1 Vsota diskretnih slučajnih spremenljivk

Naj bosta X in Y neodvisni Bernoullijevo porazdeljeni spremenljivki, $B(p)$.

- Kako je porazdeljena njuna vsota?
  Označimo $Z = X + Y$. Velja

  $$P(Z = z) = P(X + Y = z) = \sum_y P(X = z - y | Y = y) P(Y = y)$$

  Za neodvisni X in Y torej velja

  $$P(Z = z) = P(X + Y = z) = \sum_y P(X = z - y) P(Y = y)$$

  V našem primeru:

  $$\begin{aligned} P(Z = 0) &= P(X + Y = 0) = P(X = 0)P(Y = 0) = (1 - p)^2 \\ P(Z = 1) &= P(X = 0)P(Y = 1) + P(X = 1)P(Y = 0) \\ &= (1 - p)p + p(1 - p) = 2p(1 - p) \\ P(Z = 2) &= P(X = 1)P(Y = 1) = p^2 \end{aligned}$$

  Velja torej

  $$P(Z = z) = \binom{2}{z} p^z (1 - p)^{2-z}$$

- Kako pravimo porazdelitvi vsote n i.i.d. Bernoullijevih spremenljivk?

  Označimo $Z = \sum_{i=1}^{n} X_i$, $i = 1, \ldots, n$, $X_i \sim B(p)$. Z je porazdeljena binomsko, $Z \sim Bin(n, p)$.

  $$P(Z = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

2 Vsota zveznih slučajnih spremenljivk

Poleg posamičnih vrednosti, želijo pri športnikih proučevati tudi zaporedje večih meritev. Zanima nas porazdelitev vsote kvadriranih standardiziranih odmikov od povprečja pri ničelni domnevi, da športnik ni kriv. Naj bo torej Z standardizirani odmik od povprečja (po predpostavki normalno porazdeljen), zanima nas $\sum Z^2$.

- Kako se porazdeljuje vsota dveh neodvisnih zveznih spremenljivk (izpeljite formulo za dve zvezni spremenljivki, primerjajte jo s formulo za diskretne)

  Zapišemo ustrezno kumulativno porazdelitveno funkcijo kot integral večrazsežne porazdelitve v ustreznih mejah:

$$
\begin{aligned}
P(Z \leq z) &= P(X+Y \leq z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_{X,Y}(v-y,y)dvdy
\end{aligned}
$$

  pri čemer smo naredili substitucijo $x = v - y$. Sedaj lahko integrala zamenjamo in odvajamo (privzamemo, da je zunanji integral zvezen v z) ter dobimo:

$$
\begin{aligned}
P(Z \leq z) &= \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_{X,Y}(v-y,y)dydv \\
f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(z-y,y)dy \\
f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy
\end{aligned}
$$

  (zadnja vrstica velja, če sta X in Y neodvisni slučajni spremenljivki). Dobili smo rezultat, ki je analogen diskretni verziji.

- Kako se porazdeljuje $S = Z_1^2 + Z_2^2$?

  Izpeljali smo že, da je $Z^2 \sim \chi_1^2$, oziroma $Z^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$, torej

  $$f_{Z^2}(z) = \frac{1}{\sqrt{2\pi z}} \exp\left\{-\frac{z}{2}\right\};\ z > 0$$

  Izračunajmo gostoto vsote $Z_1^2 + Z_2^2$:

  $$\begin{aligned} f_S(s) &= \int_0^s f_{Z_1^2}(s - z_2) f_{Z_2^2}(z_2) dz_2 \\ &= \frac{1}{2\pi} \int_0^s \frac{1}{\sqrt{s - z_2}} \exp\left\{-\frac{s - z_2}{2}\right\} \frac{1}{\sqrt{z_2}} \exp\left\{-\frac{z_2}{2}\right\} dz_2 \\ &= \frac{1}{2\pi} \exp\left\{-\frac{s}{2}\right\} \int_0^s \frac{1}{\sqrt{s - z_2}} \frac{1}{\sqrt{z_2}} dz_2 \end{aligned}$$

  Naredimo substitucijo $z_2 = sv$, $dz_2 = sdv$, meje so torej od 0 do 1:

  $$\begin{aligned} f_S(s) &= \frac{1}{2\pi} \exp\left\{-\frac{s}{2}\right\} \int_0^1 \frac{1}{\sqrt{s - sv}} \frac{1}{\sqrt{sv}} z dv \\ &= \frac{1}{2\pi} \exp\left\{-\frac{s}{2}\right\} \int_0^1 \frac{1}{\sqrt{1 - v}} \frac{1}{\sqrt{v}} dv \\ &= \frac{1}{2\pi} \exp\left\{-\frac{s}{2}\right\} \pi \\ &= \frac{1}{2} \exp\left\{-\frac{s}{2}\right\} \end{aligned}$$

  Gostota gama porazdelitve je

  $$f_X(x) = \frac{\lambda^a x^{a-1} e^{-\lambda x}}{\Gamma(a)};\ x > 0, \lambda > 0, a > 0$$

  Dobljeni rezultat je torej porazdelitev gama z $\lambda = \frac{1}{2}$ in $a = 1$.

- Denimo, da so športnikove vrednosti na petih merjenjih naslednje: $1{,}6, 1{,}5, -1{,}6, 1{,}8, 1{,}4$. Kaj lahko sklepamo?

  Uporabimo da je vsota n neodvisnih enako porazdeljenih spremenljivk $X_i \sim \Gamma(\frac{1}{2}, \frac{1}{2})$ porazdeljena kot $\sum_{i=1}^n X_i \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ (dokazali smo le za $n = 2$).

```
> vr <- c(1.6,1.5,-1.6, 1.8, 1.4)
> rez <- sum(vr^2)
> rez
[1] 12.57
> 1-pgamma(rez, 2.5, 0.5)
[1] 0.02775943
```

Naša ničelna domneva je, da športnik ni kriv. Pod to ničelno domnevo se vsota kvadriranih odmikov porazdeljuje po gama porazdelitvi $\Gamma(\frac{5}{2}, \frac{1}{2})$. Verjetnost, da je vsota 12,57 ali več, je približno 0,03.

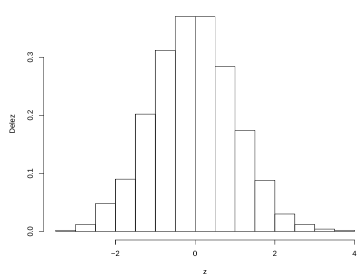
3 Vsota dveh odvisnih spremenljivk

Naj bosta X in Y neodvisni standardizirani normalni spremenljivki, Z pa enaka $|Y|$, če je $X > 0$, in $-|Y|$ če je $X < 0$.

- Kako je porazdeljena spremenljivka Z?

  Najprej narišimo simulirane vrednosti z R-om:

```
> set.seed(1)
> x <- rnorm(1000,0,1)          #1000 realizacij normalne spremenljivke, povprecje=0, sd=1
> y <- rnorm(1000,0,1)
> z <- abs(y)                   #z = |y|
> z[x<0] <- -z[x<0]             #z =-|y|, ce je x<0
> hist(z,main="",ylab="Delez",prob=T) #histogram z
```



Slika 1: *Porazdelitev spremenljivke Z.*

  Sedaj še izpeljimo porazdelitveno funkcijo. Naj bo $z < 0$, velja:

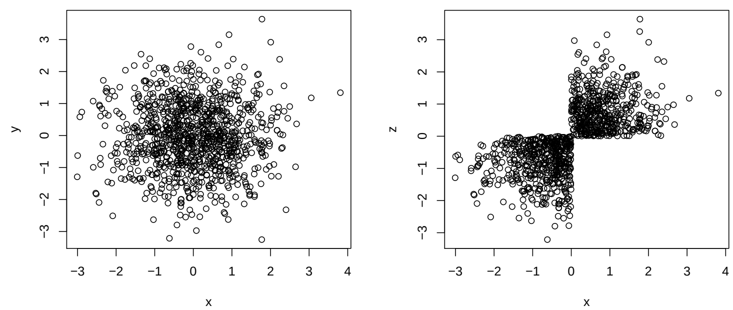
$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X < 0, |Y| \geq |z|) \\ &= P(X < 0)[P(Y \leq z) + P(Y \geq |z|)] \\ &= \frac{1}{2}[2 \cdot P(Y \leq z)] \\ &= P(Y \leq z) = F_Y(z) \end{aligned}$$

  Na enak način izpeljemo še $P(0 \leq Z \leq z) = P(0 \leq Y \leq y)$ za $z > 0$. Pokazali smo, da je porazdelitvena funkcija Z enaka porazdelitveni funkciji Y, Z je torej standardizirana normalna spremenljivka.

- Skicirajte skupno porazdelitev spremenljivk X in Z. Ali sta spremenljivki neodvisni?

```
> plot(x,y)
> plot(x,z)
```

Očitno je, da spremenljivki nista neodvisni, vedno imata enak predznak.

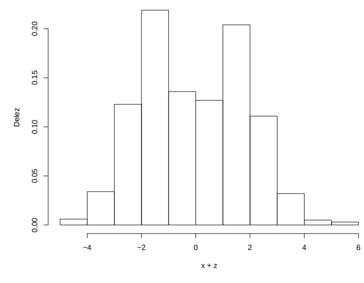


Slika 2: *Razsevni diagram realizacij X in Y ter X in Z.*

- Ali je vsota $X + Z$ porazdeljena normalno?

  Iz slike je očitno, da porazdelitev vsote ni normalna.

```
> hist(x+z,main="",prob=T,ylab="Delez")
```



Slika 3: *Porazdelitev vsote $X + Z$.*