

### 1.3 Vsota diskretnih slučajnih spremenljivk

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

- Kako je porazdeljena njuna vsota?  
Označimo  $Z = X + Y$ . Verjetnost, da je  $P(Z = z)$  za nek  $z$  zapišemo kot vsoto verjetnosti vseh kombinacij  $X = x$  in  $Y = y$ , ki dajo vsoto enako  $z$  (lahko seštevamo, saj so dogodki nezdružljivi):

$$\begin{aligned}P(Z = z) &= P(X + Y = z) = \sum_y P(X = z - y, Y = y) \\ &= \sum_y P(X = z - y | Y = y) P(Y = y)\end{aligned}$$

Za neodvisni  $X$  in  $Y$  torej velja

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

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  $n$  neodvisnih enako porazdeljenih (i.i.d.) Bernoullijevih spremenljivk?

Označimo  $Z = \sum_{i=1}^n X_i$ ,  $i = 1, \dots, n$ ,  $X_i \sim B(p)$ . V prejšnji točki smo pokazali, da je vsota dveh Bernoullijevih spremenljivk binomsko porazdeljena s parametroma 2 in  $p$ . S pomočjo matematične indukcije želimo pokazati, da je porazdelitev binomska za vsoto poljubnega

števíla Bernoullíjevih spremenljívk. Pokazali smo za prvi korak, sedaj moramo pokazati še, da če to velja za vsoto  $n$  spremenljívk, velja tudi za vsoto  $n+1$  spremenljívk. Naj bo torej  $U \sim \text{Bin}(n, p)$  in  $X \sim \text{Ber}(p)$ , pokazati moramo, da velja  $U + X \sim \text{Bin}(n + 1, p)$ .

Naj bo  $1 \leq z \leq n$ , velja

$$\begin{aligned}
 P(Z = z) &= \\
 &= \sum_x P(U = z - x)P(X = x) \\
 &= P(U = z)P(X = 0) + P(U = z - 1)P(X = 1) \\
 &= \binom{n}{z} p^z (1 - p)^{n-z} \cdot (1 - p) + \binom{n}{z-1} p^{z-1} (1 - p)^{n-z+1} \cdot p \\
 &= p^z (1 - p)^{n-z+1} \left[ \binom{n}{z} + \binom{n}{z-1} \right] \\
 &= p^z (1 - p)^{n+1-z} \left[ \frac{n!}{z!(n-z)!} + \frac{n!}{(z-1)!(n-z+1)!} \right] \\
 &= p^z (1 - p)^{n+1-z} \left[ \frac{n!(n+1)(n-z+1)}{z!(n-z)!(n-z+1)(n+1)} + \frac{n!(n+1)z}{(z-1)!(n-z+1)!(n+1)z} \right] \\
 &= p^z (1 - p)^{n+1-z} \frac{(n+1)n!}{z(z-1)!(n-z+1)(n-z)!} \left[ \frac{n-z+1}{n+1} + \frac{z}{n+1} \right] \\
 &= p^z (1 - p)^{n+1-z} \binom{n+1}{z} \frac{n-z+1+z}{n+1} \\
 &= \binom{n+1}{z} p^z (1 - p)^{n+1-z}
 \end{aligned}$$

Dokaz za  $z = 0$  in  $z = n + 1$  je prepuščen bralcu.

### Predlogi za vaje v R-u:

- S pomočjo funkcije `sample` generirajte po 100 realizacij dveh Bernoullíjevih spremenljívk in si oglejte porazdelitev njune vsote. Ponovite še za vsote 20 Bernoullíjevih spremenljívk.

## 1.4 Vsota zveznih slučajnih spremenljívk

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  $\sum Z^2$  (gledamo vsoto kvadriranih odmikov, saj so vrednosti lahko negativne ali pozitivne). Pri tem predpostavimo, da so bile meritve narejene v dovolj velikih časovnih presledkih, da so vrednosti med seboj neodvisne.

- Najprej nas zanima, kako je porazdeljena vsota dveh neodvisnih zveznih spremenljivk (izpeljite formulo za dve zvezni spremenljivki, torej  $Z = X + Y$ , primerjajte jo s formulo za diskretne)

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

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

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

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

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

- Kako se porazdeljuje  $S = Z_1^2 + Z_2^2$ , če sta spremenljivki  $Z_1$  in  $Z_2$  porazdeljeni standardno normalno in med seboj neodvisni?

Namig: Uporabite rezultat

$$\int_0^1 \frac{1}{\sqrt{(1-x)x}} dx = \pi$$

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

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

Izračunajmo gostoto vsote  $Z_1^2 + Z_2^2$ . Za dano vrednost  $s$  vemo, da mora biti vrednost  $z_2$  med 0 ( $z_1$  in  $z_2$  ne moreta biti negativni) in  $s$  (ker sta obe pozitivni in je njuna vsota enaka  $s$ ), zato morajo biti meje integracije med 0 in  $s$ .

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

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

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

Gostota gama porazdelitve je

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

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

- Denimo, da so športnikovi standardizirani odmiki (vrednosti  $Z$ ) na petih merjenjih naslednji: 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  $\sum_{i=1}^n X_i \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ .

```
> 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.

### Predlogi za vaje v R-u:

- Generirajte 10 vrednosti iz porazdelitve  $X \sim N(148,85)$ , te vrednosti naj predstavljajo 10 meritev pri enem športniku. Vrednosti standardizirajte (tako da dobite  $N(0,1)$  spremenljivko), kvadrirajte in seštejte. To naj bo vrednost za prvega športnika, na enak način generirajte vrednosti za 1000 športnikov. Narišite histogram vrednosti. S pomočjo funkcije `pgamma` poiščite mejo, ki jo porazdelitev  $\Gamma(\frac{10}{2}, \frac{1}{2})$  preseže z verjetnostjo manj kot 0,01 in izračunajte delež športnikov na vašem vzorcu, ki presežejo to mejo.
- Recimo, da ima dopingiran športnik enako povprečje, a večjo varianco (vrednosti bolj nihajo, saj manipulira s krvjo). Generirajte 1000 športnikov z večjo varianco in si oglejte, kakšen delež bo presegel meje iz prejšnje točke.

## 1.5 Vsota normalnih spremenljivk

Pokazali smo že, da je linearna transformacija normalno porazdeljene spremenljivke zopet normalno porazdeljena. V tej vaji bomo pokazali, da je vsota dveh neodvisnih (standardiziranih) normalnih spremenljivk zopet normalno porazdeljena. Nato bomo pokazali še protiprimer, ko vsota dveh odvisnih normalnih spremenljivk ni več normalna.

- Naj bosta  $X \sim N(0,1)$  in  $Y \sim N(0,1)$  in med seboj neodvisna. Kako je porazdeljena njuna vsota?

Uporabimo formulo za gostoto vsote dveh neodvisnih slučajnih spremenljivk:

$$\begin{aligned}
 f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy \\
 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(z-y)^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} dy \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \exp\{-y^2\} \exp\{zy\} dy
 \end{aligned}$$

Sedaj dele izraza, v katerih nastopa  $y$ , zapišemo kot kvadrat neke vsote:

$$\begin{aligned}
 y^2 - zy &= y^2 - 2y\frac{z}{2} + \left(\frac{z}{2}\right)^2 - \left(\frac{z}{2}\right)^2 \\
 &= \left(y - \frac{z}{2}\right)^2 - \frac{z^2}{4}
 \end{aligned}$$

Gornji integral torej prepišemo v

$$\begin{aligned}
 f_Z(z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} \exp\left\{-\left(y - \frac{z}{2}\right)^2\right\} \exp\left\{\frac{z^2}{4}\right\} dy \\
 &= \frac{1}{2\pi} \exp\left\{-\frac{z^2}{4}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{\left(y - \frac{z}{2}\right)^2}{2 \cdot \frac{1}{2}}\right\} dy \\
 &= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{4}\right\} \frac{1}{\sqrt{2}} \left[ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot \frac{1}{2}}} \exp\left\{-\frac{\left(y - \frac{z}{2}\right)^2}{2 \cdot \frac{1}{2}}\right\} dy \right] \\
 &= \frac{1}{\sqrt{2\pi \cdot 2}} \exp\left\{-\frac{z^2}{2 \cdot 2}\right\}
 \end{aligned}$$

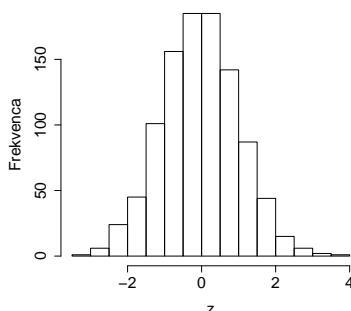
V predzadnji vrstici smo pod integralom dobili ravno gostoto normalne porazdelitve ( $N(\frac{z}{2}, (\sqrt{\frac{1}{2}})^2)$ ), njen integral je zato 1 (ne glede na vrednost  $z$ , ki je znotraj tega integrala konstanta). Spremenljivka  $Z$  je

normalno porazdeljena,  $Z \sim N(0, (\sqrt{2})^2)$ .

- Naj bosta  $X$  in  $Y$  neodvisni standardizirani normalni spremenljivki,  $Z$  pa enaka  $|Y|$ , če je  $X \geq 0$ , in  $Z = -|Y|$  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 spr.,
> y <- rnorm(1000,0,1)      #povprecje=0, sd=1
> z <- abs(y)               #z = |y|
> z[x<0] <- -z[x<0]        #z = -|y|, ce je x<0
> hist(z,main="",ylab="Frekvenca") #histogram z
```



Slika 1: Porazdelitev spremenljivke  $Z$ .

Sedaj še izpeljimo porazdelitveno funkcijo. Naj bo  $z < 0$  (torej  $X$  negativen), uporabimo, da sta spremenljivki  $X$  in  $Y$  neodvisni:

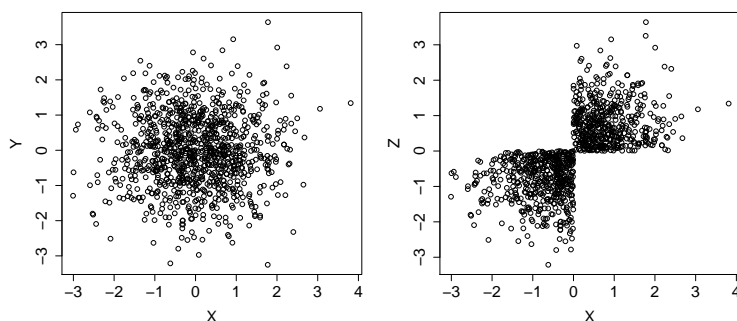
$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(X < 0, -|Y| \leq z) \\ &= P(X < 0)[P(Y \leq z) + P(Y \geq -z)] \\ &= \frac{1}{2}[2 \cdot P(Y \leq z)] \\ &= P(Y \leq z) = F_Y(z) \end{aligned}$$

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.



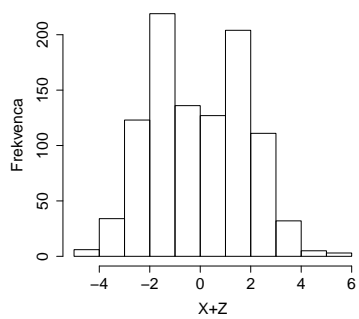
Slika 2: Razsevni diagram realizacij  $X$  in  $Y$  ter  $X$  in  $Z$ .

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

Iz slike je očitno, da porazdelitev vsote ni normalna. Vsota dveh normalnih spremenljivk torej ni nujno normalna, če spremenljivki nista neodvisni (dana naloga je protiprimer).

```
> hist(x+z,main="",ylab="Frekvenca")
```





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