

Poglavje 1

PORAZDELITVE NAKLJUČNIH SPREMENLJIVK

Porazdelitve nam predstavljajo pogostnost posameznih vrednosti. Predstavimo jih lahko s številom posameznih vrednosti (dogodkov) ali z deleži (pogostnostjo). Porazdelitve lahko prikazujemo v preglednicah ali grafih. Pri opisu opazujemo:

1. Število ekstremov: unimodalna (z enim vrhom), bimodalna (z dvema vrhovoma), več modalna...
2. Zalogo vrednosti: najmanjša in največja vrednost, zveznost ali diskretnost
3. Sploščenost: sploščena, koničasta
4. Asimetričnost: simetrična, levo asimetrična, desno asimetrična
5. Oblika: je to ena od znanih porazdelitev?

Porazdelitve z več vrhovi: doba od pripusta do pregonitve, interim obdobje (doba od odstavitve do prvega pripusta pri prašičih), servis perioda (doba od poroda do uspešnega pripusta).

1.1 Opis porazdelitve

a) s sliko

Predstavitev porazdelitve s sliko je zelo nazorna. Tega se pogosto poslužimo, ko predstavljamo rezultate ali ko porazdelitve neke lastnosti ne poznamo. Pri diskretnih spremenljivkah preštejemo dogodke pri posameznih vrednostih. Pri porazdelitvi velikosti gnezda pri prašičih enostavno preštejemo število prasitev z 0, 1, 2, 3, ..., 20 in več živorojenimi pujski v gnezdu in narišemo graf. Pri zveznih spremenljivkah ne moremo ubrati iste poti. Vrednosti, ki jih meritev lahko zasede je neskončno mnogo, natančno ena vrednost se zgodi zelo redko. Kljub temu pa so na določenem intervalu vrednosti bolj pogoste, zgoščene. Grafično zadevo naredimo tako, da spremenljivko razdelimo v "primerne" razrede in potem narišemo.

b) s parametri porazdelitve

Parametre posameznih porazdelitev in njihove značilnosti bomo obravnavali pri posameznih porazdelitvah. Pri navajanju parametrov smemo izbrati le tiste, ki dobro opisujejo porazdelitev.

1.2 Porazdelitev diskretnih naključnih spremenljivk

Diskretna naključna spremenljivka je tista, ki ima končno zalogo vrednosti ali neskončno zalogo vrednosti iz množice celih števil. Primer s končno zalogo vrednosti predstavlja število potomcev pri sesalcih, z neskončno zalogo vrednosti pa število odloženih jajčec pri čebelah.

Dogodek iz zaloge vrednosti ima pripadajočo verjetnost. Vzemimo primer za velikost gnezda pri prašičih (1.2). V gnezdu pri modernih, mesnatih pasmah je lahko od 0 do 22 živorojenih pujskov. Tako je verjetnost, da so v gnezdu samo mrtvorjeni pujski (1.1), enaka 0.0126 ali 1.26 %. Najpogostejša so gnezda z 11 (1.2) in 12 (1.3) živorojenimi pujski. Gnezda z 19 živorojenimi pujski so izredno redka (1.4).

$$P(X = 0) = 0.0126$$

[1.1]

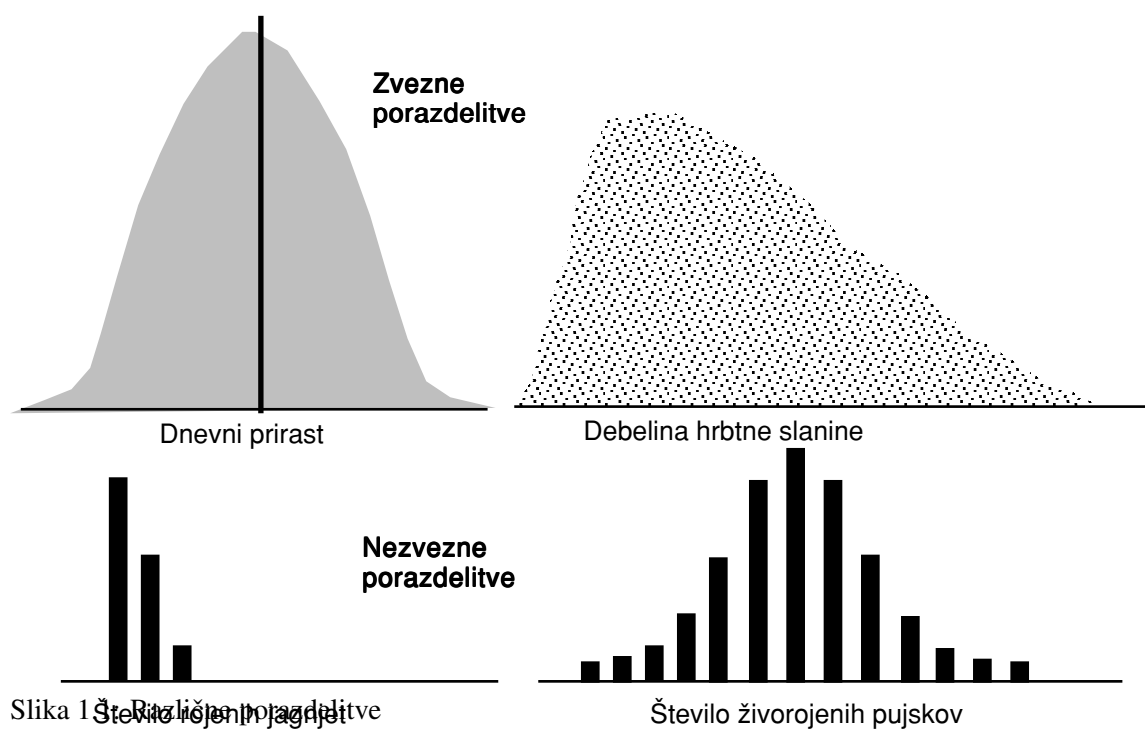
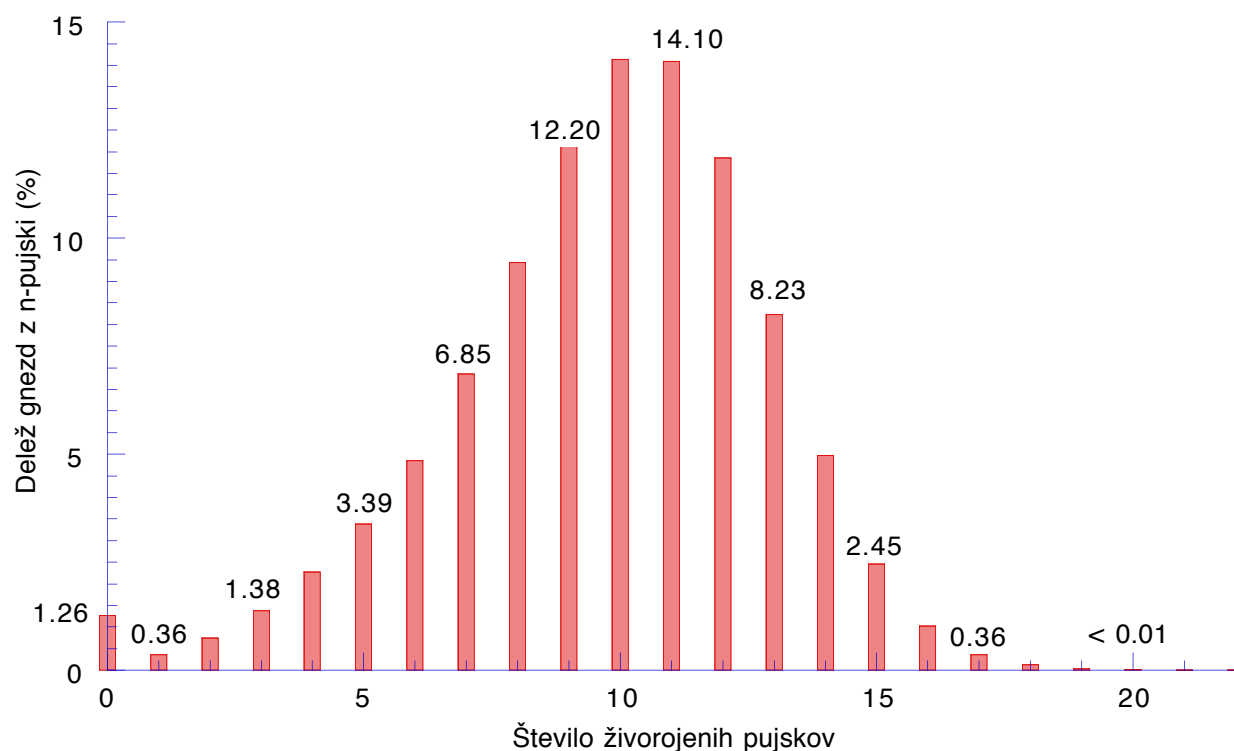


Tabela 1.1: Porazdelitev gnezd pri prašičih z ozirom na število živorojenih pujskov v gnezd

Št. živ. puj./gn.	Delež (%)	Št. živ. puj./gn.	Delež (%)	Št. živ. puj./gn.	Delež (%)
0	1.264	8	9.4310	16	1.0140
1	0.3610	9	12.195	17	0.3570
2	0.7400	10	14.146	18	0.1190
3	1.3790	11	14.098	19	0.0340
4	2.2730	12	11.847	20	0.0110
5	3.3870	13	8.2300	21	0.0010
6	4.8440	14	4.9720	22	0.0010
7	6.8470	15	2.4490	23	0.0000



Slika 1.2: Distribucija gnezd pri prašičih z ozirom na število živorojenih pujskov pri prašičih

$$P(X = 11) = 0.1415 \quad [1.2]$$

$$P(X = 12) = 0.1410 \quad [1.3]$$

$$P(X = 19) = 0.00034 \quad [1.4]$$

Oznaka $P(X = x_i)$ ali samo $P(x_i)$ predstavlja verjetnost, da naključna spremenljivka X zavzame natanko vrednost x_i . Vsekakor pa lahko izraz posplošimo. Tako oznaka $P(X < x_i)$ predstavlja verjetnost, da naključna spremenljivka X zavzame katerokoli vrednost manjšo od x_i . Za ponazoritev ponovno vzemimo primer za velikost gnezda pri prašičih. Izračunajmo verjetnost, da so v gnezdu manj kot trije živorojeni pujski (1.5). Zanima nas torej verjetnost, da je v gnezdu nič, eden ali dva živorojena pujska. Ker se dve možnosti ne moreta zgoditi hkrati, je verjetnost $P(X < 3)$ enaka kar vsoti verjetnosti za tri možne dogodke.

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0.01264 + 0.00361 + 0.00740 = 0.02365 \quad [1.5]$$

Sedaj pa izračunajmo verjetnost še za nekaj primerov. Pri tem imamo več možnosti, prikazali pa bomo le po eno za vsak primer.

$$P(X < 5) = P(X < 3) + P(X = 3) + P(X = 4) = 0.02365 + 0.01379 + 0.02273 + 0.03387 = 0.09404 \quad [1.6]$$

$$P(X \leq 10) = P(X < 5) + P(X = 6) + \dots + P(X = 10) = 0.09404 + 0.04844 + \dots + 0.14146 = 0.56867 \quad [1.7]$$

$$P(X \geq 10) = P(X = 10) + P(X = 11) + \dots = 0.14146 + 0.14098 + \dots + 0.001 = 0.57279 \quad [1.8]$$

V vseh predhodnih primerih se dogodki ne prekrivajo. V gnezdu je lahko samo 0, 1, 2, 3, ... živorojenih pujskov. Nikakor ne more biti v istem gnezdu hkrati npr. 5 ali 8 živorojenih pujskov. Dogodki so neodvisni. Verjetnosti neodvisnih dogodkov lahko kar seštejemo.

$$P(X < 5 \vee X \geq 10) = P(X < 5) + P(X \geq 10) = 0.09404 + 0.57279 = 0.66683 \quad [1.9]$$

Prav tako sta neodvisna dogodka v zadnjem primeru (1.9). Iščemo verjetnost, da je v gnezdju manj kot 5 živorojenih pujskov ali 10 oziroma več živorojenih pujskov. Če jih je manj kot 5, jih ne more biti hkrati (v istem gnezdju) več ali enako 10. Dogodka sta torej tudi neodvisna.

Sedaj pa pogledajmo še naslednji primer (enačba 1.10): poiskati želimo verjetnost, da je v gnezdju manj kot 5 (prvi dogodek) ali 10 oziroma manj pujskov (drugi dogodek). Če je v gnezdju 5, 6, 7, 8, 9 ali 10 pujskov se je gotovo zgodil drugi dogodek. Ko pa imamo v gnezdju 0, 1, 2, 3 ali 4 pujske, pa sta se zgodila tako prvi kot drugi dogodek hkrati. Prvi dogodek je torej podmnožica drugega dogodka, zato je verjetnost, da se zgodita oba dogodka enaka verjetnosti, da se je zgodil drugi dogodek. Prvi in drugi dogodek nista več neodvisna, saj kadarkoli se zgodi prvi dogodek, se zgodi hkrati tudi drugi.

$$P(X < 5 \vee X \leq 10) = P(X \leq 10) = 0.56867 \quad [1.10]$$

V enačbi 1.11 želimo ugotoviti, kolikšna je verjetnost, da je v gnezdju manj kot 5 živorojenih pujskov ali pa je v gnezdju od vključno 3 do vključno 11 pujskov.

$$P(X < 5 \vee (3 \leq X \leq 11)) \quad [1.11]$$

Problem v enačbi 1.11 je, da se pri prvem ($X < 5$) in drugem ($3 \leq X \leq 11$) dogodku prekrivajo možnosti, da bi imela svinja v gnezdju 3 ali 4 pujske. To lahko rešimo vsaj na dva načina. Če bi verjetnosti, da se zgodi prvi ali drugi dogodek samo sešteli, bi bila vsota prevelika. Dogodka, da so v gnezdju 3 ali 4 pujski, bi upoštevali dvakrat. Tako moramo verjetnost, da se ta dva dogodka zgodita, še odšteti. Enačba 1.12 nazorno pokaže, da dogodka iz enačbe 1.11 nista neodvisna.

$$= P(X < 5) + P(3 \leq X \leq 11) - P(3 \leq X < 5) \quad [1.12]$$

V drugem načinu (1.13) smo iz enega dogodka izbrisali tiste dogodke, ki so v prvem že upoštevani. Spodnji primer tudi da pravilen rezultat, je pa nekoliko skrito, da sta dogodka odvisna.

$$= P(X < 5) + P(5 \leq X \leq 11) \quad [1.13]$$

Izvednotimo še enačbo 1.14, ki zahteva, da se zgodita oba dogodka hkrati.

$$P(X < 5 \wedge (3 \leq X \leq 11)) = P(X = 3) + P(X = 4) \quad [1.14]$$

1.3 Porazdelitev zveznih naključnih spremenljivk

1.3.1 Gostota verjetnosti ali verjetnostna funkcija (probability density function)

Porazdelitvena funkcija (distribution function)

1.3.2 Porazdelitvena funkcija za slučajno spremenljivko

$$f(y_i)$$

$$f(y_1, y_2, \dots, y_n) = \prod_{i=1}^n f(y_i)$$

1.4 Bernullijeva porazdelitev

Naključna spremenljivka X ima Bernullijevo porazdelitev, če lahko zavzema vrednosti 0 in 1 z verjetnostjo p in $q = 1 - p$. Vrednosti se izključujeta. Verjetnostna funkcija je prikazana v enačbi (1.15).

$$P(X = x) = \begin{cases} Br(x|p) = p^x (1 - p)^{1-x} & \text{za } x = 1, 2 \\ 0 & \text{za ostale vrednosti } x \end{cases} \quad [1.15]$$

Verjetnost p naj bi bila poznana, v modelih pa je pogosto nepoznana in jo ocenjujemo.

Pričakovana vrednost je enaka verjetnosti za uspešni (p) dogodek.

$$\mu = E(X) = \sum_{x=0}^1 x p^x (1 - p)^{1-x} = (0)(1 - p) + (1)(p) = p \quad [1.16]$$

Varianca (1.17) je produkt verjetnosti za uspešni (p) in neuspešni ($1 - p$) dogodek.

$$\sigma^2 = var(X) = \sum_{x=0}^1 (x - p)^2 p^x (1 - p)^{1-x} = p^2 (1 - p) + (1 - p)^2 p = p(1 - p) \quad [1.17]$$

Z Bernullijevo porazdelitvijo lahko opišemo tiste meritve, ki imajo samo dve vrednosti. Sem štejemo težavnost telitve, če imamo samo dve možnosti. Tudi podatki o oplojenosti jajca, uspešnost oziroma neuspešnost pripusta. Pri tem imamo samo za en dogodek, en poskus. Če je nanizanih več Bernullijevih poskusov, dobimo binomsko porazdelitev.

1.5 Binomska porazdelitev

Binomska porazdelitev ima samo dve vrednosti, ki pa se n -krat ponovi. Opišemo jo z dvema parametroma in sicer **številom poskusov** (n) in **verjetnostjo** (p), da se je zgodil prvi možni dogodek. Vrednost p variira lahko med 0 in 1. Drugi možni dogodek se je zgodil natanko z verjetnostjo $q = 1 - p$. Verjetnostno funkcijo prikazujemo v enačbi (1.18). Funkcijo grafično ponazorimo s histogramom.

$$P(X = x) = \begin{cases} Bi(x|p, n) = \binom{n}{x} p^x (1 - p)^{n-x} & \text{za } x = 1, 2, \dots, n \\ 0 & \text{v ostale vrednosti } x \end{cases} \quad [1.18]$$

PRIMER: Poiščite verjetnostno funkcijo za naključno spremenljivko X , ki predstavlja število ženskih živali pri treh zaporednih rojstvih pri govedu. Predpostavimo, da je p znan in ima vrednost 0.49.

$$P(X = 0) = Bi(0|0.49, 3) = 0.1327 \quad [1.19]$$

$$P(X = 1) = Bi(1|0.49, 3) = 0.3823 \quad [1.20]$$

$$P(X = 2) = Bi(2|0.49, 3) = 0.3674 \quad [1.21]$$

$$P(X = 3) = Bi(3|0.49, 3) = 0.1176 \quad [1.22]$$

Binomsko porazdelitev imajo torej vse naključne spremenljivke, ki opisujejo dogodke s samo dvema možnostima, ki se ponovijo večkrat. Torej: posamezni dogodek ima Bernullijevo porazdelitev, opazujemo pa več dogodkov hkrati.

PRIMERI:

1. Poiščite verjetnostno funkcijo za naključno spremenljivko X , ki predstavlja število uspešnih osemenitev pri 500 pripustih. Predpostavimo, da je p znan in ima vrednost 0.80.

$$P(X = y) = Bi(y|0.80, 500) = \binom{500}{y} 0.80^y (1 - 0.80)^{1-y} = g(y) \quad [1.23]$$

2. Prelux-G piščance vzreajo kot nesnice za kmečko rejo. V valilnici vložijo po 1000 jajc naenkrat. Verjetnost, da se izvali piščanec, je 0.70. Kolikšna je verjetnost, da se izvali natanko 450 piščancev?
3. Med izvaljenimi piščanci je 0.50 ženskih živali. Kolikšna je verjetnost, da se bo iz 1000 vloženih jajc izvalilo 500 ženskih živali? Izračunajte srednjo vrednost, varianco in standardni odklon!
4. Narišite gostoto verjetnosti za število preživelih (odstavljenih) pujskov! Verjetnost, da pujsek do odstavitve pogine je 0.12. Za izhodišče vzemite tri možne velikosti gnezda ob pravitvi: 5, 10 in 15.
5. Narišite $Bi(16, 0.2)$, $Bi(16, 0.5)$, $Bi(16, 0.7)$ in $Bi(16, 0.90)$!
6. V jati je patentno okuženih 15 % rac. Privzemimo, da so dogodki neodvisni. Naključna spremenljivka X predstavlja število okuženih rac med 20 naključno izbranimi racami. Napišite gostoto verjetnosti in jo upodobite na grafu!

1.6 Poissonova porazdelitev

Pri nekaterih poskusih štejemo frekvenco posameznih dogodkov v danem časovnem intervalu ali na danem fizičnem objektu. Lahko štejemo število telefonskih klicev med 7. in 11. uro, število kupcev kart na blagajni zadnjo uro pred predstavo. Lahko preštujemo število obolelih živali v čredi, število živorojenih pujskov v gnezdu, število izmerjenih klavnih trupov na liniji klanja v eni uri, število žretij med 7. in 9. uro zjutraj, število agresij med 7. uro zjutraj in zvečer.

Vzemimo, da štejemo dogodke (spremembe), ki se zgodijo na danem zveznem intervalu. Pri tem dobimo približno Poissonov proces s parametrom $\lambda > 0$, če je izpolnjeno naslednje:

- (i) Število sprememb, ki se zgodijo na neprekrivajočih se intervalih, je neodvisno.
- (ii) Verjetnost natanko ene spremembe na zadosti kratkem intervalu dolžine h je približno λh .
- (iii) Verjetnost dveh ali več sprememb na zadosti kratkem intervalu je praktično nič.

Da bi našli zadostno majhen interval, celotni interval razdelimo na n podintervalov. Število podintervalov naj bo zadostno večji od števila prešteti sprememb oziroma dogodkov. Verjetnost enega dogodka na tem intervalu je približno λ/n . Verjetnost, da se je dogodek zgodil ali pa ne, je pravzaprav Bernullijev proces. Z binomsko porazdelitvijo lahko predstavimo približek za

$$P(X = x) = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Če s približkom nismo zadovoljni, povečajmo n in poiščemo limito naslednjega izraza.

$$\lim_{n \rightarrow \infty} \left(\frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \right), \quad x = 0, 1, 2, \dots$$

Brez dokaza bomo navedli gostoto verjetnosti za Poissonovo porazdelitev, pričakovano vrednost in varianco.

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$\mu = E(X) = \lambda$$

$$\sigma^2 = \text{var}(X) = \lambda$$

Tako pri opazovanju živali naredimo raster, ki je dovolj kratek, da se ne moreta zgoditi dva dogodka hkrati. Potem preštevamo število posameznih oblik obnašanja.

1. Naključna spremenljivka X ima Poissonovo porazdelitev z $\lambda = 5$. Navedite pričakovano vrednost in varianco. Narišite histogram gostote verjetnosti.
2. Narišite histograme za porazdelitveno funkcijo za Poissonove naslednje porazdelitve $\lambda = 0.7$, $\lambda = 1.3$, $\lambda = 6.5$ in $\lambda = 10.5$.
3. Naključna spremenljivka X predstavlja frekvenco pitja. Opazovanja smo opravili na 100 živalih. Preverimo, ali lahko privzamemo Poissonovo porazdelitev!

Izid (x)	Pogostnost (f)	fx	fxx
1	1	1	1
2	4	8	16
3	13	39	117
4	19	76	304
5	16	80	400
6	15	90	540
7	9	63	441
8	12	96	768
9	7	63	567
10	2	20	200
11	1	11	121
12	1	12	144
Skupaj	100	559	3619

Povprečje v tem primeru je 5.59, varianca pa znaša 4.942. Vrednosti so dovolj blizu, zato bi smeli distribucijo opisati s Poissonovo porazdelitvijo.

Poissonova porazdelitev ni pomembna samo zaradi nje same. Zelo uporabna je tudi kot približek za binomsko porazdelitev, ko je n zadosti velik in p sorazmerno majhen.

$$\frac{\lambda^x e^{-\lambda}}{x!} \approx \binom{n}{x} p^x (1-p)^{1-x}$$

$$\lambda = np$$

Približek je zadostno natančen, če je $n \geq 100$ in $p \leq 0.05$.

1.7 Enakomerna (uniformna) porazdelitev

Gostota verjetnosti

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b \quad [1.24]$$

Porazdelitvena funkcija

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \end{cases} \quad [1.25]$$

Srednja vrednost

$$\mu = \frac{a+b}{2} \quad [1.26]$$

Varianca

$$\sigma^2 = \frac{(b-a)^2}{12} \quad [1.27]$$

1.8 Normalna porazdelitev

Normalna porazdelitev je morda najpomembnejša porazdelitev v statistiki. Tudi v živinoreji je veliko lastnosti, ki jih proučujemo, vsaj približno normalno porazdeljenih. Pomembna pa je tudi vloga normalne porazdelitve v izreku centralne limite (ang. Central Limit Theorem).

Naključna (slučajna) spremenljivka y ima normalno porazdelitev, če je **gostota verjetnosti** (p. d. f) definirana z enačbo 1.28.

$$f(y) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(y-\mu)^2}{2\sigma^2} \right], \quad -\infty < y < \infty \quad [1.28]$$

Parametra μ in σ predstavljata srednjo vrednost in standardni odklon. Srednja vrednost se nahaja na intervalu od negativne neskončnosti do pozitivne neskončnosti 1.29, standardni odklon pa ima pozitivno vrednost (1.30).

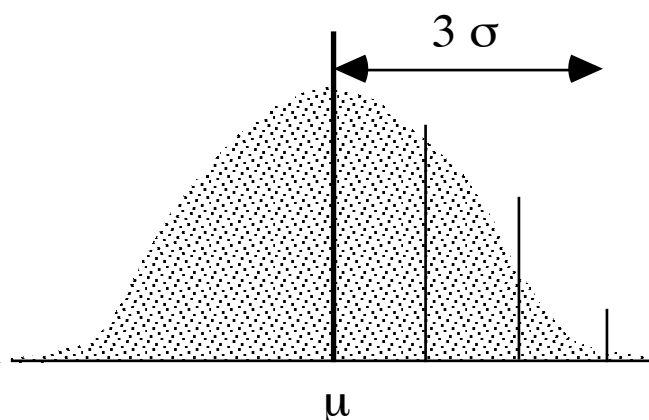
$$-\infty < \mu < \infty \quad [1.29]$$

$$0 < \sigma < \infty \quad [1.30]$$

Na kratko predstavimo normalno porazdelitev naključne spremenljivke tudi v naslednji obliki (1.31).

$$y_i \sim N(\mu, \sigma^2) \quad [1.31]$$

Normalno porazdelitev prepoznamo po zvončasti obliki Gaussove porazdelitvene funkcije. Povsem zadostno jo opišemo z dvema parametroma: s **srednjo vrednostjo** μ , tudi pričakovana vrednost) in z **varianco** (σ^2) oz. s **standardnim odklonom** (σ , standardna deviacija). Povprečje predstavlja lokacijski parameter, varianca pa opiše razpršenost. Vemo namreč, da bomo pri normalno porazdeljeni spremenljivki na intervalu med $\mu - \sigma$ in $\mu + \sigma$ našli kar 68 % meritev, na intervalu med $\mu - 2\sigma$ in $\mu + 2\sigma$ kar



Slika 1.3: Normalna porazdelitev

95 %, na intervalu med $\mu - 2.56\sigma$ in $\mu + 2.56\sigma$ pa kar 99 %. Izven intervala $\mu - 3\sigma$ in $\mu + 3\sigma$ pa skoraj ni opazovanj, čeprav je Gaussova porazdelitvena funkcija navzdol in navzgor neomejena.

Posebno vlogo ima **standardna normalna porazdelitev**. Za naključno spremenljivko z rečemo, da ima standardno normalno porazdelitev, če je srednja vrednost enaka 0 in standardni odklon 1 (1.32).

$$z \sim N(0, 1) \quad [1.32]$$

Verjetnostna funkcija

PRIMER: Napišite gostoto verjetnosti za naključno spremenljivko y , ki ima porazdelitev prikazano v enačbi 1.33.

$$y \sim N(-7, 16) \quad [1.33]$$

Gostota verjetnosti je prikazana v enačbi 1.34. Srednja vrednost porazdelitve je -7 , standardni odklon pa znaša 4.

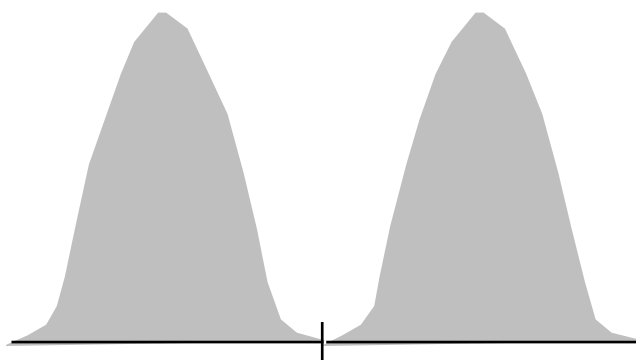
$$f(y) = \frac{1}{\sqrt{32\pi}} \exp \left[-\frac{(y+7)^2}{32} \right], \quad -\infty < y < \infty \quad [1.34]$$

Veliko gospodarsko pomembnih lastnosti je porazdeljeno normalno. To velja za maso pri določeni starosti, starost pri določeni masi, dnevni prirasti in konverzija krme na izbranem intervalu rasti, debelina hrbtna slanina, telesne mere, mase klavnih kosov, odstotek mesnatosti, klavnost, površina hrbtna mišice in pripadajoče slanina, razmerje meso:slanina, itd...

V izjemnih primerih te lastnosti niso več normalno porazdeljene. Tako npr. po uspešni selekciji na mesnatost lahko postane slanina nagnjena s strmino proti 0 mm in "tolstim repom" proti višjim vrednostim. Srednja vrednost in standardni odklon ne opisujeta porazdelitve več zadovoljivo. Primerneje je poiskati vrh oziroma **modus** porazdelitve, za razpršenost pa nimamo primernejšega parametra kot je standardni odklon. Če želimo biti natančni, uporabimo pri teh porazdelitvah merilo za **nagnjenost** - skewness in **sploščenost** - kurtosis. Da pa bi bili razumljivi, porazdelitev raje narišemo.

1.9 Ostale porazdelitve

Nekatere lastnosti pa ne moremo uvrstiti v nobeno od omenjenih porazdelitev. Lahko so zelo nenavadnih oblik: nikakor pa jih ne moremo predstaviti z nekaj parametri, kot smo to storili v prej omenjenih porazdelitvah. So pa v živinoreji vredne, da jih podrobneje spoznamo. Redkeje te lastnosti opisujejo proizvodne lastnosti, večkrat pa so v povezavi z lastnostmi obnašanja, počutja, senzoričnimi lastnostmi



Slika 1.4: Bimodalna porazdelitev

Tabela 1.2: Podatki o testiranju mladic na rast in zamaščenost

Žival	Pasma	Mesec	Masa (kg)	Debelina slanine (mm)	Dnevni prirast (g/dan)
1	SL	JAN	102	13	540
2	SL	JAN	98	16	550
3	SL	FEB	105	16	550
4	SL	FEB	102	15	580
5	LW	JAN	95	20	520
6	LW	FEB	101	24	500
7	LW	FEB	101	27	490
8	NL	JAN	97	26	560
9	NL	JAN	100	22	550
10	NL	FEB	97	23	600
11	NL	FEB	102	24	610

produktov. Tudi subjektivne ocene, kjer niso dobro postavljene skale, se lahko sprevržejo v "neurejene" porazdelitve. Take porazdelitve nikakor ne opisujemo s parametri, omenjenimi pri porazdelitvah znanih oblik. Pri naslednji porazdelitvi se srednja vrednost nahaja na mestu, kjer meritev najmanj pričakujemo, med obema vrhovoma.

1.10 Enorazsežne in večrazsežna porazdelitev

Vzemimo, da imamo dve spremenljivki $\mathbf{y}_1$ in $\mathbf{y}_2$, ki so porazdeljene $N(\mathbf{X}\boldsymbol{\beta}, \mathbf{R})$

$$p(\mathbf{y}_1, \mathbf{y}_2 | \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{R}) = \underbrace{\left(\sqrt{(2\pi)^{n_1+n_2} |\mathbf{R}|} \right)^{-1}}_{\text{integracijska konstanta}} \exp \underbrace{\left\{ -\frac{1}{2} \left[\mathbf{y}'_1 - \boldsymbol{\beta}'_1 \mathbf{X}'_1 \quad \mathbf{y}'_2 - \boldsymbol{\beta}'_2 \mathbf{X}'_2 \right] \mathbf{R}^{-1} \begin{bmatrix} \mathbf{y}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1 \\ \mathbf{y}_2 - \mathbf{X}_2 \boldsymbol{\beta}_2 \end{bmatrix} \right\}}_{\text{jedro}}$$

1.11 Vaje

Datoteka s podatki o testiranju mladic na rast in mesnatost vsebuje 11 zapisov (1.2). Izmerjenih je bilo 11 mladic (živali), treh pasem v mesecih januar in februar. Mase pri merjenju so bile med 96 in 105 kg. Slanino so merili z dvema ponovitvama, dnevni prirast pa je izračunan iz podatkov o starosti in masi pri merjenju.

Pri obdelavah bomo uporabljali tri različna porekla (1.3). V prvem poreklu (**poreklo 0**) bomo imeli primer, ko poreklo pri živalih ni znano. Običajno v takih primerih predpostavimo, da so predniki nesorodni

Tabela 1.3: Poreklo za mladice

Poreklo 0			Poreklo A			Poreklo B		
Žival	Mati	Oče	Žival	Mati	Oče	Žival	Mati	Oče
1	-	-	1	12	-	1	12	-
2	-	-	2	12	-	2	12	-
3	-	-	3	-	-	3	-	-
4	-	-	4	-	-	4	-	-
5	-	-	5	13	14	5	13	14
6	-	-	6	13	14	6	13	14
7	-	-	7	-	14	7	-	14
8	-	-	8	-	-	8	-	-
9	-	-	9	-	-	9	-	-
10	-	-	10	-	-	10	-	-
11	-	-	11	-	-	11	-	-
			12	-	-	12	-	-
			13	-	-	13	15	-
			14	-	-	14	15	-
						15	-	-

Tabela 1.4: Podatki o preizkusu mladice na rast in zamaščenost z manjkajočimi podatki

Žival	Rejec	Mesec	Masa (kg)	Debelina slanine (mm)	Dnevni prirast (g/dan)
1	1	JAN	102	13	13
2	1	JAN	98	16	
3	1	FEB	105	16	
4	1	FEB	102	15	580
5	2	JAN	95	20	520
6	2	FEB	101	24	500
7	2	FEB	101	27	490
8	3	JAN	97	26	560
9	3	JAN	100	22	550
10	3	FEB	97	23	600
11	3	FEB	102	24	610

in različni. Lahko bi imeli tudi drugačne primere. Morda vemo, da se v enem letu ali paritveni sezoni uporablja na kmetiji samo en samec. Čeprav številke nimamo, se pač za to žival izmislimo novo oznako in jo uporabljamo. V takem primeru bomo potem lahko določili, da imajo vsi mladiči rojeni po teh pripustih istega očeta, porekla očeta pa ne bomo več vedeli.

V drugem primeru (**poreklo A**) bomo imeli nekaj prednikov znanih, nekaj neznanih. Znane prednike moramo dopisati na koncu seznama in jim poiščemo prednike. Postopke ponavljamo, dokler nimamo več novih, dodanih staršev. V našem primeru smo dodali samo tri starše: dve materi in enega očeta. Vsaka žival ima samo eno vrstico. V drugem poreklu so bila izvedena vsa parjenja tako, da sorodnih živali nismo parili.

V tretje poreklo (**poreklo B**) smo dodali še eno žival, ki je skupen prednik (mati) svinji 13 in merjascu 14. Parjenje med svinjo 13 in merjascem 14, ki sta sestra in brat, je parjenje sorodnikov. Potomci so inbridirani. Takih parjenj v živinoreji praviloma ne izvajamo in so celo zakonsko prepovedana. Tule pa smo jih uporabili samo zato, da bomo lahko prikazali vpliv pri obdelavi podatkov. Če bi imeli bolj oddaljene skupne prednike, bi morali dodati preveč živali iz porekla. Mi pa želimo, da so primeri obvladljivi.

Izračunajte:

	Masa	Dnevni prirast	Debelina slanine
Število meritev			
Povprečje			
Varianca			
Standardni odklon			
Modus			
Mediana			
Minimum			
Maksimum			

Ali lahko opišete porazdelitve za naslednje lastnosti:

- "učni uspeh pri etologiji (skala 1 do 10)" samo za opravljene izpite,
- "učni uspeh pri etologiji (skala 1 do 10)" za vse poskuse,
- "telesna višina (cm)",
- "prisotnost na vajah (%)",
- "izostanek na vajah (%)",
- "prisotnost na predavanjih (%)" pri študentih drugega letnika univerzitetnega študija zootehnike?

Ali lahko opišete porazdelitve za proizvodne lastnosti domačih živali? Kjer lahko, pripišite srednjo vrednost, standardni odklon, minimalno in maksimalno vrednost? Pri tem si pomagajte z viri informacij, ki ste jih spoznali pri predmetu Informatika!

- dnevna količina mleka (kg),
- količina mleka v standardni laktaciji (kg),
- dnevni prirast pri rastočih živalih (g/dan)
- doba od pripusta do pregonitve (dni)
- velikost gnezda
- dolžina brejosti (dni)
- uspešnost pripustov (uspešni, neuspešni)
- delež uspešnih pripustov (%)
- valilnost jajc (%)
- dnevna poraba krme (kg/dan)
- konverzija krme
- količina proizvedenega medu po panju (kg)

Vajo ponovite pri različnih vrstah domačih živalih, kjer se to da! Dopolnjujte seznam lastnosti! Preverite definicije lastnosti! Katera porazdelitvena funkcija je primerna? Katere parametre je najbolje navesti pri opisovanju te lastnosti? Poiščite ocene parametrov, ki veljajo za slovenske populacije!