}5) \approx 1 - \Phi\left(\frac{2002\text{˙}5 - 1920}{\sqrt{22950}}\right) \doteq 1 - \Phi(0\text{˙}5446) \doteq \\ \doteq 0\text{˙}2930 \,. \end{aligned}$$

Vidimo pa, da nedeljski kilometri prispevajo glavnino variance, zato se morda splača pogojevati na število nedelj v prvih 12 tednih, ko Bernard naredi izlet, kar označimo z N_{12}. Velja $S_{12} := T_{12} + 100\, N_{12}$, kjer je $T_{12} := \sum_{k=1}^{12} \sum_{l=1}^{6} X_{kl}$. Slučajna spremenljivka N_{12} je porazdeljena binomsko $\mathrm{Bin}(12, 1/4)$, torej velja:

$$\begin{aligned} P(S_{12} > 2000) &= \sum_{n=0}^{12} P(N = n)\, P(S_{12} > 2000 \mid N = n) = \\ &= \sum_{n=0}^{12} \binom{12}{n} \left(\frac{1}{4}\right)^n \left(\frac{3}{4}\right)^{12-n} P(T_{12} > 2000 - 100n) \,. \end{aligned}$$

slučajna spremenljivka T_{12} pa je vsota velikega števila neodvisnih in enako porazdeljenih slučajnih spremenljivk, torej centralni limitni izrek zanjo da natančnejšo aproksimacijo. Ker je:

$$\mathbb{E}(T_{12}) = 12 \cdot 6 \cdot 22\text{˙}5 = 1620 \,, \qquad \operatorname{var}(T_{12}) = 12 \cdot 6 \cdot 6\text{˙}25 = 450 \,,$$

je približno:

$$\begin{aligned} P(S_{12} > 2000) &= P(S_{12} > 2002\text{˙}5) \\ &\approx \sum_{n=0}^{12} \binom{12}{n} \left(\frac{1}{4}\right)^n \left(\frac{3}{4}\right)^{12-n} \left[1 - \Phi\left(\frac{382\text{˙}5 - 100n}{\sqrt{450}}\right)\right] \,. \end{aligned}$$

Tabeliramo ustrezne normalne verjetnosti, zaokrožene na 4 decimalke natančno:

n	0	1	2	3	4	≥ 5
$1-\Phi\left(\frac{382.5-100n}{\sqrt{450}}\right)$	0˙0000	0˙0000	0˙0000	0˙0001	0˙7953	1˙0000

in vidimo, da je edina netrivialna normalna verjetnost pri $n = 4$, poleg tega pa še, da se bolj splača računati nasprotno verjetnost:

$$\begin{aligned} P(S_{12} \leq 2000) &\approx \sum_{n=0}^{3} \binom{12}{n} \left(\frac{1}{4}\right)^n \left(\frac{3}{4}\right)^{12-n} + \binom{12}{4} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right)^8 \cdot 0˙2047 \\ &\doteq 0˙6886\,. \end{aligned}$$

Verjetnost, da Bernard v prvih 12 tednih prevozi več kot 2000 km, je torej približno 0˙3114.

Točen rezultat: 0˙3115884.

b) Tu uporabimo manj natančni približek iz prve točke. Če označimo $Y_k := \sum_{l=1}^{7} X_{kl}$, velja:

$$\mathbb{E}(Y_k) = 6 \cdot 22˙5 + 25 = 160\,, \qquad \mathrm{var}(Y_k) = 6 \cdot 6˙25 + 1875 = 1912˙5\,,$$

torej nastavimo:

$$1 - \Phi\left(\frac{20002˙5 - 160\,n}{\sqrt{1912˙5\,n}}\right) = 0˙95$$

oziroma

$$160\,n - 20002˙5 = \Phi^{-1}(0˙95)\sqrt{1912˙5\,n}\,.$$

Če uvedemo $u = \sqrt{n}$, dobimo rešitvi $u_1 \doteq -10˙959$ in $u_2 \doteq 11˙408$. Prava je rešitev s pozitivnim korenom, ker je $\sqrt{n}$ po dogovoru nenegativno število. Iz $u_2^2 \doteq 130˙15$ dobimo, da postane verjetnost večja od 95% po približno 131 tednih.

V resnici je 131 tednov potrebnih in zadostnih: po 130 tednih verjetnost pride 0˙9480694, po 131 tednih pa 0˙9750782.

23. Najprej izračunamo, da je $\mathbb{E}(X_n) = 0$ in $\mathrm{var}(X_n) = 1$ za vse n. Da njihove delne vsote, ki imajo pričakovane vrednosti 0 in variance n, izpolnjujejo pogoje centralnega limitnega izreka, pomeni, da slučajne spremenljivke $W_n := S_n/\sqrt{n}$ šibko konvergirajo proti standardni normalni porazdelitvi, kar pomeni tudi, da za poljubna $a \leq b$ velja $\mathbb{P}(a \leq W_n \leq b) = \Phi(b) - \Phi(a)$. Torej bi moralo veljati tudi:

$$\lim_{n\to\infty} \mathbb{P}\left(-\tfrac{1}{2} \leq W_n \leq \tfrac{1}{2}\right) = 2\,\Phi\left(\tfrac{1}{2}\right) \doteq 0˙3829 < 0˙4\,. \qquad (*)$$

Toda že za $n \geq 4$ velja:

$$\begin{aligned}
\mathbb{P}\big(-\tfrac{1}{2} \leq W_n \leq \tfrac{1}{2}\big) &= \mathbb{P}\left(-\frac{\sqrt{n}}{2} \leq S_n \leq \frac{\sqrt{n}}{2}\right) \geq \\
&\geq \mathbb{P}(X_2 = X_3 = \cdots = X_n = 0) = \\
&= \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right) \cdots \left(1 - \frac{1}{n^2}\right) = \\
&= \left(1 - \frac{1}{2}\right)\left(1 + \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 + \frac{1}{3}\right) \cdots \left(1 - \frac{1}{n}\right)\left(1 + \frac{1}{n}\right) = \\
&= \frac{n+1}{2n} \geq \frac{1}{2} ,
\end{aligned}$$

torej $(*)$ ne more veljati.

24. a) Lahko si predstavljamo, da iz posode, v kateri je n kroglic, od tega r rdečih, na slepo in brez vračanja izvlečemo s kroglic, ter da je N število rdečih med izvlečenimi. Torej je $N = X_1 + X_2 + \cdots + X_s$, kjer je X_i indikator dogodka, da je bila i-ta izvlečena kroglica rdeča. Iz $\mathbb{E}(X_i) = \frac{r}{n}$ sledi $\mathbb{E}(N) = \frac{rs}{n}$. Nadalje za različna i in j velja $\mathbb{E}(X_j) = \frac{r(r-1)}{n(n-1)}$. Sledi:

$$\operatorname{var}(N) = \frac{rs}{n} + \frac{r(r-1)s(s-1)}{n(n-1)} - \frac{r^2 s^2}{n^2} = \frac{rs(n-r)(n-s)}{n^2(n-1)} .$$

b) Slučajno spremenljivko N lahko zapišemo kot vsoto indikatorjev na vsaj dva načina: kot vsoto indikatorjev 100 dogodkov, da je posamezna izvlečena kroglica rdeča, ali kot vsoto indikatorjev 200 dogodkov, da je posamezna rdeča kroglica izvlečena. Čeprav ti indikatorji niso neodvisni, normalna aproksimacija vseeno daje dobre rezultate.

Iz $\mathbb{E}(N) = 40$ in $\sigma(N) \doteq 4{\cdot}386169$ sklepamo:

$$\mathbb{P}(N > 45) \approx \frac{1}{2} - \Phi\left(\frac{45{\cdot}5 - 40}{4{\cdot}386169}\right) \doteq 0{\cdot}10493 .$$

Točen rezultat: $0{\cdot}1050956$.

25. Centralnega limitnega izreka ne moremo uporabiti neposredno na S, ker produkt slučajne spremenljivke X in normalno porazdeljene slučajne spremenljivke, četudi neodvisne od X, ni nujno normalno porazdeljen. Prav tako centralni limitni izrek ne velja za vsoto $XY_1 + XY_2 + \cdots + XY_{100}$, saj so seštevanci odvisni (čeprav se da centralni limitni izrek posplošiti tudi na vsote slučajnih spremenljivk z določeno vrsto odvisnosti, je odvisnost prej omenjenih seštevancev premočna). Pravilno pa bo iskano verjetnost računati s pomočjo pogojnih verjetnosti glede na X. Če pišemo $T := Y_1 + Y_2 + \cdots + Y_{100}$, po izreku o polni verjetnosti velja:

$$\begin{aligned}
\mathbb{P}(400 < S < 500) = \mathbb{P}(X = 1)\, \mathbb{P}(400 < S < 500 \mid X = 1) &+ \\
+ \mathbb{P}(X = 2)\, \mathbb{P}(400 < S < 500 \mid X = 2) &= \\
= \frac{2}{3}\, \mathbb{P}(400 < T < 500) + \frac{1}{3}\, \mathbb{P}(200 < T < 250) .&
\end{aligned}$$

Za slučajno spremenljivko T pa centralni limitni izrek velja: iz $\mathbb{E}(T) = 300$ in $\operatorname{var}(T) = 10000$ dobimo:

$$\mathbb{P}(a < T < b) \approx \Phi\left(\frac{b - 300}{100}\right) - \Phi\left(\frac{a - 300}{100}\right),$$

torej je:

$$\mathbb{P}(400 < S < 500) \approx \frac{2}{3}\big[\Phi(2) - \Phi(1)\big] + \frac{1}{3}\big[\Phi(1) - \Phi(0{\cdot}5)\big] \doteq 0{\cdot}141\,.$$

Oglejmo si še, koliko bi znašala iskana verjetnost za normalno slučajno spremenljivko z enako pričakovano vrednostjo in varianco kot S. Za ta namen izračunamo:

$$\mathbb{E}(S) = \mathbb{E}(X)\,\mathbb{E}(T) = 400\,.$$

Nadalje je $\mathbb{E}(T^2) = \operatorname{var}(T) + \big(\mathbb{E}(T)\big)^2 = 100000$ in zato:

$$\begin{aligned} \mathbb{E}(S^2) &= \mathbb{E}(X^2)\,\mathbb{E}(T^2) = 200000\,, \\ \operatorname{var}(S) &= \mathbb{E}(S^2) - \big(\mathbb{E}(S)\big)^2 = 40000\,. \end{aligned}$$

Za normalno slučajno spremenljivko bi bila torej iskana verjetnost enaka:

$$\Phi(0{\cdot}5) \doteq 0{\cdot}191\,,$$

kar se občutno razlikuje od pravega rezultata.

10. Zadostne in postranske statistike

1. Pogledati moramo pogojno porazdelitev slučajnega vektorja (X, Y) glede na $X+Y$. A ker je Y natančno določena z X in $X+Y$, je dovolj gledati porazdelitev slučajne spremenljivke X glede na $X+Y$. Za le-to pa smo v 15. nalogi iz 7. razdelka izračunali, da je binomska $\mathrm{Bin}(X+Y, p)$. V primeru a) torej X ni zadostna statistika, v primeru b) pa je zadostna.

2. Če želimo ugotoviti, ali je X zadostna statistika, moramo gledati pogojno porazdelitev slučajnega vektorja (X, Y) glede na X, za slednjo pa je dovolj gledati pogojno porazdelitev slučajne spremenljivke Y glede na X. Podobno kot v 34. nalogi iz 7. razdelka izračunamo, da za $x > 0$ velja:

$$f_{Y|X}(y \mid x\,;\lambda) = \begin{cases} 1/x & ;\ 0 < y < x \\ 0 & ;\ \text{sicer.} \end{cases}$$

Pogojno na X je torej Y porazdeljena zvezno enakomerno na intervalu od 0 do X in to velja ne glede na λ. Sledi, da je X zadostna statistika.

Če pa želimo ugotoviti, ali je Y zadostna statistika, moramo gledati pogojno porazdelitev slučajne spremenljivke X glede na Y. Podobno kot zgoraj izračunamo, da za $y > 0$ velja:

$$f_{X|Y}(x \mid y\,;\lambda) = \begin{cases} \lambda\, e^{-\lambda(x-y)} & ;\ x > y \\ 0 & ;\ \text{sicer.} \end{cases}$$

Ta pogojna gostota pa je odvisna od λ, zato Y ni zadostna statistika.

3. Gostoto porazdelitve opažanja X zapišemo v obliki:

$$f_X(x) = \frac{e^{-\frac{\mu^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}\, e^{-\frac{|x|^2}{2\sigma^2}}\, e^{\frac{\mu x}{\sigma^2}} .$$

Iz Fisher–Neymanovega faktorizacijskega izreka zdaj sledi, da $|X|$ v primerih a) in c) ni zadostna statistika, v primerih b) in d) pa je (primer b) je poseben primer primera d)).

4. Uspešnost poskusov ponazorimo s slučajnim vektorjem $(X_1, \ldots, X_n)$, kjer je $X_i = 1$, če i-ti poskus uspe, in 0, če ne uspe. Tedaj za $x_1, x_2, \ldots, x_n \in \{0, 1\}$ velja:

$$\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1;\theta)\, f(x_2;\theta) \cdots f(x_n;\theta)\,,$$

kjer je $f(0;\theta) = 1-\theta$ in $f(1;\theta) = \theta$. Zdaj pa opazimo, da velja tudi:

$$\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(0;\theta)^{n-x_1-x_2-\cdots-x_n} f(1;\theta)^{x_1+x_2+\cdots+x_n} .$$

Torej je $X_1 + X_2 + \cdots + X_n$ zadostna statistika, slednji izraz pa predstavlja ravno število uspelih poskusov.

5. Iz:

$$\mathbb{P}_\lambda(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{\lambda^{x_1+x_2+\cdots+x_n} e^{-n\lambda}}{x_1!\, x_2! \cdots x_n!}$$

razberemo, da je statistika $X_1 + X_2 + \cdots + X_n$ zadostna.

6. Točko a) dokažemo s popolno indukcijo. Dokazati moramo, da za vsak k velja $\mathbb{P}_\theta(X_k = 1) = \theta$, in to smo že privzeli za $k = 1$. Indukcijski korak s k na $k+1$ izpeljemo tako, da najprej zapišemo:

$$\mathbb{P}_\theta(X_{k+1} = 1 \mid X_1, X_2, \ldots, X_k) = p_{X_k,1}\,,$$

nakar izračunamo:

$$\begin{aligned}\mathbb{P}_\theta(X_{k+1} = 1) &= \mathbb{E}\big[\mathbb{P}_\theta(X_{k+1} = 1 \mid X_1, X_2, \ldots, X_k)\big] = \\ &= \mathbb{E}\big[p_{X_k,1}\big] = \\ &= \mathbb{P}(X_k = 0)\, p_{01} + \mathbb{P}(X_k = 1)\, p_{11} = \\ &= (1-\theta)\,\frac{\theta}{2} + \theta\,\frac{1+\theta}{2} = \\ &= \theta\,.\end{aligned}$$

b) Označimo z S število uspelih poskusov in pri $n = 3$ izračunajmo pogojno porazdelitev našega opažanja (X_1, X_2, X_3) glede na $S = 1$. Velja:

$$\begin{aligned}\mathbb{P}_\theta(X_1 = 0, X_2 = 0, X_3 = 1) &= \frac{\theta(1-\theta)(2-\theta)}{4} \\ \mathbb{P}_\theta(X_1 = 0, X_2 = 1, X_3 = 0) &= \frac{\theta(1-\theta)^2}{4} \\ \mathbb{P}_\theta(X_1 = 1, X_2 = 0, X_3 = 0) &= \frac{\theta(1-\theta)(2-\theta)}{4}\,.\end{aligned}$$

Ko seštejemo, dobimo $\mathbb{P}_\theta(S = 1) = \theta(1-\theta)(5-3\theta)/4$, torej je pogojna porazdelitev enaka:

$$\begin{pmatrix} (0,0,1) & (0,1,0) & (1,0,0) \\ \frac{2-\theta}{5-3\theta} & \frac{1-\theta}{5-3\theta} & \frac{2-\theta}{5-3\theta} \end{pmatrix},$$

kar je odvisno od θ, zato S ni zadostna.

c) Podobno kot v 4. nalogi opazimo, da je:

$$\begin{aligned}&\mathbb{P}_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\ &\quad = p_{x_1} p_{x_1 x_2} p_{x_2 x_3} \cdots p_{x_{n-1} x_n} = \\ &\quad = p_{x_1} p_{00}^{\tau_{00}(x_1,\ldots,x_n)} p_{01}^{\tau_{01}(x_1,\ldots,x_n)} p_{10}^{\tau_{10}(x_1,\ldots,x_n)} p_{11}^{\tau_{11}(x_1,\ldots,x_n)}\,,\end{aligned}$$

kjer je $\tau_{ij}(x_1, \ldots, x_n)$ število pojavitev sosledja i, j v zaporedju $x_1, x_2, \ldots, x_n$. Ustrezna zadostna statistika je torej $(X_1, N_{00}, N_{01}, N_{10}, N_{11})$, kjer je $N_{ij} = \tau_{ij}(X_1, \ldots, X_n)$ število pojavitev sosledja i, j v zaporedju $X_1, X_2, \ldots, X_n$.

7. a) $b = 1 - 5a - 6a^2 = (1 + a)(1 - 6a)$.

b) $0 \le a \le 1/6$.

c) Po Fisher–Neymanovem faktorizacijskem izreku je statistika $T = \tau(X)$ zadostna natanko tedaj, ko sta za poljubna x_1 in x_2, za katera je $\tau(x_1) = \tau(x_2)$, verjetnosti $\mathbb{P}(X = x_1)$ in $\mathbb{P}(X = x_2)$ *sorazmerni*, kar pomeni, da obstaja tak $k \neq 0$, da za vse a in b velja $\mathbb{P}(X = x_2) = k\,\mathbb{P}(X = x_1)$ (nobena od verjetnosti ni vselej enaka nič). Tako definirana sorazmernost je ekvivalenčna relacija in minimalna zadostna statistika bo tista, pri kateri bo $\tau(x_1) = \tau(x_2)$ natanko tedaj, ko sta verjetnosti $\mathbb{P}(X = x_1)$ in $\mathbb{P}(X = x_2)$ sorazmerni.

Opazimo, da sta $\mathbb{P}(X = 1)$ in $\mathbb{P}(X = 4)$ sorazmerni, nadalje so sorazmerne verjetnosti $\mathbb{P}(X = 2)$, $\mathbb{P}(X = 3)$ in $\mathbb{P}(X = 5)$, drugje pa ni sorazmernosti. Minimalna zadostna statistika bo torej oblike:

$$\tau(1) = \tau(4) = t_1\,, \quad \tau(2) = \tau(3) = \tau(5) = t_2\,, \quad \tau(6) = t_3\,,$$

kjer so t_1, t_2 in t_3 same različne vrednosti.

d) Zdaj pa so sorazmerne vse verjetnosti $\mathbb{P}(X = i)$, kjer je $i = 1, 2, 3, 4, 5$, drugih sorazmernosti pa ni. Minimalna zadostna statistika bo torej oblike:

$$\tau(1) = \tau(2) = \tau(3) = \tau(4) = \tau(5) = t_1\,, \quad \tau(6) = t_2\,,$$

kjer sta t_1 in t_2 različni vrednosti.

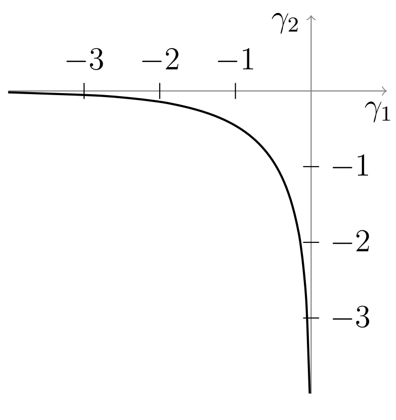
8. Najprej za $X \sim \begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}$ in $x \in \{0, 1\}$ pišimo:

$$\mathbb{P}(X = x) = (1-\theta)^{1-x}\theta^x = e^{(1-x)\ln(1-\theta) + x\ln\theta}\,,$$

kar nam da dvoparametrično eksponentno družino:

$$\gamma_1 = \ln(1-\theta)\,,\ \gamma_2 = \ln\theta\,, \quad h_1(x) = 1 - x\,,\ h_2(x) = x\,, \quad \rho(x) = 1\,, \quad g(\gamma_1, \gamma_2) = 1$$

z naravnim parametričnim prostorom $\{(\gamma_1, \gamma_2) \in \mathbb{R}^2\ ;\ e^{\gamma_1} + e^{\gamma_2} = 1\}$:



Družino pa lahko zapišemo tudi enoparametrično. Velja namreč tudi:

$$\mathbb{P}(X = x) = (1 - \theta) \exp\left(x \ln \frac{\theta}{1 - \theta}\right),$$

kar nam da:

$$\gamma = \ln \frac{\theta}{1 - \theta}, \quad h(x) = x, \quad \rho(x) = 1, \quad g(\gamma) = \frac{1}{1 + e^\gamma}$$

in naravni parametrični prostor je kar cela realna os.

Podobno, če je $X \sim \mathrm{Bin}(n, p)$, lahko za $x \in \{0, 1, \ldots, n\}$ pišemo:

$$\mathbb{P}(X = x) = \binom{n}{x}(1 - \theta)^n \exp\left(x \ln \frac{\theta}{1 - \theta}\right),$$

kar je spet zapis enoparametrične eksponentne družine z:

$$\gamma = \ln \frac{\theta}{1 - \theta}, \quad h(x) = x, \quad \rho(x) = \binom{n}{x}, \quad g(\gamma) = \frac{1}{\left(1 + e^\gamma\right)^n}$$

in naravni parametrični prostor je spet cela realna os.

9. Gre za eksponentno družino, saj lahko verjetnostno funkcijo zapišemo v obliki:

$$\mathbb{P}_\lambda(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \frac{e^{-n\lambda}}{x_1!\, x_2! \cdots x_n!}\, e^{(x_1 + x_2 + \cdots + x_n) \ln \lambda}.$$

Ker je naravni parametrični prostor cela realna os, je vsota res minimalna zadostna statistika.

10. Najprej verjetnostno gostoto za eno samo realizacijo pri splošni normalni porazdelitvi $\mathrm{N}(\mu, \sigma)$ zapišemo v obliki:

$$f(x; \mu, \sigma) = \frac{e^{-\mu^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}\, e^{-x^2/(2\sigma^2) + (\mu/\sigma^2)x},$$

iz katere razberemo, da gre za eksponentno družino z naravnima parametroma $-1/(2\sigma^2)$ in μ/σ^2 ter pripadajočo naravno zadostno statistiko (X^2, X).

a) Vzorčna vsota $X_1 + X_2 + \cdots + X_n$ je minimalna zadostna statistika.

b) Minimalna zadostna statistika je $\sum_{i=1}^n X_i^2 - 2\mu \sum_{i=1}^n X_i$ ali pa tudi $\sum_{i=1}^n (X_i - \mu)^2$.

c) Minimalna zadostna statistika je par $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

d) Minimalna zadostna statistika je $\sum_{i=1}^n X_i^2$.

e) Ker so $a \mapsto 1$, $a \mapsto 1/a$ in $a \mapsto -1/(2a^2)$ linearno neodvisne funkcije, je minimalna zadostna statistika spet par $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$, tako kot če bi bila oba parametra neznana.

11. Pri porazdelitvi Gama(λ, a) je porazdelitvena gostota za eno samo realizacijo enaka:

$$f(x;\lambda,a)=\frac{a^\lambda}{\Gamma(\lambda)}\,x^{\lambda-1}e^{-ax}=\frac{a^\lambda}{\Gamma(\lambda)}\,e^{(\lambda-1)\ln x-ax}\,,$$

torej gre za eksponentno družino z naravnima parametroma $\lambda-1$ in a ter pripadajočo naravno zadostno statistiko $(\ln X, X)$. Zaradi linearne neodvisnosti je potem minimalna zadostna statistika za vzorec lahko:

$$\big(\ln X_1+\ln X_2+\cdots+\ln X_n,\ X_1+X_2+\cdots+X_n\big)\,,$$

lahko pa recimo tudi:

$$\big(X_1X_2\cdots X_n,\ X_1+X_2+\cdots+X_n\big)\,.$$

12. To je recimo $|X_1+X_2+\cdots+X_n|$.

13. Pišemo lahko $X_i=\mu+\sigma Z_i$, kjer so $Z_1,\ldots,Z_n$ neodvisne in porazdeljene standardno normalno (torej poznamo porazdelitev slučajnega vektorja $(Z_1,Z_2,\ldots,Z_n)$).
a) Vzemimo:

$$\big(X_1-\bar X, X_2-\bar X,\ldots,X_n-\bar X\big)=\sigma\big(Z_1-\bar Z, Z_2-\bar Z,\ldots,Z_n-\bar Z\big)\,,$$

kjer je $\bar X=(X_1+X_2+\cdots+X_n)/n$ vzorčno povprečje, na enak način pa je definiran tudi $\bar Z$. Iz zapisa z Z-ji se vidi, da je to res postranska statistika, saj (kot funkcija) ni odvisna od μ.

Dimenzijo zaloge vrednosti preslikave, ki vektor $(x_1,x_2,\ldots,x_n)$ preslika v $(x_1-\bar x, x_2-\bar x,\ldots,x_n-\bar x)$, kjer je $\bar x$ definiran na enak način kot $\bar X$, lahko izračunamo tako, da izračunamo rang njene matrike:

$$\begin{bmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{bmatrix},$$

ki je enak $n-1$, ali pa dokažemo, da je ta preslikava projektor na prostor vseh vektorjev $(x_1,x_2,\ldots,x_n)$, za katere je $\bar x=0$.
b) Vzemimo:

$$\left(\frac{X_1-\mu}{\sqrt{\sum_{i=1}^n(X_i-\mu)^2}},\frac{X_2-\mu}{\sqrt{\sum_{i=1}^n(X_i-\mu)^2}},\ldots,\frac{X_n-\mu}{\sqrt{\sum_{i=1}^n(X_i-\mu)^2}}\right)=$$
$$=\left(\frac{Z_1}{\sqrt{\sum_{i=1}^n Z_i^2}},\frac{Z_2}{\sqrt{\sum_{i=1}^n Z_i^2}}\ \ldots,\frac{Z_n}{\sqrt{\sum_{i=1}^n Z_i^2}}\right)\,.$$

Spet vidimo, da je porazdelitev tega slučajnega vektorja neodvisna od σ in je zato statistika postranska. Njena zaloga vrednosti je enotska sfera v $\mathbb{R}^n$, torej $(n-1)$-dimenzionalen prostor. Še več: iz radialne simetrije gostote slučajnega vektorja

$(Z_1, \ldots, Z_n)$ se vidi, da je naša statistika porazdeljena enakomerno na enotski sferi.
c) Vzemimo:

$$\left(\frac{X_1 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \frac{X_2 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \ldots, \frac{X_n - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \right) =$$
$$= \left(\frac{Z_1 - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}, \frac{Z_2 - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}, \ldots, \frac{Z_n - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}} \right).$$

Zaloga vrednosti tega slučajnega vektorja je presek enotske sfere in množice vseh vektorjev x z $\bar{x} = 0$, kar je $(n-2)$-dimenzionalen prostor. Da se dokazati, da je porazdelitev tudi tega slučajnega vektorja enakomerna na tej množici.

14. Če so $X_1, \ldots, X_n$ neodvisne in porazdeljene normalno $\mathrm{N}(\mu, \sigma)$,

a) sta $\bar{X}$ in $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ neodvisna;

b) sta $\left(\frac{X_1 - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}, \ldots, \frac{X_n - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}} \right)$ in $\sum_{i=1}^n (X_i - \mu)^2$ neodvisna;

c) so $\left(\frac{X_1 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \ldots, \frac{X_n - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \right)$, $\sum_{i=1}^n (X_i - \bar{X})^2$ in $\bar{X}$ neodvisni.

15. Velja:

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 .$$

Iz 14. naloge vemo, da sta statistiki $\sum_{i=1}^n (X_i - \bar{X})^2$ in $n(\bar{X} - \mu)^2$ neodvisni. Poleg tega iz 56. naloge iz 4. razdelka in 28. naloge iz 5. razdelka sledi, da je:

$$\sum_{i=1}^n (X_i - \mu)^2 \sim \mathrm{Gama}\left(\frac{n}{2}, \frac{1}{2\sigma^2} \right) \quad \text{in} \quad n(\bar{X} - \mu)^2 \sim \mathrm{Gama}\left(\frac{1}{2}, \frac{1}{2\sigma^2} \right)$$

ali, ekvivalentno,

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n) \quad \text{in} \quad \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1) .$$

Vemo, da, če sta $U \sim \chi^2(n-1)$ in $V \sim \chi^2(1)$ neodvisni, mora veljati $U + V \sim \chi^2(n)$. Iz teorije momentno-rodovnih ali karakterističnih funkcij pa sledi, da velja tudi obratno: če sta U in V neodvisni ter je $V \sim \chi^2(1)$ in $U + V \sim \chi^2(n)$, mora biti $U \sim \chi^2(n-1)$. Sledi:

$$\sum_{i=1}^n (X_i - \bar{X})^2 \sim \mathrm{Gama}\left(\frac{n-1}{2}, \frac{1}{2\sigma^2} \right) \quad \text{oziroma} \quad \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1) .$$

16. Računajmo:

$$U = \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}} = \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2}} = \frac{T}{\sqrt{n(n-1) + nT^2}} ,$$

kar nam da zahtevano bijektivno korespondenco na dogodku, da niso vse slučajne spremenljivke X_i enake, le-ta pa ima verjetnost ena. Za izračun porazdelitve uporabimo neodvisnost slučajnih spremenljivk $\bar{X} - \mu$ in $\sum_{i=1}^{n}(X_i - \bar{X})^2$ ter rezultat iz 15. naloge, ki pravi, da je $\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \mathrm{Gama}((n-1)/2, 1/(2\sigma^2))$, nakar se skličemo na 30. nalogo iz 5. razdelka, ki pravi, da mora imeti zato T Studentovo porazdelitev z $n-1$ prostostnimi stopnjami.

11. Točkasto ocenjevanje in vzorčenje

1. V obeh primerih so slučajne spremenljivke X_i porazdeljene enako kot X, zato je:

$$\mathbb{E}(\bar{X}) = \frac{n\,\mathbb{E}(X)}{n} = \mathbb{E}(X) = \mu\,,$$

torej gre res za nepristransko cenilko. Srednja kvadratična napaka je torej enaka kar varianci. Če gre za vzorec s ponavljanjem, so slučajne spremenljivke $X_1, \ldots, X_n$ neodvisne, zato je:

$$\operatorname{var}(\bar{X}) = \frac{\operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n)}{n^2} = \frac{\sigma^2}{n}\,,$$

kjer je:

$$\begin{aligned}\sigma^2 = \operatorname{var}(X) &= \frac{(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_N-\mu)^2}{N} = \\ &= \mathbb{E}(X^2) - \mu^2 = \frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N} - \mu^2\end{aligned}$$

populacijska varianca.

Če gre za vzorec brez ponavljanja, pa nastavimo:

$$\begin{aligned}\operatorname{var}(\bar{X}) &= \frac{1}{n^2}\operatorname{var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \operatorname{cov}(X_i, X_j) = \\ &= \frac{1}{n^2}\left[\sum_{i=1}^n \operatorname{var}(X_i) + \sum_{i,j;i\neq j}\operatorname{cov}(X_i, X_j)\right].\end{aligned}$$

Velja $\operatorname{var}(X_i) = \operatorname{var}(X) = \sigma^2$, za $i \neq j$ pa je:

$$\begin{aligned}\mathbb{E}(X_iX_j) &= \frac{1}{N(N-1)}\sum_{k,l;k\neq l} x_k x_l = \\ &= \frac{1}{N(N-1)}\left[\sum_{k=1}^N\sum_{l=1}^N x_k x_l - \sum_{k=1}^N x_k^2\right] = \\ &= \frac{1}{N(N-1)}\big(N^2\mu^2 - N(\mu^2+\sigma^2)\big) = \\ &= \mu^2 - \frac{\sigma^2}{N-1}\,,\end{aligned}$$

torej je $\operatorname{cov}(X_i, X_j) = -\sigma^2/(N-1)$ in končno:

$$\operatorname{var}(\bar{X}) = \frac{N-n}{N-1}\,\frac{\sigma^2}{n}\,.$$

Tako pri vzorcu s ponavljanjem kot brez ponavljanja je cenilka dosledna, ker gre srednja kvadratična napaka proti nič. Pri vzorcu brez ponavljanja gre za končno zaporedje cenilk, pri čemer pri zadnji zajamemo vso populacijo, zato je kar $\bar{X} = \mu$ in seveda $\operatorname{var}(\bar{X}) = 0$.

2. a) Iz 29. naloge v 5. razdelku sledi, da je $X_{(k)} \sim \mathrm{Beta}(k, n-k+1)$. Torej je:

$$\mathbb{E}\big[X_{(k)}\big] = \frac{n!}{(k-1)!\,(n-k)!}\int_0^1 t^k(1-t)^{n-k}\,\mathrm{d}t = \frac{k}{n+1}\,.$$

b) Če je $X \sim \mathrm{Unif}(a,b)$, iz točke a) sledi:

$$\mathbb{E}\big(X_{(k)}\big) = \frac{n+1-k}{n+1}\,a + \frac{k}{n+1}\,b\,.$$

Za cenilko $\hat{q}_p$ iskanega kvantila $q_p = (1-p)a + pb$ nastavimo linearno kombinacijo $\lambda X_{(i)} + \mu X_{(j)}$. Iz nepristranskosti sledi, da mora za poljubna a in b veljati:

$$\frac{\lambda(n+1-i)+\mu(n+1-j)}{n+1}\,a + \frac{\lambda i + \mu j}{n+1}\,b = (1-p)a + pb\,,$$

od koder dobimo sistem:

$$\begin{aligned}\lambda\left(1-\frac{i}{n+1}\right) + \mu\left(1-\frac{j}{n+1}\right) &= 1-p\\ \lambda\,\frac{i}{n+1} + \mu\,\frac{j}{n+1} &= p\,,\end{aligned}$$

ki ima rešitev:

$$\lambda = \frac{j-(n+1)p}{j-i}\,,\qquad \mu = \frac{i-(n+1)p}{i-j}\,.$$

c) Če je $p = i/(n+1)$, je $\hat{q}_p = X_{(i)}$; podobno, če je $p = j/(n+1)$, je $\hat{q}_p = X_{(j)}$. Opazimo, da sta $X_{(i)}$ in $X_{(j)}$ tudi vzorčna kvantila za verjetnosti $i/(n+1)$ in $j/(n+1)$.

Cenilko za kvantil za dano verjetnost p lahko torej opišemo takole: če je $p = i/(n+1)$ ali $p = j/(n+1)$, je ocena enaka kar vzorčnemu kvantilu za verjetnost p. Sicer pa med tema dvema točkama linearno interpoliramo oz. ekstrapoliramo.

d) Če le gre, je smiselno izbrati tista i in j, pri katerih interpoliramo. Brž ko je $1/(n+1) \le p \le n/(n+1)$, lahko i izberemo tako, da je $i/(n+1) \le p \le (i+1)/(n+1)$, in postavimo $j = i+1$. Med tema dvema točkama linearno interpoliramo.

e) Velja $2/9 \le 1/4 \le 3/9$, torej bo ocena enaka:

$$\hat{q}_{1/4} = \left(3 - \frac{9}{4}\right)\cdot 6 + \left(\frac{9}{4} - 2\right)\cdot 14 = 8\,.$$

3. a) Naj bo N moč celotne statistične množice, $N_1 = Np_1, \ldots, N_r = Np_r$ pa moči posameznih stratumov (torej je $N = \sum_{i=1}^r N_i$). Nadalje naj bo x_{ij}, $i = 1, \ldots, r$, $j = 1, \ldots, N_i$, vrednost naše statistične spremenljivke na j-ti enoti i-tega stratuma. Tedaj velja:

$$\mathbb{E}\big[f(X_i)\big] = \frac{1}{N_i}\sum_{j=1}^{N_i} f(x_{ij}) = \frac{1}{Np_i}\sum_{j=1}^{N_i} f(x_{ij})\,.$$

Sledi:

$$\sum_{i=1}^{r} p_i \, \mathbb{E}\big[f(X_i)\big] = \sum_{i=1}^{r} \frac{1}{N} \sum_{j=1}^{N_i} f(x_{ij}) = \mathbb{E}\big[f(X)\big] \,.$$

b) Po prejšnjem je:

$$\mu = p_1\mu_1 + p_2\mu_2 + \cdots + p_r\mu_r \,.$$

Nadalje je:

$$\sigma^2 = \mathbb{E}\big[(X-\mu)^2\big] = \sum_{i=1}^{r} p_i \, \mathbb{E}\big[(X-\mu)^2\big] = \sum_{i=1}^{r} p_i\big(\sigma_i^2 + (\mu_i - \mu)^2\big) \,.$$

Pišemo lahko:

$$\sigma^2 = \sigma_{\mathrm{W}}^2 + \sigma_{\mathrm{B}}^2 \,,$$

kjer je:

$$\sigma_{\mathrm{B}}^2 := p_1(\mu_1 - \mu)^2 + p_2(\mu_2 - \mu)^2 + \cdots + p_r(\mu_r - \mu)^2 \,.$$

varianca *med stratumi* ali *pojasnjena* varianca (pojasnjena z različnostjo stratumov),

$$\sigma_{\mathrm{W}}^2 := p_1\sigma_1^2 + p_2\sigma_2^2 + \cdots + p_r\sigma_r^2$$

pa je varianca *znotraj stratumov* ali *nepojasnjena* varianca.

c) Iskana nepristranska cenilka vzorčnega povprečja je:

$$\bar{X}^{(\mathrm{s})} = p_1\bar{X}_1 + p_2\bar{X}_2 + \cdots + p_r\bar{X}_r \,,$$

njena varianca pa je enaka:

$$\mathrm{var}\big(\bar{X}^{(\mathrm{s})}\big) = \frac{p_1^2\sigma_1^2}{n_1} + \frac{p_2^2\sigma_2^2}{n_2} + \cdots + \frac{p_r^2\sigma_r^2}{n_r} \,.$$

d) Ko vstavimo $n_i = np_i$, dobimo $\mathrm{var}\big(\bar{X}^{(\mathrm{s})}\big) = \dfrac{\sigma_{\mathrm{W}}^2}{n} \le \dfrac{\sigma^2}{n} = \mathrm{var}\big(\bar{X}^{(\mathrm{e})}\big)$, kjer je $\bar{X}^{(\mathrm{e})}$ vzorčno povprečje na nestratificiranem enostavnem slučajnem vzorcu.

e) Poiskati moramo ekstrem variance $\mathrm{var}\big(\bar{X}^{(\mathrm{s})}\big)$, ki smo jo izračunali v točki a), kot funkcije spremenljivk $n_1, n_2, \ldots, n_r$, pri pogoju $n_1 + n_2 + \cdots + n_r = n$. Nastavimo Lagrangeovo funkcijo:

$$L = \frac{p_1^2\sigma_1^2}{n_1} + \frac{p_2^2\sigma_2^2}{n_2} + \cdots + \frac{p_r^2\sigma_r^2}{n_r} - \lambda(n_1 + n_2 + \cdots + n_r) \,.$$

Njeni parcialni odvodi so enaki:

$$\frac{\partial L}{\partial n_i} = -\frac{p_i^2\sigma_i^2}{n_i^2} - \lambda \,.$$

Ko jih postavimo na nič, še iz pogoja po nekaj računanja dobimo:

$$n_i = \frac{p_i\sigma_i}{p_1\sigma_1 + p_2\sigma_2 + \cdots + p_r\sigma_r} \, n \,.$$

Pri tej izbiri dobimo:

$$\operatorname{var}\bigl(\bar{X}^{(\mathrm{s})}\bigr) = \frac{\bigl(p_1\sigma_1 + p_2\sigma_2 + \cdots + p_r\sigma_r\bigr)^2}{n}$$

in iz Jensenove neenakosti sledi, da je to res manjše ali enako σ_{W}^2/n.

4. a) Če enote populacije označimo kar z $1, 2, \ldots, N$, lahko cenilko zapišemo v obliki $\sum_{i=1}^N a_i x_i \,\mathbf{1}(i \in S)$. Njena pričakovana vrednost je enaka $\sum_{i=1}^N a_i x_i \pi_i$ in mora biti enaka populacijskemu povprečju $\frac{1}{N}\sum_{i=1}^N x_i$ ne glede na vrednosti $x_1, \ldots, x_N$ spremenljivke X. To pa je možno le pri $a_i = 1/(N\pi_i)$. Horvitz–Thompsonova cenilka obstaja natanko tedaj, ko je $\pi_i > 0$ za vse i.

b) Pri enostavnem slučajnem vzorcu velikosti n brez ponavljanja je $\pi_i = n/N$, torej je Horvitz–Thompsonova cenilka enaka $\frac{1}{n}\sum_{i\in S} x_i$, torej gre kar za vzorčno povprečje.

c) Pri vzorcu velikosti n s ponavljanjem, zajetem iz populacije velikosti N, je $\pi_i = 1 - \bigl(1 - \frac{1}{N}\bigr)^n$. Horvitz–Thompsonova cenilka je torej enaka:

$$\frac{1}{N\Bigl[1 - \bigl(1 - \frac{1}{N}\bigr)^n\Bigr]} \sum_{i\in S} x_i$$

oziroma za $N = 10$ in $n = 3$:

$$\frac{100}{271} \sum_{i\in S} x_i \,.$$

Vredosti na naših vzorcih: $2\text{˙}21, 1\text{˙}85, 2\text{˙}21, 24\text{˙}35$.

Opazimo, da ocena povprečja pri vzorcu $(21, 22, 23)$ presega maksimum, kar ni smiselno! To je zato, ker Horvitz–Thompsonova cenilka ne da nujno konveksne kombinacije vrednosti spremenljivke in zato ne komutira s translacijami.

d) Horvitz–Thompsonova cenilka komutira s translacijami natanko tedaj, ko je $\sum_{i\in S}(1/\pi_i) = N$ za vse vzorce S. Primer takšnega vzorčnega načrta za populacijo iz treh enot, kjer verjetnosti, da je posamezna enota v vzorcu, niso vse enake:

$$\begin{pmatrix} \{1\} & \{2,3\} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} .$$

Lahko pa Horvitz–Thompsonovo cenilko tudi normaliziramo – vzamemo:

$$\frac{\sum_{i\in S} \frac{x_i}{\pi_i}}{\sum_{i\in S} \frac{1}{\pi_i}} \,.$$

Ta cenilka je robustnejša, ni pa več linearna.

e) Enote populacije indeksirajmo z dvojnimi indeksi (i, j), kar pomeni j-ta enota v i-tem stratumu. Tedaj velja:

$$\pi_{(i,j)} = \frac{k}{K}\,\frac{n_i}{N_i} \,,$$

Če z I_i označimo indikator dogodka, da je i-ti stratum izbran, velja:

$$\sum_{(i,j)\in S} \frac{1}{\pi_{(i,j)}} = \frac{K}{k} \sum_{i=1}^{K} N_i I_i \,.$$

To je za vse vzorce enako N natanko tedaj, ko so stratumi bodisi vsi izbrani bodisi vsi enako veliki, tj. $N_i = N/K$.

f) Če s T označimo Horvitz–Thompsonovo cenilko, velja:

$$\begin{aligned}
T^2 &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\mathbf{1}(x_i \in S, x_j \in S)}{\pi_i \pi_j} \, x_i x_j \,, \\
\mathbb{E}(T^2) &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} \, x_i x_j \,, \\
\big(\mathbb{E}(T)\big)^2 &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \,, \\
\operatorname{var}(T) &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) x_i x_j \,.
\end{aligned}$$

Opomba. Varianca je zagotovo enaka nič, če je $\pi_{ij} = \pi_i \pi_j$ za vse i in j. To pa pomeni, da sta poljubna dogodka $\{i \in S\}$ in $\{j \in S\}$ neodvisna. Za različne i in j to lahko dosežemo: za vsako enoto se neodvisno odločimo, ali jo vzamemo v vzorec ali ne. Za $i = j$ pa je to res le, če je $\pi_i = 0$ (kar za Horvitz–Thompsonovo cenilko ni smiselno) ali $\pi_i = 1$, kar pomeni, da v vzorec zajamemo vso populacijo.

g) Ker Horvitz–Thompsonova cenilka komutira s translacijami, velja:

$$\begin{aligned}
\operatorname{var}(T) &= \operatorname{var}(T - \mu) = \\
&= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) (x_i - \mu)(x_j - \mu) = \\
&= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} \, (x_i - \mu)(x_j - \mu) - \left(\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu) \right)^2 = \\
&= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} \, (x_i - \mu)(x_j - \mu) \,.
\end{aligned}$$

5. Ne, ker se porazdelitev vzorčnega povprečja ujema s porazdelitvijo populacije (glej 45. nalogo iz 8. razdelka) in je zato verjetnost v limiti pri pogoju za doslednost konstantna.

6. a) Iz $z_1 = \mathbb{E}(X) = a/2$ dobimo $\hat{a} = 2\bar{X}$.

b) $\hat{a} = 6{\cdot}8$, kar je nesmiselno glede na to, da je v vzorcu tudi vrednost 10.

c) Če vzamemo r-ti moment, iz $z_r = a^r/(r+1)$ dobimo $\hat{a} = \sqrt[r]{(r+1)\hat{z}_r}$. Tabela prvih nekaj ocen:

r	$\hat{z}_r$	$\hat{a}$
1	3·4	6·8
2	23	8·31
3	207·4	9·40
4	2019·8	10·02

Pri četrtem momentu torej dobimo smiselno oceno. V limiti, ko gre r proti neskončno, pa dobimo 10.

Opomba. Spreminjanje metode ocenjevanja, potem ko že imamo podatke, sicer ni dobra praksa, ker dopušča veliko prostora za manipulacije. Pri samo petih podatkih ne moremo pričakovati prav natančne ocene, tako da ni presenetljivo, če je ocena očitno napačna. Slednje se je zgodilo zaradi vrednosti, ki izstopa. Če pa bi imeli veliko podatkov, med katerimi bi jih bila večina zmerne velikosti, nekaj pa bi jih izstopalo, bi lahko upravičeno posumili, da statistični model z enakomerno porazdelitvijo ni ustrezen.

7. $\hat{\alpha} = \bar{X} = (X_1 + X_2 + \cdots + X_n)/n$. Cenilka je nepristranska. Je tudi dosledna, ker je dobljena po metodi momentov.

8. $\hat{a} = (\hat{z}_2 - 1)/2, \quad \hat{b} = (3 + \hat{z}_1 - 2\hat{z}_2)/2$.
Na našem konkretnem vzorcu iz $\hat{z}_1 = 1/2$ in $\hat{z}_2 = 13/10$ dobimo $\hat{a} = 3/20$ in $\hat{b} = 9/20$.

9. Velja $\mathbb{E}(X) = 0$ (neodvisno od α), zato iz prvega momenta ne dobimo ničesar. Iz drugega momenta dobimo $\hat{\alpha} = \sqrt{\hat{z}_2/2}$.

10. Po metodi momentov dobimo $\bar{X}$ kot cenilko za $\mu := \mathbb{E}(X)$ in:

$$\hat{z}_2 = \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}$$

kot cenilko za $\mathbb{E}(Y^2) = \mu^2 + \sigma^2$. Cenilka za σ^2 bo torej $s_0^2 = \hat{z}_2 - \bar{X}^2$. Z nekaj računanja dobimo, da lahko to zapišemo tudi v obliki:

$$s_0^2 = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

Označimo $Y := X - \mu$. Tedaj velja:

$$\begin{aligned}
s_0^2 &= \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2 = \\
&= \frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2 = \\
&= \frac{1}{n}\sum_{i=1}^n Y_i^2 - \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n Y_i Y_j \\
&= \frac{n-1}{n^2}\sum_{i=1}^n Y_i^2 - \frac{1}{n^2}\sum\sum_{\substack{1\le i,j\le n\\ i\ne j}} Y_i Y_j
\end{aligned}$$

Ker je $\mathbb{E}(Y) = 0$, je očitno:

$$\mathbb{E}(s_0^2) = \frac{n-1}{n}\,\sigma^2\,,$$

torej je s_0^2 pristranska cenilka za σ^2 (je pa *asimptotično nepristranska*). Če postavimo $k' := n/(n-1)$, dobimo, da je:

$$s^2 = \frac{n}{n-1}\,s_0^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$$

nepristranska cenilka za σ^2.

Za izračun srednje kvadratne napake pišimo:

$$\begin{aligned}
s_0^4 = \frac{(n-1)^2}{n^4}\sum_{i=1}^n\sum_{j=1}^n Y_i^2 Y_j^2 - \frac{2(n-1)}{n^4}\sum_{i=1}^n \sum\sum_{\substack{1\le j,k\le n\\ j\ne k}} Y_i^2 Y_j Y_k + \\
+ \frac{1}{n^4}\sum\sum_{\substack{1\le j,k\le n\\ i\ne j}}\ \sum\sum_{\substack{1\le k,l\le n\\ k\ne l}} Y_i Y_j Y_k Y_l = \\
= \frac{(n-1)^2}{n^4}\sum_{i=1}^n Y_i^4 + \frac{(n-1)^2}{n^4}\sum\sum_{\substack{1\le i,j\le n\\ i\ne j}} Y_i^2 Y_j^2 - \frac{2(n-1)}{n^4}\sum_{i=1}^n \sum\sum_{\substack{1\le j,k\le n\\ j\ne k}} Y_i^2 Y_j Y_k + \\
+ \frac{1}{n^4}\sum\sum_{\substack{1\le i,j\le n\\ i\ne j}}\ \sum\sum_{\substack{1\le k,l\le n\\ k\ne l}} Y_i Y_j Y_k Y_l\,.
\end{aligned}$$

Pričakovana vrednost prvih dveh členov v zadnjem izrazu je očitna. Pri tretjem členu opazimo, da je, če je $j \ne k$, $\mathbb{E}(Y_i^2 Y_j Y_k)$ enako bodisi $\mathbb{E}(Y^3)\,\mathbb{E}(Y)$ bodisi $\mathbb{E}(Y^2)\big(\mathbb{E}(Y)\big)^2$, kar je v vsakem primeru enako nič. Za četrti člen, ko je $i \ne j$ in $k \ne l$, pa dobimo, da je $\mathbb{E}(Y_i Y_j Y_k Y_l)$ različno od nič, kvečjemu če je $i = k$ in $j = l$ ali pa $i = l$ in $j = k$. Takih členov je $2n(n-1)$ in njihove pričakovane vrednosti so enake $\big(\mathbb{E}(Y^2)\big)^2 = \sigma^4$. Sledi:

$$\begin{aligned}\mathbb{E}(s_0^4) &= \frac{(n-1)^2}{n^3}\,\kappa^4 + \frac{(n-1)^3}{n^3}\,\sigma^4 + \frac{2(n-1)}{n^3}\,\sigma^4 = \\ &= \frac{(n-1)^2}{n^3}\,\kappa^4 + \frac{(n-1)(n^2-2n+3)}{n^3}\,\sigma^4\,,\end{aligned}$$

kjer je κ^4 četrti centralni moment. Od tod dobimo:

$$\mathrm{MSE}(ks_0^2) = \left(\frac{(n-1)^2\kappa^4}{n^3} + \frac{(n-1)(n^2-2n+3)\sigma^4}{n^3}\right) k^2 - \frac{2(n-1)\sigma^4}{n}\,k + \sigma^4$$

Za $k = 1$ dobimo:

$$\mathrm{MSE}(s_0^2) = \frac{(n-1)^2\kappa^4 + (-n^2+5n-3)\sigma^4}{n^3}$$

(to je vedno nenegativno, ker po Jensenovi neenakosti velja $\kappa \ge \sigma$). Za $k = n/(n-1)$ pa dobimo:

$$\mathrm{MSE}(s^2) = \frac{1}{n}\,\kappa^4 + \frac{(3-n)}{n(n-1)}\,\sigma^4\,.$$

Če je X porazdeljena normalno, je $\kappa^4 = 3\sigma^4$ in velja:

$$\mathrm{MSE}(ks_0^2) = \left(\frac{n^2-1}{n^2}\,k^2 - \frac{2(n-1)}{n}\,k + 1\right)\sigma^4$$

kar je minimalno pri $k = n/(n+1)$. Najučinkovitejša izmed cenilk oblike ks_0^2 je torej:

$$(s_0^*)^2 = \frac{1}{n+1}\sum_{i=1}^{n}(X_i - \bar{X})^2\,.$$

Srednje kvadratične napake cenilk pa so:

$$\mathrm{MSE}(s_0^2) = \frac{2n-1}{n^2}\,\sigma^4\,, \quad \mathrm{MSE}(s^2) = \frac{2}{n-1}\,\sigma^4\,, \quad \mathrm{MSE}\big((s_0^*)^2\big) = \frac{2}{n+1}\,\sigma^4\,.$$

Ni težko preveriti, da je s_0^2 vselej učinkovitejša od s^2 in da je $(s_0^*)^2$ vselej učinkovitejša od s_0^2.

11. Ker vemo, da je $\mathbb{E}(X) = \lambda$ in da je $\bar{X}$ nepristranska cenilka za $\mathbb{E}(X)$, je $\bar{X}$ res nepristranska cenilka za λ. Poleg tega je tudi $\mathrm{var}(X) = \lambda$ (glej npr. 33. nalogo iz 8. razdelka) in iz prejšnje naloge vemo, da je s^2 nepristranska cenilka za $\mathrm{var}(X) = \lambda$.

Izračun variance prve cenilke je preprost:

$$\mathrm{var}(\bar{X}) = \frac{\mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n)}{n^2} = \frac{\lambda}{n}\,.$$

Pri izračunu variance druge cenilke pa se lahko spet opremo na prejšnjo nalogo: ker je cenilka nepristranska, velja:

$$\mathrm{var}(s^2) = \mathrm{MSE}(s^2) = \frac{1}{n}\,\kappa^4 + \frac{(3-n)}{n(n-1)}\,\sigma^4\,,$$

kjer je σ^2 drugi centralni moment (tj. varianca), κ^4 pa je četrti centralni moment. Spet iz 33. naloge iz 8. razdelka poberemo $\kappa^4 = 3\lambda^2 + \lambda$. Sledi:

$$\operatorname{var}(s^2) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}\,.$$

Torej ima s^2 v vsakem primeru večjo varianco kot $\bar{X}$.

12. Naj bo A dogodek, da prvi poskus uspe, drugi ne uspe, tretji pa spet uspe. Tedaj je:

$$\mathbb{P}_\theta(A) = \theta\left(1 - \frac{1+\theta}{2}\right)\frac{\theta}{2} = \frac{\theta^2(1-\theta)}{4}\,.$$

Iščemo maksimum tega izraza za $\theta \in [0,1]$. Iz $\mathbb{P}_0(A) = \mathbb{P}_1(A) = 0$ in $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{P}_\theta(A) = \frac{1}{4}(2\theta - 3\theta^2)$, kar je enako nič pri $\theta = 2/3$, dobimo, da je lahko maksimum dosežen kvečjemu v prej omenjenih točkah. Ker je edino $\mathbb{P}_{2/3}(A) > 0$, se to zgodi pri $\theta = \hat{\theta} := 2/3$, kar je tudi ocena po metodi največjega verjetja.

13. Iz porazdelitvene gostote:

$$f(x;\alpha) = \begin{cases} \frac{1}{\alpha}\,e^{-x/\alpha} & ;\ x > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

dobimo logaritem verjetja:

$$\ln L(\alpha) = \sum_{i=1}^{n} \ln\left(\frac{1}{\alpha}\,e^{-X_i/\alpha}\right) = -n\ln\alpha - \frac{1}{\alpha}\sum_{i=1}^{n} X_i\,.$$

Ta slučajna funkcija je odvedljiva in gre ne glede na opažanje proti minus neskončno, brž ko gre α bodisi proti nič bodisi proti neskončno. Zato vedno doseže maksimum v stacionarni točki. Odvajamo:

$$\frac{\mathrm{d}\ln L}{\mathrm{d}\alpha}(\alpha) = -\frac{n}{\alpha} + \frac{1}{\alpha^2}\sum_{i=1}^{n} X_i\,.$$

in iz prej povedanega sledi, da iskana cenilka $\hat{\alpha}$ zadošča enačbi $\frac{\mathrm{d}\ln L}{\mathrm{d}\alpha}(\hat{\alpha}) = 0$. Po krajšem računu dobimo:

$$\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}\,,$$

kar je ista cenilka kot pri metodi momentov.

14. a) Formula lahko definira gostoto pri $a > 1$, to pa se zgodi, če je $c = a - 1$.

b) Logaritem verjetja pride:

$$\ln L(a) = n\ln(a-1) - a\sum_{i=1}^{n} \ln X_i\,.$$

Ta slučajna funkcija je odvedljiva in gre ne glede na opažanje proti minus neskončno, brž ko gre a bodisi proti nič bodisi proti neskončno. Zato vedno doseže maksimum v stacionarni točki. Iz odvoda:

$$\frac{\mathrm{d} \ln L}{a} = \frac{n}{a-1} - \sum_{i=1}^{n} \ln X_i \,.$$

dobimo iskano cenilko:

$$\hat{a} = \frac{n}{\ln X_1 + \ln X_2 + \cdots + \ln X_n} + 1 \,.$$

15. Verjetje je enako:

$$L = (1 - 2a - b)^2 a^5 b^3 \,.$$

Je diferenciabilno, na robu parametričnega prostora $\{(a, b) \; ; \, a \geq 0, b \geq 0, 2a + b \leq 1\}$ enako nič in v notranjosti parametričnega prostora strogo večje od nič. Zato je maksimum dosežen v stacionarni točki v notranjosti, poleg tega pa lahko tudi logaritmiramo:

$$\ln L = 2 \ln(1 - 2a - b) + 5 \ln a + 3 \ln b \,,$$

$$\frac{\partial \ln L}{\partial a} = -\frac{4}{1 - 2a - b} + \frac{5}{a} \,, \qquad \frac{\partial \ln L}{\partial b} = -\frac{2}{1 - 2a - b} + \frac{3}{b} \,.$$

Ko odvoda izenačimo z nič, po krajšem računu dobimo oceni $\hat{a} = 1/4$ in $\hat{b} = 3/10$, kar se ne ujema z ocenama po metodi momentov.

16. Postavimo se v položaj, preden smo določili genotipe. V vzorcu je n osebkov. Za opažanje lahko vzamemo podatek, da jih ima N_1 genotip RR, N_2 genotip Rr in N_3 genotip rr ($N_1 + N_2 + N_3 = n$): opažanje torej sestoji iz vektorja (N_1, N_2, N_3). Zdaj pa vzemimo širše opažanje, kjer osebke v vzorcu oštevilčimo in zabeležimo, da ima i-ti osebek genotip X_i: to opažanje sestoji iz vektorja $(X_1, X_2, \ldots, X_n)$ in verjetje je enako:

$$L = \big((1 - \theta)^2\big)^{N_1} \big(2\theta(1 - \theta)\big)^{N_2} (\theta^2)^{N_3} = (1 - \theta)^{2N_1 + N_2} \theta^{N_2 + 2N_3} \,,$$

kar je funkcija vektorja (N_1, N_2, N_3), zato je le-ta zadostna statistika. To pomeni, da lahko, če nam ustreza, raje gledamo verjetje, ki pripada vektorju $(X_1, X_2, \ldots, X_n)$, torej tisto, zapisano zgoraj.

Verjetje je odvedljiva funkcija parametra θ in brž ko je bodisi $2N_1 + N_2 > 0$ bodisi $N_2 + 2N_3 > 0$, je na robu parametričnega prostora $[0, 1]$ enako nič, zunaj njega pa strogo večje od nič, torej je maksimum dosežen v stacionarni točki v notranjosti. Lahko tudi logaritmiramo. Dobimo:

$$\ln L = (2N_1 + N_2) \ln(1 - \theta) + (N_2 + 2N_3) \ln \theta \,,$$

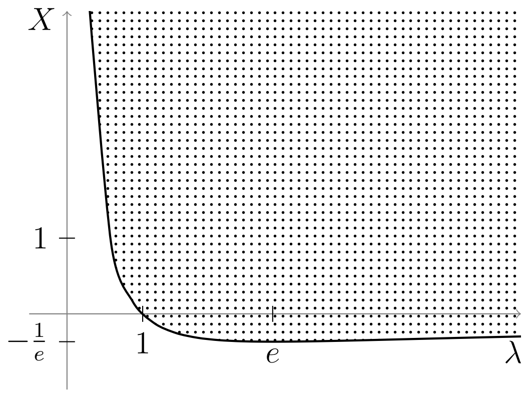
$$\frac{\mathrm{d} \ln L}{\mathrm{d}\theta} = -\frac{2N_1 + N_2}{1 - \theta} + \frac{N_2 + 2N_3}{\theta} \,.$$

Odvod izenačimo z nič in po krajšem računu dobimo cenilko:

$$\hat{\theta} = \frac{N_2 + 2N_3}{2n} .$$

V našem konkretnem primeru pride $\hat{\theta} = 0\text{'}65$.

17. Opazimo, da je verjetje kot funkcija parametra λ pri $X > 0$ strogo padajoče, pri $X > 0$ strogo naraščajoče, pri $X = 0$ pa konstantno. To seveda velja v okviru predpisane zaloge vrednosti statistične spremenljivke X, zato se splača narisati sliko:



Funkcija $\lambda \mapsto -\frac{\ln \lambda}{\lambda}$ je na intervalu $(0, e]$ padajoča, pri $\lambda = e$ doseže minimum, na intervalu $[e, \infty)$ pa je naraščajoča. Velja še $\lim_{\lambda \downarrow 0}\bigl(-\frac{\ln \lambda}{\lambda}\bigr) = \infty$ in $\lim_{\lambda \to \infty}\bigl(-\frac{\ln \lambda}{\lambda}\bigr) = 0$. Sledi:

- Pri $X > 0$ je cenilka edina rešitev enačbe $-\frac{\ln \hat{\lambda}}{\hat{\lambda}} = X$.
- Pri $X = 0$ cenilka ni natančno določena (funkcija verjetja je enaka za vse λ).
- Pri $-e^{-1} \le X < 0$ je cenilka *večja* rešitev enačbe $-\frac{\ln \hat{\lambda}}{\hat{\lambda}} = X$.
- Opažanje $X < -e^{-1}$ je v nasprotju z našim statističnim modelom.

18. Po metodi momentov dobimo $\bar{X}$ kot nepristransko cenilko za $\mathbb{E}(X) = a/2$, torej je tudi $M = 2\bar{X}$ nepristranska cenilka za a.

Poiščimo še cenilko po metodi največjega verjetja. Verjetje je enako:

$$L(a) = \begin{cases} 1/a^n & ; \, a \ge X_1, X_2, \ldots, X_n \\ 0 & ; \, \text{sicer} \end{cases}$$

in je maksimalno pri $a = V := \max\{X_1, \ldots, X_n\}$; to je iskana cenilka. Za izračun njenega matematičnega upanja izhajamo iz kumulativne porazdelitvene funkcije, ki je za $x \in [0, a]$ enaka:

$$F_V(x) = \mathbb{P}(V < x) = \mathbb{P}(X_1 < x, X_2 < x, \ldots, X_n < x) = \left(\frac{x}{a}\right)^n .$$

Od tod dobimo verjetnostno gostoto:

$$f_V(x) = \begin{cases} nx^{n-1}/a^n & ; 0 \le x \le a \\ 0 & ; \text{sicer}, \end{cases}$$

iz nje pa izračunamo $\mathbb{E}(V) = \dfrac{n}{n+1}\, a$, torej je cenilka V pristranska (je pa *asimptotično nepristranska*). Nadalje iz $\mathbb{E}(V^2) = \dfrac{n}{n+2}\, a^2$ izračunamo:

$$q(V) = \mathbb{E}\big[(V-a)^2\big] = \mathbb{E}(V^2) - 2a\,\mathbb{E}(V) + a^2 = \frac{2a^2}{(n+1)(n+2)}\,.$$

Pri $n = 1$ in $n = 2$ sta M in V enako učinkoviti, pri $n \ge 3$ pa je V učinkovitejša.

b) Cenilka $V_1 := \dfrac{n+1}{n}\, V$ je nepristranska in velja:

$$q(V_1) = \operatorname{var}(V_1) = \frac{(n+1)^2}{n^2}\operatorname{var}(V) = \frac{(n+1)^2}{n^2}\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) a^2 = \frac{a^2}{n(n+2)}\,.$$

Pri $n = 1$ sta V in V_1 obe enako učinkoviti, pri $n \ge 2$ pa je V_1 učinkovitejša.

c) Oglejmo si vse cenilke oblike kV, kjer je k konstanta. Velja:

$$q(kV) = \left[\frac{n}{n+2}\, k^2 - \frac{2n}{n+1}\, k + 1\right] a^2\,,$$

kar je minimalno pri $k = k^* := (n+2)/(n+1)$, neodvisno od a. Velja:

$$q(k^* V) = \frac{a^2}{(n+1)^2}\,,$$

kar pomeni, da je ta cenilka vselej učinkovitejša od vseh ostalih.

19. Pripadajoča zadostna statistika je število uspelih poskusov, ki ga označimo z S. Za iskano cenilko $h(S)$ bo torej moralo veljati:

$$\mathbb{E}\big[h(S)\big] = p^2\,.$$

Ker je S porazdeljena binomsko $\mathrm{Bin}(3, p)$, to pomeni:

$$h(0)\,(1-p)^3 + 3\,h(1)\,p(1-p)^2 + 3\,h(2)\,p^2(1-p) + h(3)\,p^3 = p^2$$

za vse $p \in (0, 1)$. S primerjavo koeficientov dobimo, da bo to natanko tedaj, ko bo:

$$h(0) = 0\,, \quad h(1) = 0\,, \quad h(2) = \tfrac{1}{3}\,, \quad h(3) = 1\,.$$

Za posplošitev cenilke na n poskusov le-to iščemo kot polinom pripadajoče zadostne statistike S. Iz $\mathbb{E}(S) = np$ in $\mathbb{E}(S^2) = np + (n^2 - n)p^2$ dobimo $\mathbb{E}(S^2 - S) = (n^2 - n)p^2$, torej bo iskana cenilka enaka:

$$\frac{S^2 - S}{n^2 - n}$$

in zlahka se lahko prepričamo, da se za $n = 3$ ujema s prej dobljeno cenilko.

Opomba. Ker je $\mathbb{E}(S^k)$ polinom stopnje k v θ in ker je vsako funkcijo na množici $\{0, 1, \ldots, n\}$ možno zapisati kot polinom stopnje največ n, nepristranske cenilke obstajajo le za karakteristike, ki so polinomi parametra θ stopnje največ n.

20. a) Iz generičnega eksponentnega zapisa verjetnostne funkcije:

$$P_X(x) = \\ = \exp\left(\rho_0(x)\ln\frac{1}{1+3\theta+2\theta^2} + \rho_1(x)\ln\frac{\theta}{1+3\theta+2\theta^2} + \rho_2(x)\ln\frac{\theta^2}{1+3\theta+2\theta^2}\right)$$

dobimo triparametrični zapis s pripadajočo zadostno statistiko $\big(\rho_0(X), \rho_1(X), \rho_2(X)\big)$. Iz zapisa:

$$P_X(x) = \frac{\rho_0(x) + 3\,\rho_1(x) + 2\,\rho_2(x)}{1+3\theta+2\theta^2}\, e^{x\ln\theta}$$

pa dobimo enoparametričen zapis s pripadajočo zadostno statistiko X.
b) Cenilka $\varphi(X)$ bo nepristranska natanko tedaj, ko bo za vsak θ veljalo:

$$\mathbb{E}\big[\varphi(X)\big] = \frac{\varphi(0) + 3\,\varphi(1)\,\theta + 2\,\varphi(2)\,\theta^2}{1+3\theta+2\theta^2} = \frac{1}{1+\theta}\,,$$

kar je ekvivalentno:

$$\varphi(0) + 3\,\varphi(1)\,\theta + 2\,\varphi(2)\,\theta^2 = 1 + 2\theta\,.$$

Cenilka bo torej nepristranska natanko tedaj, ko bo $\varphi(0) = 1$, $\varphi(1) = 2/3$ in $\varphi(2) = 0$. Ker je ena sama, ima seveda tudi najmanjšo možno varianco.

21. a) Seveda gre za eksponentno družino z naravnim parametrom λ in pripadajočo zadostno statististiko $S = X_1 + \cdots + X_n$. Iščemo funkcijo φ, za katero je $\mathbb{E}\big[\varphi(S)\big] = \lambda$ za vse $\lambda > 0$. Statistika S ima porazdelitev Gama(n, λ), torej je:

$$\mathbb{E}\big[\varphi(S)\big] = \frac{\lambda^n}{\Gamma(n)} \int_0^\infty \varphi(x)\, x^{n-1}\, e^{-\lambda x}\, \mathrm{d}x\,.$$

Veljati mora torej:

$$\int_0^\infty \varphi(x)\, x^{n-1}\, e^{-\lambda x}\, \mathrm{d}x = \frac{\Gamma(n)}{\lambda^{n-1}}\,.$$

Opazimo, da je izraz pod integralom Laplaceova transformiranka funkcije $x \mapsto x^{n-1}\varphi(x)$. Iz tabele Laplaceovih transformirank po nekaj računanja razberemo, da danemu pogoju ustreza funkcija $\varphi(x) = (n-1)/x$, torej bo iskana cenilka $(n-1)/S$.
b) Krajši račun pokaže:

$$\mathrm{MSE}\left(\frac{a}{S}\right) = \lambda^2\left(\frac{a^2}{(n-1)(n-2)} - \frac{2a}{n-1} + 1\right)\,.$$

od koder dobimo, da je srednja kvadratična napaka minimalna pri $a = n - 2$.

22. Seveda gre za eksponentno družino s pripadajočo zadostno statistiko $S = X_1 + \cdots + X_n$. Razmeroma lahko je opaziti, da je statistika $\mathbf{1}(X_1 = 1)$ nepristranska cenilka za p. Za $n = 1$ je to tudi iskana statistika. Pri $n > 1$ pa iz:

$$\mathbb{P}(X_1 = 1 \mid S = k) = \frac{n-1}{k-1}$$

razberemo iskano cenilko $\dfrac{n-1}{S-1}$.

23. a) Verjetnostna funkcija:

$$P_X(k; \lambda) = \frac{\lambda^k\, e^{-\lambda}}{k!}$$

nam da:

$$L(\lambda) = \frac{\lambda^X\, e^{-\lambda}}{X!}, \qquad \frac{\mathrm{d}L}{\mathrm{d}\lambda} = \frac{(X-\lambda)\lambda^{X-1}\, e^{-\lambda}}{X!},$$

od koder dobimo cenilko $\hat{\lambda} = X$, cenilka za λ^2 pa je X^2. Le-ta je pristranska, saj je $\mathbb{E}(X^2) = \lambda^2 + \lambda$.

b) Ker gre za eksponentno družino, katere naravni parametrični prostor ima neprazno notranjost, bo nepristranska cenilka imela enakomerno najmanjšo možno varianco, brž ko bo funkcija minimalne zadostne statistike X. Iščemo torej tako funkcijo φ, da bo $\mathbb{E}_\lambda\big(\varphi(X)\big) = \lambda^2$ za vse λ. To lahko naredimo z nastavkom:

$$\sum_{k=0}^{\infty} \frac{\varphi(k)\, \lambda^k\, e^{-\lambda}}{k!} = \lambda^2$$

oziroma:

$$\sum_{k=0}^{\infty} \frac{\varphi(k)\, \lambda^k}{k!} = \lambda^2\, e^{\lambda},$$

nakar z razvojem:

$$\lambda^2\, e^{\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^{n+2}}{n!} = \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} = \sum_{k=2}^{\infty} \frac{k(k-1)\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{k(k-1)\lambda^k}{k!}$$

in primerjavo koeficientov dobimo, da mora biti $\varphi(k) = k(k-1)$, torej je iskana statistika $X(X-1)$. Le-to lahko tudi kar uganemo iz prvih dveh momentov Poissonove porazdelitve.

c) Velja:

$$\begin{aligned}
\mathrm{MSE}(X^2 - aX) &= \mathbb{E}\big[(X^2 - aX - \lambda^2)^2\big] = \\
&= \mathbb{E}(X^4) + a^2\,\mathbb{E}(X^2) + \lambda^4 - 2a\,\mathbb{E}(X^3) - 2\lambda^2\,\mathbb{E}(X^2) + 2a\lambda^2\,\mathbb{E}(X) = \\
&= 4\lambda^3 + (7 - 6a + a^2)\lambda^2 + (1 - 2a + a^2)\lambda\,.
\end{aligned}$$

Najnatančnejša cenilka bi bila tista, ki bi imela pri *vseh* λ najmanjšo srednjo kvadratično napako; smiselna pa je tista, za katero ne obstaja nobena druga vrednost a, pri kateri bi bila srednja kvadratična napaka strogo manjša za vse λ. Z drugimi besedami, cenilke delno uredimo, in sicer naj bo $\hat{\lambda}_1 \preceq \hat{\lambda}_2$, če je $\mathrm{MSE}(\hat{\lambda}_1) \le \mathrm{MSE}(\hat{\lambda}_2)$ za vse λ. Najboljša cenilka je tista, ki je glede na to ureditev najmanjši element, smiselna pa je taka, ki je minimalni element.

Delno urejenost cenilk v izbrani obliki lahko izrazimo s koeficientoma $c_1(a) := 1 - 2a + a^2$ in $c_2(a) = 7 - 6a + a^2$: za $\hat{\lambda}_1 = X^2 - a_1 X$ in $\hat{\lambda}_2 = X^2 - a_2 X$ je namreč $\hat{\lambda}_1 \preceq \hat{\lambda}_2$ natanko tedaj, ko je $c_1(a_1) \le c_1(a_2)$ in $c_2(a_1) \le c_2(a_2)$. Za $a < 1$ sta oba koeficienta strogo padajoča, za $a > 3$ sta oba strogo naraščajoča, za $1 \le a \le 3$ pa je c_1 naraščajoč, c_2 pa padajoč. Od tod sledi, da najmanjšega elementa iz te družine ni, minimalne pa dobimo za $1 \le a \le 3$. To so torej smiselne cenilke, najboljše pa ni.

24. a) Če pišemo $a = e^{-3\lambda}$ in $\hat{a} = h(X)$, mora za vsak λ veljati:

$$\mathbb{E}\big[h(X)\big] = \sum_{k=0}^{\infty} f(k)\,\frac{\lambda^k\, e^{-\lambda}}{k!} = e^{-3\lambda}\,,$$

torej:

$$\sum_{k=0}^{\infty} f(k)\,\frac{\lambda^k}{k!} = e^{-2\lambda} = \sum_{k=0}^{\infty} \frac{(-2)^k \lambda^k}{k!}\,.$$

Iz enoličnosti potenčnih vrst sledi $f(k) = (-2)^k$, torej je *edina možna* nepristranska cenilka $\hat{a} = (-2)^X$. Toda ta cenilka lahko zavzame zelo velike vrednosti, čeprav je karakteristika, ki jo ocenjuje, strogo pozitivna.

b) Cenilka po metodi največjega verjetja je e^{-3X}. Iz:

$$\mathrm{MSE}\big[a^X\big] = e^{(a^2-1)\lambda} - 2\,e^{(a-4)\lambda} + e^{-6\lambda}$$

dobimo:

$$\mathrm{MSE}\big[(-2)^X\big] - \mathrm{MSE}\big[e^{-3\lambda}\big] = e^{3\lambda} - e^{(e^{-6}-1)\lambda} + 2\,e^{(e^{-3}-4)\lambda} - 2\,e^{-6\lambda} > 0$$

za vse $\lambda > 0$.

c) Cenilko iščemo kot funkcijo minimalne zadostne statistike, ki je v našem primeru recimo $S_n = X_1 + X_2 + \cdots + X_n$. Ker gre za eksponentno družino, katere parametrični prostor vsebuje odprto množico, bo imela taka cenilka tudi enakomerno najmanjšo možno varianco med vsemi nepristranskimi.

Nastavimo torej $\hat{a} = h_n(S_n)$. Ker je $S_n \sim \mathrm{Pois}(n\lambda)$, bo moralo veljati:

$$\mathbb{E}\big[h_n(S_n)\big] = \sum_{k=0}^{\infty} h_n(k)\,\frac{n^k \lambda^k\, e^{-n\lambda}}{k!} = e^{-3\lambda}\,,$$

torej:

$$\sum_{k=0}^{\infty} f(k)\,\frac{n^k \lambda^k}{k!} = e^{(n-3)\lambda} = \sum_{k=0}^{\infty} \frac{(n-3)^k \lambda^k}{k!}\,.$$

Dobimo cenilko $\hat{a} = \left(\dfrac{n-3}{n}\right)^S$, kar se za velike n bliža $e^{-3S/n}$, to pa je ravno cenilka po metodi največjega verjetja.

12. Intervali zaupanja

1. Obravnavajmo najprej primer, ko je $n = 1$. Najprej opazimo, da mora interval za $S = 0$ vsebovati določen interval oblike $[0, \varepsilon)$, $\varepsilon > 0$, sicer je lahko verjetnost pokritosti poljubno majhna. Torej mora biti $[0, b_0)$ ali $[0, b_0]$, kjer je $b_0 > 0$. Zaradi simetrije mora biti potem interval za opažanje $S = 1$ enak $(1 - b_0, 1]$ oz. $[1 - b_0, 1]$. Brž ko je $b_0 < 1/2$ ali pa je $b_0 = 1/2$ in je interval za $S = 0$ odprt pri b_0, interval pri $\theta = 1/2$ ni pokrit in je torej verjetnost pokritosti enaka nič. Pri $b_0 = 1/2$ in zaprti različici je verjetnost pokritosti enaka:

$$\begin{cases} 1 - \theta & ; 0 \le \theta \le 1/2 \\ \theta & ; 1/2 \le \theta \le 1 \end{cases},$$

torej je minimalna verjetnost pokritosti enaka $1/2$. Pri $b_0 > 1/2$ in odprti različici pa je verjetnost pokritja enaka:

$$\begin{cases} 1 - \theta & ; 0 \le \theta \le 1 - b_0 \\ 1 & ; 1 - b_0 < \theta < b_0 \\ \theta & ; b_0 \le \theta \le 1 \end{cases},$$

torej je minimalna verjetnost pokritosti enaka b_0. Pri $\beta > 1/2$ bo torej iskani interval oblike:

$$\begin{cases} [0, \beta) & ; S = 0 \\ (1 - \beta, 1] & ; S = 1 \end{cases},$$

pri $\beta \le 1/2$ pa bo oblike:

$$\begin{cases} [0, 1/2] & ; S = 0 \\ [1/2, 1] & ; S = 1 \end{cases}.$$

Da se dokazati, da je to edina rešitev, ki izpolnjuje pogoje naloge.

Oglejmo si zdaj še primer, ko je $n = 2$. Podobno kot prej ugotovimo, da mora biti interval pri opažanju $S = 0$ oblike $[0, b_0)$ ali $[0, b_0]$ $(b > 0)$, kar za $S = 2$ da $(1 - b_0, 1]$ oziroma $[1 - b_1, 1]$. Pri $S = 1$ pa je interval lahko oblike $(1 - b_1, b_1)$ ali $[1 - b_1, b_1]$ (za $b_1 > 1/2$) ali pa tudi kar $\{1/2\}$. Zaradi monotonosti mora biti $b_1 \ge b_0$, poleg tega pa mora biti tudi $b_0 + b_1 \ge 1$ (oz. $b_0 \ge 1/2$, če je interval za $S = 1$ kar $\{1/2\}$), sicer za $b_0 < \theta < 1 - b_1$ (oz. za $b_0 < \theta < 1/2$) interval ni pokrit. Naj bo najprej $b_0 > 1/2$ in naj bo za $S = 0$ interval oblike $[0, b_0)$, za $S = 1$ pa oblike $(1 - b_1, b_1)$. Tedaj je verjetnost pokritosti enaka:

$$\begin{cases} (1 - \theta)^2 & ; 0 \le \theta \le 1 - b_0 \\ 1 - \theta^2 & ; 1 - b_1 < \theta \le 1 - b_0 \\ 1 & ; 1 - b_0 < \theta < b_0 \\ 2\theta - \theta^2 & ; b_0 \le \theta < b_1 \\ \theta^2 & ; \theta \ge b_1 \end{cases}$$

in minimalna verjetnost pokritosti bo $\min\{2b_0 - b_0^2, b_1^2\}$. Za $\beta > 3/4$ je možno nastaviti $2b_0 - b_0^2 = b_1^2 = \beta$: postavimo namreč $b_0 := 1 - \sqrt{1 - \beta}$ in $b_1 := \sqrt{\beta}$.

Minimalni interval zaupanja je torej:

$$\begin{cases} [0, 1-\sqrt{1-\beta}) & ; S = 0 \\ (1-\sqrt{\beta}, \sqrt{\beta}) & ; S = 1 \\ (\sqrt{1-\beta}, 1] & ; S = 2 \end{cases} .$$

in da se dokazati, da je edini minimalni interval zaupanja te oblike.

Pri $1/2 < \beta \le 3/4$ bo še vedno $b_1 = \sqrt{\beta} > 1/2$, medtem ko bo $b_0 = 1 - \sqrt{1-\beta} \le 1/2$. Toda za $b_0 \le 1/2$ bo verjetnost pokritosti pri prejšnjem intervalu zaupanja enaka:

$$\begin{cases} (1-\theta)^2 & ; 0 \le \theta \le 1 - b_0 \\ 1-\theta^2 & ; 1 - b_1 < \theta < b_0 \\ 2\theta(1-\theta) & ; b_0 \le \theta \le 1 - b_0 \\ 2\theta - \theta^2 & ; b_0 \le \theta < b_1 \\ \theta^2 & ; \theta \ge b_1 \end{cases}$$

in minimalna verjetnost pokritosti bo $\min\{2b_0 - 2b_0^2, b_1^2\} \le 1/2$, kar ne bo v redu. Od tod dobimo, da je minimalni interval zaupanja pri $1/2 < \beta \le 3/4$ oblike:

$$\begin{cases} [0, 1/2] & ; S = 0 \\ \left(1-\sqrt{\beta}, \sqrt{\beta}\right) & ; S = 1 \\ [1/2, 1] & ; S = 2 \end{cases} .$$

Sistem enačb $2b_0 - b_0^2 = b_1^2 = \beta$ ima v okviru naših pogojev rešitev $b_0 = \frac{1}{2}\left(1 - \sqrt{1-2\beta}\right)$, $b_1 = \sqrt{\beta}$. Toda zdaj moramo paziti na pogoj $b_0 + b_1 \ge 1$. Krajši račun pokaže, da gre to le pri $\beta \ge 4/9$; pri $\beta = 4/9$ dobimo $b_0 + b_1 = 1$, kjer moramo bolj paziti, zato bomo ta primer obravnavali kasneje. Torej bo minimalni interval zaupanja pri $4/9 < \beta \le 1/2$ enak:

$$\begin{cases} \left[0, \frac{1}{2}\left(1-\sqrt{1-2\beta}\right)\right) & ; S = 0 \\ \left(1-\sqrt{\beta}, \sqrt{\beta}\right) & ; S = 1 \\ \left(\frac{1}{2}\left(1+\sqrt{1-2\beta}\right), 1\right] & ; S = 2 \end{cases} .$$

Za $\beta \le 4/9$ pa gremo lahko v dve smeri: lahko določimo $b_0 = \frac{1}{2}\left(1 - \sqrt{1-2\beta}\right)$ in $b_1 = 1 - b_0$ ali pa $b_1 = \sqrt{\beta}$ in $b_0 = 1 - b_1$. To nas pripelje do naslednjih družin intervalov zaupanja:

$$\begin{cases} [0, 1-b_1) & ; S = 0 \\ [1-b_1, b_1] & ; S = 1 \\ (b_1, 1] & ; S = 2 \end{cases}, \quad \begin{cases} [0, 1-b_1] & ; S = 0 \\ (1-b_1, b_1) & ; S = 1 \\ [b_1, 1] & ; S = 2 \end{cases} \quad \text{in} \quad \begin{cases} \left[0, \frac{1}{2}\right) & ; S = 0 \\ \left\{\frac{1}{2}\right\} & ; S = 1 \\ \left(\frac{1}{2}, 1\right] & ; S = 2 \end{cases}$$

Vsak interval zaupanja iz te družine je minimalen, pogoje pa izpolnjuje, brž ko je njegova verjetnost pokritja vsaj β: za prvi dve družini to pomeni $b_1^2 \le \beta$ in $2b_1(1-b_1) \ge \beta$ (rešitev tega sistema neenačb na b_1 je prepuščena bralcu), tretji interval zaupanja pa izpolnjuje pogoje, brž ko je $\beta \le 1/4$. Pri $\beta < 4/9$ torej dobimo bistveno različne minimalne intervale zaupanja.

2. Označimo dano statistično spremenljivko z X. Najprej opazimo, da je X pri $\theta = 1$ z verjetnostjo 1 enak 0, kar pomeni, da mora interval zaupanja pri opažanju $X = 0$ obvezno vsebovati tudi 1. Zaradi predpisane oblike mora biti torej to kar interval $[0, 1]$. Preostala dva parametra, b_1 in b_2, lahko nastavimo strogo manjša od 1.

Če nastavimo $b_1 > b_2$, velja:

$$\mathbb{P}_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ \mathbb{P}_\theta(X = 0) + \mathbb{P}_\theta(X = 1) & ; b_2 \le \theta < b_1 \\ \mathbb{P}_\theta(X = 0) & ; b_1 \le \theta \le 1 \end{cases}$$

oziroma:

$$\mathbb{P}_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ \dfrac{\theta}{\theta^2 - \theta + 1} & ; b_2 \le \theta < b_1 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & ; b_1 \le \theta \le 1 \end{cases} .$$

Ni se težko prepričati, da sta funkciji $\theta \mapsto \frac{\theta}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0, 1]$ obe naraščajoči, torej je:

$$\inf_{0\le\theta\le1} \mathbb{P}_\theta(\theta \in I) = \min\left\{ \frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_1^2 - b_1 + 1} \right\} .$$

Interval bo minimalen, če bo kar:

$$\frac{b_2}{b_2^2 - b_2 + 1} = \frac{b_1^2}{b_1^2 - b_1 + 1} = \beta .$$

Ker sta funkciji $\theta \mapsto \frac{\theta}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0, 1]$ obe naraščajoči in ker za $\theta \in (0, 1)$ velja $\frac{\theta}{\theta^2-\theta+1} > \frac{\theta^2}{\theta^2-\theta+1}$, je pri zgornji izbiri tudi $b_1 > b_2$. Velja:

$$b_1 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)}, \quad b_2 = \frac{1 + \beta - \sqrt{(1 + 3\beta)(1 - \beta)}}{2\beta} .$$

Pri $\beta = 0{\cdot}95$ dobimo $b_1 \doteq 0{\cdot}9523$ in $b_2 \doteq 0{\cdot}7954$ (obakrat smo zaokrožili navzgor).

Ni pa to edina možnost. Če postavimo $b_1 < b_2$, dobimo:

$$\mathbb{P}_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ \mathbb{P}_\theta(X = 0) + \mathbb{P}_\theta(X = 2) & ; b_1 \le \theta < b_2 \\ \mathbb{P}_\theta(X = 0) & ; b_2 \le \theta \le 1 \end{cases}$$

oziroma:

$$\mathbb{P}_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_1 \\ \dfrac{2\theta^2 - 2\theta + 1}{\theta^2 - \theta + 1} & ; b_1 \le \theta < b_2 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & ; b_2 \le \theta \le 1 \end{cases} .$$

Kot smo že omenili, je funkcija $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0,1]$ naraščajoča. Funkcija $\theta \mapsto \frac{2\theta^2-2\theta+1}{\theta^2-\theta+1}$ pa je na intervalu $\left[0,\frac{1}{2}\right]$ padajoča, na intervalu $\left[\frac{1}{2},1\right]$ pa naraščajoča; v $\frac{1}{2}$ ima minimum $\frac{2}{3}$. Zato mora biti $b_1 \geq \frac{1}{2}$, brž ko je $\beta \geq \frac{1}{2}$. V tem primeru velja:

$$\inf_{0\leq\theta\leq 1} \mathbb{P}_\theta(\theta \in I) = \min\left\{\frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_2^2 - b_2 + 1}\right\}.$$

Interval bo minimalen, če bo kar:

$$\frac{2b_1^2 - 2b_1 + 1}{b_1^2 - b_1 + 1} = \frac{b_2^2}{b_2^2 - b_2 + 1} = \beta\,.$$

Spet ker sta funkciji $\theta \mapsto \frac{2\theta^2-2\theta+1}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $\left[\frac{1}{2},1\right]$ obe naraščajoči in ker za $\theta \in \left(\frac{1}{2},1\right)$ velja $\frac{2\theta^2-2\theta+1}{\theta^2-\theta+1} > \frac{\theta^2}{\theta^2-\theta+1}$, je pri zgornji izbiri tudi $b_1 < b_2$. Velja:

$$b_1 = \frac{2 - \beta + \sqrt{(3\beta - 2)(2 - \beta)}}{2(2 - \beta)}\,, \quad b_2 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)}\,.$$

Pri $\beta = 0{\cdot}95$ dobimo $b_1 \doteq 0{\cdot}9499$ in $b_2 \doteq 0{\cdot}9523$ (obakrat smo zaokrožili navzgor). Ta izbira je torej slabša glede na seštevek dolžin pri vseh opazovanjih. Spet je podobno kot prej dovolj, da je velja le ena enakost, drugi parameter pa se lahko poveča (a potem dobimo še daljši interval zaupanja za enako stopnjo zaupanja).

Za $b_1 = b_2$ pa bi moralo veljati:

$$b_1 = b_2 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)}$$

(torej $0{\cdot}9523$ pri $\beta = 0{\cdot}95$), kar ni minimalno, saj ta interval strogo vsebuje katero koli izmed prejšnjih dveh možnosti.

3. Po definiciji intervala zaupanja mora biti verjetnost, da je prava vrednost parametra θ res v intervalu zaupanja, pri vseh θ enaka najmanj β. Vrednost $\theta < \theta^*$ se vedno znajde v intervalu zaupanja, vrednost $\theta \geq \theta^*$ pa le, če se vsaj eden od poskusov ponesreči. Verjetnost za to je $1 - (1 - \theta)^n$. Le-ta bo za vse $\theta \geq \theta^*$ enaka vsaj β natanko tedaj, ko bo $1 - (1 - \theta^*)^n \geq \beta$. Najmanjša vrednost θ^*, ki temu še ustreza, je:

$$\theta^* = 1 - (1 - \beta)^{1/n}\,.$$

Ko gre n proti neskončno, je to asimptotično enako:

$$\theta^* = 1 - e^{(\ln(1-\beta))/n} \sim \frac{-\ln(1 - \beta)}{n}\,.$$

Pri $\beta = 95\%$ je to približno enako $3/n$.

4. Pri $N = 0$ moramo za interval vzeti kar $[0,1]$, sicer za $\theta = 1$ dobimo ničelno verjetnost pokritosti. Pri $N = n > 0$ pa zadošča vzeti $[0, b_n)$, saj za zaprto različico dobimo enako minimalno verjetnost pokritosti.

Če fiksiramo $b_n \le \theta < b_{n+1}$, je dejanska vrednost θ v intervalu zaupanja natanko tedaj, ko je $N \le n$, torej natanko tedaj, ko se v prvih $n+1$ poskusih vsaj eden ponesreči; verjetnost tega dogodka je $1-(1-\theta)^{n+1}$. Pri vseh $\theta \in [b_n, b_{n+1})$ mora biti torej $1-(1-\theta)^{n+1} \ge \beta$. To drži natanko tedaj, ko je $1-(1-b_n)^n \ge \beta$ oziroma $b_n \ge 1-(1-\beta)^{1/n}$. Minimalnost dobimo pri $b_n = 1-(1-\beta)^{1/n}$.

Če primerjamo s prejšnjo nalogo, dobimo isti rezultat za primer iz prejšnje naloge, ko vemo, da se je prvih n poskusov posrečilo, in primer iz te naloge, ko vemo, da se je prvih n poskusov posrečilo, naslednji pa se je ponesrečil.

5. Očitno mora biti $b_0 > 0$, sicer dobimo poljubno majhno ali celo ničelno verjetnost pokritosti. Poleg tega mora biti interval zaupanja pri opažanju $S = n$ kar $[0,1]$, sicer pri $\theta = 1$ nimamo pokritosti. Za $k = 1, 2, \ldots$ pa lahko pri $S = k$ nastavimo kar $[0, b_k)$, saj za zaprto različico dobimo enako minimalno verjetnost pokritosti.

Če fiksiramo $b_k < \theta < b_{k+1}$, je torej dejanska vrednost θ v intervalu zaupanja natanko tedaj, ko je $S > k$. Pri vseh $\theta \in (b_k, b_{k+1})$ mora biti torej $\mathbb{P}(S > k) \ge \beta$, kar drži natanko tedaj, ko je to res pri $\theta = b_k$. Minimalnost dobimo natanko tedaj, ko za b_k postavimo tisti θ, pri katerem je $\mathbb{P}(S > k) = \beta$.

Zadevo pa si lahko predstavljamo tudi v duhu 29. naloge iz 5. razdelka (glej tudi 2. nalogo iz 11. razdelka): naj torej n gostov pride na zabavo, vsak z zamudo, porazdeljeno enakomerno na $[0,1]$, neodvisno od drugih gostov. Če z S_θ označimo število gostov, ki pridejo na zabavo do časa θ, je $S_\theta \sim \mathrm{Bin}(n, \theta)$. V tem duhu mora torej veljati $\mathbb{P}(S_{b_k} > k) = \beta$. Toda če s T_k označimo prihod gosta, ki pride k-ti po vrsti, je $\{S_\theta > k\} = \{T_{k+1} \le \theta\}$, torej mora veljati $\mathbb{P}(T_{k+1} \le b_k) = \beta$. Zgornja meja b_k mora biti torej natanko kvantil porazdelitve $\mathrm{Beta}(k+1, n-k)$ za verjetnost β.

Podobno dobimo tudi nasprotni enostranski interval zaupanja, tj. $(a_k, 1]$ pri $S = k$, $k = 1, 2, \ldots, n$. Lahko postavimo $a_k = 1 - b_{n-k}$, kar je kvantil porazdelitve $\mathrm{Beta}(k, n-k+1)$ za verjetnost $1-\beta$.

Vzemimo sedaj Bernoullijevo zaporedje poskusov, pri katerem se posamezen poskus ponesreči z verjetnostjo θ, in si oglejmo dogodek, da se prvih $n-1$ poskusov posreči, n-ti pa ponesreči. V skladu s konstrukcijo iz te naloge dobimo interval $[0, b)$, kjer je $b = b_1$ določen s pogojem $(1-b)^n + nb(1-b)^{n-1} = 1-\beta$, torej $\bigl(1+(n-1)b\bigr)(1-b)^{n-1} = 1-\beta$. V skladu s konstrukcijo iz prejšnje naloge pa dobimo interval $[0, b')$, kjer je b' določen s pogojem $(1-b')^{n-1} = 1-\beta$.

Za isto opažanje torej dobimo različna intervala zaupanja. To se lahko zgodi zato, ker moramo opažanje vedno gledati v kontekstu drugih opažanj. Pri obeh konstrukcijah (in tipično je tako) se izkaže, da se lahko glede tega, ali določena vrednost parametra pade v interval zaupanja ali ne, odločimo na naslednji način: opažanju priredimo določen dogodek, ki ga zajema. Če je verjetnost tega dogodka pri dani vrednosti parametra strogo večja od $1-\beta$, smo v intervalu zaupanja, sicer pa nismo.

Toda dogodek, ki ga priredimo opažanju, je lahko od konstrukcije do konstrukcije različen. Pri konstrukciji iz te naloge gre za dogodek, da se v prvih n poskusih ponesreči največ en poskus, medtem ko gre pri konstrukciji iz prejšnje naloge za dogodek, da se prvih $n-1$ poskusov posreči. To sta različna dogodka in od tod različna intervala zaupanja za isto opažanje.

6. Dovolj je izračunati zgornje krajišče, in sicer le za primer, ko ne uspe noben poskus ali pa uspe en poskus. Če ne uspe noben poskus, je zgornje krajišče pri Clopper–Pearsonovem intervalu kvantil porazdelitve $\mathrm{Beta}(2,1)$ za verjetnost $(1+\beta)/2$. Ta porazdelitev ima gostoto $f(t) = 2(1-t)$ in kumulativno porazdelitveno funkcijo $F(t) = 1-(1-t)^2$, kar nam da krajišče $1-\sqrt{(1-\beta)/2}$. Če pa uspe natanko en poskus, dobimo kvantil porazdelitve $\mathrm{Beta}(2,1)$; le-ta ima gostoto $f(t) = 2t$ in kumulativno porazdelitveno funkcijo $F(t) = t^2$, kar nam da krajišče $\sqrt{(1+\beta)/2}$. Popoln zapis je torej:

$$\begin{cases} \left[0, 1-\sqrt{\frac{1-\beta}{2}}\right) & ; S = 0 \\ \left(1-\sqrt{\frac{1+\beta}{2}}, \sqrt{\frac{1+\beta}{2}}\right) & ; S = 1 \\ \left(\sqrt{\frac{1-\beta}{2}}, 1\right] & ; S = 2 \end{cases},$$

kar je pri vseh izidih strogo več kot pri intervalu zaupanja iz 1. naloge.

7. Za mejo $a_{\max}$ lahko brez škode za splošnost postavimo funkcijo minimalne zadostne statistike. Iz gostote porazdelitve je razvidno, da gre za eksponentno družino z minimalno zadostno statistiko $S_n := X_1 + \cdots + X_n$. Iz centralnega limitnega izreka pa sledi, da je le-ta porazdeljena približno normalno $\mathrm{N}\big(n\mu, \sqrt{n}\,\sigma\big)$, kjer je:

$$\mu = \mathbb{E}(X_i) = a\,, \qquad \sigma = \sqrt{\mathrm{var}(X_i)} = a\sqrt{3}/2\,.$$

Postavimo torej $a_{\max} := h(S_n)$ in poskusimo, ali konstrukcija intervala zaupanja deluje, če je h strogo naraščajoča funkcija. Tedaj je namreč:

$$\mathbb{P}\big(a \in [0, a_{\max}]\big) = \mathbb{P}\big(a \le h(S_n)\big) = \mathbb{P}\big(S_n \ge h^{-1}(a)\big) \approx \frac{1}{2} - \Phi\left(\frac{h^{-1}(a) - na}{a\sqrt{3n}/2}\right).$$

Desna stran bo enaka $\beta = 0{\dot{}}95$, če bo:

$$\frac{h^{-1}(a) - na}{a\sqrt{3n}/2} = -z\,,$$

kjer je $z := \Phi^{-1}\big(\beta/2\big) \doteq 1{\dot{}}645$. Sledi:

$$h^{-1}(a) = \left(n - z\frac{\sqrt{3n}}{2}\right) a$$

oziroma:

$$h(s) = \frac{s}{n - z\sqrt{3n}/2}\,.$$

Za velike n je to dobro definirana strogo naraščajoča funkcija, torej lahko postavimo:

$$a_{\max} = \frac{S_n}{n - z\sqrt{3n}/2}\,.$$

Za funkcijo h pa ne bi mogli vzeti strogo padajoče funkcije. V tem primeru bi namreč veljalo:

$$\mathbb{P}\big(a \in [0, a_{\max}]\big) = \mathbb{P}\big(a \le h(S_n)\big) = \mathbb{P}\big(S_n \le h^{-1}(a)\big) \approx \frac{1}{2} + \Phi\left(\frac{h^{-1}(a) - na}{a\sqrt{3n}/2}\right)$$

in desna stran bi bila enaka β za

$$\frac{h^{-1}(a) - na}{a\sqrt{3n}/2} = z\,,$$

od koder bi sledilo $h^{-1}(a) = \big(n + z\sqrt{3n}/2\big)a$ oziroma $h(s) = s/\big(n + z\sqrt{3n}/2\big)$. Ta funkcija pa je spet strogo naraščajoča, kar je v protislovju s prvotno zahtevo.

8. $\bar{X} = 97$, $\Delta \doteq 3{\cdot}27$ (zaokroženo navzgor),
$93{\cdot}73 < \mu < 100{\cdot}27$.

9. $s = 5$, $t_{0.025}(8) \doteq 2{\cdot}306$, $\Delta \doteq 3{\cdot}85$ (zaokroženo navzgor),
$93{\cdot}15 < \mu < 100{\cdot}85$.

10. $\bar{X} \doteq 1{\cdot}5477$, $s_0 \doteq 1{\cdot}303$, $s \doteq 1{\cdot}304$. Interval zaupanja za μ: $1{\cdot}460 < \mu < 1{\cdot}635$.

11. Definirajmo slučajne spremenljivke $X_1, \ldots, X_n$, in sicer naj bo $X_i = 1$, če je i-ti poskus uspel, sicer pa 0. Tedaj je $\bar{X} = S/n =: \hat{\theta}$. Nadalje velja:

$$s_0^2 = \frac{S(1-\hat{\theta})^2 + (n-S)\hat{\theta}^2}{n} = \hat{\theta}(1-\hat{\theta})\,.$$

Dobimo *Waldov*[74] *interval zaupanja*:

$$\hat{\theta} - c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\,,$$

kjer je $c = z_{(1-\beta)/2} = \Phi^{-1}(\beta/2)$.

Je lahko izračunljiv, a ne dosega nominalne verjetnosti pokritosti niti v povprečju. Dejanska verjetnost pokritosti je često daleč pod nominalno. Res pa je, da se dejanska minimalna verjetnost pokritosti, ko θ preteče interval $[a, b]$ $(0 < a \le b < 1)$, bliža nominalni β, ko gre n proti neskončno.

[74] Abraham Wald (1902–1950), transilvanski matematik judovskega rodu

12. Clopper–Pearsonov interval: $(0{\cdot}0713, 0{\cdot}4011)$.
Waldov interval: $(0{\cdot}0528, 0{\cdot}3472)$.
Wilsonov interval: $(0{\cdot}0806, 0{\cdot}4161)$.
Agresti–Coullov interval: $(0{\cdot}0749, 0{\cdot}4218)$.
Wilsonov interval s popravkom za zveznost: $(0{\cdot}0661, 0{\cdot}4427)$.
Agresti–Coullov interval s popravkom za zveznost: $(0{\cdot}0499, 0{\cdot}4468)$.
Spodnja krajišča so zaokrožena navzdol, zgornja pa navzgor.

13. Clopper–Pearsonov interval: $0{\cdot}1266, 0{\cdot}2919$.
Waldov interval: $(0{\cdot}1216, 0{\cdot}2784)$.
Wilsonov interval: $(0{\cdot}1333, 0{\cdot}2889)$.
Agresti–Coullov interval: $(0{\cdot}1326, 0{\cdot}2896)$.
Wilsonov interval s popravkom za zveznost: $(0{\cdot}1292, 0{\cdot}2944)$.
Agresti–Coullov interval s popravkom za zveznost: $(0{\cdot}1276, 0{\cdot}2946)$.
Spodnja krajišča so zaokrožena navzdol, zgornja pa navzgor.

14. Clopper–Pearsonov interval: $(0{\cdot}496525849771,\ 0{\cdot}5130720831478)$.
Waldov interval: $(0{\cdot}49657611084,\ 0{\cdot}51302388916)$.
Wilsonov interval: $(0{\cdot}496575924714,\ 0{\cdot}513021478666)$.
Agresti–Coullov interval: $(0{\cdot}496575924612,\ 0{\cdot}513021478769)$.
Wilsonov interval s popravkom za zveznost: $(0{\cdot}496525930387,\ 0{\cdot}513071457209)$.
Agresti–Coullov interval s pop. za zveznost: $(0{\cdot}496525924612,\ 0{\cdot}513071478769)$.
Spodnja krajišča so zaokrožena navzdol, zgornja pa navzgor.

Agresti–Coullov interval je sicer nasploh odličen približek Wilsonovega, a tu je razlika še posebej majhna, ker je opaženi delež grbov blizu 1/2. Clopper–Pearsonov interval je potrebno določiti z računalnikom.

15. $\chi^2_{0.975}(8) \doteq 2{\cdot}180$, $\quad \chi^2_{0.025}(8) \doteq 17{\cdot}53$,
$3{\cdot}37 < \sigma < 9{\cdot}58$ (spodnja meja zaokrožena navzdol, zgornja pa navzgor).

16. $\bar{X} = 125$, $\quad s \doteq 2{\cdot}646$, $\quad \chi^2_{0.95}(4) \doteq 0{\cdot}7107$, $\quad \chi^2_{0.05}(4) \doteq 9{\cdot}488$, $\quad 1{\cdot}72 < \sigma < 6{\cdot}28$.

17. $\bar{X} \doteq 45{\cdot}51$, $\quad s \doteq 3{\cdot}71$,
$t_{0.005}(74) \doteq 2{\cdot}644$, $\quad \Delta \doteq 1{\cdot}14$ (zaokroženo navzgor), $\quad 44{\cdot}37 < \mu < 46{\cdot}65$,
$\chi^2_{0.995}(74) \doteq 46{\cdot}41$, $\quad \chi^2_{0.005}(74) \doteq 109{\cdot}1$, $\quad 3{\cdot}05 < \sigma < 4{\cdot}69$.
Pri obeh intervalih zaupanja je spodnja meja zaokrožena navzdol, zgornja pa navzgor.

18. Če s κ^4 ($\kappa \geq 0$) označimo četrti centralni moment spremenljivke X, po centralnem limitnem izreku približno velja:

$$\sum_{i=1}^{n} (X_i - \mu)^2 \sim \mathrm{N}\big(n\sigma^2, \sqrt{n(\kappa^4 - \sigma^4)}\big)$$

Iz Jensenove neenakosti sledi $\kappa \geq \sigma$, pri čemer enakost velja natanko tedaj, ko je spremenljivka X bodisi skoncentrirana v eni točki bodisi skoncentrirana v dveh točkah, pri čemer ima v vsaki verjetnost 1/2. Privzemimo najprej, da to ni res, torej da je $\kappa > \sigma$. V tem primeru lahko zapišemo:

$$\frac{1}{\sqrt{n(\kappa^4 - \sigma^4)}} \left[\sum_{i=1}^{n} (X_i - \mu)^2 - n\sigma^2 \right] \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)\,.$$

V nadaljevanju moramo iz spremenljivke na levi odstraniti vse neopazljive parametre razen σ, pri tem pa moramo paziti, da se ohrani asimptotična normalnost. Najprej velja:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)^2 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \bar{X})^2 - \sqrt{n}(\bar{X} - \mu)^2\,.$$

Ker gre $\operatorname{var}\left[\sqrt[4]{n}(\bar{X} - \mu)\right] = \sigma^2/\sqrt{n} \xrightarrow[n\to\infty]{} 0$, gre po neenačbi Čebiševa tudi $\sqrt[4]{n}(\bar{X} - \mu) \xrightarrow[n\to\infty]{\mathbb{P}} 0$ in zato tudi $\sqrt{n}(\bar{X} - \mu)^2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Po izreku Sluckega potem velja tudi:

$$\frac{1}{\sqrt{n(\kappa^4 - \sigma^4)}} \left[\sum_{i=1}^{n} (X_i - \bar{X})^2 - n\sigma^2 \right] \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)\,.$$

Naj bosta $\hat{\sigma}$ in $\hat{\kappa}$ cenilki za κ in σ po metodi momentov, tj.:

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}\,, \quad \hat{\kappa} = \sqrt[4]{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^4}\,.$$

Ker so cenilke po metodi momentov dosledne, gre $\hat{\sigma} \xrightarrow[n\to\infty]{\mathbb{P}} \sigma$ in $\hat{\kappa} \xrightarrow[n\to\infty]{\mathbb{P}} \kappa$. Iz izreka Sluckega za deljenje sedaj sledi:

$$\frac{1}{\sqrt{n(\hat{\kappa}^4 - \hat{\sigma}^4)}} \left[\sum_{i=1}^{n} (X_i - \bar{X})^2 - n\sigma^2 \right] \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)$$

oziroma:

$$\frac{\sqrt{n}}{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}} (\hat{\sigma}^2 - \sigma^2) \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)\,.$$

Seveda se lahko zgodi, da pride do deljenja z nič, a asimptotično prihaja do tega vse redkeje. V primeru deljenja z nič lahko rezultat poljubno nastavimo (lahko tudi na vrednost “nedefinirano”) in bo konvergenca v porazdelitvi še vedno veljala.

Sedaj si oglejmo še primer, ko je $\kappa = \sigma$, torej ko je X bodisi skoncentrirana v eni točki bodisi v dveh točkah z enakima verjetnostma. Če je X skoncentrirana v eni točki, vemo, da je $\sigma = 0$, velja pa tudi $\hat{\sigma} = \hat{\kappa} = 0$. Torej zgornja konstrukcija še vedno velja, če pri neenačajih vključimo še enakost. Preostane še primer, ko je X skoncentrirana v dveh enako verjetnih točkah. Zaradi invariantnosti leve strani za translacije in raztege lahko brez škode za splošnost privzamemo, da je $X \sim \mathrm{Ber}(1/2)$. Če spet s $\hat{\theta}$ označimo delež enic, velja:

$$\begin{aligned} \hat{\sigma}^2 &= \hat{\theta}(1-\hat{\theta})\,, \\ \hat{\kappa}^4 &= \hat{\theta}(1-\hat{\theta})(1-3\hat{\theta}+3\hat{\theta}^2)\,, \\ \hat{\kappa}^4-\hat{\sigma}^4 &= \hat{\theta}(1-\hat{\theta})(1-2\hat{\theta}^2)\,, \\ \frac{\sqrt{n}}{\sqrt{\hat{\kappa}^4-\hat{\sigma}^4}}(\hat{\sigma}^2-\sigma^2) &= \sqrt{n}\,\frac{\hat{\theta}(1-\hat{\theta})-\frac{1}{4}}{\sqrt{\hat{\theta}(1-\hat{\theta})(1-2\hat{\theta})^2}} = \sqrt{n}\,\frac{2\hat{\theta}-1}{4\sqrt{\hat{\theta}(1-\hat{\theta})}}\,. \end{aligned}$$

Po Laplaceovi integralski formuli gre $\sqrt{n}\,(2\hat{\theta}-1) \xrightarrow[n\to\infty]{\mathbb{P}} \mathrm{N}(0,1)$. Ker gre $\hat{\theta} \xrightarrow[n\to\infty]{\mathbb{P}} \frac{1}{2}$, spet iz izreka Sluckega za deljenje dobimo:

$$\frac{\sqrt{n}}{\sqrt{\hat{\kappa}^4-\hat{\sigma}^4}}(\hat{\sigma}^2-\sigma^2) \xrightarrow[n\to\infty]{\mathbb{P}} \mathrm{N}\big(0,\tfrac{1}{2}\big)\,,$$

od koder dobimo naslednjo konstrukcijo intervala zaupanja:

$$\hat{\sigma}^2 - c\,\frac{\hat{\kappa}^4-\hat{\sigma}^4}{2\sqrt{n}} < \sigma^2 < \hat{\sigma}^2 + c\,\frac{\hat{\kappa}^4-\hat{\sigma}^4}{2\sqrt{n}}\,,$$

kar pomeni, da prvotna konstrukcija za predpisano stopnjo zaupanja še vedno velja, pravkar napisana pa tudi, a je natančnejša. Toda to pomeni le, da nam pri konstrukciji intervalov zaupanja za standardni odklon ni treba skrbeti za primer, ko je $\kappa = \sigma$: če to vemo, lahko preprosto zatrdimo, da je kar $\sigma = (\max_i X_i - \min_i X_i)/2$. Če bo dobljena vrednost večja od nič, bo ta izjava z verjetnostjo ena pravilna, sicer pa bo pravilna z verjetnostjo β, brž ko bo vzorec dovolj velik.

19. $\bar{X} \doteq 1{\cdot}55$, $\quad \hat{\sigma} \doteq 1{\cdot}30$, $\quad \hat{\kappa} \doteq 2{\cdot}09$.
Interval zaupanja za σ: $1{\cdot}19 < \sigma < 1{\cdot}41$.

Opomba. Zanimiva je tudi ocena za presežno sploščenost (kurtozis):

$$\frac{\hat{\kappa}^4}{\hat{\sigma}^4} - 3 \doteq 3{\cdot}56\,,$$

kar namiguje na to, da je dejanska porazdelitev daleč od normalne. Vendar pa največji relativni delež v κ^4 prispeva edina ženska z desetimi otroki. Le-ta nam torej znatno poveča širino intervala zaupanja.

13. Preizkusi značilnosti

1. Če z D označimo število dobitnih srečk med kupljenimi, je $D \sim \mathrm{Bin}(n, \theta)$, kjer je $n = 8$, θ pa delež dobitnih med vsemi srečkami v seriji. Verjetnostna funkcija je torej enaka:

$$\mathbb{P}(D = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \,,$$

kar nam da verjetje $L(\theta) = \binom{n}{D} \theta^D (1-\theta)^{n-D}$. Ničelna domneva trdi, da je $\theta \geq 1/2$, alternativna pa, da je $\theta < 1/2$. Torej je $\Theta = [0,1]$ in $\Theta_{H_0} = [0, 1/2]$. Po krajšem računu dobimo:

$$\sup_{\theta \in \Theta} L(\theta) = \binom{n}{D} \left(\frac{D}{n}\right)^D \left(1 - \frac{D}{n}\right)^{n-D}$$

Za $D \geq n/2$ velja tudi:

$$\sup_{\theta \in \Theta_{H_0}} L(\theta) = \binom{n}{D} \left(\frac{D}{n}\right)^D \left(1 - \frac{D}{n}\right)^{n-D} \,,$$

medtem ko za $D \leq n/2$ velja:

$$\sup_{\theta \in \Theta_{H_0}} L(\theta) = \binom{n}{D} \left(\frac{1}{2}\right)^n \,.$$

Torej velja:

$$\Lambda = \begin{cases} 2^{-n} \left(\frac{D}{n}\right)^{-D} \left(1 - \frac{D}{n}\right)^{-(n-D)} & ; \; D \leq n/2 \\ 1 & ; \; D \geq n/2 \end{cases} .$$

Za $D \leq n/2$ velja $\ln \Lambda = g(D/n)$, kjer je $g(x) = -n\bigl(\ln 2 + x \ln x + (1-x)\ln(1-x)\bigr)$. Z odvajanjem dobimo, da je funkcija g strogo naraščajoča na $(0, 1/2]$. Torej je vzorec toliko ustreznejši, kolikor večji je D. Naj bo D' kopija statistike D. Ker je:

$$\sup_{\theta \in \Theta_{H_0}} \mathbb{P}(D' \leq 2) \doteq 0{\cdot}145 \,, \qquad \sup_{\theta \in \Theta_{H_0}} \mathbb{P}(D' < 2) \doteq 0{\cdot}035 \,,$$

brez randomizacije domneve ne moremo zavrniti. Sicer pa jo zavrnemo z verjetnostjo 0˙135 (zaokrožili smo jo navzdol).

2. Označimo s θ verjetnost po modelu, da pade grb, z S pa število opaženih grbov (za naš konkretni primer bo torej $S = 5$, a za konstrukcijo preizkusa potrebujemo slučajno spremenljivko). Označimo še $n = 20$. Iz:

$$\mathbb{P}_\theta(S = k) = \binom{n}{k} \theta^k (1-\theta)^{n-k} \,,$$

$$\mathbb{P}_{1/2}(S = k) = \binom{n}{k} 2^{-n} \,,$$

$$\sup_{0 \leq \theta \leq 1} \mathbb{P}_\theta(S = k) = \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}$$

dobimo razmerje verjetij:

$$\Lambda = \frac{n^n}{2^n\, S^S\, (n-S)^{n-S}}\,.$$

Velja $\ln\Lambda = g(S)$, kjer je $g(s) = n\ln\frac{n}{2} - s\ln s - (n-s)\ln(n-s)$. Z odvajanjem dobimo, da je g strogo naraščajoča na $(0, n/2]$. Brž ko je torej $S \le n/2$, je opažanje toliko ustreznejše za ničelno domnevo, kolikor večji je S. Toda ker se razmerje verjetij ohrani, če S zamenjamo z $n-S$, velja, da je opažanje toliko ustreznejše, kolikor bližje je S številu $n/2$. Ničelno domnevo bomo torej zavrnili, če bo vrednost $\left|S - \frac{n}{2}\right|$ prevelika.

Naj bo S' kopija statistike S. Za naše opažanje dobimo:

$$\begin{aligned}
\mathbb{P}_{1/2}(S' \le 5 \text{ ali } S' \ge 15) &\doteq 0{\dot{}}041\,,\\
\mathbb{P}_{1/2}(S' < 5 \text{ ali } S' > 15) &\doteq 0{\dot{}}012\,.
\end{aligned}$$

Pri $\alpha = 0{\dot{}}05$ bomo torej našo domnevo zavrnili, pri $\alpha = 0{\dot{}}01$ pa ne. Do randomizacije tukaj ne pride.

3. Ničelno domnevo lahko interpretiramo kot to, da je verjetnost, da je posameznik, ki dobi nagrado, ženska z verjetnostjo $\theta_0 = 0{\dot{}}7$. Pri alternativni domnevi pa je posameznik, ki dobi nagrado, ženska z verjetnostjo $\theta \ne 0{\dot{}}7$.

V splošnem je torej verjetnost, da nagrado dobi k žensk in $n-k$ moških, enaka:

$$f_n(k;\theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}\,.$$

Če velja ničelna domneva, v ta izraz samo vstavimo $\theta = \theta_0$. V splošnem pa to zoptimiziramo po θ. Velja:

$$\frac{\partial \ln f(k;\theta)}{\partial\theta} = \frac{k}{\theta} - \frac{n-k}{1-\theta}$$

in maksimum je dosežen pri $\theta = k/n$. Če n fiksiramo, število žensk, ki dobijo nagrado (doslej k), pa naredimo za slučajno spremenljivko S, je torej razmerje verjetij enako:

$$\Lambda = \left(\theta_0\,\frac{n}{S}\right)^S \left((1-\theta_0)\,\frac{n}{n-S}\right)^{n-S}\,.$$

Ničelno domnevo bomo zavrnili, če bo to razmerje premajhno, natančneje, če bo verjetnost, da bo tako majhno ali še manjše, kot smo ga opazili, največ $0{\dot{}}05$. Za $n = 12$ in $\theta_0 = 0{\dot{}}7$ tabelirajmo:

S	Λ
0	$5{'}314410 \cdot 10^{-7}$
1	$3{'}875139 \cdot 10^{-5}$
2	$6{'}449463 \cdot 10^{-4}$
3	$0{'}005754585$
4	$0{'}03270222$
5	$0{'}1273437$
6	$0{'}3512980$
7	$0{'}6933160$
8	$0{'}9693585$
9	$0{'}9286985$
10	$0{'}5666800$
11	$0{'}1853767$
12	$0{'}01384129$

Razmerje verjetij je manjše ali enako opaženemu pri $S \in \{0, 1, 2, 3, 4, 5, 12\}$. Če je torej S' porazdeljena tako kot S pri H_0, torej $S' \sim \mathrm{Bin}(12, 0{'}7)$, je p-vrednost našega opažanja enaka:

$$p(5) = \mathbb{P}\big(S' \in \{0, 1, 2, 3, 4, 5, 12\}\big) \doteq 0{'}0524\,.$$

Ker je to večje od α, izračunamo še:

$$p^-(5) = \mathbb{P}\big(S' \in \{0, 1, 2, 3, 4, 12\}\big) \doteq 0{'}0233\,.$$

Ker je to manjše od α, ničelno domnevo zavrnemo z verjetnostjo $0{'}916$.

4. Izvedemo dvostranski preizkus. Če je $S' \sim \mathrm{Bin}(12, 0{'}7)$, je $\mathbb{P}(S' \le 5) \doteq 0{'}0386$ in $\mathbb{P}(S' \ge 5) \doteq 0{'}9905$, torej je $p(5) \doteq 0{'}0772$. Domneve ne moremo zavrniti. Zavrnili pa bi jo, če bi za alternativno domnevo postavili, da so izjemni študijski dosežki pristranski v korist moških.

Enostranska preizkusa se ujemata s preizkusom na podlagi razmerja verjetij. Če npr. preizkušamo ničelno domnevo, da je $\theta = \theta_0$, proti alternativni, da je $\theta < \theta_0$, je za $S \le n\theta_0$ razmerje verjetij enako:

$$\Lambda = n^n \left(\frac{\theta_0}{S}\right)^S \left(\frac{1 - \theta_0}{n - S}\right)^{n-S}\,.$$

Z odvajanjem po S dobimo, da je razmerje verjetij naraščajoča funkcija spremenljivke S, kar pomeni, da sta razmerje verjetij in S ekvivalentni preizkusni statistiki.

Dvostranski preizkus pa ni primerljiv s preizkusom na podlagi razmerja verjetij. Lahko pa rečemo, da je pretirano konservativen. Za $k \le n\theta_0$ je namreč $\rho(k) = 2\,\mathbb{P}(S' \le k)$, medtem ko je verjetnost značilnosti za opažanje $S = k$ glede na preizkus, ki temelji na preizkusni statistiki ρ, enaka $p(k) = \mathbb{P}(S' \le k) + \mathbb{P}(S' \ge l)$, kjer je l najmanjše naravno število, za katero je $\mathbb{P}(S' \ge l) \le \mathbb{P}(S' \le k)$. Seveda je

$\mathbb{P}(S' \geq l) \leq \mathbb{P}(S' \leq k)$, a tipično $\mathbb{P}(S' \geq l) < \mathbb{P}(S' \leq k)$. Torej je $p(k) \leq \rho(k)$, a tipično $p(k) < \rho(k)$. Enako velja tudi za primer, ko je $k \geq n\theta_0$. Preizkus, ki zavrne ničelno domnevo na podlagi preizkusne statistike ρ in verjetnosti značilnosti, še vedno ustreza predpisani stopnji značilnosti, a ničelno domnevo zavrne večkrat. Njegova računska zahtevnost pa je primerljiva s preizkusom na podlagi razmerja verjetij.

5. Označimo s θ verjetnost po modelu, da bo dobitek izžreban. Označimo $\theta_0 = 1/50$ in naj bo N število iger, po katerih je dobitek prvič izžreban. Tedaj velja $\mathbb{P}_\theta(N = k) = \theta(1-\theta)^{k-1}$, torej je $L(\theta) = \theta(1-\theta)^{N-1}$ in:

$$\sup_{\theta \leq \theta_0} L(\theta) = \begin{cases} \theta_0(1-\theta_0)^{N-1} & ; N \leq 1/\theta_0 \\ \frac{1}{N}\left(1-\frac{1}{N}\right)^{N-1} & ; N \geq 1/\theta_0 \end{cases}.$$

Razmerje verjetij je zato enako:

$$\Lambda = \begin{cases} 1 & ; N \leq 1/\theta_0 \\ \theta_0 N \left(\dfrac{1-\theta_0}{1-\frac{1}{N}}\right)^{N-1} & ; N \geq 1/\theta_0 \end{cases}.$$

Za $N > 1/\theta_0$ velja $\ln \Lambda = g(N)$, kjer je $g(x) = \ln\theta_0 + \ln x + (x-1)\ln(1-\theta_0) - (x-1)\ln\left(1-\frac{1}{x}\right)$. Z odvajanjem dobimo, da je g na intervalu $[1/\theta_0, \infty)$ strogo padajoča funkcija. Preizkus bo torej potekal tako, da bomo ničelno domnevo zavrnili, če bo preizkusna statistika $\max\{\frac{1}{\theta_0}, N\}$ prevelika.

Poiskati moramo torej najmanjši tak k, za katerega bo $\mathbb{P}_{\theta_0}\left(\max\left\{\frac{1}{\theta_0}, N\right\} > k\right) \leq \alpha$. Velja $\mathbb{P}_{\theta_0}(N > k) = (1-\theta_0)^k$. Ker je $\mathbb{P}_{1/50}(\max\{50, N\} > 50) = \mathbb{P}_{1/50}(N > 50) \doteq 0{\cdot}364$, bo to najmanjši k, za katerega bo $(49/50)^k < 0{\cdot}1$. Iz $\log_{49/50} 0{\cdot}1 \doteq 113{\cdot}97$ dobimo, da je ta k enak 114 – velja:

$$\mathbb{P}_{1/50}\big(\max\{50, N\} > 114\big) = \mathbb{P}_{1/50}(N > 114) \doteq 0{\cdot}09994766\,,$$
$$\mathbb{P}_{1/50}\big(\max\{50, N\} \geq 114\big) = \mathbb{P}_{1/50}(N \geq 114) \doteq 0{\cdot}1019874\,.$$

Preizkus bo torej potekal tako, da bomo igro igrali, dokler ne bo izžreban dobitek, vendar pa bomo odigrali največ 114 iger. Če bo dobitek izžreban pred zadnjo, 114. igro, domneve ne bomo zavrnili. Če dobitek tudi po 114 igrah še ne bo izžreban, bomo domnevo zavrnili. Če pa bo dobitek (prvič) izžreban v zadnji, 114. igri, bomo domnevo zavrnili z verjetnostjo:

$$\frac{0{\cdot}1 - 0{\cdot}09994766}{0{\cdot}1019874 - 0{\cdot}09994766} \doteq 0{\cdot}025$$

(zaokrožili smo navzdol).

6. Glede na to, da sta ničelna in alternativna domneva obe enostavni, je preizkus, ki temelji na razmerju verjetij, najmočnejši. Če z L_0 označimo verjetje pri ničelni, z L_1 pa pri alternativni domnevi, velja:

$$\frac{L_0}{L_1} = \frac{\frac{1}{500}\,e^{-X/500}}{\frac{1}{100}\,e^{-X/100}} = \frac{1}{5}\,e^{4X/500}\,,$$

kjer je X opažena življenjska doba žarnice. Ničelno domnevo bomo torej zavrnili, če bo preizkusna statistika $\frac{1}{5}\,e^{4X/500}$ premajhna, to pa bo tedaj, ko bo življenjska doba X premajhna. Ker je X pri ničelni domnevi porazdeljena zvezno, lahko H_0 zavrnemo, brž ko je $X \le c$, kjer je $\mathbb{P}_{H_0}(X \le c) = \alpha$. Iz:

$$\mathbb{P}_{H_0}(X \le c) = \int_0^c \frac{1}{500}\,e^{-x/500}\,\mathrm{d}x = 1 - e^{-c/500}$$

dobimo $c = -500\ln(0{\cdot}95) \doteq 25{\cdot}64$ (zaokroženo navzdol). Moč tega preizkusa je:

$$\mathbb{P}_{H_1}(X \le c) = \int_0^c \frac{1}{100}\,e^{-x/100}\,\mathrm{d}x = 1 - e^{-c/100} \doteq 0{\cdot}226\,.$$

7. a) Tinetovo povprečje je porazdeljeno normalno $\mathrm{N}(0, 1/10)$, torej ima gostoto:

$$f_{\mathrm{Ti}}(x) = \frac{10}{\sqrt{2\pi}}\,e^{-50x^2}$$

Tonetovo povprečje pa je porazdeljeno normalno $\mathrm{N}(0, 1/100)$, torej ima gostoto:

$$f_{\mathrm{To}}(x) = \frac{100}{\sqrt{2\pi}}\,e^{-5000x^2}\,.$$

Ker sta ničelna in alternativna domneva obe enostavni, je preizkus, ki temelji na razmerju verjetij, najmočnejši. Če torej z X označimo opaženo povprečje, ničelno domnevo torej zavrnemo, če je razmerje:

$$\frac{f_{\mathrm{To}}(X)}{f_{\mathrm{Ti}}(X)} = 10\,e^{-4950X^2}$$

premajhno, to pa je takrat, ko je vrednost $|X|$ prevelika. Če je $\mathbb{P}_{\mathrm{To}}$ verjetnost pri Tonetovem povprečju, za $c \ge 0$ velja:

$$\mathbb{P}_{\mathrm{To}}(|X| \ge c) = 1 - 2\Phi(100c)$$

Vrednost c moramo nastaviti tako, da bo to enako α, torej:

$$c = \frac{\Phi^{-1}\left(\frac{1-\alpha}{2}\right)}{100} \doteq 0{\cdot}0196\,.$$

Moč preizkusa pa je:

$$\mathbb{P}_{\mathrm{Ti}}(|X| \ge c) = 1 - 2\Phi(10c) \doteq 0{\cdot}845\,.$$

8. Iz zapisa gostote:

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n\,;\,\mu, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right]\,.$$

je očitno, da gre za eksponentno družino. Širši model ima dimenzijo 2, ožji model pa dimenzijo 1 (in je podmnogoterost, ker ga lahko celo globalno karakteriziramo z zvezo $\mu - \sigma^2 = 0$ ali $\mu/\sigma^2 = 1$). Maksimum verjetja je pri širšem modelu dosežen pri:

$$\mu = \hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad \sigma^2 = \hat{\sigma}_1^2 := \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 .$$

V ožjem modelu pa je:

$$L = f_{X_1,\ldots,X_n}(X_1,\ldots,X_n\,;\sigma^2,\sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n}\exp\left[-\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} X_i - \frac{n\sigma^2}{2}\right]$$

in iz:

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}\ln L = \frac{1}{\sigma^3}\sum_{i=1}^{n} x_i^2 - n\sigma - \frac{n}{\sigma}$$

dobimo, da je maksimum dosežen pri:

$$\sigma^2 = \hat{\sigma}_2^2 := \hat{\mu}_2 := \frac{1}{2}\left[\sqrt{1 + \frac{4}{n}\sum_{i=1}^{n} X_i^2} - 1\right] .$$

Če z Λ označimo razmerje verjetij:

$$\Lambda = \frac{\sup_{\Theta_{H_0}} L}{\sup_\Theta L} = \frac{f_{X_1\ldots X_n}(X_1,\ldots,X_n\,;\hat{\mu}_2,\hat{\sigma}_2)}{f_{X_1,\ldots,X_n}(X_1,\ldots,X_n\,;\hat{\mu}_1,\hat{\sigma}_1)}\,,$$

ima $-2\ln\Lambda$ približno porazdelitev hi kvadrat z eno prostostno stopnjo. Za naše opažanje dobimo $-2\ln\Lambda \doteq 5{\cdot}61$, kritična vrednost pa je $\chi^2_{0.01}(1) \doteq 6{\cdot}635$. Domneve torej ne moremo zavrniti.

9. a) $Z = 1 < 1{\cdot}645$, domneve ne moremo zavrniti.
b) $Z \doteq 1{\cdot}732 \geq 1{\cdot}645$, domnevo zavrnemo.
c) $Z \doteq 1{\cdot}732 < 2{\cdot}326$, domneve ne moremo zavrniti.

10. $n = 10000$, $S = 5090$, H_0: $\theta = 0{\cdot}5$: $H_1^{\pm}$: $\theta \neq 0{\cdot}5$: $1{\cdot}79 < 1{\cdot}960$, domneve ne moremo zavrniti.
$n = 10000$, $S = 5090$, H_0: $\theta = 0{\cdot}5$: H_1^{+}: $\theta > 0{\cdot}5$: $1{\cdot}79 > 1{\cdot}645$, domnevo zavrnemo.

11. $\bar{X} = 97$, $Z = -1{\cdot}8$.
Pri alternativni domnevi $\mu \neq 100$ upoštevamo $z_{0.025} \doteq 1{\cdot}960$ in domneve ne moremo zavrniti.
Pri alternativni domnevi $\mu < 100$ upoštevamo $z_{0.05} \doteq 1{\cdot}645$ in domnevo zavrnemo.
Pri alternativni domnevi $\mu > 100$ tudi upoštevamo $z_{0.05} \doteq 1{\cdot}645$, a domneve ne zavrnemo.

12. Ker je preizkusna statistika $\frac{3}{5}(\bar{X}-100)$ porazdeljena normalno $\mathrm{N}\big(\frac{3}{5}(\mu-100),1\big)$, je moč preizkusa enaka:

$$1-\Phi\big(z_{0.025}-\tfrac{3}{5}(\mu-100)\big)+\Phi\big(-z_{0.025}-\tfrac{3}{5}(\mu-100)\big)\,.$$

Pri $\mu = 99{\cdot}9$ je to enako približno $0{\cdot}0504$, pri $\mu = 97$ približno $0{\cdot}4356$, pri $\mu = 90$ pa približno $0{\cdot}99997$.

13. $\bar{X} = 107, \quad s = 10, \quad T = 2{\cdot}1, \quad t_{0.025}(8) \doteq 2{\cdot}306.$
Domneve ne moremo zavrniti.
Če bi vedeli, da je $\sigma = 10$, pa bi upoštevali $z_{0.025} \doteq 1{\cdot}960$ in domnevo bi zavrnili.

14. $\bar{X} = 48, \quad s \doteq 2{\cdot}58, \quad T = -2{\cdot}45, \quad df = 9.$
Če za H_1 vzamemo, da je $\mu \neq 50$, upoštevamo $t_{0.025}(9) \doteq 2{\cdot}262$ in domnevo zavrnemo.
Če za H_1 vzamemo, da je $\mu < 50$, pa upoštevamo $t_{0.05}(9) \doteq 1{\cdot}833$ in domnevo prav tako zavrnemo.

15. Označimo z Δ razliko v teži posamezne osebe (teža po dieti minus teža pred dieto). Tedaj je $\Delta \sim \mathrm{N}(\mu,\sigma)$ in preizkušamo ničelno domnevo, da je $\mu = 0$, proti alternativni domnevi, da je $\mu < 0$.
Velja $\bar{\Delta} \doteq -5{\cdot}6, \quad s \doteq 6{\cdot}433, \quad T \doteq -2{\cdot}75, \quad t_{0.05}(9) \doteq 1{\cdot}833.$
Ničelno domnevo zavrnemo, torej sprejmemo domnevo, da dieta deluje.

16. Uporabimo Welchev preizkus. Če z X označimo teže deklic, z Y pa teže dečkov, dobimo $\bar{X} \doteq 3022, \quad \bar{Y} \doteq 3364, \quad s_X \doteq 762, \quad s_Y \doteq 317, \quad \widehat{\mathrm{SE}} \doteq 287, \quad \widehat{df} \doteq 8{\cdot}939,$ $T \doteq -1{\cdot}192, \quad F^{-1}_{\mathrm{Student}(8)}(0{\cdot}975) \doteq 2{\cdot}306, \quad F^{-1}_{\mathrm{Student}(9)}(0{\cdot}975) \doteq 2{\cdot}262.$
Ničelne domneve torej ne moremo zavrniti.

17. $\bar{X} = 100, \quad \bar{Y} = 96, \quad \widehat{\mathrm{SE}} \doteq 1{\cdot}057, \quad T = 3{\cdot}783, \quad df \doteq 15{\cdot}22,$
$t_{0.01}(15) \doteq 2{\cdot}602, \quad t_{0.01}(16) \doteq 2{\cdot}583, \quad t_{0.01}(df) \doteq 2{\cdot}598.$
Domnevo zavrnemo.

18. $\bar{X}_1 = 4, \quad \bar{X}_2 = 3, \quad \bar{X}_3 \doteq 1{\cdot}667,$
$S^2_{\mathrm{B}} \doteq 9{\cdot}333, \quad S^2_{\mathrm{W}} \doteq 8{\cdot}667, \quad F \doteq 4{\cdot}31, \quad F_{0.05}(2,8) \doteq 4{\cdot}459.$
Domneve ne moremo zavrniti.

19. $s = 7{\cdot}45.$

a) $\chi^2 = 20, \quad \chi^2_{0.975}(9) \doteq 2{\cdot}700, \quad \chi^2_{0.025}(9) \doteq 19{\cdot}02,$ domnevo zavrnemo.
b) $\chi^2 = 5, \quad \chi^2_{0.95}(9) \doteq 3{\cdot}325,$ domneve ne moremo zavrniti.

20. Pri frekvencah 21, 42, 77 in 116 je: $\chi^2 \doteq 81{\cdot}3 \quad \chi^2_{0.01}(3) \doteq 11{\cdot}34.$
Domnevo zavrnemo.
Če so frekvence 21, 37, 53 in 25, pa je $\chi^2 \doteq 18{\cdot}2$ in domnevo še vedno zavrnemo.

21. Če so frekvence 2, 5 in 4, je: $\chi^2 = 1, \quad \chi^2_{0.05}(2) \doteq 5{\cdot}991.$
Domneve ne moremo zavrniti.
Pri frekvencah 20, 50 in 40 pa je $\chi^2 = 10$. Ker je kritično območje isto, domnevo zdaj zavrnemo.

22. Najprej moramo deležem dodati še delež preostalih, torej tistih, ki se niso opredelili za nobeno od strank, ki so prišle v Državni zbor. Pričakovane frekvence so dokončni deleži, pomnoženi z n. Opažene frekvence pa žal niso znane v celoti, jih pa rekonstruiramo tako, da objavljene deleže pomnožimo z n in zaokrožimo: tiste z manjšimi ulomljenimi deli navzdol, tiste z večjimi ulomljenimi deli navzgor. Če deleže, pomnožene z n, imenujemo *preliminarne frekvence*, le-te za anketo agencije Parsifal pridejo:

$$202{\cdot}720\,,\ 191{\cdot}860\,,\ 52{\cdot}852\,,\ 58{\cdot}644\,,\ 44{\cdot}888\,,\ 173{\cdot}036\,.$$

Vsota navzdol zaokroženih preliminarnih frekvenc pride 720, torej bo treba štiri zaokrožiti navzgor, dve pa navzdol. Glede na ulomljene dele četrto in šesto preliminarno frekvenco zaokrožimo navzdol, preostale štiri pa navzgor. Skupaj s pričakovanimi frekvencami (ki jih ne zaokrožimo) torej dobimo:

	GS	SDS	NSi	SD	Levica	Ostali
Opažene	203	192	53	58	45	173
Pričakovane	249˙418	169˙995	49˙666	48˙436	32˙290	174˙194

Pričakovane frekvence so vse večje od 5, zato ni treba združevati. Vrednost preizkusne statistike pride $\chi^2 \doteq 18{\cdot}61$, če vzamemo preliminarne opažene frekvence brez zaokrožanja, pa pride 18˙83. Ker je oboje večje od $\chi^2_{0.01}(5) \doteq 15{\cdot}09$, ničelno domnevo tudi pri stopnji značilnosti 0˙01 zavrnemo. Če torej sprejmemo 1% tveganje, lahko rečemo, da so se določeni volivci premislili.

Za vzporedne volitve pa dobimo:

	GS	SDS	NSi	SD	Levica	Ostali
Opažene	5447	3424	1035	1004	669	3637
Pričakovane	5241˙91	3572˙72	1043˙82	1017˙95	678˙63	3660˙97

in vrednost preizkusne statistike pride $\chi^2 \doteq 14{\cdot}77$, če vzamemo preliminarne opažene frekvence brez zaokrožanja, pa pride 14˙82 . Ker je oboje večje od $\chi^2_{0.05}(5) \doteq 11{\cdot}07$, a manjše od $\chi^2_{0.01}(5) \doteq 15{\cdot}09$, ničelno domnevo pri stopnji značilnosti 0˙05 zavrnemo, medtem ko je pri stopnji značilnosti 0˙01 ne zavrnemo.

23. Diskretizirana porazdelitev iz ničelne domneve:

pod -1	od -1 do 0	od 0 do 1	nad 1
0˙1839	0˙3161	0˙3161	0˙1839

$\chi^2 \doteq 21{\cdot}9, \quad \chi^2_{0.05}(3) \doteq 7{\cdot}815.$
Domnevo zavrnemo.

24. Diskretizirana porazdelitev iz ničelne domneve:

pod 10	10–20	20–30	30–40	nad 40
0˙0668	0˙2417	0˙3829	0˙2417	0˙0668

$\chi^2 \doteq 11\dot{}8, \quad df = 4, \quad \chi^2_{0.05}(4) \doteq 9\dot{}488.$
Domnevo zavrnemo.

25. a) Že v 16. nalogi iz 11. razdelka smo izračunali oceno po metodi največjega verjetja $\hat{\theta} = 0\dot{}65$. To nam da oceno celotne porazdelitve:

$$\begin{pmatrix} RR & Rr & rr \\ 0\dot{}1225 & 0\dot{}455 & 0\dot{}4225 \end{pmatrix}.$$

Od tod dobimo:

$$\chi^2 = \frac{(20 - 12\dot{}25)^2}{12\dot{}25} + \frac{(30 - 45\dot{}5)^2}{45\dot{}5} + \frac{(50 - 42\dot{}25)^2}{42\dot{}25} \doteq 11\dot{}60\,,$$

kar primerjamo s $\chi^2_{0.05}(1) \doteq 3\dot{}841$. Domnevo torej zavrnemo.

b) Naj bodo p, q in r populacijski deleži genotipov RR, Rr in rr v širšem statističnem modelu. V ožjem modelu pa je $p = (1-\theta)^2$, $q = 2\theta(1-\theta)$ in $r = \theta^2$ za neki $\theta \in (0,1)$. To lahko spravimo v okvir Wilksovega izreka tako, da upoštevamo $q = 1 - p - r$ in postavimo:

$$\Theta := \{(p,r)\ ;\ p > 0, r > 0, p + r < 1\}\,, \qquad \Theta_{H_0} := \left\{\left((1-\theta)^2, \theta^2\right)\ ;\ 0 < \theta < 1\right\}.$$

Ni težko preveriti, da je preslikava $\theta \mapsto \left((1-\theta)^2, \theta^2\right)$ neskončnokrat zvezno odvedljiva topološka vložitev intervala $(0,1)$ v množico Θ. Tako je ožji model Θ_{H_0} podmnogoterost razreda $\mathcal{C}^{(\infty)}$ dimenzije 1 v prostoru Θ dimenzije 2.

Tako kot v 16. nalogi iz 11. razdelka tudi tu priredimo opažanje. Postavimo se v položaj, preden smo določili genotipe. V vzorcu je n osebkov. Glede na besedilo naloge je opažanje vektor (N_1, N_2, N_3), kjer je N_1 število osebkov z genotipom RR, N_2 število osebkov z genotipom Rr in N_3 število osebkov z genotipom rr ($N_1 + N_2 + N_3 = n$). Zdaj pa vzemimo širše opažanje, kjer osebke v vzorcu oštevilčimo in zabeležimo, da ima i-ti osebek genotip X_i: to opažanje sestoji iz vektorja $(X_1, X_2, \ldots, X_n)$ in verjetje je enako:

$$L = p^{N_1} q^{N_2} r^{N_3}\,,$$

kar je funkcija vektorja (N_1, N_2, N_3), zato je le-ta zadostna statistika. Zato lahko opažanje (N_1, N_2, N_3) nadomestimo z opažanjem $(X_1, X_2, \ldots, X_n)$ in delamo z verjetjem glede nanj.

Za oceno parametrov v širšem modelu Θ najprej opazimo, da je verjetje na parametričnem prostoru Θ strogo pozitivno. Še več, brž ko je $N_1, N_2, N_3 > 0$, gre verjetje proti nič, če se bližamo njegovemu robu. Zato je maksimum dosežen v stacionarni točki v notranjosti, kjer lahko logaritmiramo. Glede na to, da imamo vez $p + q + r = 1$, nastavimo Lagrangeovo funkcijo za logaritem verjetja:

$$F = N_1 \ln p + N_2 \ln q + N_3 \ln r - \lambda(p + q + r)\,.$$

Iz nje dobimo, da cenilke v širšem modelu zadoščajo sistemu:

$$\frac{N_1}{\hat{p}} = \frac{N_2}{\hat{q}} = \frac{N_3}{\hat{r}}\,, \qquad \hat{p} + \hat{q} + \hat{r} = 1\,,$$

ki ima rešitev:

$$\hat{p} = \frac{N_1}{n}\,, \qquad \hat{q} = \frac{N_2}{n}\,, \qquad \hat{r} = \frac{N_3}{n}\,.$$

V našem primeru pride $\hat{p} = 0{\cdot}2$, $\hat{q} = 0{\cdot}3$ in $\hat{r} = 0{\cdot}5$, torej je:

$$\sup_{\Theta} L = 0{\cdot}2^{20}\, 0{\cdot}3^{30}\, 0{\cdot}5^{50}\,.$$

Za ožji model pa smo že prej dobili oceno $\hat{\theta} = 0{\cdot}65$, torej je:

$$\sup_{\Theta_{H_0}} L = 0{\cdot}4225^{20}\, 0{\cdot}455^{30}\, 0{\cdot}1225^{50}\,.$$

Razmerje verjetij je torej enako:

$$\Lambda = \frac{\sup_{\Theta_{H_0}} L}{\sup_{\Theta} L} = \left(\frac{0{\cdot}4225}{0{\cdot}2}\right)^{20} \left(\frac{0{\cdot}455}{0{\cdot}3}\right)^{30} \left(\frac{0{\cdot}1225}{0{\cdot}2}\right)^{50}\,,$$

od koder dobimo numerično $-2\ln\Lambda \doteq 11{\cdot}45$, kar se ne razlikuje prav dosti od Pearsonove statistike. To primerjamo z isto kritično vrednostjo kot prej, namreč $\chi^2_{0.05}(1) \doteq 3{\cdot}841$, in domnevo spet zavrnemo.

26. Gre za isti statistični model kot v 15. nalogi iz 11. razdelka, torej dobimo oceni $\hat{a} = 1/4$ in $\hat{b} = 3/10$, od koder dobimo $\chi^2 = 2$. To spet primerjamo s $\chi^2_{0.05}(1) \doteq 3{\cdot}841$ in dobimo, da domneve ne moremo zavrniti.

27. Porazdelitev najprej diskretiziramo:

pod -1	od -1 do 0	od 0 do 1	nad 1
$\theta/2$	$(1-\theta)/2$	$(1-\theta)/2$	$\theta/2$

,

kjer je $\theta = e^{-\lambda}$. Iz vzorca dobimo $\hat{\theta} = 0{\cdot}59$ in pripadajočo ocenjeno diskretizirano porazdelitev:

pod -1	od -1 do 0	od 0 do 1	nad 1
0·295	0·205	0·205	0·295

kar nam da $\chi^2 \doteq 0{\cdot}448$. Ker je $\chi^2_{0.05}(2) \doteq 5{\cdot}991$, domneve ne zavrnemo.

28. Velja $S_X = 2$ in $S_Y = 8$. Ker za slučajno spremenljivko $S' \sim \mathrm{Bin}(10, 1/2)$ velja $\mathbb{P}(S' \le 2) \doteq 0{\cdot}0547$, ničelne domneve o enakosti ustreznih verjetnosti s preizkusom z znaki pri tej stopnji značilnosti ne moremo zavrniti.

Pač pa lahko zavrnemo domnevo, da imata enaki povprečji:
$\bar{X} = 17{\cdot}4$, $\quad \bar{Y} = 27{\cdot}1$, $\quad s \doteq 11{\cdot}27$, $\quad T \doteq -2{\cdot}72$, $\quad t_{0.05}(9) \doteq 1{\cdot}833$.

29. Velja $S_X = 10$ in $S_Y = 2$. Ker za slučajno spremenljivko $S' \sim \mathrm{Bin}(12, 1/2)$ velja $\mathbb{P}(S' \le 2) \doteq 0{\cdot}0193$, ničelno domnevo o enakosti porazdelitev zavrnemo: sprejmemo domnevo, da je dogodek $\{X > Y\}$ verjetnejši od dogodka $\{Y > X\}$.

Pač pa ne moremo zavrniti domneve, da imata X in Y enaki povprečji:
$\bar{X} = 27{\cdot}25, \quad \bar{Y} = 26{\cdot}5, \quad s \doteq 4{\cdot}575, \quad T \doteq 0{\cdot}57, \quad t_{0.05}(11) \doteq 1{\cdot}796.$

30. Velja $S_X = 12$ in $S_Y = 4$; če je $S' \sim \mathrm{Bin}(16, 1/2)$, velja $\mathbb{P}(S' \le 4) \doteq 0{\cdot}0384$ in $\mathbb{P}(S' \ge 4) \doteq 0{\cdot}9894$.

a) Domneve ne moremo zavrniti.

b) Ničelno domnevo zavrnemo proti domnevi, da je dogodek $\{X > Y\}$ verjetnejši od dogodka $\{Y > X\}$. Ne zavrnemo pa je proti domnevi, da sta verjetnosti teh dveh dogodkov različni.

c) Iz:

$$\bar{X} = 25{\cdot}5\,, \quad \bar{Y} = 48{\cdot}375\,, \quad s \doteq 44{\cdot}36\,, \quad T \doteq 2{\cdot}06\,, \quad t_{0.05}(15) \doteq 1{\cdot}753$$

dobimo, da lahko ničelno domnevo zavrnemo proti domnevi, da ima Y ves čas večje povprečje od X. Tako smo pri preizkusu povprečij zavrnili ničelno domnevo ravno v nasprotno smer kot pri preizkusu z znaki.

Opomba. Kontradiktorna rezultata smo dobili, ker preizkus z znaki gleda le, ali vplivi večajo določeno spremenljivko na račun druge, preizkus povprečij pa tudi, koliko jo večajo. Tako lahko vpliv, ki določeno spremenljivko močno poveča, prevlada nad več vplivi, ki delujejo v nasprotno smer, a le malo.

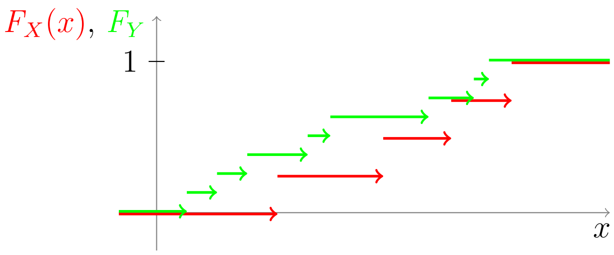
31. Ko preštejemo, dobimo, da se je 12 ljudi pred ogledom počutilo boljše kot po ogledu, 25 ljudi pa po ogledu boljše kot pred ogledom; 13 ljudi se je pred in po ogledu počutilo enako. Ker je:

$$Z = \frac{|12 - 25| - 1}{\sqrt{37}} \doteq 1{\cdot}97 > 1{\cdot}960\,,$$

domnevo zavrnemo.

32. $S_{\text{prej}} = 4, \quad S_{\text{potem}} = 30, \quad Z \doteq 4{\cdot}29 > 2{\cdot}576.$
Domnevo zavrnemo – sprejmemo domnevo, da se je kakovost hrane spremenila.

33. Grafa:



Porazdelitev slučajne spremenljivke je stohastično strogo večja od porazdelitve slučajne spremenljivke Y.

34. $m = 9$, $n = 11$, $\sum_{i=1}^{9} R_i^M = 70$, $Z_M \doteq -1{\cdot}82$, $z_{0.025} \doteq 1{\cdot}960$.
Domneve ne moremo zavrniti.

35. Pri preizkusu z znaki bo na eni tretjini enot veljalo $X > Y$, na dveh tretjinah pa $X < Y$. Za dovolj velik vzorec bo torej preizkus z znaki zavrnil ničelno domnevo o uravnovešenosti vplivov, in sprejel domnevo, da vplivi niso uravnovešeni oz. da vplivi, ki večajo Y na račun X, prevladujejo nad vplivi, ki večajo X na račun Y.

Pri inverzijskem preizkusu pa bodo vse vrednosti enako zastopane, zato bo vrednost preizkusne statistike enaka nič. Ta preizkus torej nikoli ne bo sprejel domneve, da je ena od spremenljivk stohastično strogo večja od druge.

36. Pripravljalo se jih je 12, 8 pa se jih ni pripravljalo. Vsota rangov tistih, ki se niso pripravljali, je 110. Preizkusna statistika v korist teh rangov pride $Z \doteq 1{\cdot}97$, kar je več od kritične vrednosti $1{\cdot}645$. Ničelno domnevo tako zavrnemo in sprejmemo domnevo, da so tisti, ki so se pripravljali, tekli bolje od tistih, ki se niso.

Pri T-preizkusu pa povprečje tistih, ki so se pripravljali, pride $8{\cdot}083$, povprečje tistih, ki se niso pripravljali, pa $8{\cdot}362$. Velja še $s \doteq 0{\cdot}4066$, preizkusna statistika pa pride $T \doteq 1{\cdot}50$, kar primerjamo s kritično vrednostjo $t_{0.05}(18) \doteq 1{\cdot}734$. Ničelne domneve tokrat ne moremo zavrniti.

37. Če mnenja uredimo navzdol, od 'odlična' proti 'grozna', so vezani rangi razvidni iz naslednje tabele:

mnenje	ženske	moški	skupaj	kum.	vez. rang
odlična	3	0	3	3	2
dobra	21	13	34	37	20·5
povprečna	28	20	48	85	61·5
slaba	18	20	38	123	104·5
grozna	22	41	63	186	155
Skupaj	92	94			

Vsota vezanih rangov za ženske je:

$$3 \cdot 2 + 21 \cdot 20{\cdot}5 + 28 \cdot 61{\cdot}5 + 18 \cdot 104{\cdot}5 + 22 \cdot 155 = 7449{\cdot}5\,,$$

preizkusna statistika pa pride:

$$\sqrt{\frac{3}{92 \cdot 94 \cdot 187}}\bigl(2 \cdot 7449{\cdot}5 - 92 \cdot 187\bigr) \doteq -3{\cdot}14\,.$$

Ker je preizkusna statistika negativna, so bili rangi pri ženskah v povprečju nižji, kar pri urejenosti, za katero smo se dogovorili, pomeni, da so imele na našem vzorcu ženske o premierki boljše mnenje. Kritična vrednost je $z_{0.025} \doteq 2{\cdot}326$, torej ničelno domnevo zavrnemo: sprejmemo, da imajo ženske in moški različno dobro mnenje o premierki.

Opomba. Če ignoriramo urejenost odgovorov, se da izvesti tudi kontingenčni preizkus hi kvadrat. Pričakovane frekvence pridejo:

mnenje	ženske	moški
odlična	1˙48	1˙52
dobra	16˙82	17˙18
povprečna	23˙74	24˙26
slaba	18˙80	19˙20
grozna	31˙16	31˙84

Vrednost preizkusne statistike pa pride $\chi^2 \doteq 12˙03$, kar ne presega kritične vrednosti $\chi^2_{0.01}(4) \doteq 13˙28$, zato ničelne domneve ne moremo zavrniti. Je pa res, da so pričakovane frekvence tistih, ki imajo odlično mnenje, tako pri moških kot tudi pri ženskah pod 5. Zato odlična in dobra mnenja združimo in dobimo:

Opažene frekvence:

mnenje	ženske	moški
odlična, dobra	24	13
povprečna	28	20
slaba	18	20
grozna	22	41

Pričakovane frekvence:

mnenje	ženske	moški
odlična, dobra	18˙30	18˙70
povprečna	23˙74	24˙26
slaba	18˙80	19˙20
grozna	31˙16	31˙84

Vrednost preizkusne statistike zdaj pride $\chi^2 \doteq 10˙42$, kar spet ne presega kritične vrednosti $\chi^2_{0.01}(3) \doteq 11˙34$, zato ničelne domneve ne moremo zavrniti.

Kakor koli, brez urejenosti odgovorov torej različnost empirične porazdelitve med ženskami in moškimi ni tako prepričljiva.

14. Povezanost dveh številskih spremenljivk

1. Označimo z X sistolični, z Y pa diastolični pritisk. Izračunajmo:
$\bar{X} = 126$, $\bar{Y} \doteq 79{\cdot}67$, $C_x \doteq 30{\cdot}17$, $C_y \doteq 19{\cdot}32$, $C_{xy} = 130$, $R \doteq 0{\cdot}223$, $Z \doteq 0{\cdot}227$, $c \doteq 1{\cdot}960$.
Interval zaupanja: $-0.327 < \rho < 0.656$.

2. $T \doteq 0{\cdot}825$, $df = 13$, $t_{0.05}(13) \doteq 1{\cdot}771$.
Domneve ne moremo zavrniti.

3. Teoretične frekence:

oči\lasje	rdeči, blond	rjavi, črni	Skupaj
modre	6˙29	6˙71	13
zelene	11˙12	11˙87	23
rjave	12˙58	13˙42	26
Skupaj	30	32	62

so vse višje od 5. Velja $\chi^2 \doteq 22{\cdot}8$, kar primerjamo z $\chi^2_{0.01}(2) \doteq 9{\cdot}210$.
Domnevo o neodvisnosti zavrnemo: barva las in barva oči sta bili statistično zelo značilno povezani.

4. Najprej opazimo, da so glede na to, da so vse opažene frekvence vsaj 5, tudi vse pričakovane frekvence vsaj 5. Nadalje pride $\chi^2 \doteq 0{\cdot}0089$, kar primerjemo s $\chi^2_{0.05}(1) \doteq 3{\cdot}841$. Domneve o neodvisnosti ne moremo zavrniti.

5. Vse opažene in zato tudi vse pričakovane frekvence so enake vsaj 5.
$\chi^2 \doteq 169$, $\chi^2_{0.05}(8) \doteq 15{\cdot}51$. Domnevo zavrnemo.

15. Linearna regresija

1. $\bar{X} = 3$, $\bar{Y} = 2{\cdot}6$, $C_{xx} = 14$, $C_{xy} = 7$, $C_{yy} = 5{\cdot}2$.
Regresijski premici: $y = 1{\cdot}1 + 0{\cdot}5\,x$ in $x = -\frac{1}{2} + \frac{35}{26}\,y \doteq -0{\cdot}5 + 1{\cdot}346\,y$.

2. $t_{0.025}(3) \doteq 3{\cdot}182$, $S \doteq 0{\cdot}753$,
$\Delta_a \doteq 2{\cdot}19$, $-1{\cdot}10 < a < 3{\cdot}30$,
$\Delta_b \doteq 0{\cdot}64$, $0{\cdot}14 < b < 1{\cdot}14$.

3. Pri $X = 10$ je $\hat{Y} = 6{\cdot}1$.
$t_{0.025}(3) \doteq 3{\cdot}182$, $S \doteq 0{\cdot}753$, $\Delta \doteq 5{\cdot}2$.
Napovedni interval: $0{\cdot}9 < Y < 11{\cdot}3$.

4. a) V modelu $\Phi = a + bP + \varepsilon$ pride $\hat{a} \doteq -525{\cdot}76$ in $\hat{b} \doteq 18{\cdot}98$. Ocena svetilnosti 75-wattne žarnice pride 898 lm, ocena svetilnosti 15-wattne žarnice pa pride -241 lm, kar je zgrešeno.

b) Na enostavno linearno regresijo prevedemo tako, da zvezo logaritmiramo:

$$\ln \Phi = \ln k + \alpha \ln P\,.$$

Z napako vred se regresijski model zdaj glasi:

$$\ln \Phi = \ln k + \alpha \ln P + \varepsilon \quad \text{oziroma} \quad \Phi = k\,P^{\alpha} e^{\varepsilon}\,.$$

Za oceni koeficientov dobimo $\hat{\alpha} \doteq 1{\cdot}1876$, $\widehat{\ln k} \doteq 1{\cdot}7048$ in $\hat{k} \doteq 5{\cdot}5000$. Ocenjena svetilnost 75-wattne žarnice pride 927 lm, kar je blizu ocene iz direktnega linearnega modela, ocenjena svetilnost 15-wattne žarnice pa pride 137 lm, kar je vsekakor bolj realistično kot prej.

5. a) Ocenjuje se, da se segreva se s stopnjo $0{\cdot}0539$ stopinje na leto, torej približno vsakih 18 let za eno stopinjo.

b) Ocenjena standardna napaka strmine znaša $0{\cdot}00879$, vrednost preizkusne statistike pa $6{\cdot}140$. Glede na to, da je $t_{0.005}(34) \doteq 2{\cdot}728$, ni dvoma, da se podnebje segreva.

c) Točkasta napoved: $14{\cdot}20$°C.
$t_{0.025}(34) \doteq 2{\cdot}032$, napovedni interval: ($12{\cdot}79$°C, $15{\cdot}62$°C).

6. Poiskati moramo, pri katerem b je izraz:

$$Q = (1-a)^2 + (3-2a)^2 + (2-3a)^2 + (3-3a)^2 + (4-6a)^2$$

minimalen. Iz:

$$\frac{\mathrm{d}Q}{\mathrm{d}a} = -92 + 118a$$

dobimo točkasto oceno $\hat{a} = 92/118 \doteq 0{\cdot}780$. Napoved pri $X = 10$ je torej
$\hat{Y} = 920/118 \doteq 7{\cdot}80$.