University of Primorska

FAMNIT, Mathematics

Probability

Written examination

June 26th, 2023

Instructions

Read carefully the text of the problems before attempting to solve them. Five problems out of six count for 100%. You are allowed one A4 sheet with formulae and theorems, and a handbook of mathematics. Time allowed: 120 minutes.

Question	a.	b.	c.	d.	Total
1.			●	●	
2.			●	●	
3.			●	●	
4.				●	
5.			●	●	
6.			●	●	
Total					

1. (20) We toss n balls into r boxes where $n \geq 2r$. Tosses are independent and we hit every box with the same probability $1/r$. Let A_i be the event that the i-th box contains exactly two balls for $i = 1, 2, \ldots, r$.

a. (10) Compute the probability $P(A_1 \cap A_2 \cap \cdots \cap A_i)$ for $i \leq r$.

Solution: first, we select the pairs of tosses that will land in boxes $1, 2, \ldots, i$. We can do this in

$$\binom{n}{2}\binom{n-2}{2}\cdots\binom{n-2(i-1)}{2} = \frac{n!}{2^i \cdot (n-2i)!}$$

ways, where $0! = 1$. The remaining $n - 2i$ tosses must land in the other $r - i$ boxes. By independence

$$P(A_1 \cap A_2 \cap \cdots \cap A_i) = \frac{n!}{2^i \cdot (n-2i)!} \cdot \left(\frac{1}{r}\right)^{2i} \cdot \left(\frac{r-i}{r}\right)^{n-2i},$$

where $0^0 = 1$.

b. (10) What is the probability that no box will contain exactly two balls? You do not need to simplify sums and binomial coefficients.

Solution: the event that at least one box will contain exactly 2 balls is $\cup_{i=1}^r A_i$. By the inclusion-exclusion formula and symmetry we have

$$P\left(\bigcup_{i=1}^r A_i\right) = \sum_{i=1}^r (-1)^{i-1} \binom{r}{i} \frac{n!}{2^i \cdot (n-2i)!} \cdot \left(\frac{1}{r}\right)^{2i} \cdot \left(\frac{r-i}{r}\right)^{n-2i},$$

while the desired probability is

$$1 - P\left(\bigcup_{i=1}^r A_i\right) = \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{n!}{2^i \cdot (n-2i)!} \cdot \left(\frac{1}{r}\right)^{2i} \cdot \left(\frac{r-i}{r}\right)^{n-2i}.$$

2. (20) An urn contains a white and $b \geq 2$ black balls. We draw balls at random. Once a ball is drawn, it is replaced by a white ball irrespective of its color. The draws are independent. Let X be the number of draws up to and including the first black ball, and let Y be the number of draws after the first black ball up to and including the second black ball.

a. (10) Find the joint distribution of X and Y.

Solution: the possible values of the pair (X, Y) are all integer pairs (k, l) with $k, l \geq 1$. The event $\{X = k, Y = l\}$ happens if we get $k - 1$ white balls, a black ball, then $l - 1$ white balls and then a black ball. We have

$$P(X = k, Y = l) = \left(\frac{a}{a+b}\right)^{k-1} \cdot \frac{b}{a+b} \cdot \left(\frac{a+1}{a+b}\right)^{l-1} \cdot \frac{b-1}{a+b}.$$

The expression simplifies to

$$P(X = k, Y = l) = \frac{a^{k-1}(a+1)^{l-1}b(b-1)}{(a+b)^{k+l}}.$$

In other words, X and Y are independent with $X \sim \text{Geom}\left(\frac{b}{a+b}\right)$ and $Y \sim \text{Geom}\left(\frac{b+1}{a+b}\right)$.

b. (10) Find the distribution of $Z = X + Y$.

Solution: for $n \geq 2$ we compute

$$
\begin{aligned}
P(Z = n) &= \sum_{k=1}^{n-1} P(X = k, Y = n - k) \\
&= \frac{b(b-1)}{(a+b)^n} \sum_{k=1}^{n-1} a^{k-1}(a+1)^{n-k-1} \\
&= \frac{b(b-1)(a+1)^{n-2}}{(a+b)^n} \sum_{k=1}^{n-1} \left(\frac{a}{a+1}\right)^{k-1} \\
&= \frac{b(b-1)(a+1)^{n-2}}{(a+b)^n} \cdot \frac{1 - \left(\frac{a}{a+1}\right)^{n-1}}{1 - \frac{a}{a+1}} \\
&= \frac{b(b-1)(a+1)^{n-2}}{(a+b)^n} \cdot \frac{(a+1)^{n-1} - a^{n-1}}{(a+1)^{n-2}} \\
&= \frac{b(b-1)}{(a+b)^n} \left((a+1)^{n-1} - a^{n-1}\right).
\end{aligned}
$$

3. (20) Let X and Y be independent with

$$X \sim \Gamma(a, 1) \quad \text{and} \quad Y \sim \Gamma\left(a + \frac{1}{2}, 1\right).$$

Define

$$(U, V) = \left(2\sqrt{\frac{Y}{X}}, \, 2\sqrt{XY}\right).$$

a. (10) Compute the density of the vector (U, V).

Solution: the map

$$\Phi\left(2\sqrt{\frac{y}{x}}, \, 2\sqrt{xy}\right)$$

is bijective on $(0, \infty)^2$ with

$$\Phi^{-1}(u, v) = \left(\frac{v}{u}, \, \frac{uv}{4}\right).$$

The maps Φ and Φ^{-1} both have continuous partial derivatives. We compute

$$J_{\Phi^{-1}}(u, v) = \det\begin{pmatrix} -\frac{v}{u^2} & \frac{1}{u} \\ \frac{1}{4}v & \frac{1}{4}u \end{pmatrix} = -\frac{v}{2u}.$$

The transformation formula gives

$$f_{U,V}(u, v) = \frac{1}{\Gamma(a)\,\Gamma\left(a + \frac{1}{2}\right)} \left(\frac{v}{u}\right)^{a-1} e^{-\frac{v}{u}} \left(\frac{uv}{4}\right)^{a-\frac{1}{2}} e^{-\frac{uv}{4}} \cdot \frac{v}{2u}$$

for $u, v > 0$; elsewhere, one can set $f_{U,V}(u, v) = 0$. The density simplifies to

$$f_{U,V}(u, v) = \frac{1}{4^a\,\Gamma(a)\,\Gamma\left(a + \frac{1}{2}\right)} u^{-\frac{1}{2}} \cdot v^{2a-\frac{1}{2}} \cdot e^{-\frac{uv}{4} - \frac{v}{u}}.$$

b. (10) Find the distribution of V, naming it explicitly.

Hint: you can assume as known that

$$\int_0^\infty \frac{1}{\sqrt{s}} e^{-\alpha s - \frac{\beta}{s}}\, ds = \sqrt{\frac{\pi}{\alpha}}\, e^{-2\sqrt{\alpha\beta}}$$

for all $\alpha, \beta > 0$.

Solution: for $v > 0$, we compute

$$f_V(v) = \frac{v^{2a-\frac{1}{2}}}{4^a\,\Gamma(a)\,\Gamma\left(a + \frac{1}{2}\right)} \int_0^\infty \frac{1}{\sqrt{u}} e^{-\frac{uv}{4} - \frac{v}{u}}\, du$$

$$= \frac{v^{2a-\frac{1}{2}}}{4^a\,\Gamma(a)\,\Gamma\left(a + \frac{1}{2}\right)} \cdot \sqrt{\frac{4\pi}{v}}\, e^{-v}$$

$$= \frac{\sqrt{\pi}}{2^{2a-1}\,\Gamma(a)\,\Gamma\left(a + \frac{1}{2}\right)} \cdot v^{2a-1} e^{-v};$$

for $v \leq 0$, we can, of course, set $f_V(v) = 0$. Noting that we can ignore a constant factor, we infer that $V \sim \Gamma(2a, 1)$.

Remark: one implication of the above calculation is the Legendre duplication formula

$$\Gamma(2a) = \frac{1}{\sqrt{\pi}} \, 2^{2a-1} \, \Gamma(a) \, \Gamma\left(a + \frac{1}{2}\right).$$

4. (20) An urn contains a white and $b \geq 2$ black balls. We draw balls at random. Once a ball is drawn, it is replaced by a white ball irrespective of its color. The draws are independent. Let X be the number of draws up to and including the first black ball, and let Y be the number of draws after the first black ball up to and including the second black ball. Denote $e_{a,b} = E(X_{a,b})$ and $v_{a,b} = \text{var}(X_{a,b})$.

a. (5) Let Z be the number of draws until we select the first black ball, including the first black ball. Show that Z and $X_{a,b} - Z$ are independent and that $X_{a,b} - Z$ has the same distribution as $X_{a+1,b-1}$.

Solution: when the first black ball has been drawn, we are left with $a+1$ white and $b-1$ black balls. Given $\{Z = k\}$, $X_{a,b} - Z = X_{a,b} - k$ is the number of remaining draws until we select the last black ball. By independence, its conditional distribution of $X_{a,b} - Z$ given $\{X = k\}$ is the same as the distribution of $X_{a+1,b-1}$, irrespective of k. However, this means that $X_{a,b} - Z$ and Z are independent, and $X_{a,b} - Z$ has the same distribution as $X_{a+1,b-1}$.

b. (10) Compute $e_{a,b}$. The solution is a sum that you do not need to simplify.

Solution: write
$$e_{a,b} = E(X_{a,b}) = E(Z) + E(X_{a,b} - Z).$$
Because $Z \sim \text{Geom}(b/(a+b))$, we have
$$E(Z) = \frac{a+b}{b}.$$
The second expectation can be deduced from part a., transforming the above formula to
$$e_{a,b} = \frac{a+b}{b} + e_{a+1,b-1}.$$
Iterating we get
$$e_{a,b} = \frac{a+b}{b} + \frac{a+b}{b-1} + \cdots + \frac{a+b}{2} + e_{a+b-1,1}.$$
We have that
$$X_{a+b-1,1} \sim \text{Geom}\left(\frac{1}{a+b}\right)$$
and hence
$$e_{a+b-1,1} = E(X_{a+b-1,1}) = a+b.$$
Finally,
$$e_{a,b} = (a+b)\sum_{k=1}^{b}\frac{1}{k}.$$

c. (5) Let $v_{a,b} = \text{var}(X_{a,b})$. Show that
$$v_{a,b} = \frac{a(a+b)}{b^2} + v_{a+1,b-1}$$

and compute $v_{a,b}$. The solution is a sum that you do not need to simplify.

Solution: since the variance of the sum of independent random variables is the sum of variances, we have

$$v_{a,b} = \text{var}(Z) + \text{var}(X_{a,b} - Z) = \frac{a(a+b)}{b^2} + v_{a+1,b-1}.$$

Iterating we get

$$v_{a,b} = \frac{a(a+b)}{b^2} + \frac{(a+1)(a+b)}{(b-1)^2} + \cdots + \frac{(a+b-2)(a+b)}{2^2} + v_{a+b-1,1}.$$

Recalling the distribution of $X_{a+b-1,1}$, we find that

$$v_{a+b-1,1} = \text{var}(X_{a+b-1,1}) = (a+b)(a+b-1),$$

concluding that

$$v_{a,b} = (a+b) \sum_{k=1}^{b} \frac{a+b-k}{k^2} = (a+b)^2 \sum_{k=1}^{b} \frac{1}{k^2} - (a+b) \sum_{k=1}^{b} \frac{1}{k}.$$

5. (20) Let the random variables $I \sim$ Bernoulli(p) and W_1 be independent, integer valued and nonnegative. Assume that the variable

$$W = 1 + I \cdot W_1$$

has the same distributions as W_1.

a. (10) Show that
$$G_W(s) = qs + psG_W(s),$$

where $q = 1 - p$.

Solution: we have
$$E\left(s^W | I = 0\right) = s$$

and
$$E\left(s^W | I = 1\right) = sE\left(s^{W_1}\right) = sG_W(s).$$

The formula follows from the law for total expectations.

b. (10) What is the distribution of W?

Solution: the first part gives

$$G_W(s) = \frac{qs}{1 - ps} = \sum_{k=1}^{\infty} qp^{k-1}s^k.$$

We have
$$P\left(W = k\right) = qp^{k-1}$$

or $W \sim$ Geom(q).

6. (20) A bored statistician draws tickets from one of the boxes below. She denotes the numbers on the tickets by $X_1, X_2, \ldots, X_n$, and their sum by S_n.

(i) $\boxed{-1} \quad \boxed{0} \quad \boxed{1}$

(ii) $\boxed{-1} \quad \boxed{0} \quad \boxed{0} \quad \boxed{0} \quad \boxed{0} \quad \boxed{0} \quad \boxed{0} \quad \boxed{0} \quad \boxed{1}$

a. (10) The statistician computes

$$P(-30 \leq S_{1000} \leq 30) \approx 0.96 \, .$$

Which box is she drawing tickets from? Justify your answer.

Solution: for the first box we get $\operatorname{var}(X_1) = 2/3$, *and for the second* $\operatorname{var}(X_1) = 2/9$. *Using the central limit theorem and letting* $Z \sim N(0,1)$, *we get*

$$
\begin{aligned}
P(-30 \leq S_{1000} \leq 30) &= P\left(-\frac{30}{\sqrt{1000}\sqrt{2/3}} \leq \frac{S_{1000}}{\sqrt{\operatorname{var}(S_{1000})}} \leq \frac{30}{\sqrt{1000}\sqrt{2/3}}\right) \\
&= P\left(-1.16 \leq \frac{S_{1000}}{\sqrt{\operatorname{var}(S_{1000})}} \leq 1.16\right) \\
(CLT) \quad &\approx P\left(-1.16 \leq Z \leq 1.16\right) \\
&\doteq 0.75 \, .
\end{aligned}
$$

for the first box, and

$$
\begin{aligned}
P(-30 \leq S_{1000} \leq 30) &= P\left(-\frac{30}{\sqrt{1000}\sqrt{2/9}} \leq \frac{S_{1000}}{\sqrt{\operatorname{var}(S_{1000})}} \leq \frac{30}{\sqrt{1000}\sqrt{2/9}}\right) \\
&= P\left(-2.01 \leq \frac{S_{1000}}{\sqrt{\operatorname{var}(S_{1000})}} \leq 2.01\right) \\
(CLT) \quad &\approx P\left(-2.01 \leq Z \leq 2.01\right) \\
&\doteq 0.96 \, .
\end{aligned}
$$

for the second. The statistician draws from the second box.

b. (10) The statistician computes $P(S_{100} = 0) \approx 0.049$. For which of the two boxes is the above approximation. Justify your answer.

Solution: using the central limit theorem we approximate for the first box

$$
\begin{aligned}
P(S_{100} = 0) &= P\left(-\frac{1}{2} \leq S_{100} \leq \frac{1}{2}\right) \\
&= P\left(-\frac{1}{20\sqrt{2/3}} \leq \frac{S_{100}}{\sqrt{\operatorname{var}(S_{100})}} \leq \frac{1}{20\sqrt{2/3}}\right) \\
(CLT) \quad &\approx P\left(-0.061 \leq Z \leq 0.061\right) \\
&\doteq 0.049 \, .
\end{aligned}
$$

Similarly, for the second box we get the approximation $P(S_{100} = 0) \approx 0.085$. The statistician draws from the first box.