# Verjetnostni račun in statistika

Goran Turk

Ljubljana, 16. januar 2011

TO JE DELOVNA RAZLIČICA UČBENIKA.

VSEM, KI ME BODO OPOZORILI NA TIPKARSKE, RAČUNSKE IN DRUGE NAPAKE, SE LEPO ZAHVALJUJEM. HVALEŽEN BOM TUDI VSEM, KI MI BODO POSLALI PRIPOMBE IN KOMENTARJE.

NOVEJŠE VERZIJE LAHKO DOBITE NA SPLETNI STRANI:
`http://www.km.fgg.uni-lj.si/predmeti/sei/vrs1.pdf`

VERZIJA NA SPLETNI STRANI JE ZAPISANA NA VEČ DATOTEKAH. ZA PREHOD MED DATOTEKAMI LAHKO UPORABITE ,,TIPKE" NA SPODNJEM ROBU STRANI, S KATERIMI SE LAHKO PREMIKATE ZA ENO STRAN (ENOJNA PUŠČICA) ALI PO POGLAVJIH (DVOJNA PUŠČICA).

LEP POZDRAV, GORAN TURK

# Vsebina

# 1 Statistika

*Če bi mi dejali, da bom živel le še en dan, bi ga želel preživeti pri pouku statistike – ta dan bi se mi zdel toliko daljši.*

*Zapis iz študentskega koledarja*

*Statistiko uporabljamo tako, kot pijanec uporablja obcestno svetilko: da bi se podprli, ne pa, da bi se razsvetlili.*

*Andrew Lang*

*Najprej zberi podatke, potem pa jih spreminjaj po mili volji!*

*Mark Twain*

**Statistika** je veda o obravnavanju množice podatkov. S statistiko podatke zberemo, razvrstimo, uredimo in na osnovi določenih pravil sprejemamo zaključke. Matematična osnova statistike je verjetnostni račun.

Inženir se pri svojem delu nenehno srečuje z veliko množico podatkov, ki so bolj ali manj urejeni. Vsak mora biti sposoben te podatke urediti in se na osnovi izsledkov odločiti za pravilen ukrep.

Drugi pomemben dejavnik v inženirstvu je dejstvo, da je večina količin, ki jih obravnava, naključne narave. Inženir se mora včasih odločiti na osnovi zelo omejenih podatkov. Oglejmo si zanimiv problem.

**Primer 1.1:** *Avtobusno podjetje je lastnik $N$ avtobusov, ki so vsi označeni z zaporednimi številkami od 1 do $N$. Na cesti naletimo na avtobus s številko 22. Koliko avtobusov ima to podjetje?*

**Rešitev:** Pravilnega oziroma zanesljivega odgovora seveda ni, razen v izjavi, da ima avtobusno podjetje vsaj 22 avtobusov. Ali smo lahko bolj določni? Ker je verjetnost za srečanje s katerimkoli avtobusom enaka, bi se pri mnogoterih srečanjih v povprečju srečali z avtobusom, označenim s številko na sredini zaporedja števil. Tako lahko izračunamo, da je število $N$, ki ga lahko v povprečju pričakujemo, enako

$$N = 21 + 1 + 21 = 43. \tag{1.1}$$

Če nas ne zanima, kateri odgovor bi bil v povprečju najbolj pravilen, temveč bi radi s čim večjo verjetnostjo določili točen odgovor, bi bil pravi odgovor $N = 22$.

Ta primer morda zgleda nenavadno, vendar se s podobnimi problemi inženir zelo pogosto srečuje, ko mora na osnovi pomanjkljivih podatkov sprejeti odločitve.

Podrobneje se bomo s tem primerom spoznali v poglavju 9 oziroma natančneje v primeru 9.6.

# 2 Opisna statistika

Če obravnavamo večje število podatkov, jih posamično ne moremo opazovati, saj nam posamezni podatek premalo pove. Te podatke lahko bolje predstavimo grafično ali pa številčno z določenimi značilnimi vrednostmi vzorca podatkov.

Skupino podatkov $x_i, i = 1, \ldots, n$ imenujemo **vzorec**, posamezni podatek $x_i$ je **element vzorca**, število $n$ pa je število elementov v vzorcu oziroma **velikost vzorca**.

## 2.1 Grafične predstavitve vzorcev

Obravnavajmo najprej grafične predstavitve. Vzemimo, da imamo en stolpec s številčnimi podatki. Prikažemo jih lahko s črtnim diagramom, s katerim grafično prikažemo vse posamezne vrednosti. Pri velikem številu podatkov nam tak diagram ne pojasni značilnosti vzorcev. Bolje je, da podatke združimo oziroma uredimo v razrede in za vsak razred izračunamo relativno frekvenco, ki je kvocient med številom podatkov v določenem razredu ter vsemi podatki. Vsota relativnih frekvenc vseh razredov je seveda enaka ena. Te rezultate običajno grafično prikažemo s **histogramom** oziroma **stolpčnim diagramom**.

Drugi način je, da podatke razvrstimo po velikosti in po vrsti narišemo njihove vrednosti. Pri tem podatkov ni potrebno razporejati v razrede. Tako dobimo **diagram kumulativnih frekvenc** oziroma **graf empirične porazdelitvene funkcije** .

Histogram relativnih frekvenc in diagram kumulativnih frekvenc sta dobra pokazatelja lastnosti vzorca podatkov. Slabost histograma je, da moramo izbrati število oziroma velikost razredov in pri razporeditvi elementov v razrede potrebujemo več podatkov.

Včasih nas zanima, kolikšna je odvisnost ene količine od druge. Grafično to najlažje prikažemo s točkovnim diagramom, kjer vsak par vrednosti dveh količin prikažemo kot točko v koordinatnem sistemu, pri čemer na absciso nanašamo eno količino, na ordinato pa drugo.

**Primer 2.1:** *Obdelajmo mesečne podatke o padavinah, odtoku vode, temperature reke in izhlapevanju za obdobje med letoma 1953 in 1972 v kraju Clayton na reki Chattooga v zvezni državi Giorgia (ZDA).*[1] *Ti podatki so zajeti v preglednici (2.1) s štirimi stolpci podatkov. V vsakem stolpcu je 216 števil.*

Preglednica 2.1: Podatki o padavinah, odtoku, temperaturi in izhlapevanju

| Leto | Mesec | Padavine [cm] | Odtok [cm] | Temp. [°C] | Izhlap. [cm] |
|---|---|---|---|---|---|
| 1953 | 10 | 8.9 | 2.5 | 15.4 | 10.0 |
| | 11 | 12.0 | 3.2 | 9.3 | 5.3 |
| | 12 | 26.2 | 10.2 | 4.8 | 7.1 |
| 1954 | 1 | 30.3 | 16.2 | 5.9 | 5.4 |
| | 2 | 12.3 | 9.2 | 7.4 | 7.9 |
| | 3 | 18.8 | 10.9 | 9.1 | 8.5 |
| | 4 | 11.6 | 11.0 | 16.2 | 12.5 |
| | 5 | 6.7 | 7.9 | 15.6 | 13.2 |
| | 6 | 10.8 | 5.8 | 21.9 | 19.0 |
| | 7 | 9.3 | 3.6 | 25.0 | 20.9 |
| | 8 | 10.7 | 2.9 | 24.4 | 16.2 |
| | 9 | 1.1 | 1.7 | 22.2 | 15.6 |
| | 10 | 0.8 | 1.4 | 16.3 | 11.6 |
| | 11 | 9.0 | 2.1 | 8.1 | 4.8 |
| | 12 | 21.4 | 5.2 | 4.6 | 3.3 |
| 1955 | 1 | 6.8 | 4.3 | 5.4 | 2.9 |
| | 2 | 22.8 | 9.0 | 6.3 | 6.0 |
| | 3 | 11.8 | 7.8 | 10.7 | 8.5 |
| | 4 | 17.0 | 11.4 | 16.2 | 15.1 |
| | 5 | 21.6 | 12.9 | 19.4 | 17.0 |
| | 6 | 13.4 | 7.3 | 19.7 | 17.7 |
| | 7 | 24.2 | 7.5 | 24.0 | 16.2 |
| | 8 | 8.6 | 6.0 | 24.3 | 15.8 |
| | 9 | 7.4 | 3.2 | 21.3 | 14.7 |
| | 10 | 3.7 | 3.0 | 14.5 | 10.0 |
| | 11 | 7.1 | 2.6 | 9.5 | 5.9 |
| | 12 | 5.0 | 2.6 | 4.4 | 3.6 |
| 1956 | 1 | 5.7 | 2.2 | 4.4 | 4.1 |
| | 2 | 32.4 | 10.7 | 8.0 | 6.5 |
| | 3 | 13.5 | 8.8 | 9.7 | 10.9 |
| | 4 | 24.6 | 12.0 | 13.3 | 14.7 |
| | 5 | 7.4 | 7.1 | 19.2 | 16.1 |
| | 6 | 6.9 | 4.4 | 21.3 | 19.1 |
| | 7 | 20.6 | 5.0 | 22.9 | 17.4 |
| | 8 | 6.9 | 2.7 | 23.3 | 16.5 |
| | 9 | 15.2 | 2.4 | 19.1 | 13.7 |
| | 10 | 7.7 | 2.9 | 15.7 | 8.3 |
| | 11 | 6.5 | 3.2 | 8.8 | 4.7 |
| | 12 | 18.3 | 7.1 | 10.4 | 3.8 |
| 1957 | 1 | 26.9 | 7.9 | 6.2 | 4.3 |
| | 2 | 19.2 | 13.7 | 9.4 | 4.6 |
| | 3 | 8.8 | 11.2 | 9.1 | 8.1 |
| | 4 | 21.2 | 16.2 | 15.8 | 13.5 |
| | 5 | 14.5 | 7.8 | 18.7 | 16.1 |
| | 6 | 18.6 | 9.3 | 22.2 | 15.4 |
| | 7 | 4.9 | 6.0 | 23.3 | 19.2 |
| | 8 | 5.5 | 3.6 | 22.4 | 18.0 |
| | 9 | 21.1 | 4.5 | 20.6 | 10.6 |
| | 10 | 16.7 | 7.5 | 12.6 | 7.0 |
| | 11 | 30.0 | 12.7 | 10.1 | 4.7 |
| | 12 | 17.2 | 13.2 | 6.3 | 3.4 |
| 1958 | 1 | 12.5 | 11.5 | 2.2 | 3.7 |
| | 2 | 16.1 | 10.7 | 1.4 | 6.1 |
| | 3 | 15.7 | 11.6 | 7.4 | 7.1 |
| | 4 | 23.1 | 14.9 | 14.1 | 11.2 |
| | 5 | 12.8 | 15.4 | 18.7 | 13.8 |
| | 6 | 6.1 | 7.3 | 21.9 | 18.4 |
| | 7 | 32.4 | 11.8 | 23.2 | 10.7 |
| | 8 | 14.5 | 6.6 | 23.1 | 15.0 |
| | 9 | 3.1 | 3.9 | 19.7 | 13.7 |
| | 10 | 1.0 | 3.5 | 15.1 | 8.4 |
| | 11 | 8.5 | 2.9 | 12.5 | 5.5 |
| | 12 | 7.8 | 3.5 | 4.4 | 2.7 |
| 1959 | 1 | 15.1 | 6.7 | 4.5 | 3.3 |
| | 2 | 15.0 | 6.4 | 7.1 | 4.0 |
| | 3 | 18.5 | 7.8 | 8.6 | 9.1 |
| | 4 | 18.5 | 10.9 | 12.7 | 11.3 |
| | 5 | 30.2 | 13.3 | 19.2 | 15.1 |
| | 6 | 3.0 | 13.0 | 21.6 | 14.7 |
| | 7 | 24.1 | 8.2 | 22.8 | 18.5 |
| | 8 | 8.2 | 5.1 | 23.9 | 15.0 |
| | 9 | 20.8 | 7.6 | 19.7 | 10.0 |
| | 10 | 35.4 | 14.8 | 16.0 | 8.1 |
| | 11 | 7.4 | 8.8 | 8.3 | 4.3 |
| | 12 | 14.2 | 9.7 | 5.6 | 4.1 |

se nadaljuje...

[1] McCuen, R. H., Snyder, W. M., Hydrologic Modeling: Statistical Methods and Applications, Prentice-Hall, 1986.

Preglednica 2.1: Podatki o padavinah, odtoku, temperaturi in izhlapevanju (nadaljevanje)

| Leto | Mesec | Padavine [cm] | Odtok [cm] | Temp. [°C] | Izhlap. [cm] |
|---|---|---|---|---|---|
| 1960 | 1 | 19.7 | 11.4 | 4.7 | 3.3 |
| | 2 | 20.7 | 16.7 | 4.3 | 5.9 |
| | 3 | 21.4 | 14.6 | 2.7 | 10.0 |
| | 4 | 8.3 | 15.4 | 14.3 | 14.1 |
| | 5 | 7.3 | 8.8 | 14.8 | 15.9 |
| | 6 | 11.1 | 6.8 | 20.8 | 16.8 |
| | 7 | 12.6 | 6.1 | 22.5 | 19.9 |
| | 8 | 21.3 | 7.8 | 23.5 | 14.0 |
| | 9 | 22.5 | 6.1 | 20.5 | 11.3 |
| | 10 | 15.0 | 8.6 | 15.9 | 6.8 |
| | 11 | 3.6 | 5.5 | 9.3 | 4.8 |
| | 12 | 8.2 | 5.0 | 2.7 | 6.9 |
| 1961 | 1 | 9.0 | 5.9 | 2.5 | 5.1 |
| | 2 | 27.9 | 12.5 | 7.6 | 4.1 |
| | 3 | 16.3 | 13.8 | 10.4 | 7.9 |
| | 4 | 14.2 | 13.1 | 11.2 | 12.4 |
| | 5 | 8.2 | 8.6 | 16.3 | 12.2 |
| | 6 | 24.0 | 9.0 | 20.6 | 14.6 |
| | 7 | 16.1 | 7.5 | 22.2 | 16.1 |
| | 8 | 22.8 | 8.6 | 22.5 | 13.2 |
| | 9 | 6.7 | 6.5 | 21.0 | 12.7 |
| | 10 | 4.0 | 4.3 | 13.8 | 10.4 |
| | 11 | 14.8 | 5.7 | 11.7 | 6.9 |
| | 12 | 37.7 | 19.2 | 5.8 | 5.8 |
| 1962 | 1 | 20.0 | 14.8 | 4.7 | 5.0 |
| | 2 | 17.5 | 12.3 | 9.3 | 6.7 |
| | 3 | 20.0 | 15.0 | 8.1 | 8.8 |
| | 4 | 15.4 | 18.2 | 11.8 | 10.6 |
| | 5 | 5.1 | 8.9 | 21.1 | 17.8 |
| | 6 | 17.2 | 8.9 | 21.9 | 14.7 |
| | 7 | 11.6 | 4.8 | 23.6 | 16.5 |
| | 8 | 11.5 | 3.8 | 22.9 | 16.9 |
| | 9 | 14.1 | 3.3 | 19.7 | 11.8 |
| | 10 | 11.2 | 5.2 | 15.2 | 8.8 |
| | 11 | 12.2 | 4.5 | 8.8 | 5.2 |
| | 12 | 10.5 | 4.7 | 3.6 | 5.3 |
| 1963 | 1 | 13.0 | 6.7 | 2.7 | 3.9 |
| | 2 | 7.9 | 6.2 | 2.8 | 5.3 |
| | 3 | 27.9 | 18.0 | 10.8 | 11.5 |
| | 4 | 17.8 | 9.0 | 15.6 | 14.9 |
| | 5 | 11.2 | 8.8 | 17.9 | 16.1 |
| | 6 | 25.5 | 7.5 | 20.6 | 15.8 |
| | 7 | 18.1 | 10.0 | 22.2 | 15.6 |
| | 8 | 6.9 | 6.1 | 22.6 | 15.8 |
| | 9 | 8.0 | 4.4 | 19.9 | 12.9 |
| | 10 | 0.8 | 3.5 | 14.9 | 10.8 |
| | 11 | 16.9 | 5.0 | 9.1 | 6.3 |
| | 12 | 11.2 | 6.4 | 1.6 | 6.3 |

| Leto | Mesec | Padavine [cm] | Odtok [cm] | Temp. [°C] | Izhlap. [cm] |
|---|---|---|---|---|---|
| 1964 | 1 | 26.5 | 13.8 | 4.6 | 8.5 |
| | 2 | 15.2 | 10.4 | 2.9 | 3.3 |
| | 3 | 31.6 | 18.7 | 8.8 | 13.4 |
| | 4 | 29.6 | 22.4 | 14.2 | 12.6 |
| | 5 | 8.5 | 14.1 | 18.6 | 15.0 |
| | 6 | 11.5 | 7.2 | 22.3 | 19.0 |
| | 7 | 17.3 | 6.1 | 22.4 | 14.4 |
| | 8 | 20.8 | 6.5 | 22.2 | 13.9 |
| | 9 | 15.6 | 8.0 | 19.6 | 12.1 |
| | 10 | 27.2 | 21.6 | 12.3 | 9.1 |
| | 11 | 10.8 | 8.3 | 11.2 | 6.5 |
| | 12 | 17.8 | 12.9 | 6.3 | 6.1 |
| 1965 | 1 | 10.2 | 11.7 | 4.9 | 8.5 |
| | 2 | 18.2 | 14.1 | 5.3 | 11.7 |
| | 3 | 20.9 | 15.9 | 8.3 | 9.2 |
| | 4 | 12.5 | 13.6 | 15.8 | 12.5 |
| | 5 | 11.8 | 11.2 | 19.8 | 17.9 |
| | 6 | 16.6 | 11.1 | 20.9 | 14.2 |
| | 7 | 8.8 | 8.1 | 23.5 | 14.8 |
| | 8 | 8.7 | 6.6 | 22.9 | 13.2 |
| | 9 | 15.6 | 5.4 | 20.7 | 11.7 |
| | 10 | 11.5 | 10.5 | 13.7 | 8.4 |
| | 11 | 7.9 | 4.5 | 10.3 | 6.0 |
| | 12 | 2.4 | 3.9 | 6.3 | 5.0 |
| 1966 | 1 | 17.3 | 6.5 | 3.0 | 5.5 |
| | 2 | 26.9 | 17.9 | 5.1 | 5.8 |
| | 3 | 10.9 | 15.6 | 8.8 | 10.5 |
| | 4 | 23.7 | 10.8 | 13.1 | 11.4 |
| | 5 | 21.4 | 13.9 | 17.8 | 12.7 |
| | 6 | 8.0 | 8.3 | 20.9 | 16.3 |
| | 7 | 8.4 | 5.9 | 23.8 | 16.5 |
| | 8 | 19.9 | 6.5 | 23.1 | 13.6 |
| | 9 | 11.9 | 4.7 | 19.2 | 10.6 |
| | 10 | 15.6 | 6.6 | 13.8 | 8.5 |
| | 11 | 14.2 | 12.0 | 9.7 | 6.7 |
| | 12 | 11.0 | 7.9 | 5.3 | 4.9 |
| 1967 | 1 | 10.7 | 10.4 | 5.4 | 5.1 |
| | 2 | 12.3 | 8.7 | 4.1 | 5.7 |
| | 3 | 10.1 | 10.1 | 12.0 | 10.4 |
| | 4 | 9.5 | 6.7 | 16.1 | 15.0 |
| | 5 | 17.2 | 7.4 | 17.6 | 15.3 |
| | 6 | 19.3 | 16.3 | 20.7 | 14.2 |
| | 7 | 23.3 | 12.4 | 21.7 | 13.9 |
| | 8 | 40.7 | 13.3 | 21.6 | 12.8 |
| | 9 | 8.9 | 10.9 | 17.4 | 10.8 |
| | 10 | 14.7 | 8.4 | 13.6 | 9.2 |
| | 11 | 15.3 | 9.1 | 7.2 | 7.3 |
| | 12 | 28.0 | 16.9 | 7.3 | 7.4 |

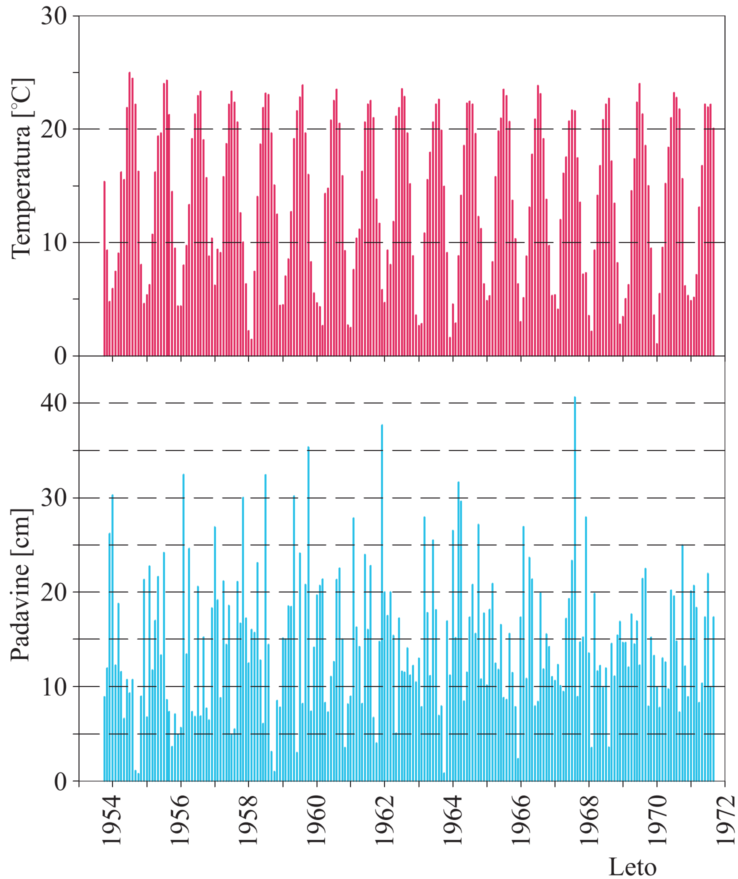
se nadaljuje...

Preglednica 2.1: Podatki o padavinah, odtoku, temperaturi in izhlapevanju (nadaljevanje)

| Leto | Mesec | Padavine [cm] | Odtok [cm] | Temp. [°C] | Izhlap. [cm] |
|---|---|---|---|---|---|
| 1968 | 1 | 13.6 | 15.0 | 3.6 | 6.1 |
| | 2 | 3.6 | 8.2 | 2.2 | 6.4 |
| | 3 | 19.9 | 12.0 | 9.3 | 13.6 |
| | 4 | 11.7 | 10.8 | 14.2 | 10.2 |
| | 5 | 12.2 | 8.2 | 16.8 | 15.4 |
| | 6 | 9.9 | 8.7 | 20.8 | 14.0 |
| | 7 | 12.0 | 5.2 | 22.2 | 14.7 |
| | 8 | 3.6 | 3.9 | 22.7 | 16.8 |
| | 9 | 14.6 | 3.9 | 17.2 | 10.9 |
| | 10 | 11.1 | 4.3 | 13.4 | 8.2 |
| | 11 | 15.5 | 5.3 | 8.2 | 4.7 |
| | 12 | 16.9 | 7.6 | 2.8 | 5.6 |
| 1969 | 1 | 14.7 | 8.8 | 3.4 | 6.9 |
| | 2 | 14.7 | 12.4 | 5.1 | 5.4 |
| | 3 | 12.1 | 10.2 | 6.3 | 8.2 |
| | 4 | 17.7 | 13.3 | 14.6 | 13.3 |
| | 5 | 14.5 | 11.3 | 17.4 | 14.7 |
| | 6 | 17.0 | 13.3 | 22.4 | 16.1 |
| | 7 | 12.3 | 6.4 | 24.0 | 17.9 |
| | 8 | 21.4 | 10.6 | 21.3 | 12.4 |
| | 9 | 22.5 | 9.9 | 18.6 | 10.9 |
| | 10 | 7.9 | 7.8 | 15.0 | 8.1 |
| | 11 | 15.2 | 9.2 | 9.5 | 4.6 |
| | 12 | 13.3 | 8.7 | 3.6 | 6.7 |

| Leto | Mesec | Padavine [cm] | Odtok [cm] | Temp. [°C] | Izhlap. [cm] |
|---|---|---|---|---|---|
| 1970 | 1 | 10.0 | 8.6 | 1.1 | 4.7 |
| | 2 | 7.8 | 7.7 | 5.5 | 7.8 |
| | 3 | 13.1 | 8.7 | 9.6 | 10.5 |
| | 4 | 12.6 | 9.0 | 15.2 | 10.6 |
| | 5 | 9.8 | 6.2 | 18.4 | 13.3 |
| | 6 | 20.2 | 7.6 | 21.0 | 13.4 |
| | 7 | 19.6 | 5.1 | 23.2 | 16.3 |
| | 8 | 14.8 | 5.5 | 22.8 | 13.3 |
| | 9 | 7.3 | 3.4 | 21.8 | 13.4 |
| | 10 | 25.0 | 6.2 | 15.6 | 8.0 |
| | 11 | 12.2 | 7.3 | 6.2 | 4.2 |
| | 12 | 8.9 | 5.2 | 5.3 | 5.6 |
| 1971 | 1 | 20.1 | 8.6 | 4.9 | 4.8 |
| | 2 | 20.7 | 12.4 | 5.2 | 8.7 |
| | 3 | 18.4 | 13.5 | 7.2 | 8.9 |
| | 4 | 8.3 | 9.3 | 13.1 | 13.9 |
| | 5 | 10.4 | 8.2 | 16.8 | 13.4 |
| | 6 | 17.4 | 5.6 | 22.2 | 15.6 |
| | 7 | 22.0 | 6.9 | 21.9 | 11.6 |
| | 8 | 9.9 | 11.1 | 22.2 | 13.3 |
| | 9 | 17.4 | 5.6 | 20.1 | 9.7 |

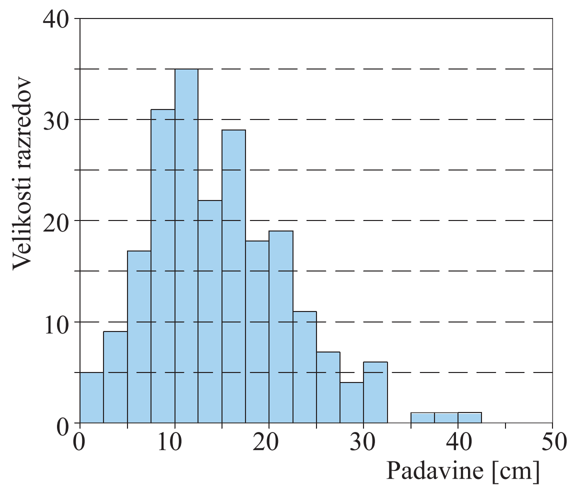
Podatke o temperaturi in padavinah lahko grafično prikažemo s črtnim diagramom (slika 2.1). Iz črtnega diagrama lahko razberemo, da se temperatura zelo značilno periodično spreminja, medtem ko za padavine tako izrazite periodičnosti ni.



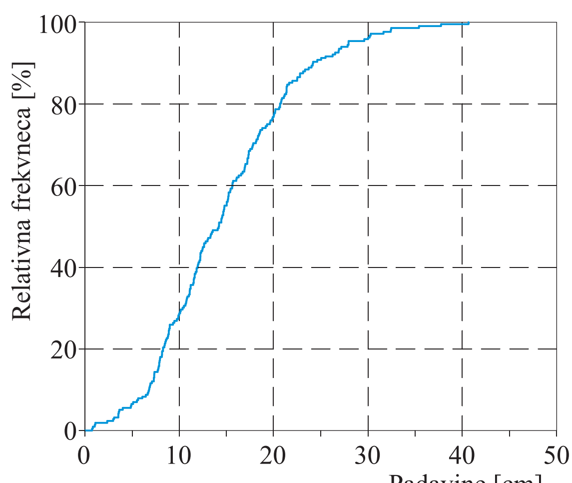

Slika 2.1: Črtni diagram podatkov o temperaturi in padavinah

S histogramom frekvenc na sliki 2.2 in kumulativnim frekvenčnim diagramom na sliki 2.3 prikazujemo podatke o padavinah. Pri pripravi histograma podatke razvrstimo v 18 enako velikih razredov: $0-2.5$, $2.5-5, \ldots, 42.5-45$. Na slikah lahko vidimo, da iz črtnega diagrama ne moremo sklepati o značilnostih podatkov, medtem ko nam druga dva diagrama o padavinah povesta že nekaj več.

Iz histograma frekvenc po razredih (slika 2.2) lahko vidimo, da so bile mesečne padavine večinoma med 10 in 20 cm. Iz te slike vidimo tudi, da so padavine okoli 40 cm zelo redek pojav. Iz kumulativnega diagrama relativnih frekvenc (slika 2.3) lahko zaključimo, da je bila polovica mesečnih padavin manjša od približno 15 cm. Kot bomo prikazali pozneje, je ta količina ena izmed mer pričakovane vrednosti (mediana). Kumulativni diagram relativnih frekvenc je primeren tudi za ugotavljanje, ali podatki ustrezajo eni izmed poznanih porazdelitev, kar bomo spoznali v Poglavju o preizkušanju domnev.

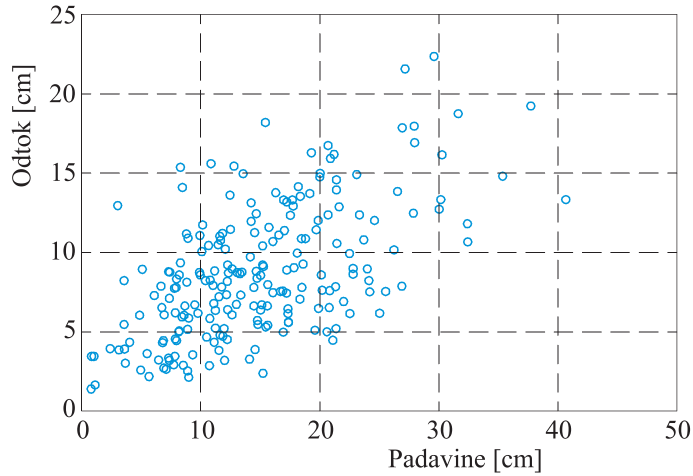



Slika 2.2: Histogram frekvenc podatkov o padavinah




Slika 2.3: Kumulativni diagram relativnih frekvenc podatkov o padavinah

Odvisnost med padavinami in odtokom padavin lahko prikažemo v točkovnem diagramu, ki ga prikazujemo na sliki 2.4.




Slika 2.4: Točkovni diagram podatkov o padavinah in odtoku

## 2.2 Numerične predstavitve vzorcev

Numerične predstavitve vzorcev so mere pričakovane vrednosti (imenujemo jih tudi mere centralne tendence, mere srednje vrednosti ali mere matematičnega upanja), mere razpršenosti oziroma variabilnosti, mere asimetričnosti in mere sploščenosti.

Najbolj običajna mera pričakovane vrednosti je aritmetična sredina ali povprečje $\bar{X}$, ki jo izračunamo po enačbi

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \tag{2.1}$$

Druga možna mera srednje vrednosti je **mediana** $X_M$ ali $Me_X$, ki predstavlja tisto vrednost količine $X$, da je 50 % elementov vzorca manjših od nje. Mediano vzorca določimo tako, da elemente $x_i$ najprej razvrstimo po velikosti. Če je velikost vzorca liho število $n = 2m + 1$, je mediana kar element $x_{m+1}$, če pa je velikost vzorca sodo število $n = 2m$, mediano določimo kot povprečje med srednjima dvema elementoma $x_m$ in $x_{m+1}$

$$X_M = Me_X = \begin{cases} x_{m+1} & \text{za liho velikost vzorca.} \\ \frac{1}{2}(x_m + x_{m+1}) & \text{za sodo velikost vzorca} \end{cases} \tag{2.2}$$

Mera srednje vrednosti je tudi **geometrijska sredina** $\bar{X}_g$ ali $M_{GX}$, ki jo izračunamo po enačbi

$$\bar{X}_g = M_{GX} = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} . \tag{2.3}$$

Tudi mer razpršenosti je več. Najbolj preprosta je srednja vrednost absolutnih odklonov od srednje vrednosti,

$$d_X = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{X}|. \tag{2.4}$$

Verjetno najpogosteje uporabljena je **varianca** $S_X^2$, ki predstavlja srednjo vrednost kvadratov odklonov od povprečja

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2. \tag{2.5}$$

Za izračun variance večjega vzorca je bolje, da zadnjo enačbo nekoliko preuredimo:

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{X} + \bar{X}^2) = \frac{1}{n}\left(\sum_{i=1}^{n}x_i^2 - 2\bar{X}\sum_{i=1}^{n}x_i + \bar{X}^2\sum_{i=1}^{n}1\right). \tag{2.6}$$

Če upoštevamo (2.1), lahko zadnjo enačbo zapišemo v obliki

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{X}^2. \tag{2.7}$$

Kot bomo pokazali v poglavju Vzorčenje, varianca $S_X^2$ v povprečju nekoliko podcenjuje varianco celotne populacije, zato moramo vpeljati varianco $S_X^{*2}$, ki je nepristranska

$$S_X^{*2} = \frac{n}{n-1}S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2. \tag{2.8}$$

**Standardni deviaciji** $S_X$ oz. $S_X^*$ vzorca imata bistveno prednost pred variancama v tem, da ju merimo z istimi enotami kot vrednosti same. Zato je pomen standardne deviacije bolj nazoren. Standardni deviaciji izračunamo po enačbah

$$S_X = \sqrt{S_X^2} \text{ oziroma } S_X^* = \sqrt{S_X^{*2}}. \tag{2.9}$$

Naslednja mera disperzije je brezdimenzionalna, izračunamo jo po enačbi

$$V_X = \frac{S_X}{\bar{X}}, \tag{2.10}$$

imenujemo pa **koeficient variacije**.

Razpršenost lahko opišemo tudi z območjem, v katerem ležijo vse vrednosti vzorca. Spodnja meja območja je najmanjši člen vzorca, zgornja meja pa največji člen vzorca. To območje imenujemo **variacijski** ali **totalni razmik** $VR_X$ in prikažemo z dvema mejama

$$VR_X = \left[\min_{i=1}^{n}(x_i), \max_{i=1}^{n}(x_i)\right]. \tag{2.11}$$

Mero asimetrije izračunamo po enačbi

$$g_{X1} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^3}{S_X^3}. \tag{2.12}$$

Mera asimetrije $g_{X1}$ je enaka nič, če so podatki razporejeni simetrično okoli srednje vrednosti $\bar{X}$.

Mero za koničastost oziroma sploščenost imenujemo koeficient kurtosis[2] in jo izračunamo po enačbi

$$g_{X2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^4}{S_X^4}. \tag{2.13}$$

Običajno $g_{X2}$ primerjamo s številko 3, ki je značilna za **normalno porazdelitev** podatkov, ki jo bomo spoznali v naslednjih poglavjih. Zato lahko definiramo mero za sploščenost tudi z drugim koeficientom kurtosis po naslednji enačbi:

$$\tilde{g}_{X2} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^4}{S_X^4} - 3 = g_{X2} - 3. \tag{2.14}$$

Povezavo med dvema skupinama podatkov lahko določimo s kovarianco vzorca, ki jo izračunamo po enačbi

$$S_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n}. \tag{2.15}$$

Tudi to enačbo lahko zapišemo v drugi obliki:

$$S_{XY} = \frac{\sum_{i=1}^{n} x_i y_i}{n} - \bar{X}\bar{Y}. \tag{2.16}$$

Brezdimenzionalna mera povezanosti med dvema skupinama podatkov je korelacijski koeficient $R_{XY}$

$$R_{XY} = \frac{S_{XY}}{S_X S_Y}. \tag{2.17}$$

Vrednost korelacijskega koeficienta je med $-1$ in 1, pri čemer vrednosti blizu ena pomenijo močno linearno povezavo (oziroma korelacijo) med dvema skupinama podatkov, kar pomeni, da večja vrednost ene

[2] Izraz izvira iz grške besede *kyrtosis*, ki pomeni konveksnost.

količine pogojuje večjo vrednost druge. Negativna korelacija pomeni, da večje vrednosti ene količine pogojujejo manjše vrednosti druge. Če je vrednost $R_{XY}$ okoli nič, pomeni, da linearne povezave med podatki ni.

**Primer 2.2:** *Določimo vse naštete numerične mere za podatke iz primera 2.1. Izračunajmo tudi kovariance med vsemi količinami!*

**Rešitev:** Z računalniškim programom EXCEL smo določili numerične predstavitve podatkov za vse štiri obravnavane podatke, kar prikazujemo v naslednji preglednici. Pri tem smo uporabili naslednje ukaze:

| | |
|---|---|
| AVERAGE | povprečje, |
| MEDIAN | mediana, |
| GEOMEAN | geometrijska sredina, |
| MIN in MAX | najmanjša in največja vrednost, |
| STDEV | standardni odklon oziroma standardna deviacija, |
| GEOMEAN | geometrijska sredina, |
| SKEW | koeficient asimetrije, |
| KURT | koeficient sploščenosti, |
| COVAR | kovarianca, |
| CORREL | korelacijski koeficient. |

Preglednica 2.2: Numerični opis podatkov

| Mere za opis podatkov | Padavine [cm] | Odtok [cm] | Temperatura [°C] | Izhlapevanje [cm] |
|---|---|---|---|---|
| povprečje | 14.76 | 8.72 | 13.89 | 10.62 |
| mediana | 14.20 | 8.20 | 14.53 | 10.66 |
| geometrijska sredina | 12.66 | 7.71 | 11.55 | 9.52 |
| najmanjša vrednost | 0.79 | 1.40 | 1.06 | 2.67 |
| največja vrednost | 40.67 | 22.35 | 25.00 | 20.88 |
| standardna deviacija | 7.34 | 4.13 | 7.03 | 4.55 |
| koeficient variacije | 0.497 | 0.473 | 0.506 | 0.429 |
| srednji absolutni odklon | 5.81 | 3.30 | 6.24 | 3.96 |
| asimetričnost $g_1$ | 0.68 | 0.64 | −0.13 | 0.07 |
| sploščenost $\tilde{g}_2$ | 0.51 | 0.12 | −1.39 | −1.16 |

Iz preglednice lahko odčitamo mere srednje vrednosti, razpršenosti, asimetričnosti in sploščenosti.

Tudi korelacijski koeficient izračunamo s programom EXCEL in prikazujemo v naslednji preglednici.

Preglednica 2.3: Korelacijski koeficienti

| | Padavine | Odtok | Temperatura | Izhlapevanje |
|---|---|---|---|---|
| Padavine | / | 0.590 | −0.068 | −0.088 |
| Odtok | 0.590 | / | −0.243 | −0.071 |
| Temperatura | −0.068 | −0.243 | / | 0.858 |
| Izhlapevanje | −0.088 | −0.071 | 0.858 | / |

Očitno je, da sta temperatura in izhlapevanje močno linearno povezana, relativno močno sta povezana tudi količina padavin in odtok vode. Vidimo, da linearne povezave med drugimi pari podatkov skoraj ni, oziroma obstaja majhna negativna linearna povezava. Ali je ta povezava statistično pomembna oziroma značilna, bomo bomo obravnavali v Poglavju o preizkušanju domnev.

# 3 Verjetnostni račun

## 3.1 Uvod

Verjetnostni račun predstavlja matematične temelje za statistiko. Osnovni pojmi verjetnostnega računa so **poskus**, **dogodek** in **verjetnost dogodka**. V nadaljevanju pa spoznamo tudi pojme **slučajna spremenljivka**, **porazdelitvena funkcija** in drugo.

Zakaj verjetnostni račun in statistika? Večina stvari, ki jih obravnavamo, ni gotovih. Verjetnostni račun je veja matematike, s katero obravnavamo dogodke, ki se bolj ali manj **verjetno** zgodijo.

Pokažimo na nekaj primerov iz vsakodnevnega (televizijskega) življenja, kjer je naključnost pojavov premalo poudarjena:

**Vremensko napoved** napovedovalci običajno podajo z izjavo, kot je na primer: „*Jutri bo sončno.*“ Naslednji dan smo vsi razočarani, če je vreme slabo. Če bi napovedovalec izjavil „*Verjetno bo jutri sončno*“ ali pa „*Jutri bo sončno z verjetnostjo 90 %,*“ bi bila izjava ustreznejša, saj bi s tem poudaril, da je vreme v veliki meri naključen pojav, ki ga je s popolno gotovostjo nemogoče napovedati.

**Športni napovedovalci** pogosto izjavijo: „*Zapravil je 100-odstotno priložnost!*“ Če bi bila priložnost 100-odstotna, bi je športnik ne zapravil. Natančnejša analiza bi pokazala, da športniki tako imenovano 100-odstotno priložnost morda izkoristijo le v 90 % primerov. Športni novinar bi torej moral izjaviti „*Zapravil je 90-odstotno priložnost!*“ ali pa vzklik spremeniti: „*Zapravil je zelo lepo priložnost!*“

Tudi v gradbeništvu je mnogo primerov, kjer se premalo zavedamo, da so pojavi naključni. En tak primer je projektiranje konstrukcij, pri čemer projektant vedno uporablja deterministične podatke o konstrukciji in obtežbi, čeprav so predvsem podatki o obtežbi in trdnosti materiala slučajne spremenljivke s sorazmerno veliko razpršenostjo.

## 3.2 Dogodek

Poskus je splet pojavov, ki ga sprožimo in opazujemo ali pa le opazujemo. Rezultat poskusa je dogodek. Enaki poskusi oziroma ponovitve poskusa lahko privedejo do različnih dogodkov. Dogodke označujemo z velikimi tiskanimi črkami $A$, $B$ ...

Dogodek je **gotov dogodek**, če se zgodi ob vsaki ponovitvi poskusa. Tak dogodek označimo z $G$. Dogodek je **nemogoč**, če se nikoli ne zgodi. Tak dogodek označimo z $N$. Vsi drugi dogodki so **naključni**.

Zapišimo nekaj zvez med dogodki.

**Način dogodka:** Če se dogodek $B$ zgodi vedno, ko se zgodi dogodek $A$, potem je dogodek $A$ način dogodka $B$. To zvezo označimo takole:

$$A \subset B.$$

Poglejmo primer: če pri metanju kocke vržemo tri pike, je ta dogodek način dogodka, da vržemo liho število pik.

**Enakost dogodkov:** Če se dogodek $A$ zgodi vedno, ko se zgodi dogodek $B$, in se dogodek $B$ zgodi vedno, ko se zgodi dogodek $A$, potem sta dogodka enaka

$$A \subset B \quad \text{in} \quad A \supset B \quad \Longleftrightarrow \quad A = B.$$

**Vsota dogodkov:** Dogodek, da se zgodi vsaj eden od dogodkov $A$ in $B$, imenujemo vsota dogodkov in označimo takole:

$$A \cup B \ \text{ ali } \ A + B.$$

Na primeru meta kocke lahko ilustriramo pomen vsote dogodkov takole: dogodek $A$ pomeni lihi rezultat meta kocke $\{1, 3, 5\}$, dogodek $B$ pa primer, ko vržemo manj kot štiri pike $\{1, 2, 3\}$. Vsota $A \cup B$ je dogodek, ki ustreza naslednjim rezultatom meta kocke: $\{1, 2, 3, 5\}$.

Poleg komutativnosti in asociativnosti dogodkov,

$$A \cup B = B \cup A, \qquad (A \cup B) \cup C = A \cup (B \cup C),$$

veljata še ti zvezi:

$$A \cup N = A, \qquad A \cup G = G.$$

**Produkt dogodkov:** Dogodek, da se sočasno zgodita dva dogodka, imenujemo produkt dogodkov in označimo takole:

$$A \cap B \ \text{ ali } \ A\,B.$$

Na primeru meta kocke lahko ilustriramo pomen produkta dogodkov takole: dogodek $A$ pomeni lihi rezultat meta kocke $\{1, 3, 5\}$, dogodek $B$ pa primer, ko vržemo manj kot štiri pike $\{1, 2, 3\}$. Produkt $A \cap B$ predstavlja dogodek, ki vključuje rezultata ena in tri.

Poleg komutativnosti in asociativnosti dogodkov,

$$A \cap B = B \cap A, \qquad (A \cap B) \cap C = A \cap (B \cap C),$$

veljajo še naslednje zveze:

$$A \cap N = N, \qquad A \cap G = A$$

in

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C), \qquad (A \cup B) \cap C = (A \cap C) \cup (B \cap C).$$

**Nezdružljivost dogodkov:** Dva dogodka sta **nezdružljiva**, če se ne moreta zgoditi sočasno. Na primer: če mečemo kocko, je dogodek, da vržemo število tri, nezdružljiv z dogodkom, da vržemo število pet. To preprosto pomeni, da ne moremo hkrati vreči števila tri in pet. Za nezdružljiva dogodka velja, da je njun produkt nemogoč dogodek,

$$A \cap B = N.$$

**Nasprotni dogodek:** Dogodek $\bar{A}$ je nasproten dogodku $A$, če se zgodi vedno, kadar se dogodek $A$ ne zgodi. To pomeni, da se vedno zgodi vsaj eden od dogodkov $A$ in $\bar{A}$. Na primer: če mečemo kocko, je dogodek, da vržemo manj kot šest pik, nasproten dogodku, da vržemo šest pik. Za nasprotna dogodka velja, da je njun produkt nemogoč, njuna vsota pa gotov dogodek,

$$A \cap \bar{A} = N, \qquad A \cup \bar{A} = G. \tag{3.1}$$

Zapišimo še vsoto in produkt dveh nasprotnih dogodkov:

$$\bar{A} \cap \bar{B} = \overline{A \cup B}, \qquad \bar{A} \cup \bar{B} = \overline{A \cap B},$$

kar z besedami povemo takole: nasprotni dogodek vsote dveh dogodkov $A$ in $B$ je produkt nasprotnih dogodkov $\bar{A}$ in $\bar{B}$ in obratno, nasprotni dogodek produkta dveh dogodkov $A$ in $B$ je vsota nasprotnih dogodkov $\bar{A}$ in $\bar{B}$.

**Elementarni dogodek:** Dogodek, ki ga ne moremo zapisati kot vsoto dogodkov, je elementarni dogodek, na primer pri metu kocke dogodek, da pade določeno število pik.

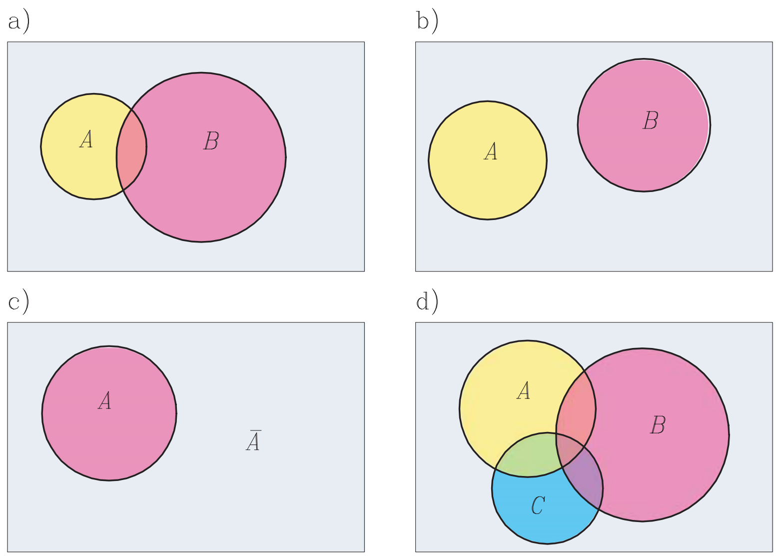
**Popolni sistem dogodkov:** Množica $n$ dogodkov je popolni sistem dogodkov, če se pri vsaki ponovitvi dogodka zgodi natanko en iz te množice. To pomeni, da so vsi dogodki v popolnem sistemu nezdružljivi, njihova vsota pa je gotov dogodek,

$$A_i \cap A_j = N, \quad i \neq j, \qquad A_1 \cup A_2 \cup \cdots \cup A_n = G.$$

Enega od značilnih popolnih sistemov dogodkov sestavljata poljubni dogodek $A$ in njemu nasprotni dogodek $\bar{A}$ (glej enačbi (3.1)).

Za prikaz dogodkov v množici dogodkov pogosto uporabimo diagram Johna Venna[1] (Vennov diagram). Primere Venovega diagrama za dva nezdružljiva, dva združljiva dogodka, za nasprotni dogodek in tri združljive dogodke prikazujemo na sliki 3.1.

Vsoto vseh možnih dogodkov, ki je seveda gotov dogodek, imenujemo tudi **verjetnostni prostor**.




Slika 3.1: Diagrami Johna Venna: a) dva združljiva dogodka
b) dva nezdružljiva dogodka
c) dogodek in njemu nasprotni dogodek
d) trije združljivi dogodki

## 3.3 Verjetnost dogodka

Poznamo več definicij verjetnosti dogodka.

**Statistična definicija verjetnosti dogodka**: Opazujemo zaporedje enakih poskusov in zapisujemo število ponovitev dogodka $n_A$ glede na vse ponovitve dogodka $n$. Kvocient $n_A/n$ se pri velikem številu ponovitev poskusa ustali pri nekem številu, ki ga definiramo kot verjetnost dogodka $A$ in označimo s $P[A]$:

$$P[A] = \lim_{n\to\infty} \frac{n_A}{n}. \tag{3.2}$$

[1] John Venn, angleški matematik (1834–1923).

Najbolj znan je poskus z zaporedjem metanjem kovanca. Naredili in dokumentirali so ga znani matematiki, začetniki verjetnostnega računa in statistike: Buffon[2], Morgan[3] in Pearson.[4] Rezultate njihovih poskusov prikazujemo v preglednici 3.1.

Preglednica 3.1: Rezultati metanja (in lovljenja) kovanca

| Izvajalec poskusov | Število metov | Število grbov | Število številk | Relativna frekvenca grbov | Relativna frekvenca številk |
|---|---|---|---|---|---|
| Buffon (1707–1788) | 4 040 | 2 060 | 1 980 | 0.5100 | 0.4900 |
| Morgan (1806–1871) | 4 090 | 2 047 | 2 043 | 0.5005 | 0.4995 |
| Pearson (1857–1936) | 12 000 | 6 019 | 5 981 | 0.5016 | 0.4984 |
| Pearson (1857–1936) | 24 000 | 12 012 | 11 988 | 0.5005 | 0.4995 |
| Mathematica (1998) | 1 000 000 | 500 274 | 499 726 | 0.5003 | 0.4997 |

S programom Mathematica[5] smo simulirali metanje kovanca. Opravili smo 1 000 000 ponovitev poskusa in vsakih 1000 ponovitev izračunali relativno frekvenco dogodka, da smo vrgli številko, in dogodka, da smo vrgli grb. Na sliki 3.2 vidimo, da se relativna frekvenca bliža 0.5, vendar tudi po tolikšnem številu ponovitev ni natančno 0.5. Čeprav je statistična definicija verjetnosti pomembna za razumevanje verjetnosti, ni primerna za osnovo verjetnostnega računa.

Tako definirana verjetnost dogodka, da smo vrgli številko, ni določeno število, saj predstavlja $n$ ponovitev tega poskusa le **vzorec** iz neskončno velike **populacije**. Količine, ki jih določimo iz vzorca, ki ne zajema celotne populacije, so tudi naključne.

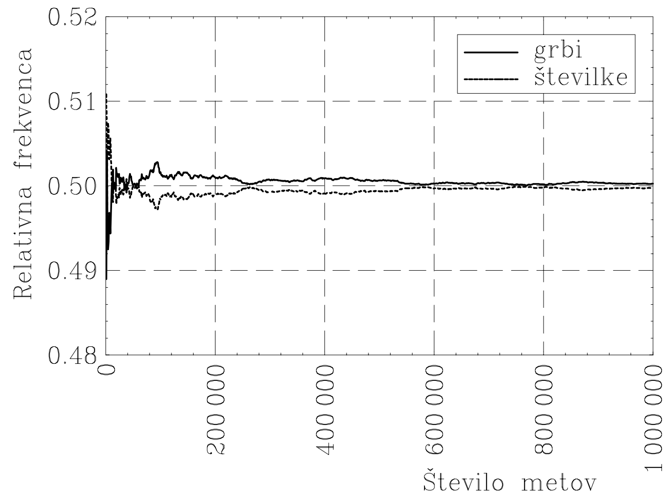
Drugače bi bilo, če bi bila populacija končno velika, kot je na primer končno število $n$ izdelkov neke tovarne. Če vse te izdelke pregledamo in ugotovimo, da je $n_A$ pokvarjenih, lahko zapišemo verjetnost dogodka $A$, da je poljubni izdelek te tovarne pokvarjen,

$$P[A] = \frac{n_A}{n}. \tag{3.3}$$

**Klasična definicija verjetnosti dogodka:** Opazujemo poskus, ki ima lahko $n$ izidov, ki so vsi enako verjetni in nezdružljivi elementarni dogodki. Vsi možni izidi poskusa sestavljajo popolni sistem dogodkov.

[2] Georges Louis Leclerc Comte de Buffon, francoski matematik (1707–1788).
[3] Augustus De Morgan, angleški matematik (1806–1871).
[4] Karl Pearson, angleški matematik (1857–1936).
[5] Simulacije, izvedene s programom MATHEMATICA:

```
stevilk=0;  grbov=0;  vseh = 1000000;
Do[dogodek=Random[Integer]; If[dogodek==1,stevilk++,grbov++],{i,vseh}]
Print["Stevilk -> ", stevilk,"          Grbov -> ", grbov];
Print["P[stevilk] = ",N[stevilk/vseh],"       P[grbov] =
",N[grbov/vseh]]

Stevilk -> 500274           Grbov -> 499726
P[stevilk] = 0.500274       P[grbov] = 0.499726
```




Slika 3.2: Relativna frekvenca pojava številke in grba

Vzemimo, da dogodek $A$ predstavlja vsoto $n_A$ izidov. V tem primeru definiramo verjetnost dogodka $A$ z enačbo

$$P[A] = \frac{n_A}{n}.$$

Rečemo lahko, da je verjetnost dogodka kvocient za dogodek $A$ ugodnih izidov $n_A$ in vseh možnih izidov $n$. Pomanjkljivost te definicije verjetnosti je v predpostavki, da so izidi enako verjetni in nezdružljivi dogodki. V nekaterih primerih je to sicer upravičeno predpostaviti, v vseh pa ne.

**Aksiomatična definicija verjetnosti dogodka:** Za verjetnostni račun je ta definicija najpomembnejša. Definicija vključuje sistem treh aksiomov:

1. Verjetnost poljubnega dogodka leži med nič in ena:

$$0 \leq P[A] \leq 1.$$

2. Verjetnost gotovega dogodka je enaka ena:

$$P[G] = 1.$$

3. Verjetnost vsote dveh nezdružljivih dogodkov $A$ in $B$ je vsota njunih verjetnosti:

$$P[A \cup B] = P[A] + P[B].$$

Na osnovi teh aksiomov, lahko dokažemo vrsto izrekov, povezanih z verjetnostjo dogodkov. Na primer: Verjetnost vsote poljubnih dveh dogodkov $A$ in $B$ zapišemo z naslednjo enačbo

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]. \tag{3.4}$$

To zvezo dokažemo tako, da vsoto dogodkov $A$ in $B$ zapišemo kot vsoto dveh nezdružljivih dogodkov takole:

$$A \cup B = A \cup (\bar{A} \cap B) \quad \to \quad P[A \cup B] = P[A] + P[\bar{A} \cap B]. \tag{3.5}$$

Nadalje zapišimo dogodek $B$ kot vsoto nezdružljivih dogodkov:

$$\begin{aligned} &B = G \cap B = (A \cup \bar{A}) \cap B = (A \cap B) \cup (\bar{A} \cap B) \\ &\quad \to \quad P[B] = P[A \cap B] + P[\bar{A} \cap B]. \end{aligned} \tag{3.6}$$

S primerjavo enačb (3.5) in (3.6) dobimo končno enačbo (3.4).

Enačbo lahko razširimo za verjetnost vsote treh dogodkov $P[A \cup B \cup C]$. Enačbo izpeljemo tako, da vsoto dogodkov $A$ in $B$ označimo z $D = A \cup B$ ter zapišemo verjetnost vsote dogodkov $D$ in $C$:

$$P[A \cup B \cup C] = P[D \cup C] = P[D] + P[C] - P[D \cap C]. \tag{3.7}$$

Sedaj zapišimo še verjetnosti dogodkov $D = A \cup B$ in $D \cap C = (A \cap C) \cup (B \cap C)$:

$$\begin{aligned} &P[D] = P[A \cup B] = P[A] + P[B] - P[A \cap B], \\ &P[D \cap C] = P[(A \cap C) \cup (B \cap C)] = P[A \cap C] + P[B \cap C] - P[A \cap B \cap C]. \end{aligned} \tag{3.8}$$

Enačbi (3.8) vstavimo v (3.7) in dobimo končni izraz

$$\begin{aligned} P[A \cup B \cup C] = {} & P[A] + P[B] + P[C] \\ & - P[A \cap B] - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]. \end{aligned} \tag{3.9}$$

Na podoben način lahko izpeljemo izraz za verjetnost vsote poljubnega števila dogodkov.

Pokažimo še, kako verjetnost nasprotnega dogodka $P[\bar{A}]$ izrazimo z verjetnostjo $P[A]$. Upoštevamo, da je vsota dogodka $A$ in nasprotnega dogodka $\bar{A}$ gotov dogodek $G$ in da sta dogodka $A$ in $\bar{A}$ nezdružljiva, ter zapišemo

$$P[G] = P[A \cup \bar{A}] = P[A] + P[\bar{A}] = 1 \quad \to \quad P[\bar{A}] = 1 - P[A]. \tag{3.10}$$

Zadnja definicija verjetnosti je **izkustvena** oziroma **subjektivna**. Čeprav je s stališča verjetnostnega računa najmanj uporabna, je zelo pogosto uporabljena. Pogosto rečemo, da je en dogodek bolj ali manj verjeten od drugega. Večkrat je taka izjava čisto subjektivna, včasih pa temelji na analizi podobnih dogodkov v preteklosti.

**Primer 3.1:** *Hkrati mečemo dve pošteni igralni kocki. Določimo verjetnost, da je vsota pik na kockah enaka štiri po vseh štirih definicijah verjetnosti.*

**Rešitev:** Najprej določimo verjetnost po *statistični definiciji*. Kocki vržemo mnogokrat (na primer milijonkrat). Preštejemo, kolikokrat je bila vsota enaka štiri, in uporabimo enačbo (3.2). Pri zelo velikem številu ponovitev meta in z zelo poštenimi kockami dobimo verjetnost $\mathrm{P}[A] \approx 1/12$. To smo naredili s programom MATHEMATICA[6] in za 1 000 000 ponovitev poskusa meta dveh kock izračunali, da je verjetnost, da je bila vsota enaka štiri, enaka 0.0838, medtem ko je teoretična vrednost enaka $1/12 = 0.083\dot{3}$.

*Klasična definicija:* Vseh možnih izidov metanja dveh kock je $N = 6^2 = 36$. Vsi izidi, pri katerih je vsota enaka 4, so $(1, 3)$, $(2, 2)$ in $(3, 1)$. Število izidov z iskanim rezultatom je torej $N_A = 3$. Po enačbi (3.3) izračunamo verjetnost $\mathrm{P}[A] = 3/36 = 1/12$.

*Aksiomatična definicija:* Dogodek $A$, da je vsota pik na obeh kockah enak štiri, je sestavljen iz treh med seboj nezdružljivih dogodkov: $A_1$ – na obeh kockah sta dve piki, $A_2$ – na prvi je ena pika, na drugi pa tri pike in $A_3$ – na prvi kocki so tri pike in na drugi ena. Verjetnost vsakega od treh dogodkov je enaka $\mathrm{P}[A_i] = 1/36$. Po tretjem aksiomu velja, da je verjetnost sestavljenega dogodka $\mathrm{P}[A] = \mathrm{P}[A_1 \cup A_2 \cup A_3] = \mathrm{P}[A_1] + \mathrm{P}[A_2] + \mathrm{P}[A_3] = 3/36 = 1/12$.

*Subjektivna definicija:* Izkušen kockar bo brez obotavljanja in računanja dejal: „Verjetnost je nekaj več kot 8 %."

### 3.3.1 Pogojna verjetnost

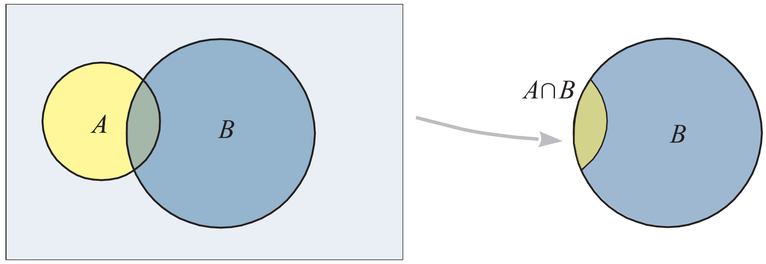
Vzemimo, da imamo dva dogodka $A$ in $B$, ki se zgodita z verjetnostima $P[A]$ in $P[B] > 0$. Verjetnost, da se zgodi dogodek $A$, ob pogoju, da se je zgodil dogodek $B$, imenujemo **pogojna verjetnost** in jo označimo z izrazom $P[A|B]$. Podobno $P[B|A]$ predstavlja pogojno verjetnost dogodka $B$ ob pogoju, da se je zgodil dogodek $A$.

Izraz za določitev pogojne verjetnosti $P[A|B]$ lahko izpeljemo na več načinov. Pri uporabi aksiomatične definicije verjetnosti, rečemo, da če se je zgodil dogodek $B$, potem ni več naključen, ampak gotov, torej ima verjetnost ena. Lahko rečemo, da se je verjetnostni prostor skrčil na dogodek $B$, kar grafično prikazujemo na sliki 3.3.

[6] Simulacije, izvedene s programom MATHEMATICA:

```
Vsotaje4 = 0;  Vsehposkusov = 1000000;
Do[Prvimet = Random[Integer,5]+1; Drugimet = Random[Integer,5]+1;
   Vsota = Prvimet + Drugimet; If[Vsota==4,Vsotaje4++],{i,Vsehposkusov}];
Print["P[Vsota je 4] = ", N[Vsotaje4/Vsehposkusov]];
```

```
P[Vsota je 4] = 0.083795
```




Slika 3.3: Verjetnostni prostor se je skrčil na dogodek $B$

Zato verjetnost, da se zgodi dogodek $A$, ob pogoju, da se je zgodil dogodek $B$, določimo tako, da verjetnost, da se zgodi produkt dogodkov $A$ in $B$, delimo z verjetnostjo dogodka $B$:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad \rightarrow \quad P[A|B]\, P[B] = P[A \cap B]. \tag{3.11}$$

Pri uporabi klasične definicije verjetnosti rečemo, da je možnih $n$ izidov. Denimo, da se dogodek $B$ zgodi $n_B$-krat, produkt dogodkov $A \cap B$ pa $n_{AB}$-krat. Verjetnosti dogodkov $B$ in $A \cap B$ sta

$$P[B] = \frac{n_B}{n}, \qquad P[A \cap B] = \frac{n_{AB}}{n}. \tag{3.12}$$

Verjetnost, da se je zgodil dogodek $A$ pri pogoju, da se je zgodil dogodek $B$ po klasični definiciji verjetnosti lahko zapišemo z izrazom

$$P[A|B] = \frac{n_{AB}}{n_B} = \frac{\frac{n_{AB}}{n}}{\frac{n_B}{n}},$$

ob upoštevanju (3.12) pa s končno enačbo

$$P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

**Neodvisnost dogodkov**. Dva dogodka sta medsebojno neodvisna, če je njuna pogojna verjetnost enaka brezpogojni verjetnosti

$$P[A] = P[A|B], \qquad P[B] = P[B|A].$$

Iz izraza (3.11) in pogoja za neodvisnost dogodkov sledi, da je verjetnost produkta dveh neodvisnih dogodkov produkt verjetnosti teh dveh dogodkov,

$$P[A] = P[A|B] = \frac{P[A \cap B]}{P[B]} \quad \rightarrow \quad P[A \cap B] = P[A]\, P[B].$$

**Primer 3.2:** *Vzemimo, da je verjetnost dogodka $A$ enaka $0.6$ ($P[A] = 0.6$), verjetnost vsote dogodkov $A \cup B$ pa $0.8$ ($P[A \cup B] = 0.8$).*

*a) Določimo verjetnost dogodka $B$, če sta dogodka $A$ in $B$ nezdružljiva.*

*b) Določimo verjetnost dogodka $B$, če sta dogodka $A$ in $B$ neodvisna.*

**Rešitev**:

a) Zaradi nezdružljivosti dogodkov $A$ in $B$ lahko verjetnost vsote dveh dogodkov izračunamo kot vsoto verjetnosti teh dveh dogodkov (3. aksiom),

$$P[A \cup B] = P[A] + P[B] \quad \rightarrow \quad P[B] = P[A \cup B] - P[A] = 0.8 - 0.6 = 0.2.$$

b) Za poljubne neodvisne dogodke moramo verjetnost vsote dveh dogodkov zapisati po enačbi (3.4), pri čemer pa upoštevamo, da je zaradi neodvisnosti dogodkov verjetnost produkta dveh dogodkov enaka produktu verjetnosti,

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = P[A] + P[B] - P[A]\, P[B]$$

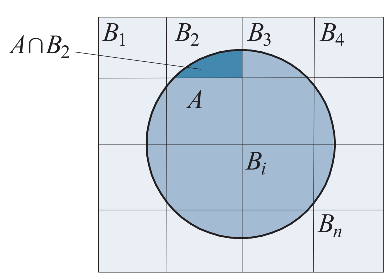
$$\rightarrow \quad P[B] = \frac{P[A \cup B] - P[A]}{1 - P[A]} = \frac{0.8 - 0.6}{1 - 0.6} = 0.5.$$

Vidimo, da je verjetnost dogodka $B$ v prvem in drugem primeru drugačna. Pomembno je, da vemo, ali so dogodki neodvisni ali pa nezdružljivi in to pravilno upoštevamo.

### 3.3.2 Popolna verjetnost dogodka

Vzemimo, da $B_i, (i = 1, \ldots, n)$ sestavljajo popolni sistem dogodkov. Dogodek $A$ lahko zapišemo kot vsoto nezdružljivih dogodkov (glej sliko 3.4):

$$A = \sum_{i=1}^{n} A \cap B_i \quad \rightarrow \quad P[A] = \sum_{i=1}^{n} P[A \cap B_i].$$




Slika 3.4: Dogodek $A$ lahko sestavimo iz posameznih dogodkov $A \cap B_i$

Ob upoštevanju izraza (3.11) lahko enačbo za določitev verjetnosti dogodka $A$ zapišemo z izrazom za popolno verjetnost dogodka,

$$P[A] = \sum_{i=1}^{n} P[A|B_i]\, P[B_i]. \tag{3.13}$$

### 3.3.3 Bayesov obrazec

Vzemimo, da $B_i, (i = 1, \ldots, n)$ sestavljajo popolni sistem dogodkov. Ugotovimo, kolikšna je verjetnost dogodka $B_j$, pri pogoju, da se je zgodil dogodek $A$. Pogojna verjetnost $P[B_j|A]$ je

$$P[B_j|A] = \frac{P[B_j \cap A]}{P[A]}$$

in ob upoštevanju izrazov (3.11) in (3.13)

$$P[B_j|A] = \frac{P[A|B_j]\,P[B_j]}{\sum_{i=1}^{n} P[A|B_i]\,P[B_i]}. \tag{3.14}$$

Enačbo (3.14) imenujemo **Bayesov obrazec**.[7]

**Primer 3.3: Okuženost s HIV** *(human immunodeficiency virus). Testiranje okuženosti je na prvi pogled izredno zanesljivo. Le v enem izmed* $20\,000$ *testiranj se zgodi, da je rezultat pozitiven, bolnik pa ni okužen. Podobno se zgodi le enkrat v* $30\,000$ *primerih, da je test negativen, bolnik pa je okužen. Vzemimo, da je v testirani populaciji okužen eden na* $10\,000$ *ljudi. Ugotovimo, kolikšna je verjetnost, da je bolnik okužen, če je test pozitiven.*

**Rešitev**: Zapišimo dogodke in njihove verjetnosti:

| | | |
|---|---|---|
| dogodek $A$: | bolnik je okužen | $P[A] = 1/10\,000 = 0.000100$ |
| dogodek $\bar{T}\|A$: | test je negativen ob tem, da je bolnik okužen | $P[\bar{T}\|A] = 1/30\,000 = 0.000033$ |
| dogodek $T\|A$: | test je pozitiven ob tem, da je bolnik okužen | $P[T\|A] = 1 - 1/30\,000 = 0.999967$ |
| dogodek $T\|\bar{A}$: | test je pozitiven ob tem, da bolnik ni okužen | $P[T\|\bar{A}] = 1/20\,000 = 0.000050$ |
| dogodek $T$: | test je pozitiven | $P[T] = ?$ |
| dogodek $A\|T$: | bolnik je okužen, če je test pozitiven | $P[A\|T] = ?$ |

Najprej izračunajmo verjetnost, da je test pozitiven $P[T]$ ob predpostavki, da je verjetnost okuženosti v testirani populaciji le 0.000100. Uporabimo izraz za popolno verjetnost (3.13), kjer popolni sistem dogodkov predstavljata dogodka $A$ in $\bar{A}$:

$$\begin{aligned} P[T] &= P[T|A]\,P[A] + P[T|\bar{A}]\,P[\bar{A}] = P[T|A]\,P[A] + P[T|\bar{A}]\,(1 - P[A]) \\ &= 0.999967 \times 0.000100 + 0.000050 \times (1 - 0.000100) = 0.000150. \end{aligned}$$

[7] Thomas Bayes, angleški matematik (1702–1761).

Vidimo, da je verjetnost, da je test pozitiven (dogodek $T$), večja od verjetnosti, da je bolnik okužen (dogodek $A$).

Verjetnost, da je bolnik okužen, če je test pozitiven (dogodek $A|T$), izračunamo z Bayesovim obrazcem (3.14):

$$P[A|T] = \frac{P[T|A]\,P[A]}{P[T|A]\,P[A] + P[T|\bar{A}]\,(1 - P[A])}$$

$$= \frac{0.999967 \times 0.000100}{0.999967 \times 0.000100 + 0.000050 \times (1 - 0.000100)} = 0.667.$$

Dobili smo relativno majhno verjetnost. Velja torej, da je verjetnost, da je bolnik okužen, če je test pozitiven, le 0.667. Ta rezultat je posledica verjetnosti, da je bolnik okužen (dogodek $A$), ki je zelo majhna. Če bi bila ta verjetnost večja, bi se tudi verjetnost dogodka $A|T$ povečala. Če bi bila ta verjetnost $P[A] = 0.5$, bi bila verjetnost, da je bolnik okužen, če je test pozitiven, enaka $P[A|T] = 0.99995$.

# 4 Slučajne spremenljivke

**Slučajna spremenljivka** lahko zavzame različne vrednosti. Katero vrednost v obravnavanem dogodku oziroma trenutku zavzame, je odvisno od naključja. Slučajne spremenljivke označujemo z velikimi tiskanimi črkami $X$, $Y$, $Z$,..., vrednosti, ki jih le-te zavzamejo, pa z malimi črkami $x$, $y$, $z$,... Rečemo lahko na primer, da slučajna spremenljivka $X$ zavzame vrednost $x$.

Primer: Vzemimo, da je število pik, ki jih lahko vržemo s kocko, slučajna spremenljivka $X$. Ta slučajna spremenljivka ima lahko vrednosti $x =$ 1, 2, 3, 4, 5 ali 6.

Vse vrednosti, ki jih slučajna spremenljivka lahko zavzame, sestavljajo **zalogo vrednosti** slučajne spremenljivke. Zaloga vrednosti je lahko končno ali neskončno število diskretnih vrednosti ali pa končni ali neskončni interval realnih števil. V prvem primeru govorimo o **diskretni slučajni spremenljivki**, v drugem pa o **zvezni slučajni spremenljivki**. V nekaterih primerih je slučajna spremenljivka kombinirana, torej diskretno-zvezna.

Primeri diskretne slučajne spremenljivke so:

- število pik pri metu kocke,
- število vozil, ki peljejo v določenem časovnem obdobju po določeni cesti,
- število poplav v nekem kraju na določeno časovno obdobje,
- število parcel v naključno izbrani katastrski občini in drugo.

V vseh primerih diskretne slučajne spremenljivke vrednosti **štejemo**.

Primeri zvezne slučajne spremenljivke pa so:

- količina padavin v nekem kraju v nekem daljšem časovnem obdobju (na primer leto),
- čas, ki mine med dvema zaporednima poplavama,
- največja hitrost vetra na nekem mestu v določenem časovnem obdobju,
- tlačna trdnost betona,
- razdalja, izmerjena z razdaljemerom in drugo.

V vseh primerih zvezne slučajne spremenljivke vrednosti **merimo**.

Primer diskretno-zvezne spremenljivke je količina padavin v nekem kraju v določenem kratkem časovnem obdobju (na primer en dan). Zaloga vrednosti te slučajne spremenljivke zajema diskretno vrednost nič za dneve, ko ni padavin, in zvezno območje količine padavin za dneve, ko so padavine.

Podoben je primer škode, ki v nekem obdobju nastane zaradi naravne nesreče (poplava, zemeljski plaz, potres, ...). Verjetnost, da škode ni, je na srečo večja od nič. Če se škoda zaradi naravne nesreče zgodi, seveda lahko zavzame poljubno zvezno vrednost, večjo od nič, če pa škode ni, slučajna spremenljivka zavzame diskretno vrednost nič.

Vsaki vrednosti ali območju vrednosti slučajne spremenljivke lahko priredimo verjetnost, da bo slučajna spremenljivka imela to vrednost. Predpis, ki določa te verjetnosti, imenujemo **porazdelitveni zakon** oziroma **porazdelitev** slučajne spremenljivke.

## 4.1 Diskretne slučajne spremenljivke

Zaloga vrednosti diskretne slučajne spremenljivke $X$ so določene diskretne vrednosti $x_i$, $i = 1, \ldots, n$, lahko pa so tudi neštevilčne oznake, na primer barve, tipi vozila in drugo. Porazdelitev te spremenljivke opišemo z verjetnostmi $p_i$, da slučajna spremenljivka $X$ zavzame vrednost $x_i$:

$$p_X(x_i) = p_i = P[X = x_i], \qquad i = 1, \ldots, n. \tag{4.1}$$

Funkcijo $p_X(x_i)$ imenujemo **verjetnostna funkcija** ali **verjetnostna masna funkcija** (angl. probability mass function – PMF). Ker slučajna spremenljivka ne more imeti sočasno dveh vrednosti in ker ne more imeti nobene druge vrednosti razen ene od $x_i$, so dogodki, da ima slučajna spremenljivka določene vrednosti, medsebojno nezdružljivi. Vsota dogodkov, da slučajna spremenljivka zavzame katerokoli vrednost iz zaloge vrednosti, je gotov dogodek.

1. Iz 1. aksioma sledi, da je $p_i$ večja ali enaka nič in manjša ali enaka ena,

   $$0 \leq p_i \leq 1.$$

2. Verjetnost vsote nezdružljivih dogodkov je enaka vsoti verjetnosti nezdružljivih dogodkov (3. aksiom). Verjetnost vsote vseh možnih dogodkov je enaka ena (2. aksiom),

   $$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} P[X = x_i] = 1.$$

Posledica 3. aksioma je tudi zveza med verjetnostjo, da $X$ zavzame vrednost med $a$ in $b$, in verjetnostno funkcijo:

$$P[a < X \leq b] = \sum_{a < x_i \leq b} p_i.$$

Grafično lahko verjetnosti v enačbi (4.1) prikažemo s črtnim diagramom, kar bomo prikazali v primerih 4.2 in 4.3.

**Porazdelitveno funkcijo** $F_X(x)$ oziroma **kumulativno porazdelitveno funkcijo** (angl. cumulative distribution function – CDF) definiramo kot verjetnost, da slučajna spremenljivka zavzame vrednost, ki je manjša ali enaka $x$:

$$F_X(x) = P[X \leq x]. \tag{4.2}$$

Pri diskretnih slučajnih spremenljivkah je vrednost porazdelitvene funkcije $F_X(x)$ enaka vsoti verjetnosti, da slučajna spremenljivka $X$ zavzame vrednost, manjšo ali enako $x$,

$$F_X(x) = \sum_{x_i \leq x} p_i. \tag{4.3}$$

Graf porazdelitvene funkcije diskretne slučajne spremenljivke je stopničasta črta, kar bomo prikazali v primerih 4.2 in 4.3.

Za vse porazdelitvene funkcije velja, da so nepadajoče funkcije, ki so za $x \to -\infty$ enake nič, za $x \to \infty$ pa ena,

$$\lim_{x \to -\infty} F_X(x) = P[X < -\infty] = 0, \qquad \lim_{x \to \infty} F_X(x) = P[X < \infty] = 1. \tag{4.4}$$

**Primer 4.1:** *Izrazimo verjetnost, da slučajna spremenljivka $X$ zavzame vrednost med $a$ in $b$, kjer je $a < b$, s porazdelitveno funkcijo slučajne spremenljivke $F_X(x)$. Nato dokažimo, da je porazdelitvena funkcija nepadajoča.*

**Rešitev:** Zapišimo verjetnost $P[a < X \leq b] = P[(a < X) \cap (X \leq b)]$ z verjetnostjo nasprotnega dogodka:

$$P[a < X \leq b] = 1 - P\left[\overline{(a < X) \cap (X \leq b)}\right] = 1 - P[(a \geq X) \cup (X > b)].$$

Ker sta dogodka $(a \geq X)$ in $(X > b)$ nezdružljiva, lahko uporabimo tretji aksiom in zapišemo

$$P[a < X \leq b] = 1 - P[X \leq a] - P[X > b] = 1 - P[X \leq a] - (1 - P[X \leq b])\,.$$

Ob upoštevanju definicije kumulativne funkcije sedaj napišemo končno enačbo:

$$P[a < X \leq b] = P[X \leq b] - P[X \leq a] = F_X(b) - F_X(a). \tag{4.5}$$

Z enačbo (4.5) lahko pokažemo, da je porazdelitvena funkcija nepadajoča. Vzemimo, da je $b = a + h$, kjer je $h \geq 0$, in to vstavimo v enačbo (4.5). Upoštevamo še prvi aksiom verjetnosti, ki zagotavlja, da je verjetnost poljubnega dogodka nenegativna,

$$P[a < X \leq a + h] = F_X(a + h) - F_X(a) \geq 0.$$

Iz zadnjega izraza sledi neenačba

$$F_X(a + h) \geq F_X(a) \quad \text{za} \quad h \geq 0,$$

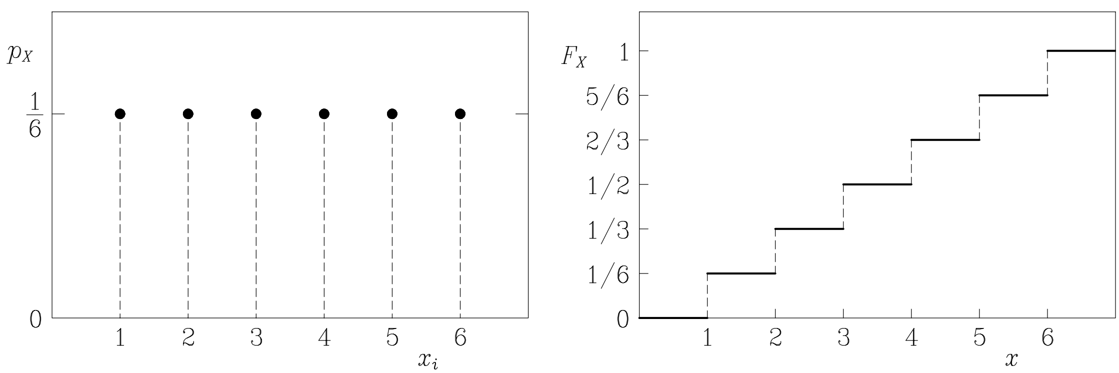
ki nam pove, da je porazdelitvena funkcija nepadajoča.

**Primer 4.2:** *Vzemimo, da slučajna spremenljivka $X$ predstavlja izide metanja poštene kocke. Določimo porazdelitev te slučajne spremenljivke in narišimo grafa verjetnostne in porazdelitvene funkcije.*

**Rešitev:** Kot smo že omenili, zalogo vrednosti te slučajne spremenljivke sestavljajo naravna števila 1, 2, 3, 4, 5 in 6. Ker so vsi izidi poštene kocke enako verjetni, lahko porazdelitev zapišemo z verjetnostno funkcijo takole:

$$p_i = \frac{1}{6}, \qquad i = 1, \dots, 6,$$

porazdelitvena funkcija je stopničasta funkcija, ki ima vrednost 0 za $x < 1$ in vrednost 1 za $x \geq 6$. Grafa verjetnostne in porazdelitvene funkcije prikazujemo na sliki 4.1.




Slika 4.1: Verjetnostna funkcija $p_X(x_i)$ in porazdelitvena funkcija $F_X(x)$

**Primer 4.3:** *Vzemimo, da slučajna spremenljivka $X$ predstavlja število avtomobilov, ki se zaradi rdeče luči semaforja ustavijo v določenem križišču. Opazovanja so pokazala, da je ustavljenih avtomobilov med 0 in 7. Tudi verjetnosti teh dogodkov so ocenili na podlagi opazovanj:*

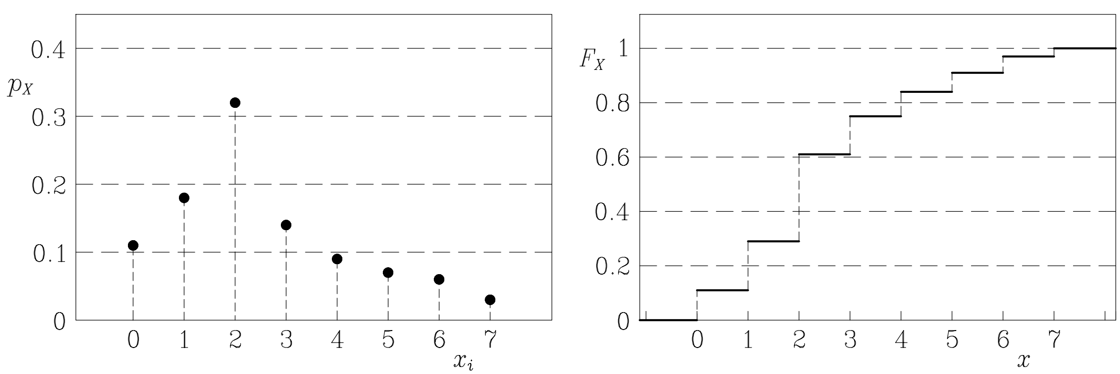
$$p_X(x_i) = \begin{cases} 0.11 & x = 0 \\ 0.18 & x = 1 \\ 0.32 & x = 2 \\ 0.14 & x = 3 \\ 0.09 & x = 4 \\ 0.07 & x = 5 \\ 0.06 & x = 6 \\ 0.03 & x = 7. \end{cases}$$

*Določimo porazdelitveno funkcijo te slučajne spremenljivke in narišimo graf verjetnostne in porazdelitvene funkcije.*

**Rešitev:** Določimo najprej vrednosti porazdelitvene funkcije po enačbi (4.3):

$$F_X(x) = \begin{cases} 0.00 & x < 0 \\ 0.11 & 0 \le x < 1 \\ 0.29 & 1 \le x < 2 \\ 0.61 & 2 \le x < 3 \\ 0.75 & 3 \le x < 4 \\ 0.84 & 4 \le x < 5 \\ 0.91 & 5 \le x < 6 \\ 0.97 & 6 \le x < 7 \\ 1.00 & 7 \le x. \end{cases}$$

Grafa verjetnostne in porazdelitvene funkcije prikazujemo na sliki 4.2. S slike 4.2 lahko na primer preberemo, da je verjetnost, da slučajna spremenljivka $X$ zavzame vrednost 2, enaka 0.32 ($P[X = 2] = p_X(2) = 0.32$) in da je verjetnost, da je slučajna spremenljivka $X$ manjša od 2.6, enaka 0.61 ($P[X \le 2.6] = F_X(2.6) = 0.61$).




Slika 4.2: Verjetnostna funkcija $p_X(x_i)$ in porazdelitvena funkcija $F_X(x)$

## 4.2 Zvezne slučajne spremenljivke

Porazdelitveno funkcijo $F_X(x)$ zvezne slučajne spremenljivke lahko definiramo na enak način kot v enačbi (4.2), ki smo jo zapisali za diskretne slučajne spremenljivke:

$$F_X(x) = P[X \le x].$$

Zaloga vrednosti zvezne slučajne spremenljivke je omejeno ali neomejeno območje oziroma območja števil. Zato je verjetnost, da slučajna spremenljivka $X$ zavzame eno točno določeno vrednost, enaka nič,

$$P[X = x] = 0,$$

in verjetnostne funkcije ne moremo definirati kot v primeru diskretnih slučajnih spremenljivk.

Verjetnost, da slučajna spremenljivka zavzame vrednost znotraj nekega območja, je lahko različna od nič. Izrazimo jo s porazdelitveno funkcijo $F_X(x)$, kot smo izpeljali v primeru 4.1 (enačba (4.5)):

$$P\left[x - \frac{\Delta x}{2} < X \leq x + \frac{\Delta x}{2}\right] = F_X\left(x + \frac{\Delta x}{2}\right) - F_X\left(x - \frac{\Delta x}{2}\right).$$

Če zadnjo enačbo delimo z $\Delta x$ in limitiramo z $\Delta x \to 0$, lahko zapišemo

$$\lim_{\Delta x \to 0} \frac{F_X\left(x + \frac{\Delta x}{2}\right) - F_X\left(x - \frac{\Delta x}{2}\right)}{\Delta x} = \frac{dF_X(x)}{dx} = F'_X(x) = f_X(x), \tag{4.6}$$

kjer odvod porazdelitvene funkcije označimo s $f_X(x)$ in imenujemo **gostota verjetnosti** ali tudi funkcija gostote verjetnosti (angl. probability density function – PDF). Iz enačbe (4.6) lahko ob upoštevanju enačbe (4.4) izračunamo obratno zvezo med porazdelitveno funkcijo in gostoto verjetnosti:

$$dF_X(x) = f_X(x)\,dx \quad \to \quad \int\limits_{F_X(-\infty)}^{F_X(x)} dF_X = \int\limits_{-\infty}^{x} f_X(\tilde{x})\,d\tilde{x}.$$

Za spodnjo mejo integriranja v zadnji enačbi smo izbrali $-\infty$, ker vemo, da gre vrednost porazdelitvene funkcije za $x \to -\infty$ proti nič. Sedaj lahko zapišemo končno enačbo

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(\tilde{x})\,d\tilde{x}. \tag{4.7}$$

Analogno s pogoji, ki jim mora ustrezati verjetnostna funkcija, lahko na osnovi (4.4) in (4.7) zapišemo pogoje, ki jim mora ustrezati gostota verjetnosti.

1. Iz 1. aksioma sledi, da je $f_X(x)$ večja ali enaka nič,

   $$0 \leq f_X(x).$$

2. Verjetnost, da slučajna spremenljivka zavzame poljubno vrednost, je enaka verjetnosti vsote vseh možnih dogodkov, ki jo določimo z integriranjem:

   $$\int\limits_{-\infty}^{\infty} f_X(x)\,dx = 1. \tag{4.8}$$

Podajmo še zvezo med verjetnostjo $P[a < X \leq b]$ in gostoto verjetnosti ter porazdelitveno funkcijo,

$$P[a < X \leq b] = \int\limits_{a}^{b} f_X(x)\,dx = F_X(b) - F_X(a),$$

kar sledi iz enačb (4.5) in (4.7).

**Primer 4.4:** *Določimo gostoto verjetnosti enakomerno porazdeljene slučajne spremenljivke na območju od $a$ do $b$. Narišimo še graf gostote verjetnosti, določimo porazdelitveno funkcijo in narišimo graf porazdelitvene funkcije.*

**Rešitev:** Gostoto verjetnosti enakomerno porazdeljene slučajne spremenljivke lahko zapišemo z enačbo

$$f_X(x) = \begin{cases} 0 & x < a \\ c & a \leq x < b \\ 0 & b \leq x. \end{cases} \tag{4.9}$$

Konstanto $c$ v zadnji enačbi določimo iz 2. lastnosti za gostoto verjetnosti, da je integral gostote verjetnosti po celotnem območju enak ena:

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{-\infty}^{a} 0\,dx + \int_{a}^{b} c\,dx + \int_{b}^{\infty} 0\,dx = \int_{a}^{b} c\,dx = c(b-a) = 1.$$

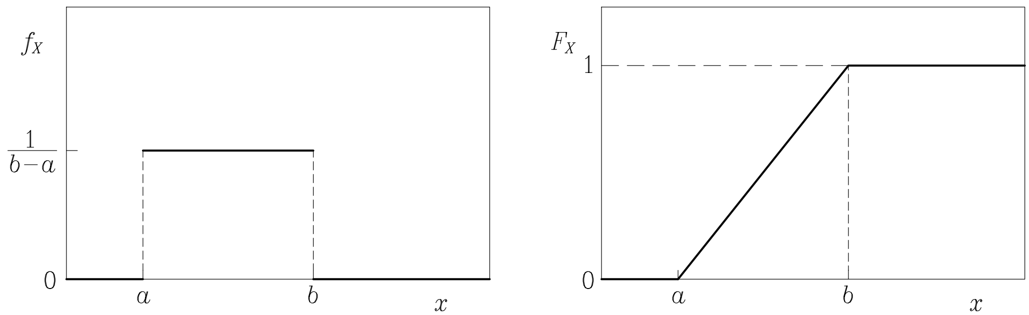
Iz zadnje enačbe sledi, da je $c = 1/(b-a)$. Gostoto verjetnosti lahko sedaj zapišemo takole:

$$f_X(x) = \begin{cases} 0 & x < a \\ \dfrac{1}{b-a} & a \leq x < b \\ 0 & b \leq x. \end{cases}$$

Porazdelitveno funkcijo določimo iz enačbe (4.7)

$$F_X(x) = \begin{cases} \displaystyle\int_{-\infty}^{x} 0\,d\tilde{x} = 0 & x < a \\ \displaystyle\int_{-\infty}^{a} 0\,d\tilde{x} + \int_{a}^{x} \frac{1}{b-a} d\tilde{x} = \frac{x-a}{b-a} & a \leq x < b \\ \displaystyle\int_{-\infty}^{a} 0\,d\tilde{x} + \int_{a}^{b} \frac{1}{b-a} d\tilde{x} + \int_{b}^{x} 0\,d\tilde{x} = 1 & b \leq x. \end{cases} \tag{4.10}$$

Grafa gostote verjetnosti in porazdelitvene funkcije za enakomerno porazdeljeno slučajno spremenljivko podajamo na sliki 4.3.




Slika 4.3: Gostota verjetnosti in porazdelitvena funkcija

## 4.3 Slučajni vektorji

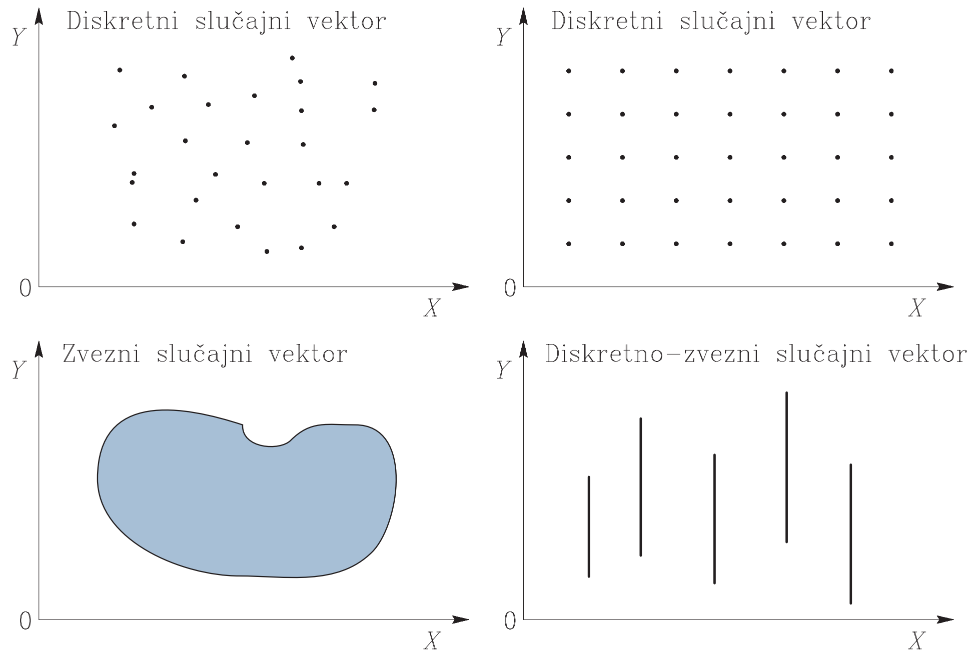
Če hkrati obravnavamo dve ali več slučajnih spremenljivk, jih lahko obravnavamo kot **slučajni vektor**. Slučajni vektor $\mathbf{X}$ $n$ slučajnih spremenljivk zapišemo takole:

$$\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$$

Tako kot slučajna spremenljivka je lahko tudi slučajni vektor diskreten ali zvezen. Možno je tudi, da je nekaj količin v slučajnem vektorju zveznih, druge pa so diskretne slučajne spremenljivke.

Porazdelitveni zakon slučajnega vektorja podajamo podobno kot porazdelitev slučajnih spremenljivk z verjetnostno funkcijo, porazdelitveno funkcijo ali z gostoto verjetnosti. V nadaljevanju bomo obravnavali slučajne vektorje dveh spremenljivk $X, Y$.

Zalogo vrednosti za slučajne vektorje z dvema spremenljivkama lahko grafično prikažemo v ravnini. V primeru, da je slučajna spremenljivka diskretna, predstavljajo zalogo vrednosti točke v ravnini, če pa je slučajna spremenljivka zvezna, zalogo vrednosti predstavlja območje v ravnini.




Slika 4.4: Zaloge vrednosti različnih slučajnih vektorjev

### 4.3.1 Diskretni slučajni vektor

Diskretni slučajni vektor, ki ga sestavljata slučajni spremenljivki $X$ in $Y$ lahko opišemo z verjetnostno funkcijo, ki je definirana kot verjetnost, da ima slučajna spremenljivka $X$ vrednost $x_i$, $i = 1, \ldots, n_X$, in slučajna spremenljivka $Y$ vrednost $y_j$, $j = 1, \ldots, n_Y$,

$$p_{XY}(x_i, y_j) = P[(X = x_i) \cap (Y = y_j)].$$

Podobno kot pri verjetnostni funkciji ene slučajne spremenljivke mora tudi verjetnostna funkcija slučajnega vektorja ustrezati pogojem aksiomov verjetnosti.

1. Verjetnost posameznega dogodka je nenegativna in manjša ali enaka ena (1. aksiom),

$$0 \leq p_{XY}(x_i, y_j) \leq 1. \tag{4.11}$$

2. Ker so dogodki, da slučajni vektor zavzame določene vrednosti, nezdružljivi in je vsota dogodkov, da slučajni vektor zavzame poljubno vrednost iz zaloge vrednosti, gotov dogodek, sledi po 2. in 3. aksiomu, da je vsota vseh $p_{XY}(x_i, y_j)$ enaka ena,

$$\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} p_{XY}(x_i, y_j) = 1. \tag{4.12}$$

Porazdelitveno funkcijo definiramo kot verjetnost, da je slučajna spremenljivka $X$ manjša od vrednosti $x$ in slučajna spremenljivka $Y$ manjša od vrednosti $y$,

$$F_{XY}(x,y) = P[(X \leq x) \cap (Y \leq y)]. \tag{4.13}$$

Ker je verjetnost, da je $X$ ali $Y$ manjša od minus neskončno, enaka nič in je verjetnost, da sta obe manjši od neskončno, enaka ena, sledijo naslednje lastnosti porazdelitvene funkcije:

$$\begin{aligned} \lim_{x\to-\infty} F_{XY}(x,y) &= P[X < -\infty \cap Y \leq y] = 0, \\ \lim_{y\to-\infty} F_{XY}(x,y) &= P[X \leq x \cap Y < -\infty] = 0, \\ \lim_{\substack{x\to\infty \\ y\to\infty}} F_{XY}(x,y) &= P[X < \infty \cap Y < \infty] = 1. \end{aligned} \tag{4.14}$$

Zvezo med verjetnostno in porazdelitveno funkcijo izpeljemo iz definicije porazdelitvene funkcije (4.13) in z upoštevanjem, da so dogodki, da slučajna spremenljivka zavzame določeno vrednost, nezdružljivi:

$$F_{XY}(x,y) = P[(X \leq x) \cap (Y \leq y)] = \sum_{\substack{x_i \leq x \\ y_j \leq y}} p_{XY}(x_i, y_j). \tag{4.15}$$

**Robne porazdelitve**

Robna ali delna porazdelitev slučajnega vektorja $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$ je porazdelitev poljubnega dela vektorja, na primer porazdelitev slučajnega vektorja $\{X_1, X_2\}$, vektorja $\{X_1, X_2, \ldots, X_{n-1}\}$ ali pa porazdelitev ene od slučajnih spremenljivk $X_i$.

Robna porazdelitev slučajnega vektorja $X, Y$ je porazdelitev ene od slučajnih spremenljivk, ki sestavljata slučajni vektor. Robni verjetnostni funkciji slučajnega vektorja $X, Y$ sta $p_X(x_i)$ in $p_Y(y_j)$, robni porazdelitveni funkciji pa sta $F_X(x)$ in $F_Y(y)$.

Robno verjetnostno funkcijo $p_X(x_i)$ določimo tako, da dogodek $X = x_i$ zapišemo kot vsoto nezdružljivih dogodkov:

$$\begin{aligned} p_X(x_i) = P[X = x_i] &= P[\bigcup_{j=1}^{n_Y}((X = x_i) \cap (Y = y_j))] \\ &= \sum_{j=1}^{n_Y} P[(X = x_i) \cap (Y = y_j)] = \sum_{j=1}^{n_Y} p_{XY}(x_i, y_j). \end{aligned} \tag{4.16}$$

Robna verjetnostna funkcija $p_Y(y_j)$ je

$$p_Y(y_j) = \sum_{i=1}^{n_X} p_{XY}(x_i, y_j). \tag{4.17}$$

Na podoben način izpeljemo tudi enačbo za robno porazdelitveno funkcijo. Upoštevamo, da je produkt poljubnega dogodka z gotovim dogodkom enak samemu dogodku. Gotov dogodek, da je $Y$ manjši od neskončno, pomnožimo z dogodkom $X \leq x$ in dobimo:

$$F_X(x) = P[X \leq x] = P[(X \leq x) \cap (Y < \infty)] = \lim_{y \to \infty} F_{XY}(x, y). \tag{4.18}$$

Robna porazdelitvena funkcija $F_Y(y)$ je

$$F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y). \tag{4.19}$$

Sedaj je morda razumljivejši tudi izraz **robna porazdelitev**, saj robni porazdelitveni funkciji $F_X(x)$ in $F_Y(y)$ predstavljata vrednosti porazdelitvene funkcije $F_{XY}(x, y)$, ko gre $x$ oziroma $y$ proti neskončnosti, kar pravzaprav predstavlja robova ravnine.

### Pogojne porazdelitve

Vzemimo, da nas zanima verjetnost, da je $X = x_i$ ob pogoju, da je $Y = y_j$. To opišemo s pogojno verjetnostjo, s katero definiramo tudi pogojno verjetnostno funkcijo:

$$P[X = x_i | Y = y_j] = \frac{P[X = x_i \cap Y = y_j]}{P[Y = y_j]} = \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)} = p_{X|Y}(x_i, y_j). \tag{4.20}$$

Podobno lahko zapišemo tudi obrnjeno pogojno verjetnostno funkcijo:

$$P[Y = y_j | X = x_i] = \frac{p_{XY}(x_i, y_j)}{p_X(x_i)} = p_{Y|X}(x_i, y_j). \tag{4.21}$$

Slučajni spremenljivki $X$ in $Y$ sta medsebojno neodvisni, če je pogojna verjetnostna funkcija enaka robni,

$$p_{X|Y}(x_i, y_j) = \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)} = p_X(x_i) \tag{4.22}$$

oziroma

$$p_{Y|X}(x_i, y_j) = \frac{p_{XY}(x_i, y_j)}{p_X(x_i)} = p_Y(y_j). \tag{4.23}$$

Iz enačb (4.22) in (4.23) sledi pomembna lastnost slučajnih vektorjev, sestavljenih iz neodvisnih slučajnih spremenljivk, ki pravi, da je verjetnostna funkcija slučajnega vektorja enaka produktu robnih verjetnostnih funkcij,

$$p_{XY}(x_i, y_j) = p_X(x_i)\, p_Y(y_j).$$

V tem primeru velja tudi, da je porazdelitvena funkcija $F_{XY}(x, y)$ produkt robnih porazdelitvenih funkcij $F_X(x)$ in $F_Y(y)$,

$$F_{XY}(x, y) = F_X(x)\, F_Y(y).$$

**Primer 4.5:** *Avtomatski števec prometa šteje vozila, ki se v časovnem intervalu 3 sekund peljejo mimo števca po določeni cesti. Pretekla opazovanja so pokazala, da dejansko število vozil ni enako izmerjenemu. Vzemimo, da je dejansko število vozil slučajna spremenljivka $X$, izmerjeno število vozil pa slučajna spremenljivka $Y$. Verjetnostno funkcijo slučajnega vektorja $X, Y$ podamo s preglednico 4.1.*

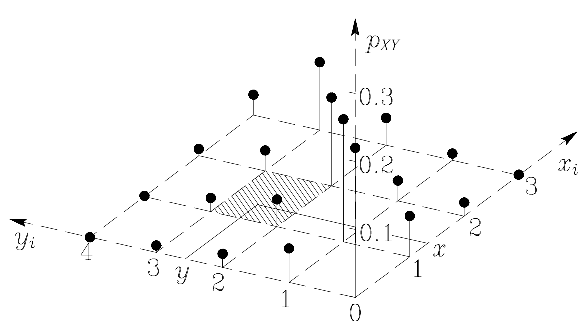
Preglednica 4.1: Verjetnostna funkcija slučajnega vektorja $X, Y$

| | $x_1 = 0$ | $x_2 = 1$ | $x_3 = 2$ | $x_4 = 3$ |
|---|---|---|---|---|
| $y_1 = 0$ | 0.22 | 0.06 | 0.02 | 0.00 |
| $y_2 = 1$ | 0.05 | 0.18 | 0.03 | 0.01 |
| $y_3 = 2$ | 0.02 | 0.04 | 0.13 | 0.04 |
| $y_4 = 3$ | 0.01 | 0.02 | 0.03 | 0.10 |
| $y_5 = 4$ | 0.00 | 0.00 | 0.01 | 0.03 |

*Preverimo, ali vrednosti v preglednici 4.1 ustrezajo pogojem verjetnostne funkcije. Narišimo graf verjetnostne funkcije $p_{XY}(x_i, y_j)$ v aksonometrični projekciji. Določimo še robni verjetnostni funkciji $p_X(x)$ in $p_Y(y)$. Določimo pogojni verjetnostni funkciji $p_{X|Y}(x_i, y_j)$ in $p_{Y|X}(x_i, y_j)$. Izračunajmo vrednosti porazdelitvene funkcije $F_{XY}(x, y)$, narišimo njen graf v aksonometrični projekciji. Izračunajmo verjetnost, da je avtomatični števec pravilno izmeril število avtov $P[X = Y]$. Izračunajmo tudi verjetnost, da je bila napaka števca večja od ena $P[|X - Y| > 1]$.*

**Rešitev:** Zelo primeren računalniški program za reševanje te naloge je EXCEL. V skupini preglednic 4.2 prikazujemo reševanje naloge, kot smo ga izvedli s tem računalniškim programom.

Verjetnostno funkcijo prikazujemo na sliki 4.5.




Slika 4.5: Verjetnostna funkcija $p_{XY}(x_i, y_j)$

V preglednici *Verjetnostna funkcija* skupine preglednic 4.2 smo izračunali, da je vsota vseh vrednosti verjetnostne funkcije $p_{XY}(x_i, y_i)$ enaka 1. Vidimo tudi, da so vse vrednosti pozitivne, zato lahko zaključimo, da funkcija $p_{XY}$ ustreza pogojem za verjetnostno funkcijo (pogoja (4.11) in (4.12)). Tudi robni porazdelitveni funkciji smo izračunali v isti preglednici po enačbi (4.16) oziroma (4.17):

$$p_X(x_i) = \begin{cases} 0.30 & x = 0 \\ 0.30 & x = 1 \\ 0.22 & x = 2 \\ 0.18 & x = 3 \end{cases}$$

in

$$p_Y(y_i) = \begin{cases} 0.30 & y = 0 \\ 0.27 & y = 1 \\ 0.23 & y = 2 \\ 0.16 & y = 3 \\ 0.04 & y = 4. \end{cases}$$

Preglednica 4.2: Določitev porazdelitvene in pogojnih verjetnostnih funkcij

Verjetnostna funkcija $p_{XY}(x_i, y_j)$ ter robni verjetnostni funkciji $p_X(x_i)$ in $p_Y(y_j)$

| | 0 | 1 | 2 | 3 | $p_Y$ |
|---|---|---|---|---|---|
| 0 | 0.22 | 0.06 | 0.02 | 0.00 | 0.30 |
| 1 | 0.05 | 0.18 | 0.03 | 0.01 | 0.27 |
| 2 | 0.02 | 0.04 | 0.13 | 0.04 | 0.23 |
| 3 | 0.01 | 0.02 | 0.03 | 0.10 | 0.16 |
| 4 | 0.00 | 0.00 | 0.01 | 0.03 | 0.04 |
| $p_X$ | 0.30 | 0.30 | 0.22 | 0.18 | 1.00 |

Porazdelitvena funkcija $F_{XY}(x, y)$ ter robni porazdelitveni funkciji $F_X(x)$ in $F_Y(y)$

| | $< 0$ | 0–1 | 1–2 | 2–3 | $> 3$ | $F_Y$ |
|---|---|---|---|---|---|---|
| $< 0$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0–1 | 0.00 | 0.22 | 0.28 | 0.30 | 0.30 | 0.30 |
| 1–2 | 0.00 | 0.27 | 0.51 | 0.56 | 0.57 | 0.57 |
| 2–3 | 0.00 | 0.29 | 0.57 | 0.75 | 0.80 | 0.80 |
| 3–4 | 0.00 | 0.30 | 0.60 | 0.81 | 0.96 | 0.96 |
| $> 4$ | 0.00 | 0.30 | 0.60 | 0.82 | 1.00 | 1.00 |
| $F_X$ | 0.00 | 0.30 | 0.60 | 0.82 | 1.00 | |

Pogojna verjetnostna funkcija $p_{Y|X}(x_i, y_j)$

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 0.73 | 0.20 | 0.09 | 0.00 |
| 1 | 0.17 | 0.60 | 0.14 | 0.06 |
| 2 | 0.07 | 0.13 | 0.59 | 0.22 |
| 3 | 0.03 | 0.07 | 0.14 | 0.56 |
| 4 | 0.00 | 0.00 | 0.05 | 0.17 |
| | 1.00 | 1.00 | 1.00 | 1.00 |

Pogojna verjetnostna funkcija $p_{X|Y}(x_i, y_j)$

| | 0 | 1 | 2 | 3 | |
|---|---|---|---|---|---|
| 0 | 0.73 | 0.20 | 0.07 | 0.00 | 1.00 |
| 1 | 0.19 | 0.67 | 0.11 | 0.04 | 1.00 |
| 2 | 0.09 | 0.17 | 0.57 | 0.17 | 1.00 |
| 3 | 0.06 | 0.13 | 0.19 | 0.63 | 1.00 |
| 4 | 0.00 | 0.00 | 0.25 | 0.75 | 1.00 |

Pogojni porazdelitveni funkciji izračunamo v preglednicah *Pogojna verjetnostna funkcija* $p_{X|Y}(x_i, y_j)$ in *Pogojna verjetnostna funkcija* $p_{Y|X}(x_i, y_j)$ v skupini preglednic 4.2 po enačbah (4.20) in (4.21). Na primer: pogojna verjetnostna funkcija $p_{X|Y=2}(x_i)$ je

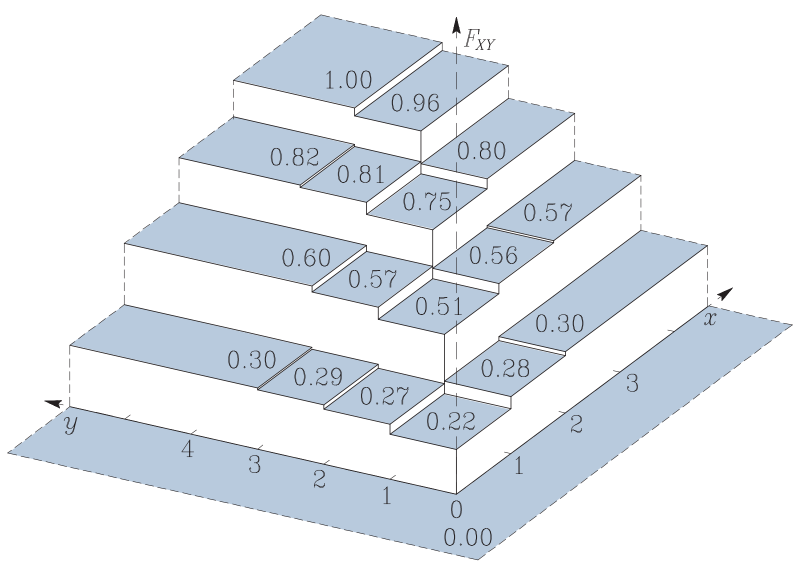
$$p_{X|Y=2}(x_i) = \begin{cases} 0.09 & x = 0 \\ 0.17 & x = 1 \\ 0.57 & x = 2 \\ 0.17 & x = 3, \end{cases}$$

kar predstavlja tretjo vrstico preglednice *Pogojna verjetnostna funkcija* $p_{X|Y}(x_i, y_j)$ v skupini preglednic 4.2.

V preglednici *Porazdelitvena funkcija* v skupini preglednic 4.2 smo izračunali porazdelitveno funkcijo $F_{XY}(x, y)$ po enačbi (4.15). Obravnavajmo poljubno točko $(x, y)$ v kvadratnem območju, ki je označeno na sliki 4.5. Verjetnost, da je $X \leq x$ in $Y \leq y$, je vsota vseh verjetnosti $p_{XY}(x_i, y_j)$, za katere velja, da je $x_i \leq x$ in $y_j \leq y$. V tem primeru je

$$\begin{aligned} F_{XY}(x, y) &= P[X \leq x \cap Y \leq y] \\ &= p_{XY}(0,0) + p_{XY}(0,1) + p_{XY}(0,2) + p_{XY}(1,0) + p_{XY}(1,1) + p_{XY}(1,2) \\ &= 0.22 + 0.05 + 0.02 + 0.06 + 0.18 + 0.04 = 0.57. \end{aligned}$$

Graf te funkcije v aksonometrični projekciji prikazujemo na sliki 4.6.




Slika 4.6: Porazdelitvena funkcija $F_{XY}(x, y)$

Iz preglednice 4.1 lahko izpišemo verjetnosti, da je števec pravilno izmeril število vozil:

$$\begin{aligned} P[X = Y] &= P[(X=0 \cap Y=0) \cup (X=1 \cap Y=1) \cup (X=2 \cap Y=2) \cup (X=3 \cap Y=3)] \\ &= P[X=0 \cap Y=0] + P[X=1 \cap Y=1] + P[X=2 \cap Y=2] + P[X=3 \cap Y=3] \\ &= 0.22 + 0.18 + 0.13 + 0.10 = 0.63. \end{aligned}$$

Na podoben način izračunamo tudi verjetnost, da se je števec zmotil za več kot ena:

$$
\begin{aligned}
P[|X-Y|>1] &= P[(X=0\cap Y=2)\cup(X=0\cap Y=3)\cup(X=0\cap Y=4)\\
&\quad\cup(X=1\cap Y=3)\cup(X=1\cap Y=4)\cup(X=2\cap Y=0)\\
&\quad\cup(X=2\cap Y=4)\cup(X=3\cap Y=0)\cup(X=3\cap Y=1)]\\
&= P[X=0\cap Y=2]+P[X=0\cap Y=3]+P[X=0\cap Y=4]\\
&\quad+P[X=1\cap Y=3]+P[X=1\cap Y=4]+P[X=2\cap Y=0]\\
&\quad+P[X=2\cap Y=4]+P[X=3\cap Y=0]+P[X=3\cap Y=1]\\
&= 0.02+0.01+0.00+0.02+0.00+0.02+0.01+0.00+0.01=0.09.
\end{aligned}
$$

### 4.3.2 Zvezni slučajni vektorji

Porazdelitveno funkcijo $F_{XY}(x,y)$ definiramo enako kot za diskretne slučajne vektorje (enačba (4.13))

$$F_{XY}(x,y)=P[(X\le x)\cap(Y\le y)] \tag{4.24}$$

in tudi lastnosti, opisane z enačbami (4.14) so enake.

Ker je verjetnost, da zvezni slučajni vektor zavzame točno določeno vrednost, enaka nič,

$$P[(X=x)\cap(Y=y)]=0,$$

podobno kot pri zvezni slučajni spremenljivki ne moremo definirati verjetnostne funkcije, temveč definiramo gostoto verjetnosti $f_{XY}(x,y)$. Ta funkcija mora ustrezati pogojema, ki sledita iz aksiomov verjetnostnega računa.

1. Verjetnost posameznega dogodka je nenegativna (1. aksiom),

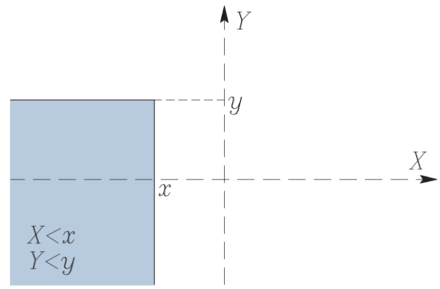
   $$0\le f_{XY}(x,y).$$

2. Ker so dogodki, da slučajni vektor zavzame določene vrednosti, nezdružljivi in je vsota dogodkov, da slučajni vektor zavzame poljubno vrednost iz zaloge vrednosti, gotov dogodek, sledi po 2. in 3. aksiomu, da je integral funkcije $f_{XY}(x,y)$ po celotni ravnini enak ena:

   $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dx\,dy=1.$$

Zvezo med gostoto verjetnosti in porazdelitveno funkcijo slučajnega vektorja izpeljemo iz enačbe (4.24). Na sliki 4.7 prikazujemo območje, ki mu ustreza pogoj $(X\le x)\cup(Y\le y)$ in ki predstavlja območje, po katerem moramo integrirati gostoto verjetnosti,

$$F_{XY}(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(\tilde{x},\tilde{y})\,d\tilde{x}\,d\tilde{y}. \tag{4.25}$$

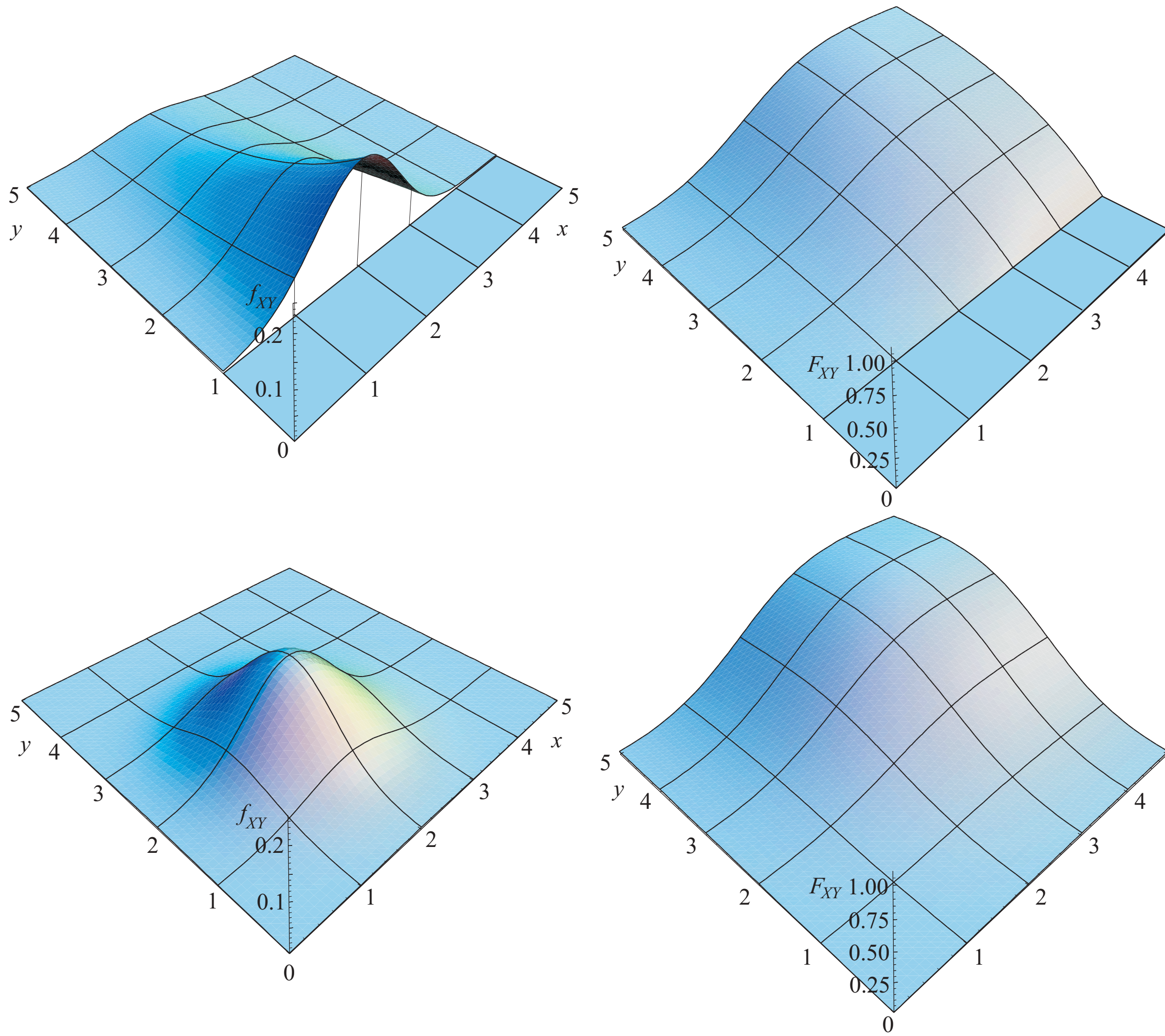



Slika 4.7: Območje integriranja pri določitvi porazdelitvene funkcije

Obratno zvezo dobimo z odvajanjem porazdelitvene funkcije

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x\, \partial y}.$$

Dva primera gostote verjetnosti in ustrezne porazdelitvene funkcije prikazujemo na sliki 4.8.




Slika 4.8: Primer gostote verjetnosti in porazdelitvene funkcije

### Robne porazdelitve

Robno porazdelitev dobimo na podoben način kot pri diskretnih slučajnih vektorjih. Robno gostoto verjetnosti določimo z integriranjem ene izmed slučajnih spremenljivk, ki sestavljajo slučajni vektor:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dy, \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y)dx. \end{aligned} \tag{4.26}$$

Robne porazdelitvene funkcije določimo na enak način kot pri diskretnih slučajnih spremenljivkah (enačbi (4.18) in (4.19)):

$$F_X(x) = \lim_{y\to\infty} F_{XY}(x,y), \qquad F_Y(y) = \lim_{x\to\infty} F_{XY}(x,y). \tag{4.27}$$

Enačbi (4.26) lahko izpeljemo tudi iz enačb (4.27) ter zveze med porazdelitveno funkcijo $F_{XY}(x,y)$ in gostoto verjetnosti $f_{XY}(x,y)$ (4.25). Porazdelitveno funkcijo $F_X(x)$,

$$F_X(x) = \lim_{y\to\infty} F_{XY}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{XY}(\tilde{x},\tilde{y})\,d\tilde{x}\,d\tilde{y},$$

odvajamo po $x$ in dobimo

$$f_X(x) = \frac{dF_X(x)}{dx} = \int_{-\infty}^{\infty} f_{XY}(x,\tilde{y})\ d\tilde{y}.$$

Na enak način izpeljemo tudi enačbo za robno gostoto verjetnosti $f_Y(y)$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(\tilde{x},y)\ d\tilde{x}.$$

### Pogojne porazdelitve

Pogojno gostoto verjetnosti $f_{X|Y=y}$ definiramo kot gostoto slučajnega vektorja $f_{XY}$ pri neki določeni oziroma izbrani vrednosti $Y = y$,

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{C},$$

kjer je $C$ konstanta, s katero moramo deliti vrednosti $f_{XY}(x,y)$, če želimo, da bo funkcija $f_{X|Y=y}(x)$ ustrezala pogoju za gostoto verjetnosti (4.8):

$$\int_{-\infty}^{\infty} f_{X|Y=y}(x)\,dx = \int_{-\infty}^{\infty} \frac{f_{XY}(x,y)}{C}\,dx = 1 \qquad \rightarrow \qquad C = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = f_Y(y).$$

Sedaj lahko zapišemo enačbo za določitev pogojne gostote verjetnosti $f_{X|Y=y}(x)$,

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)}, \tag{4.28}$$

in na podoben način še za pogojno gostoto verjetnosti $f_{Y|X=x}(y)$,

$$f_{Y|X=x}(y) = \frac{f_{XY}(x,y)}{f_X(x)}. \tag{4.29}$$

Vidimo, da sta enačbi (4.28) in (4.29) analogni enačbama za pogojno verjetnostno funkcijo (4.20) in (4.21).

Slučajni spremenljivki $X$ in $Y$ sta neodvisni, če je pogojna gostota verjetnosti enaka robni gostoti verjetnosti,

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

Če ta pogoj vstavimo v enačbi (4.28) ali (4.29), lahko zapišemo pogoj za neodvisnost dveh slučajnih spremenljivk tudi takole:

$$f_{XY}(x,y) = f_X(x)\, f_Y(y). \tag{4.30}$$

**Primer 4.6:** *Vodoravno razdaljo od epicentra potresa označimo z $R$, magnitudo potresa pa z $M$. Vzemimo, da je gostota verjetnosti razdalje $R$ linearna funkcija,*

$$f_R(r) = a\, r \qquad 0 \le r \le r_0,$$

*gostota verjetnosti magnitude potresa $M$ pa je kvadratna enačba,*

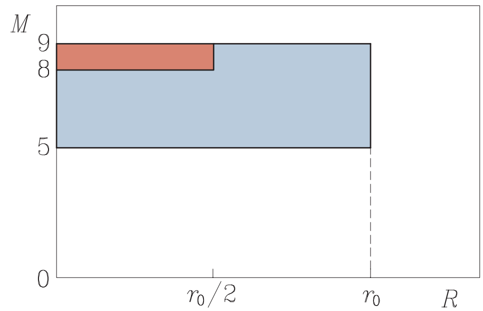
$$f_M(m) = b\,(9-m)^2 \qquad 5 \le m \le 9.$$

*Predpostavimo lahko, da je razdalja do epicentra potresa neodvisna od njegove magnitude. Določimo konstanti $a$ in $b$. Določimo še gostoto verjetnosti slučajnega vektorja $R, M$ in verjetnost, da se zgodi katastrofa: magnituda potresa je večja od $8$, epicenter pa bližje kot $r_0/2$. Narišimo tudi zalogo vrednosti tega slučajnega vektorja in območje, za katerega velja katastrofa.*

**Rešitev:** Konstanti $a$ in $b$ določimo na osnovi enačbe (4.8)

$$\int_{-\infty}^{\infty} f_R(r)\, dr = 1 = \int_0^{r_0} a\, r\, dr = a\frac{r_0^2}{2} \qquad \rightarrow \qquad a = \frac{2}{r_0^2}$$

$$\int_{-\infty}^{\infty} f_M(m)\, dm = 1 = \int_5^9 b\,(9-m)^2 dm = -b\, \frac{(9-m)^3}{3}\bigg|_5^9 = \frac{64\, b}{3} \qquad \rightarrow \qquad b = \frac{3}{64}.$$




Slika 4.9: Zaloga vrednosti slučajnega vektorja $R, M$

Gostoto verjetnosti slučajnega vektorja $R, M$ zaradi neodvisnosti slučajnih spremenljivk $R$ in $M$ izračunamo s produktom posameznih gostot verjetnosti:

$$f_{RM}(r,m) = \frac{3\,r\,(9-m)^2}{32\,r_0^2} \qquad 0 \le r \le r_0, \;\; 5 \le m \le 9.$$

Verjetnost katastrofe izračunamo z integralom (glej sliko 4.9):

$$P[(R < r_0/2) \cap (M \ge 8)] = \int_0^{r_0/2} \int_8^9 \frac{3\,r\,(9-m)^2}{32\,r_0^2} dr\,dm =$$

$$= \frac{3}{32\,r_0^2} \int_0^{r_0/2} r\,dr \int_8^9 (9-m)^2 dm = \frac{3}{32\,r_0^2}\, \frac{r^2}{2}\Big|_0^{r_0/2} \left[ -\frac{(9-m)^3}{3}\Big|_8^9 \right] = \frac{1}{256} = 0.00391.$$

**Primer 4.7:** *Podana je gostota verjetnosti slučajnega vektorja $X, Y$:*

$$f_{XY}(x,y) = \frac{x}{\pi a^2} \qquad 0 \le x \le a, \;\; 0 \le y \le 2\,\pi.$$

*Določimo pogojni verjetnosti $f_{X|Y}$ in $f_{Y|X}$. Ali sta slučajni spremenljivki $X$ in $Y$ neodvisni?*

**Rešitev:** Pogojni gostoti verjetnosti $f_{X|Y}$ in $f_{Y|X}$ določimo po enačbah (4.28) in (4.29). Zato moramo najprej izračunati robni gostoti verjetnosti po enačbah (4.26)

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_0^{2\pi} \frac{x}{\pi\,a^2}\,dy = \frac{2\,x}{a^2} \qquad 0 \le x \le a \tag{4.31}$$

in

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_0^{a} \frac{x}{\pi\,a^2}\,dx = \frac{1}{2\,\pi} \qquad 0 \le y \le 2\,\pi. \tag{4.32}$$

Zadnji dve enačbi vstavimo v enačbi (4.28)–(4.29) in dobimo pogojni gostoti verjetnosti

$$f_{X|Y}(x) = \frac{x\,2\,\pi}{\pi\,a^2} = \frac{2\,x}{a^2} = f_X(x) \qquad 0 \le x \le a$$

in

$$f_{Y|X}(y) = \frac{x\,a^2}{\pi\,a^2\,2\,x} = \frac{1}{2\,\pi} = f_Y(y) \qquad 0 \le y \le 2\,\pi.$$

Vidimo tudi, da velja preprosta zveza med robnima in skupno gostoto verjetnosti

$$f_{XY}(x,y) = f_X(x)\,f_Y(y).$$

Ker sta pogojni gostoti verjetnosti enaki robnima gostotama verjetnosti, lahko zaključimo, da sta slučajni spremenljivki $X$ in $Y$ neodvisni. Do istega zaključka lahko pridemo tudi ob ugotovitvi, da je skupna gostota verjetnosti $f_{XY}$ enaka produktu robnih porazdelitev.

**Primer 4.8:** *Slučajni vektor $X, Y$ je enakomerno porazdeljen znotraj kroga z radijem $R$:*

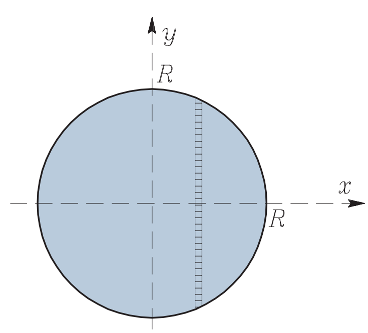
$$f_{XY}(x,y) = \frac{1}{\pi R^2} \qquad x^2 + y^2 \le R^2. \tag{4.33}$$

*Sta slučajni spremenljivki $X$ in $Y$ neodvisni?*

**Rešitev:** Določimo najprej robno gostoto verjetnosti $f_X(x)$ po enačbi (4.26):

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi\,R^2}\,dy = \frac{2}{\pi\,R^2}\sqrt{R^2-x^2} \qquad -R \le x \le R.$$

Zalogo vrednosti slučajnega vektorja $X, Y$ in način integriranja pri določitvi robnih gostot verjetnosti prikazujemo na sliki 4.10.




Slika 4.10: Zaloga vrednosti slučajnega vektorja $X, Y$

Na enak način določimo tudi robno gostoto verjetnosti $f_Y(y)$:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_{-\sqrt{R^2-y^2}}^{\sqrt{R^2-y^2}} \frac{1}{\pi\,R^2}\,dx = \frac{2}{\pi\,R^2}\sqrt{R^2-y^2} \qquad -R \le y \le R.$$

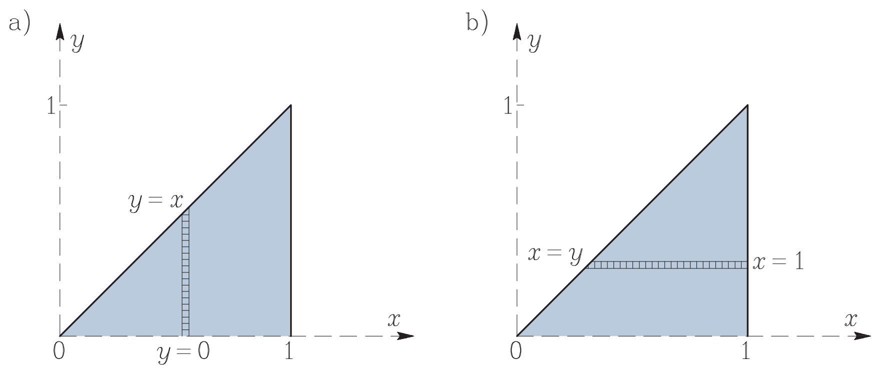
Produkt teh dveh robnih porazdelitev očitno ni enak skupni porazdelitvi. Zato lahko zaključimo, da slučajni spremenljivki $X$ in $Y$ nista neodvisni.

**Primer 4.9:** *Velikost sile na steber je slučajna spremenljivka $X$, velikost sile na isti steber zaradi koristne obtežbe pa označimo z $Y$. Podana je gostota verjetnosti slučajnega vektorja $X, Y$:*

$$f_{XY}(x,y) = 3\,x \qquad 0 \leq x \leq 1, \;\; 0 \leq y \leq x. \tag{4.34}$$

*Narišimo zalogo vrednosti slučajnega vektorja $X, Y$. Ali funkcija $f_{XY}$ ustreza pogojem gostote verjetnosti? Določimo še robni gostoti verjetnosti $f_X$, $f_Y$ in pogojni gostoti verjetnosti $f_{X|Y}$, $f_{Y|X}$. Sta slučajni spremenljivki $X$ in $Y$ neodvisni?*

**Rešitev:** Zalogo vrednosti slučajnega vektorja $X, Y$ predstavlja trikotni del ravnine $x, y$, ki ga prikazujemo na sliki 4.11.




Slika 4.11: Zaloga vrednosti slučajnega vektorja $X, Y$

Da bi ugotovili, ali funkcija ustreza pogojem gostote verjetnosti, moramo preveriti, ali je $f_{XY}$ pozitivna v območju, kjer je definirana, in ali je integral po celotnem območju, kjer je funkcija definirana, enak ena. Ker je v izrazu (4.34) določena zaloga vrednosti le za pozitivne $X$, je funkcija $f_{XY}$ pozitivna. Integral po celotnem območju izračunamo z enačbo (glej sliko 4.11)

$$\int_0^1 \left( \int_0^x 3x\,dy \right) dx = \int_0^1 \left( 3x\,y \Big|_0^x \right) dx = \int_0^1 3x^2 dx = \frac{3x^3}{3} \Big|_0^1 = 1.$$

Integral po celotnem območju je enak ena, zato funkcija $f_{XY}$ ustreza pogojem za gostoto verjetnosti slučajnega vektorja.

Robni porazdelitvi izračunamo z integriranjem

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(xy)\,dy = \int_0^x 3x\,dy = 3x\,y \Big|_0^x = 3x^2 \qquad 0 \leq x \leq 1$$

in (glej sliko 4.11)

$$f_Y(y) = \int\limits_{-\infty}^{\infty} f_{XY}(xy)\,dx = \int\limits_{y}^{1} 3x\,dx = \left.\frac{3x^2}{2}\right|_y^1 = \frac{3}{2}(1-y^2) \qquad 0 \le y \le 1.$$

Sedaj lahko izračunamo pogojni gostoti verjetnosti $f_{X|Y}$ in $f_{Y|X}$ z uporabo enačb (4.28)–(4.29):

$$f_{X|Y}(x,y) = \frac{2x}{1-y^2} \qquad 0 \le x \le 1, \quad 0 \le y \le x,$$

in

$$f_{Y|X}(x,y) = \frac{1}{x} \qquad 0 \le x \le 1, \quad 0 \le y \le x.$$

Ker pogojni gostoti verjetnosti $f_{X|Y}$ in $f_{Y|X}$ nista enaki robnim gostotam verjetnosti $f_X$ in $f_Y$, sta slučajni spremenljivki $X$ in $Y$ med seboj odvisni. Vidimo tudi, da produkt obeh robnih porazdelitev $f_X$ in $f_Y$ ni enak gostoti verjetnosti $f_{XY}$, kar je tudi zadostni pogoj za odvisnost dveh slučajnih spremenljivk.

# 5 Transformacije slučajnih spremenljivk

Vzemimo, da poznamo porazdelitev slučajne spremenljivke $X$. Določiti moramo porazdelitev slučajne spremenljivke $Y$, ki je znana funkcija slučajne spremenljivke $X$

$$Y = g(X). \tag{5.1}$$

Če je zveza (5.1) monotona funkcija, lahko zapišemo tudi obratno zvezo

$$X = g^{-1}(Y), \tag{5.2}$$

kjer je $g^{-1}(Y)$ tudi monotona funkcija.

## 5.1 Diskretne slučajne spremenljivke

Določimo najprej porazdelitev diskretne slučajne spremenljivke $Y$, če poznamo porazdelitev diskretne slučajne spremenljivke $X$ in zvezo med $Y$ in $X$. Verjetnostna funkcija slučajne spremenljivke $Y$ je

$$p_Y(y_i) = P[Y = y_i] = P[g(X) = y_i] = \sum_{g(x_j)=y_i} p_X(x_j) \qquad i = 1, \ldots, n_Y. \tag{5.3}$$

Če je funkcija $g(X)$ monotona, lahko iz enačbe (5.3) izpeljemo

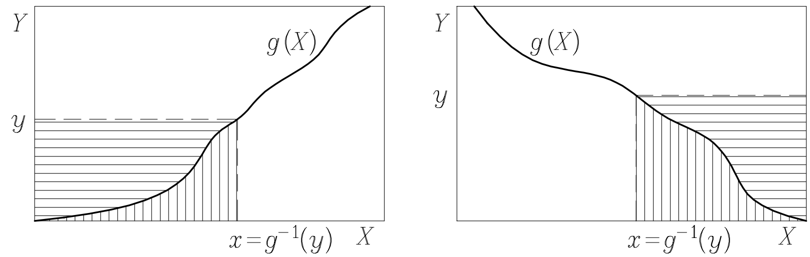
$$p_Y(y_i) = P[Y = y_i] = P[g(X) = y_i] = P[X = g^{-1}(y_i)] = p_X(g^{-1}(y_i)). \tag{5.4}$$

Na podoben način lahko izpeljemo tudi porazdelitveno funkcijo slučajne spremenljivke $Y$, če poznamo porazdelitveno funkcijo slučajne spremenljivke $X$ in zvezo med $X$ in $Y$. Zapišemo enačbo

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = \sum_{g(x_i)\leq y} p_X(x_i).$$

Če je zveza monotono naraščajoča, lahko iz zadnje enačbe izpeljemo še (glej sliko 5.1)

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y)).$$




Slika 5.1: Monotono naraščajoča in monotona padajoča zveza med spremenljivkama $X$ in $Y$

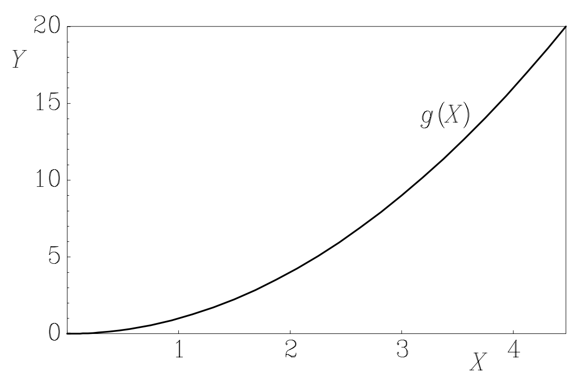
Za monotono padajočo zvezo moramo zapisati nekoliko spremenjeno enačbo (glej sliko 5.1):

$$
\begin{aligned}
F_Y(y) &= P[Y \leq y] = P[g(X) \leq y] = P[X \geq g^{-1}(y)] = 1 - P[X < g^{-1}(y)] \\
&= 1 - F_X(g^{-1}(y)) + P[X = g^{-1}(y)].
\end{aligned}
$$

**Primer 5.1:** *Določimo verjetnostno funkcijo slučajne spremenljivke $Y$, če vemo, da je slučajna spremenljivka $X$ enakomerno porazdeljena med vrednosti* 0, 1, 2, 3, *in* 4 *in je zveza med $Y$ in $X$ kvadratna enačba $Y = X^2$.*

**Rešitev:** Verjetnostno funkcijo slučajne spremenljivke $X$ lahko zapišemo z enačbo

$$p_X(x_i) = 1/5 \qquad x_i = 0, 1, 2, 3, 4. \tag{5.5}$$



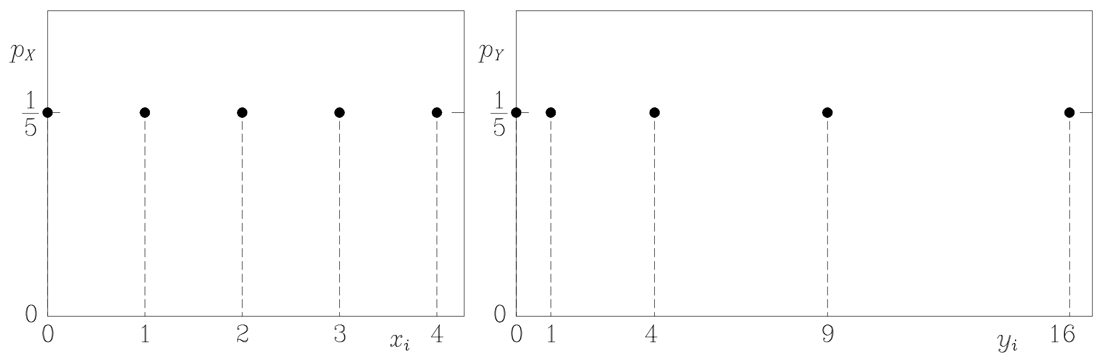

Slika 5.2: Kvadratna zveza med slučajnima spremenljivkama $X$ in $Y$

Na sliki 5.2 lahko vidimo, da je funkcija $g(X)$ na območju, ki sovpada z zalogo vrednosti slučajne spremenljivke $X$, monotona. Zato lahko zapišemo obratno zvezo $X = g^{-1}(Y) = \sqrt{Y}$ in uporabimo enačbo (5.4),

$$p_Y(y_i) = p_X(g^{-1}(y_i)) = p_X(\sqrt{y_i}). \tag{5.6}$$

S slike 5.2 vidimo, da zaloga vrednosti slučajne spremenljivke $Y$ zajema števila 0, 1, 4, 9 in 16. Če upoštevamo zalogo vrednosti $Y$ in enačbi (5.5) in (5.6), lahko zapišemo verjetnostno funkcijo slučajne spremenljivke $Y$,

$$p_Y(y_i) = 1/5 \qquad y_i = 0, 1, 4, 9, 16.$$

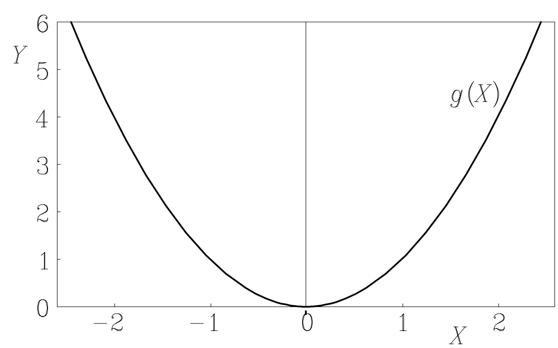



Slika 5.3: Verjetnostni funkciji slučajnih spremenljivk $X$ in $Y$

**Primer 5.2:** *Določimo verjetnostno funkcijo slučajne spremenljivke $Y$, če vemo, da je slučajna spremenljivka $X$ enakomerno porazdeljena med vrednosti $-2$, $-1$, 0, 1, 2 in je zveza med $Y$ in $X$ kvadratna enačba $Y = X^2$.*

**Rešitev:** Verjetnostno funkcijo slučajne spremenljivke $X$ lahko zapišemo z enačbo

$$p_X(x_i) = 1/5 \qquad x_i = -1, -2, 0, 1, 2. \tag{5.7}$$



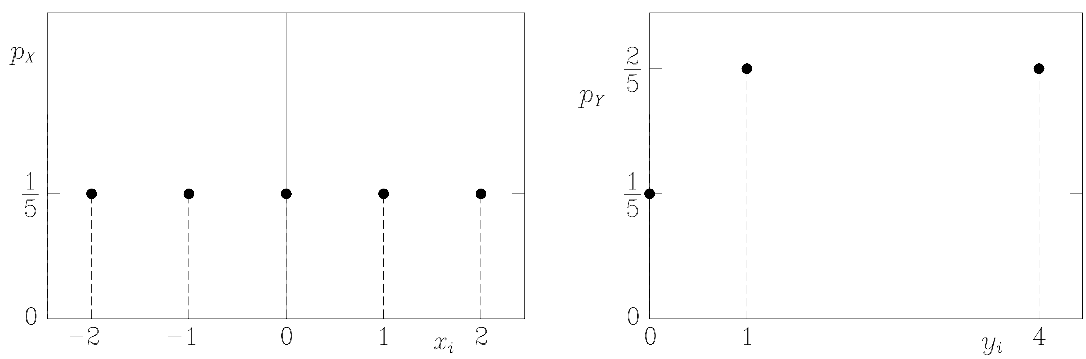

Slika 5.4: Kvadratna zveza med slučajnima spremenljivkama $X$ in $Y$

Na sliki 5.4 lahko vidimo, da zveza $g(X)$ na območju, ki vključuje zalogo vrednosti slučajne spremenljivke $X$, ni monotona, zato ne moremo zapisati enolične obratne zveze tako kot v primeru 5.1. S slike 5.4 lahko ugotovimo, da zaloga vrednosti slučajne spremenljivke $Y$ zajema le števila 0, 1 in 4. Ugotovimo verjetnosti, da slučajna spremenljivka zavzame eno izmed teh vrednosti:

$$
\begin{aligned}
p_Y(0) &= P[Y=0] = P[X^2=0] = P[X=0] = p_X(0) = 1/5,\\
p_Y(1) &= P[Y=1] = P[X^2=1] = P[(X=1)\cup(X=-1)]\\
&= P[X=1] + P[X=-1] = p_X(1) + p_X(-1) = 2/5,\\
p_Y(4) &= P[Y=4] = P[X^2=4] = P[(X=2)\cup(X=-2)]\\
&= P[X=2] + P[X=-2] = p_X(2) + p_X(-2) = 2/5.
\end{aligned}
$$

Verjetnostna funkcija slučajne spremenljivke $Y$ je torej enaka

$$
p_Y(y_i) = \begin{cases} 1/5 & y=0,\\ 2/5 & y=1,\\ 2/5 & y=4. \end{cases}
$$




Slika 5.5: Verjetnostni funkciji slučajnih spremenljivk $X$ in $Y$

## 5.2 Zvezne slučajne spremenljivke

Če poznamo porazdelitev zvezne slučajne spremenljivke $X$ ter zvezo med slučajno spremenljivko $X$ in $Y$, lahko določimo porazdelitev slučajne spremenljivke $Y$. Izpeljimo pravilo za transformacijo zveznih slučajnih spremenljivk, če je funkcija $Y = g(X)$ monotona. V tem primeru obstaja tudi enolična funkcija $X = g^{-1}(Y)$. Zapišimo najprej porazdelitveno funkcijo slučajne spremenljivke $Y$ za monotono naraščajočo funkcijo $Y = g(X)$:

$$
F_Y(y) = P[Y \le y] = P[g(X) \le y] = P\left[X \le g^{-1}(y)\right] = F_X\left(g^{-1}(y)\right). \tag{5.8}
$$

Gostoto verjetnosti slučajne spremenljivke $Y$ dobimo z odvajanjem porazdelitvene funkcije:

$$
\begin{aligned}
f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{dF_X\left(g^{-1}(y)\right)}{dy} = \left.\frac{dF_X(x)}{dx}\right|_{x=g^{-1}(y)} \frac{dg^{-1}(y)}{dy}\\
&= f_X\left(g^{-1}(y)\right)\frac{dg^{-1}(y)}{dy}.
\end{aligned} \tag{5.9}
$$

Podobno izpeljemo za monotono padajočo zvezo $g(X)$

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[g(X) \leq y] = P[X \geq g^{-1}(y)] = 1 - P[X < g^{-1}(y)] \\ &= 1 - F_X(g^{-1}(y)) + P[X = g^{-1}(y)] = 1 - F_X(g^{-1}(y)), \end{aligned} \tag{5.10}$$

saj pri zveznih slučajnih spremenljivkah velja, da je $P[X = g^{-1}(y)] = 0$.

Gostoto verjetnosti slučajne spremenljivke $Y$ dobimo z odvajanjem porazdelitvene funkcije:

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} = \frac{d}{dy}\left(1 - F_X(g^{-1}(y))\right) = - \left.\frac{dF_X(x)}{dx}\right|_{x=g^{-1}(y)} \frac{dg^{-1}(y)}{dy} \\ &= -f_X\left(g^{-1}(y)\right) \frac{dg^{-1}(y)}{dy}. \end{aligned} \tag{5.11}$$

Ker je odvod $dg^{-1}(y)/dy$ za monotono padajočo funkcijo $g(X)$ negativen, za monotono naraščajočo pa pozitiven, lahko enačbi (5.9) in (5.11) združimo v eno enačbo, ki velja za poljubno monotono funkcijo:

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \left|\frac{dg^{-1}(y)}{dy}\right|. \tag{5.12}$$

Za splošne (nemonotone) funkcijske zveze med zveznima slučajnima spremenljivkama $X$ in $Y$ lahko zapišemo porazdelitveno funkcijo slučajne spremenljivke $Y$ takole:

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = \int\limits_{g(x) \leq y} f_X(x)\, dx. \tag{5.13}$$

**Primer 5.3: Linearna transformacija.** *Vzemimo, da je zveza med $X$ in $Y$ linearna funkcija,*

$$Y = a + b\, X.$$

*Slučajna spremenljivka $X$ se porazdeljuje eksponentno z gostoto verjetnosti,*

$$f_X(x) = \lambda\, e^{-\lambda\, x}, \qquad x \geq 0.$$

*Določimo gostoto verjetnosti slučajne spremenljivke $Y$.*

**Rešitev:** Določimo najprej zalogo vrednosti slučajne spremenljivke $Y$:

$$0 \leq x \quad \rightarrow \quad g(0) \leq g(x) \quad \rightarrow \quad a + b\, 0 = a \leq y.$$

Na zgornji strani zaloga vrednosti slučajne spremenljivke $Y$ ni omejena, saj limita

$$\lim_{x \to \infty} g(x) = \lim_{x \to \infty} a + b\, x$$

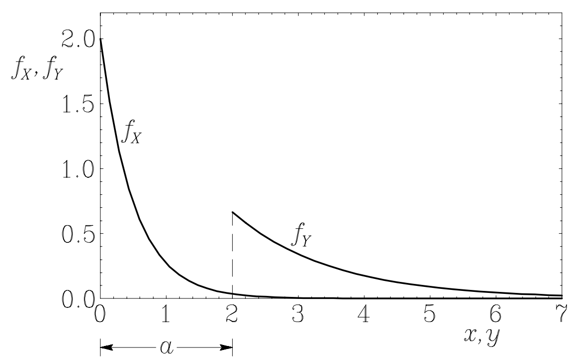
nima končne vrednosti. Ko gredo vrednosti $x$ proti neskončnosti, gredo tudi vrednosti $y$ proti neskončnosti.

Ker je linearna funkcija $y = a + b\,x$ za $b \neq 0$ monotona, lahko zapišemo inverzno funkcijo $g^{-1}(y)$ in njen odvod:

$$g^{-1}(y) = \frac{y-a}{b}, \qquad \frac{dg^{-1}(y)}{dy} = \frac{1}{b}.$$

Zadnja izraza uporabimo v enačbi (5.12) in dobimo gostoto verjetnosti slučajne spremenljivke $Y$:

$$f_Y(y) = \frac{\lambda}{b} e^{-\lambda\,\frac{y-a}{b}}, \qquad y \geq a.$$




Slika 5.6: Gostota verjetnosti dveh linearno povezanih slučajnih spremenljivk

Na sliki 5.6 prikazujemo gostote verjetnosti eksponentno porazdeljene slučajne spremenljivke $X$ z $\lambda = 2$ ter gostoto verjetnosti slučajne spremenljivke $Y = a + b\,X$ za $a = 2$ in $b = 3$. Vidimo, da parameter $a$ vpliva na premik gostote verjetnosti v levo ali desno, medtem ko parameter $b$ vpliva na razpršenost. Večji kot je $b$, bolj razpršena je porazdelitev.

**Primer 5.4:** *Vzemimo, da je zveza med slučajnima spremenljivkama $X$ in $Y$ določena z enačbo*

$$Y = g(X) = \sqrt{\frac{X}{A}}.$$

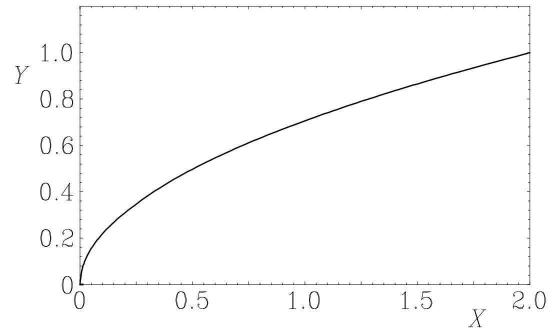
*Gostota verjetnosti slučajne spremenljivke $X$ je*

$$f_X(x) = \frac{1}{r}\left(1 - \cos\frac{2\pi x}{r}\right) \qquad 0 \leq x \leq r. \tag{5.14}$$

*Določimo gostoto verjetnosti slučajne spremenljivke $Y$.*

**Rešitev:** Ker je zveza med $X$ in $Y$ monotona (glej sliko 5.7), lahko gostoto verjetnosti $f_Y(y)$ določimo z enačbo (5.12). Zato moramo določiti inverzno funkcijo $g^{-1}(Y)$ in njen odvod,

$$x = g^{-1}(y) = A\,y^2, \qquad \frac{dg^{-1}(y)}{dy} = 2\,A\,y. \tag{5.15}$$




Slika 5.7: Zveza med slučajnima spremenljivkama $X$ in $Y$

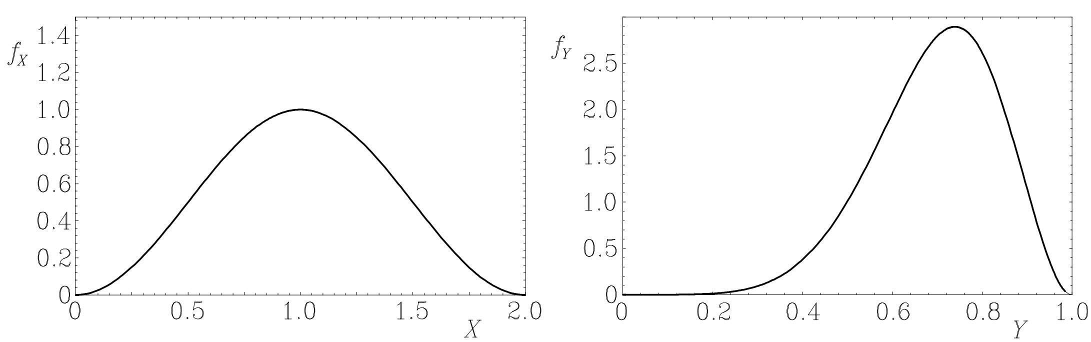
Zalogo vrednosti slučajne spremenljivke $Y$ določimo tako, da neenačbo v (5.14) transformiramo s funkcijo $g(X)$,

$$0 \leq x \leq r \qquad \rightarrow \qquad g(0) \leq g(x) \leq g(r) \qquad \rightarrow \qquad 0 \leq y \leq \sqrt{\frac{r}{A}}. \tag{5.16}$$

Uporabimo oznake (5.14) in (5.15) v enačbi (5.12) in dobimo

$$f_Y(y) = \frac{2\,A\,y}{r}\left(1 - \cos\frac{2\pi\,A\,y^2}{r}\right) \qquad 0 \leq y \leq \sqrt{\frac{r}{A}}. \tag{5.17}$$

Gostoti verjetnosti obeh slučajnih spremenljivk prikazujemo na sliki 5.8. Vidimo, da je slučajna spremenljivka $Y$ porazdeljena nesimetrično, slučajna spremenljivka $X$ pa simetrično.




Slika 5.8: Gostoti verjetnosti $f_X(x)$ in $f_Y(y)$ za $r = 2$ in $A = 2$

**Primer 5.5:** *Vzemimo, da med slučajnima spremenljivkama $X$ in $Y$ velja sinusna zveza*

$$Y = g(X) = \sin X.$$

*Gostota verjetnosti slučajne spremenljivke $X$ je*

$$f_X(x) = \frac{1}{\pi} \qquad 0 \leq x \leq \pi.$$

*Določimo porazdelitveno funkcijo in gostoto verjetnosti slučajne spremenljivke $Y$.*

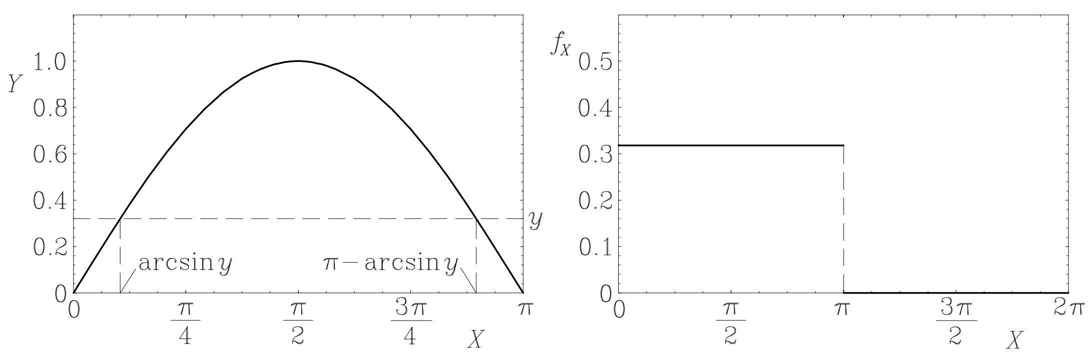
**Rešitev:** Na sliki 5.9 vidimo, da funkcija na obravnavanem območju ni monotona. Zato moramo najprej določiti porazdelitveno funkcijo $F_Y(y)$ po enačbi (5.13) ter nato z odvajanjem še gostoto verjetnosti $f_Y(y)$.

Inverzna funkcija funkcije $\sin x$ ni enolična, lahko pa definiramo glavne vrednosti inverzne funkcije, ki obsegajo le vrednosti od $-\pi/2$ do $\pi/2$. Zapišemo lahko

$$x = \arcsin y.$$

V našem primeru je $0 \leq x \leq \pi/2$ in $0 \leq y \leq 1$. Sedaj lahko uporabimo enačbo (5.13) (glej tudi sliko 5.9)

$$\begin{aligned} F_Y(y) &= \int\limits_{g(x)\leq y} f_X(x)\,dx = \int\limits_0^{\arcsin y} \frac{1}{\pi}dx + \int\limits_{\pi-\arcsin y}^{\pi} \frac{1}{\pi}dx \\ &= \frac{1}{\pi}(\arcsin y + \pi - \pi + \arcsin y) = \frac{2\,\arcsin y}{\pi}. \end{aligned} \tag{5.18}$$



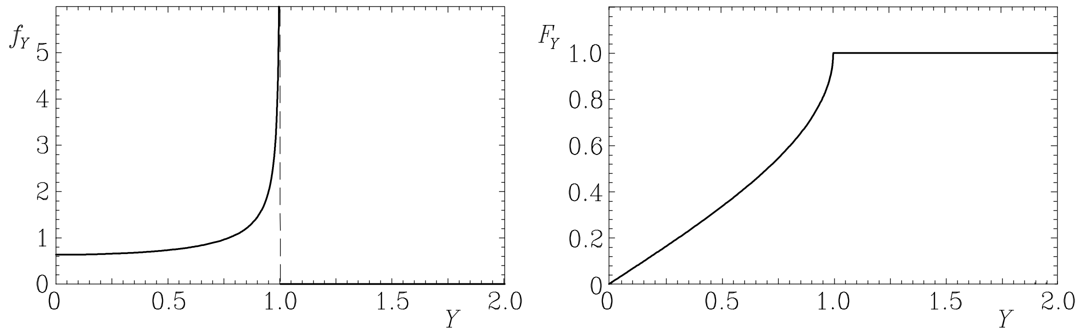

Slika 5.9: Sinusna zveza med spremenljivkama $X$ in $Y$ ter gostota verjetnosti $f_X(x)$

Zalogo vrednosti slučajne spremenljivke $Y$ najlažje določimo s slike 5.9, na kateri lahko vidimo, da območju od 0 do $\pi$ za slučajno spremenljivko $X$ ustreza območje od 0 do 1 za slučajno spremenljivko $Y$.

Z odvajanjem enačbe (5.18) določimo še gostoto verjetnosti slučajne spremenljike $Y$,

$$f_Y(y) = \frac{2}{\pi\sqrt{1-y^2}} \qquad 0 \leq y \leq 1.$$

Na sliki (5.10) prikazujemo gostoto verjetnosti $f_Y(y)$ in porazdelitveno funkcijo $F_Y(y)$ slučajne spremenljivke $Y$.




Slika 5.10: Gostota verjetnosti $f_Y(y)$ in porazdelitvena funkcija $F_Y(y)$

## 5.3 Funkcije več spremenljivk

Vzemimo, da je slučajna spremenljivka $Z$ funkcija slučajnih spremenljivk $X$ in $Y$,

$$Z = g(X, Y).$$

Porazdelitveno funkcijo slučajne spremenljivke $Z$ določimo z enačbo

$$F_Z(z) = P[Z \leq z] = P[g(X, Y) \leq z] = \iint\limits_{g(x,y)\leq z} f_{XY}(x, y)\, dx\, dy. \tag{5.19}$$

Najpreprostejša, a tudi najpomembnejša zveza med slučajno spremenljivko $Z$ in slučajnima spremenljivkama $X$ in $Y$ je linearna funkcija:

$$Z = g(X, Y) = X + Y.$$

Porazdelitveno funkcijo $F_Z(z)$ določimo po enačbi (5.19). Narišimo najprej območje, po katerem moramo integrirati za neki določeni $Z = z$ (slika 5.11),

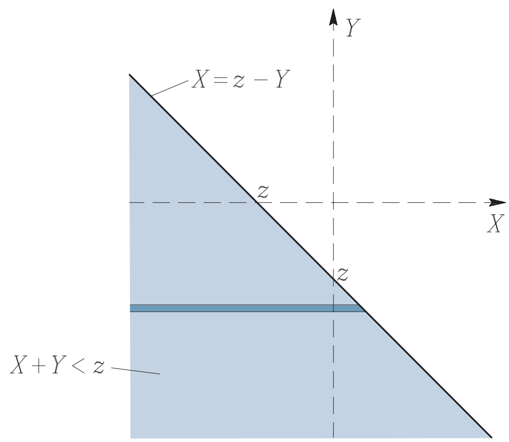
$$F_Z(z) = P[Z \leq z] = P[X + Y \leq z] = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{z-y} f_{XY}(\tilde{x}, \tilde{y})\, d\tilde{x}\, d\tilde{y}.$$

Gostoto verjetnosti $f_Z(z)$ določimo tako, da odvajamo porazdelitveno funkcijo po $z$. Pri odvajanju funkcije, podane z integralom, kjer je zgornja meja odvisna od spremenljivke, po kateri odvajamo, upoštevamo izrek (glej Bronštejn, Semendjajev, Musiol, Mühlig, Matematični priročnik, Tehniška založba Slovenije, 1997, str. 341)

$$\frac{d}{dy} \int\limits_{\alpha(y)}^{\beta(y)} f(x, y)\, dx = \int\limits_{\alpha(y)}^{\beta(y)} \frac{df(x, y)}{dy}\, dx + \beta'(y) \cdot f(\beta(y), y) - \alpha'(y) \cdot f(\alpha(y), y),$$

pri čemer upoštevamo, da je le zgornja meja odvisna od spremenljivke, po kateri odvajamo:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int\limits_{-\infty}^{\infty} f_{XY}(z - \tilde{y}, \tilde{y})\, d\tilde{y}.$$




Slika 5.11: Območje integracije za neki določeni $Z = z$

Če zamenjamo vrstni red integriranja, dobimo podobno enačbo:

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int\limits_{-\infty}^{\infty} f_{XY}(\tilde{x}, z - \tilde{x})\, d\tilde{x}.$$

Zgornji dve enačbi predstavljata izraz za določitev gostote verjetnosti slučajne spremenljivke $Z$, ki je vsota dveh slučajnih spremenljivk $X$ in $Y$. Ta proces imenujemo **konvolucija**, integrala pa **konvolucijska integrala**. V primeru, da sta slučajni spremenljivki medsebojno neodvisni, dobita konvolucijska integrala naslednjo obliko:

$$f_Z(z) = \int\limits_{-\infty}^{\infty} f_X(z - \tilde{y})\, f_Y(\tilde{y})\, d\tilde{y} = \int\limits_{-\infty}^{\infty} f_X(\tilde{x})\, f_Y(z - \tilde{x})\, d\tilde{x}. \tag{5.20}$$

Uporabo konvulucijskega integrala bomo spoznali v poglavju Verjetnostne porazdelitve – porazdelitev gama in normalna porazdelitev.

# 6 Momenti porazdelitev

Do sedaj smo slučajne spremenljivke v celoti opisovali z njihovimi porazdelitvenimi zakoni, ki jih matematično zapišemo z verjetnostno funkcijo ali gostoto verjetnosti oziroma porazdelitveno funkcijo.

Mnogokrat se zgodi, da imamo na voljo premalo podatkov, da bi lahko v celoti določili porazdelitveni zakon. Poleg tega je za inženirja pogosto dovolj, da pozna le določene lastnosti porazdelitvenega zakona.

Tem lastnostim pravimo **momenti porazdelitve**. Poimenovanje je nastalo po analogiji z mehaniko, kjer neko telo lahko v celoti opišemo tako, da za vsako točko, ki jo telo zavzema, zapišemo gostoto, lego, hitrost, temperaturo ... Po drugi strani pa lahko isto telo zelo dobro opišemo tudi z njegovimi lastnostmi, kot so masa, statični momenti (lega težišča), vztrajnostni momenti ...

## 6.1 Momenti in centralni momenti slučajne spremenljivke

Definirajmo moment $r$-tega reda za diskretno in zvezno slučajno spremenljivko $X$:

$$m_X^{(r)} = \sum_{i=1}^{n} x_i^r \, p_X(x_i), \qquad m_X^{(r)} = \int_{-\infty}^{\infty} x^r \, f_X(x) \, dx. \tag{6.1}$$

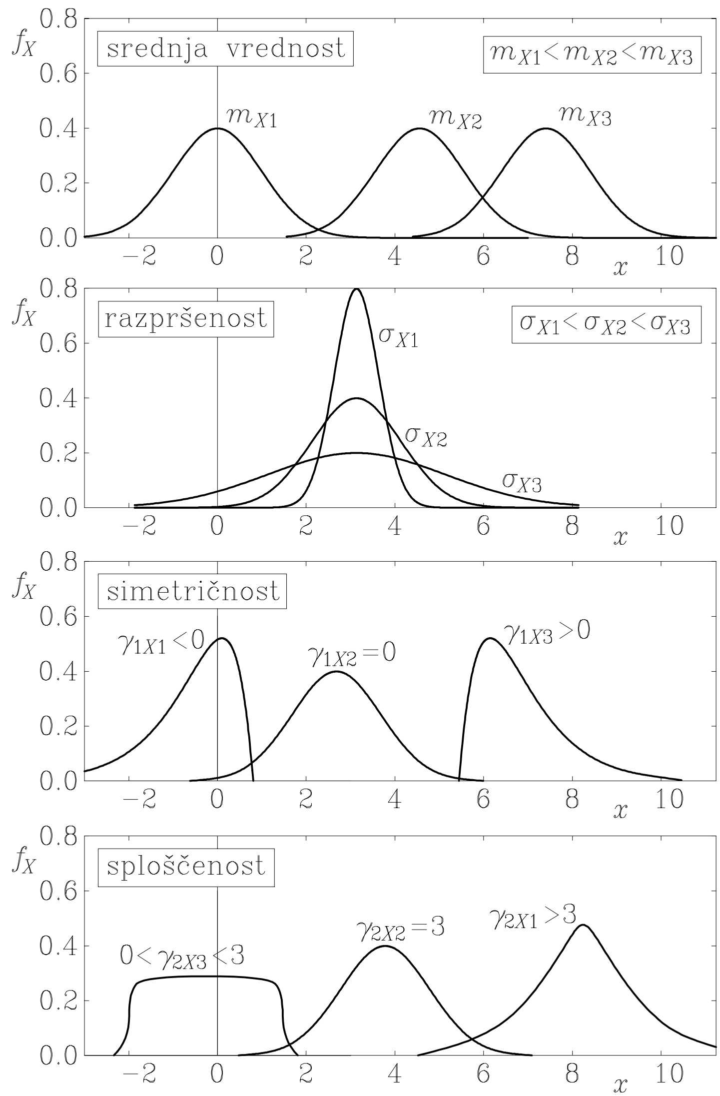
Iz lastnosti verjetnostne funkcije in gostote verjetnosti sledi, da je moment ničtega reda vedno enak 1.

$$m_X^{(0)} = \sum_{i=1}^{n} p_X(x_i) = 1, \qquad m_X^{(0)} = \int_{-\infty}^{\infty} f_X(x) \, dx = 1. \tag{6.2}$$

V analogiji z mehaniko moment ničtega reda ustreza masi telesa. Zelo pomemben je moment prvega reda, ki predstavlja **pričakovano vrednost** ali **matematično upanje** ali **srednjo vrednost** slučajne spremenljivke $X$. Moment prvega reda označimo z $m_X$ ali $\mathrm{E}[X]$

$$m_X^{(1)} = m_X = \mathrm{E}[X] = \sum_{i=1}^{n} x_i \, p_X(x_i), \qquad m_X^{(1)} = m_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx. \tag{6.3}$$

Oznaka $\mathrm{E}[X]$ predstavlja pričakovano vrednost slučajne spremenljivke $X$ (angl. expected value). V analogiji z mehaniko pričakovana vrednost slučajne spremenljivke predstavlja lego težišča telesa. Pomen pričakovane vrednosti pri neki zvezni slučajni spremenljivki prikazujmo na sliki 6.1.




Slika 6.1: Pomen različnih vrednosti $m_X$, $\sigma_X$, $\gamma_{1X}$ in $\gamma_{2X}$

Momenti višjega reda po enačbi (6.1) nimajo neposrednega pomena, zato jih tukaj ne obravnavamo. Sedaj, ko poznamo pričakovano vrednost $m_X$ slučajne spremenljivke $X$, lahko definiramo centralni moment $r$-tega reda

$$\mu_X^{(r)} = \sum_{i=1}^{n} (x_i - m_X)^r p_X(x_i), \qquad \mu_X^{(r)} = \int_{-\infty}^{\infty} (x - m_X)^r f_X(x) dx. \tag{6.4}$$

Centralni moment ničtega reda je enak ena, saj velja enačba (6.2) tudi za centralni moment $\mu_X^{(0)}$. Določimo centralni moment prvega reda! Zaradi preglednosti izpeljujemo le za primer zvezno porazdeljene slučajne spremenljivke. Iz enačbe (6.4) sledi

$$\begin{aligned} \mu_X^{(1)} = \mu_X &= \int_{-\infty}^{\infty} (x - m_X)\, f_X(x)\, dx = \int_{-\infty}^{\infty} x\, f_X(x)\, dx - \int_{-\infty}^{\infty} m_X\, f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} x\, f_X(x)\, dx - m_X \int_{-\infty}^{\infty} f_X(x)\, dx. \end{aligned} \tag{6.5}$$

Če v enačbi (6.5) upoštevamo enačbi (6.2) in (6.3), lahko zaključimo, da je centralni moment prvega reda enak nič:

$$\mu_X^{(1)} = m_X - m_X = 0.$$

Najpogosteje uporabljeni centralni moment je centralni moment drugega reda, s katerim opišemo razpršenost oziroma variabilnost slučajne spremenljivke. Centralni moment drugega reda imenujemo **varianca** in označimo z $\mathrm{var}[X]$, $\mathrm{D}[X]$ ali $\sigma_X^2$:

$$\begin{aligned} \mu_X^{(2)} &= \mathrm{var}[X] = \mathrm{D}[X] = \sigma_X^2 = \sum_{i=1}^{n} (x_i - m_X)^2\, p_X(x_i), \\ \mu_X^{(2)} &= \mathrm{var}[X] = \mathrm{D}[X] = \sigma_X^2 = \int_{-\infty}^{\infty} (x - m_X)^2\, f_X(x)\, dx. \end{aligned} \tag{6.6}$$

**Standardni odklon** ali **standardna deviacija** je definiran kot pozitivni kvadratni koren variance $\sigma_X = \sqrt{\mathrm{var}[X]}$. Pomen $\sigma_X$ pri zvezni slučajni spremenljivki prikazujmo na sliki 6.1. Pogosto želimo razpršenost opisati z brezdimenzijsko oziroma relativno mero razpršenosti, ki jo definiramo z enačbo

$$V_X = \frac{\sigma_X}{m_X} \tag{6.7}$$

in imenujemo **koeficient variacije**.

Slučajno spremenljivko $X$ lahko povsem opišemo le s porazdelitvenim zakonom ali pa z momenti vseh redov. Za praktično uporabo je običajno dovolj že podatek o pričakovani vrednosti $\mathrm{E}[X] = m_X$ in razpršenosti $\mathrm{var}[X] = \sigma_X^2$ slučajne spremenljivke $X$. Geometrijski pomen imata še dva višja centralna momenta slučajne spremenljivke $X$. Centralni moment tretjega reda $\mu_X^{(3)}$ je mera za simetričnost, centralni moment četrtega reda $\mu_X^{(4)}$ pa je mera za sploščenost oziroma koničastost. Koeficient asimetrije $\gamma_{1X}$ definiramo z enačbo

$$\gamma_{1X} = \frac{\mu_X^{(3)}}{\sigma_X^3}. \tag{6.8}$$

Pri simetričnih porazdelitvah je $\gamma_{1X}$ enak 0. Gostoto verjetnosti dveh nesimetričnih in ene simetrične zvezne slučajne spremenljivke prikazujemo na sliki 6.1.

Koeficient sploščenosti oziroma koničastosti $\gamma_{2X}$, ki ga imenujemo tudi koeficient kurtosis, definiramo z enačbo

$$\gamma_{2X} = \frac{\mu_X^{(4)}}{\sigma_X^4}. \tag{6.9}$$

Sploščenost obravnavane porazdelitve primerjamo s sploščenostjo normalno porazdeljene slučajne spremenljivke, ki jo bomo podrobneje obravnavali v nadaljevanju in katere koeficient kurtosis je enak 3. Za porazdelitve, ki imajo koeficient kurtosis večji od 3, velja, da so bolj koničaste, vrednosti, manjše od 3, pa predstavljajo bolj sploščene porazdelitve. Glej sliko 6.1.

## 6.2 Pričakovana vrednost funkcije slučajne spremenljivke

Vzemimo, da poznamo porazdelitev slučajne spremenljivke $X$ in vemo, da je slučajna spremenljivka $Y$ funkcija slučajne spremenljivke $X$

$$Y = g(X).$$

Določiti želimo pričakovano vrednost $\mathrm{E}[Y]$ slučajne spremenljivke $Y$, ne da bi določili porazdelitev slučajne spremenljivke $Y$. Obravnavamo zvezno slučajno spremenljivko, čeprav iste lastnosti veljajo tudi za diskretno. Pričakovano vrednost slučajne spremenljivke $Y$ določimo po naslednji enačbi

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx = \mathrm{E}[g(X)]. \tag{6.10}$$

Vidimo, da lahko pričakovano vrednost slučajne spremenljivke $Y$ določimo tako, da integriramo funkcijo $g(x)\, f_X(x)$ in nam ni treba določiti gostote verjetnosti slučajne spremenljivke $Y$. Izpeljimo nekaj lastnosti pričakovanih vrednosti funkcij slučajnih spremenljivk. Pričakovana vrednost konstante $c$ je

$$\mathrm{E}[c] = \int_{-\infty}^{\infty} c\, f_X(x)\, dx = c \int_{-\infty}^{\infty} f_X(x)\, dx = c, \tag{6.11}$$

kar je seveda logičen rezultat, saj konstanta $c$ ni slučajna spremenljivka in ne more zavzeti druge vrednosti kot $c$. Pričakovana vrednost linearne funkcije $a + b\,X$ slučajne spremenljivke $X$ je

$$\mathrm{E}[a + b\,X] = \int\limits_{-\infty}^{\infty}(a + b\,x)\,f_X(x)\,dx = a\int\limits_{-\infty}^{\infty} f_X(x)\,dx + b\int\limits_{-\infty}^{\infty} x\,f_X(x)\,dx = a + b\,\mathrm{E}[X]. \tag{6.12}$$

Pričakovano vrednost vsote dveh funkcij slučajne spremenljivke $X$ določimo na podoben način

$$\begin{aligned}\mathrm{E}[g_1(X) + g_2(X)] &= \int\limits_{-\infty}^{\infty}(g_1(x) + g_2(x))\,f_X(x)\,dx \\ &= \int\limits_{-\infty}^{\infty} g_1(x)\,f_X(x)\,dx + \int\limits_{-\infty}^{\infty} g_2(x)\,f_X(x)\,dx = \mathrm{E}[g_1(X)] + \mathrm{E}[g_2(X)].\end{aligned} \tag{6.13}$$

Če primerjamo enačbi (6.6) in (6.10), vidimo, da lahko varianco slučajne spremenljivke $\mathrm{var}[X]$ upoštevamo kot pričakovano vrednost funkcije $g(X) = (X - m_X)^2$

$$\mathrm{var}[X] = \mathrm{E}[(X - m_X)^2]. \tag{6.14}$$

Iz enačbe (6.14) lahko izpeljemo obrazec, ki nam v določenih primerih poenostavi računanje variance slučajnih spremenljivk. Ob upoštevanju enačb (6.3), (6.11) in (6.12) lahko iz enačbe (6.14) dobimo

$$\begin{aligned}\mathrm{var}[X] &= \mathrm{E}[(X - m_X)^2] = \mathrm{E}[X^2 - 2\,X\,m_X + m_X^2] \\ &= \mathrm{E}[X^2] - 2\,m_X\mathrm{E}[X] + m_X^2 = \mathrm{E}[X^2] - \mathrm{E}[X]^2.\end{aligned} \tag{6.15}$$

Prednost uporabe enačbe (6.15) pred (6.14) bomo prikazali v računskem primeru 6.1. Izpeljimo nekaj lastnosti variance slučajne spremenljivke. Varianca konstante $c$ je ob upoštevanju (6.11) enaka

$$\mathrm{var}[c] = \mathrm{E}[(c - c)^2] = \mathrm{E}[0] = 0, \tag{6.16}$$

kar je pričakovan rezultat, saj konstanta $c$ ni slučajna spremenljivka, ampak deterministična vrednost, katere razpršenost je enaka nič. Varianca linearne funkcije $a + b\,X$ slučajne spremenljivke $X$ je enaka (upoštevajmo tudi enačbi (6.12) in (6.14))

$$\mathrm{var}[a + b\,X] = \mathrm{E}[(a + b\,X - (a + b\,m_X))^2] = \mathrm{E}[b^2\,(X - m_X)^2] = b^2\,\mathrm{var}[X]. \tag{6.17}$$

**Primer 6.1:** *Določimo pričakovano vrednost, varianco ter koeficienta asimetrije in sploščenosti slučajne spremenljivke $X$, ki predstavlja rezultat metanja poštene kocke!*

**Rešitev:** Verjetnostno funkcijo te slučajne spremenljivke smo določili že v primeru 4.2.

$$p_X(x_i) = \frac{1}{6}, \qquad i = 1, \ldots, 6.$$

Pričakovano vrednost in varianco določimo po enačbah (6.3) in (6.6):

$$m_X = \sum_{i=1}^{6} x_i \, p_X(x_i) = \frac{1+2+3+4+5+6}{6} = 3.5$$

in

$$\begin{aligned}
\sigma_X^2 &= \sum_{i=1}^{6} (x_i - m_X)^2 \, p_X(x_i) \\
&= \frac{(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2}{6} \\
&= \frac{35}{12} = 2.9167
\end{aligned}$$

Izračunajmo varianco te slučajne spremenljivke še z uporabo enačbe (6.15):

$$\sigma_X^2 = \sum_{i=1}^{6} x_i^2 \, p_X(x_i) - m_X^2 = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} - 3.5^2 = 2.9167.$$

Standardna deviacija $\sigma_X$ je enaka

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{2.9167} = 1.7078.$$

Določimo še koeficienta asimetrije $\gamma_{1X}$ in sploščenosti $\gamma_{2X}$. Glede na to, da je enakomerna porazdelitev simetrična, lahko pričakujemo, da bo koeficient asimetrije enak nič

$$\begin{aligned}
\gamma_{1X} &= \frac{1}{\sigma_X^3} \sum_{i=1}^{6} (x_i - m_X)^3 \, p_X(x_i) \\
&= \frac{(1-3.5)^3 + (2-3.5)^3 + (3-3.5)^3 + (4-3.5)^3 + (5-3.5)^3 + (6-3.5)^3}{1.7078^3 \cdot 6} = 0.
\end{aligned}$$

Koeficient sploščenosti oziroma koeficient kurtosis pa je enak

$$\begin{aligned}
\gamma_{2X} &= \frac{1}{\sigma_X^4} \sum_{i=1}^{6} (x_i - m_X)^4 \, p_X(x_i) \\
&= \frac{(1-3.5)^4 + (2-3.5)^4 + (3-3.5)^4 + (4-3.5)^4 + (5-3.5)^4 + (6-3.5)^4}{1.7078^4 \cdot 6} = 1.7314.
\end{aligned}$$

Po pričakovanju smo izračunali koeficient kurtosis, ki je manjši od 3, kar je značilno za sploščene porazdelitve.

**Primer 6.2:** *Določimo pričakovano vrednost, varianco ter koeficienta asimetrije in sploščenosti zvezne slučajne spremenljivke $X$, porazdeljene po eksponentni porazdelitvi! Gostota verjetnosti je*

$$f_X(x) = 2\,e^{-2x} \qquad\qquad x \geq 0.$$

**Rešitev:** Pričakovano vrednost izračunamo po enačbi (6.3)

$$m_X = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_{0}^{\infty} 2\, x\, e^{-2x}\, dx = \lim_{c\to\infty} \int_{0}^{c} 2\, x\, e^{-2x}\, dx.$$

Gornji integral določimo z integriranjem po delih in dobimo:

$$m_X = \lim_{c\to\infty} \left(-x\, e^{-2x} - \frac{1}{2}\, e^{-2x}\right)\Bigg|_0^c = \lim_{c\to\infty} \left[\left(-c\, e^{-2c} - \frac{1}{2}\, e^{-2c}\right) - \left(-0\, e^{-2\cdot 0} - \frac{1}{2}\, e^{-2\cdot 0}\right)\right].$$

Za prvi izraz v limiti lahko z L'Hôpitalovim[1] pravilom pokažemo, da je enak nič. Očitno sta nič tudi drugi in tretji člen. Tako lahko zaključimo, da je pričakovana vrednost te slučajne spremenljivke enaka

$$m_X = \frac{1}{2}.$$

Na podoben način z upoštevanjem enačbe (6.15) in zaporedno uporabo integriranja po delih lahko izračunamo tudi višje momente. Rezultati, ki smo jih dobili s programom MATHEMATICA so:[2]

$$\mathrm{var}[X] = \frac{1}{4}, \qquad \gamma_{1X} = 2, \qquad \gamma_{2X} = 9.$$

Ker je eksponentna porazdelitev nesimetrična smo lahko pričakovali, da je koeficient asimetrije različen od nič. Eksponentna porazdelitev je tudi očitno bolj koničasta od normalne, zato je koeficient sploščenosti večji od 3.

---

[1] Guillaume Francois Antoine Marquis de L'Hôpital, francoski matematik, avtor prvega učbenika o diferencialnem računu, (1661–1704).

[2] Integracija, izvedena s programom MATHEMATICA:

```
fX[x_] := 2 E^(-2 x);
mx = Integrate[x fX[x],{x,0,Infinity}];
varx = Integrate[(x-mx)^2 fX[x],{x,0,Infinity}];
sigmax = Sqrt[varx];
gama1 = Integrate[(x-mx)^3 fX[x],{x,0,Infinity}]/sigmax^3;
gama2 = Integrate[(x-mx)^4 fX[x],{x,0,Infinity}]/sigmax^4;
Print["mx = ",mx];  Print["Var[X] = ",varx];
Print["gama1 = ",gama1];  Print["gama2 = ",gama2];

     1
mx = -
     2

         1
Var[X] = -
         4

gama1 = 2
gama2 = 9
```

## 6.3 Momenti in centralni momenti slučajnega vektorja

Obravnavamo slučajni vektor z dvema slučajnima spremenljivkama $X$ in $Y$. Momente slučajnih vektorjev večjih razsežnosti izračunamo na enak način in jih zato ne bomo obravnavali. Moment diskretnega in zveznega slučajnega vektorja reda $p, r$ definiramo z naslednjima enačbama

$$m_{XY}^{(p,r)} = \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} x_i^p\, y_j^r\, p_{XY}(x_i, y_j), \qquad m_{XY}^{(p,r)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p\, y^r\, f_{XY}(x,y)\, dx\, dy. \tag{6.18}$$

Enako kot pri momentih slučajne spremenljivke iz lastnosti verjetnostne funkcije in gostote verjetnosti sledi, da sta ničta momenta slučajnega vektorja enaka ena

$$m_{XY}^{(0,0)} = 1.$$

Zaradi krajšega zapisa bomo v nadaljevanju obravnavali le zvezne slučajne vektorje. Vse ugotovitve, do katerih bomo prišli, lahko brez težav pokažemo tudi za diskretne slučajne vektorje. Pri slučajnem vektorju z dvema elementoma imamo dva momenta prvega reda, ki ju izračunamo po naslednjih enačbah:

$$\begin{aligned} m_{XY}^{(1,0)} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f_{XY}(x,y)\, dx\, dy, \\ m_{XY}^{(0,1)} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f_{XY}(x,y)\, dx\, dy. \end{aligned}$$

Ob upoštevanju enačbe (4.26) za določitev delnih gostot verjetnosti $f_X(x)$ in $f_Y(y)$ in definicije pričakovane vrednosti slučajne spremenljivke (enačba (6.3)) lahko zapišemo

$$\begin{aligned} m_{XY}^{(1,0)} &= \int_{-\infty}^{\infty} x \left( \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy \right) dx = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = m_X, \\ m_{XY}^{(0,1)} &= \int_{-\infty}^{\infty} y \left( \int_{-\infty}^{\infty} f_{XY}(x,y)\, dx \right) dy = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = m_Y. \end{aligned} \tag{6.19}$$

Vidimo, da smo določili pričakovani vrednosti slučajnega vektorja $X, Y$. Ker momenti višjega reda nimajo neposrednega pomena, nadaljujemo z definicijo centralnega momenta reda $p, r$

$$\mu_{XY}^{(p,r)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_X)^p\, (y - m_Y)^r\, f_{XY}(x,y)\, dx\, dy.$$

Podobno kot pri momentih slučajne spremenljivke velja, da so $\mu_{XY}^{(0,0)} = 1$, $\mu_{XY}^{(1,0)} = 0$ in $\mu_{XY}^{(0,1)} = 0$. Najpomembnejši so momenti drugega reda, ki so v primeru dvorazsežnega slučajnega vektorja trije.

Upoštevajmo enačbi (4.26) in (6.6) in dobimo

$$\begin{aligned}\mu_{XY}^{(2,0)} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_X)^2 f_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}(x-m_X)^2\int_{-\infty}^{\infty}f_{XY}(x,y)\,dy\;dx\\ &= \int_{-\infty}^{\infty}(x-m_X)^2 f_X(x)dx\; = \mathrm{var}[X] = \sigma_X^2,\end{aligned} \tag{6.20}$$

$$\begin{aligned}\mu_{XY}^{(0,2)} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-m_Y)^2 f_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}(y-m_Y)^2\int_{-\infty}^{\infty}f_{XY}(x,y)\,dx\;dy\\ &= \int_{-\infty}^{\infty}(y-m_Y)^2 f_Y(y)dy\; = \mathrm{var}[Y] = \sigma_Y^2.\end{aligned} \tag{6.21}$$

Zgornji enačbi opisujeta varianci slučajnih spremenljivk $X$ in $Y$ in predstavljata meri za razpršenost slučajnih spremenljivk $X$ in $Y$. Tretji moment drugega reda pa označimo s $\mathrm{cov}[X,Y]$ oziroma $\sigma_{XY}$ in imenujemo **kovarianca** med slučajnima spremenljivkama $X$ in $Y$ slučajnega vektorja $X, Y$

$$\mu_{XY}^{(1,1)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-m_X)\,(y-m_Y)\,f_{XY}(x,y)\,dx\,dy = \mathrm{cov}[X,Y] = \sigma_{XY}. \tag{6.22}$$

Zapišimo sedaj še definicije pričakovanih vrednosti, varianc in kovariance za diskretni slučajni vektor dveh slučajnih spremenljivk. Podobno kot v enačbah (6.19), (6.20), (6.21) in (6.22) lahko zapišemo:

$$\begin{aligned}m_X &= m_{XY}^{(1,0)} = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} x_i\,p_{XY}(x_i,y_j)\\ m_Y &= m_{XY}^{(0,1)} = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} y_j\,p_{XY}(x_i,y_j)\\ \mathrm{var}[X] &= \sigma_X^2 = \mu_{XY}^{(2,0)} = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y}(x_i-m_X)^2\,p_{XY}(x_i,y_j)\\ \mathrm{var}[Y] &= \sigma_Y^2 = \mu_{XY}^{(0,2)} = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y}(y_j-m_Y)^2\,p_{XY}(x_i,y_j)\\ \mathrm{cov}[X,Y] &= \sigma_{XY} = \mu_{XY}^{(1,1)} = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y}(x_i-m_X)\,(y_j-m_Y)\,p_{XY}(x_i,y_j)\end{aligned} \tag{6.23}$$

Kovarianca med dvema slučajnima spremenljivka je mera za njuno linearno odvisnost. Za lažjo primerjavo vpeljemo brezdimenzionalno mero $\rho_{XY}$, ki jo imenujemo **korelacijski koeficient** in določimo po naslednji enačbi

$$\rho_{XY} = \frac{\mathrm{cov}[X,Y]}{\sigma_X\,\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\,\sigma_Y}. \tag{6.24}$$

Vrednost korelacijskega koeficienta leži na območju od $-1$ do $1$ ($-1 \leq \rho_{XY} \leq 1$), kar bomo dokazali v primeru 6.6. Vrednosti okoli ničle pomenijo, da slučajni spremenljivki nista linearno odvisni. Vrednosti blizu $1$ pomenijo močno pozitivno linearno odvisnost, vrednosti blizu $-1$ pa močno negativno linearno odvisnost. Če je korelacijski koeficient med dvema slučajnima spremenljivkama $X$ in $Y$ blizu ena, pomeni, da velikim vrednostim slučajne spremenljivke $X$ pripadajo velike vrednosti spremenljivke $Y$. Naštejmo nekaj primerov takih parov: človekova višina in teža, moč avtomobilskega motorja in največji pospeški, ki jih zmore avtomobil, obtežba in pomik (pri dovolj majhni obtežbi, tako da je vpliv nelinearnosti majhen), natezna trdnost betona in kvadratni koren njegove tlačne trdnosti[3] in še mnogo drugih. Za slučajni spremenljivki s korelacijskim koeficientom blizu $-1$ pa velja, da manjše vrednosti slučajne spremenljivke $X$ pogojujejo velike vrednosti slučajne spremenljivke $Y$ in obratno. Tudi takih parov je mnogo: teža človeka in njegov najboljši dosežek v skoku v višino, največja hitrost avtomobila v zavoju in ukrivljenost zavoja, tlačna trdnost betona pri ciklični obremenitvi in število ponovitev obtežbe (ciklov).[4] Korelacijski koeficient okoli nič pomeni, da velike vrednosti ene slučajne spremenljivke ne pomenijo nujno tudi velike ali majhne vrednosti druge. Taki pari slučajnih spremenljivk so: količina padavin, ki pade v nekem območju v dveh zaporednih letih, stalna obtežba na konstrukcijo in obtežba z vetrom in mnogo drugih. Kot smo pri zadnjih primerih nakazali, je neodvisnost med slučajnima spremenljivkama razlog zato, da je korelacija med njima enaka nič. Dokažimo!

Za dve neodvisni slučajni spremenljivki velja, da je gostota verjetnosti vektorja $X, Y$ produkt gostote verjetnosti slučajnih spremenljivk $X$ in $Y$ (enačba (4.30))

$$f_{XY}(x,y) = f_X(x)\,f_Y(y).$$

Ob upoštevanju zadnje enačbe v (6.22) in dejstvu, da sta centralna momenta $\mu_X^{(1)}$ in $\mu_Y^{(1)}$ enaka nič, dobimo

$$\begin{aligned}
\mathrm{cov}[X,Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-m_X)\,(y-m_Y)\,f_{XY}(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} (x-m_X)\,f_X(x)\,dx \cdot \int_{-\infty}^{\infty} (y-m_Y)\,f_Y(y)\,dy = \mu_X^{(1)}\,\mu_Y^{(1)} = 0.
\end{aligned}$$

[3] F.K. Kong, R.H. Evans, E. Cohen, F. Roll (eds.), Handbook of Structural Concrete, McGraw-Hill, Book Company, 1983, str. 5–11.

[4] F.K. Kong, R.H. Evans, E. Cohen, F. Roll (eds.), Handbook of Structural Concrete, McGraw-Hill, Book Company, 1983, str. 16–8.

Dokazali smo, da imata dve neodvisni slučajni spremenljivki kovarianco in s tem tudi korelacijski koeficient enako nič. **Toda pazite!** Obratno ne velja vedno. To pomeni, da sta lahko dve slučajni spremenljivki stohastično odvisni tudi v primeru, ko je korelacija med njima enaka nič (glej primer 6.4).

**Primer 6.3:** *Določimo kovarianco slučajnega vektorja $X, Y$ iz primera 4.7. Gostota verjetnosti je*

$$f_{XY}(x,y) = \frac{x}{\pi a^2} \qquad 0 \le x \le a, \;\; 0 \le y \le 2\pi.$$

**Rešitev:** Najprej moramo določiti pričakovani vrednosti slučajnih spremenljivk po enačbah (6.19), pri čemer uporabimo rezultata iz primera 4.7 (enačbi(4.31 in (4.32)):

$$m_X = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_0^a x\, \frac{2\,x}{a^2}\, dx = \left.\frac{2\,x^3}{3\,a^2}\right|_0^a = \frac{2\,a}{3},$$

$$m_Y = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_0^{2\pi} y\, \frac{1}{2\,\pi}\, dy = \left.\frac{y^2}{4\,\pi}\right|_0^{2\pi} = \pi.$$

Kovarianco slučajnega vektorja $X, Y$ izračunamo po enačbi (6.22)

$$\begin{aligned}
\mathrm{cov}[X,Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - m_X)\,(y - m_Y)\, f_{XY}(x,y)\, dx\, dy \\
&= \int_0^a \int_0^{2\pi} \left(x - \frac{2\,a}{3}\right)\,(y-\pi)\,\frac{x}{\pi\,a^2}\, dx\, dy \\
&= \int_0^a \left(x - \frac{2\,a}{3}\right)\,\frac{x}{\pi\,a^2}\,\left.\left(\frac{y^2}{2} - \pi\,y\right)\right|_0^{2\pi} dx = \int_0^a \left(x - \frac{2\,a}{3}\right)\,\frac{x}{\pi\,a^2}\; 0\; dx = 0
\end{aligned}$$

Kovarianca med tema dvema slučajnima spremenljivkama je enaka nič. Glede na to, da smo v primeru 4.7 pokazali, da sta slučajni spremenljivki $X$ in $Y$ v tem primeru neodvisni, je tak rezultat pričakovan.

**Primer 6.4:** *Določimo kovarianco slučajnega vektorja $X, Y$ iz primera 4.8. Gostota verjetnosti je določena z enačbo (4.33)*

$$f_{XY}(x,y) = \frac{1}{\pi R^2} \qquad\qquad x^2 + y^2 \le R^2.$$

**Rešitev:** Izračunati moramo pričakovani vrednosti slučajnih spremenljivk po enačbah (6.19) in kovarianco po enačbi (6.22). Za vajo izračunajmo še varianci slučajnih spremenljivk $X$ in $Y$. Integrirati moramo po naslednjih enačbah:

$$m_X = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} x \, \frac{1}{\pi R^2} \, dy \, dx$$

$$m_Y = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} y \, \frac{1}{\pi R^2} \, dy \, dx$$

$$\mathrm{var}[X] = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} (x - m_X)^2 \, \frac{1}{\pi R^2} \, dy \, dx$$

$$\mathrm{var}[Y] = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} (y - m_Y)^2 \, \frac{1}{\pi R^2} \, dy \, dx$$

$$\mathrm{cov}[X,Y] = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} (x - m_X)(y - m_Y) \, \frac{1}{\pi R^2} \, dy \, dx$$

Integracijo opravimo s programom MATHEMATICA.[5] Tako dobljeni rezultati so:

$$m_X = m_Y = 0, \qquad \mathrm{var}[X] = \mathrm{var}[Y] = \frac{R^2}{4}, \qquad \mathrm{cov}[X,Y] = 0.$$

[5] Integracija, izvedena s programom MATHEMATICA:

```
fxy = 1/Pi/R^2;
mx = Integrate[Integrate[x fxy,{y,-Sqrt[R^2-x^2],Sqrt[R^2-x^2]}],{x,-R,R}];
my = Integrate[Integrate[x fxy,{y,-Sqrt[R^2-x^2],Sqrt[R^2-x^2]}],{x,-R,R}];
Varx = Integrate[Integrate[(x-mx)^2 fxy,{y,-Sqrt[R^2-x^2],Sqrt[R^2-x^2]}],{x,-R,R}];
Vary = Integrate[Integrate[(y-my)^2 fxy,{y,-Sqrt[R^2-x^2],Sqrt[R^2-x^2]}],{x,-R,R}];
Covxy = Integrate[Integrate[(x-mx)(y-my) fxy,{y,-Sqrt[R^2-x^2],Sqrt[R^2-x^2]}],{x,-R,R}];
Print["mX = ",mx];  Print["mY = ",my];
Print["Var[X] = ",Varx];  Print["Var[Y] = ",Vary];  Print["Cov[X,Y] = ",Covxy];

mX = 0
mY = 0
          2
         R
Var[X] = --
         4
          2
         R
Var[Y] = --
         4
Cov[X,Y] = 0
```

Izkazalo se je torej, da je tudi korelacija med tema dvema slučajnima spremenljivkama enaka nič, čeprav smo v primeru 4.8 pokazali, da slučajni spremenljivki nista neodvisni. To je dokaz, da nekoreliranost (korelacija med dvema spremenljivkama je enaka nič) še ne pomeni tudi neodvisnost dveh slučajnih spremenljivk.

## 6.4 Pričakovana vrednost funkcije slučajnega vektorja

Če poznamo porazdelitev slučajnega vektorja $X, Y$ in nas zanima pričakovana vrednost funkcije tega vektorja

$$Z = g(X, Y),$$

jo izračunamo na podoben način kot pričakovano vrednost funkcije slučajne spremenljivke (enačba (6.10))

$$\mathrm{E}[Z] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f_{XY}(x, y)\, dx\, dy = \mathrm{E}[g(X, Y)]. \tag{6.25}$$

Če primerjamo enačbe, s katerimi definiramo varianci in kovarianco slučajnega vektorja $X, Y$ (enačbe (6.20), (6.21) in (6.22), vidimo, da predstavljajo varianci in kovarianca pričakovane vrednosti funkcij slučajnih vektorjev

$$\mathrm{var}[X] = \mathrm{E}[(X - m_X)^2] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$$
$$\mathrm{var}[Y] = \mathrm{E}[(Y - m_Y)^2] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2$$
$$\mathrm{cov}[X, Y] = \mathrm{E}[(X - m_X)(Y - m_Y)]$$

Morda najpomembnejša lastnost pričakovane vrednosti funkcij slučajnih vektorjev je, da je pričakovana vrednost vsote funkcij slučajnega vektorja enaka vsoti pričakovanih vrednosti funkcij slučajnega vektorja (linearnost):

$$\mathrm{E}[g_1(X, Y) + g_2(X, Y)] = \mathrm{E}[g_1(X, Y)] + \mathrm{E}[g_2(X, Y)], \tag{6.26}$$

kar lahko bralec sam preveri na enak način, kot smo to naredili za isto lastnost pričakovane vrednosti funkcije slučajne spremenljivke. Z uporabo enačbe (6.26) lahko izpeljemo nekatere uporabne zveze za pričakovane vrednosti in druge momente slučajnih vektorjev in njihovih funkcij. Izpeljimo alternativni obrazec za račun kovariance slučajnega vektorja $X, Y$

$$\begin{aligned}\mathrm{cov}[X, Y] &= \mathrm{E}[(X - m_X)(Y - m_Y)] = \mathrm{E}[X\,Y - X\,m_Y - m_X\,Y + m_X\,m_Y] \\ &= \mathrm{E}[XY] - m_Y\,\mathrm{E}[X] - m_X\,\mathrm{E}[Y] + m_X\,m_Y \\ &= \mathrm{E}[XY] - m_X\,m_Y = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y].\end{aligned} \tag{6.27}$$

**Primer 6.5:** *Določimo pričakovano vrednost in varianco slučajne spremenljivke $Z$, ki je linearna kombinacija slučajnih spremenljivk $X$ in $Y$*

$$Z = a\,X + b\,Y.$$

**Rešitev:** Pričakovano vrednost slučajne spremenljivke $Z$ določimo z enačbo (6.26)

$$m_Z = \mathrm{E}[Z] = \mathrm{E}[a\,X + b\,Y] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y] = a\,m_X + b\,m_Y. \tag{6.28}$$

Vidimo, da je pričakovana vrednost vsote dveh slučajnih spremenljivk vsota pričakovanih vrednosti. Tudi varianco določimo z uporabo iste enačbe

$$\begin{aligned}\mathrm{var}[Z] = \mathrm{var}[a\,X + b\,Y] &= \mathrm{E}[Z^2] - \mathrm{E}[Z]^2 = \mathrm{E}[(a\,X + b\,Y)^2] - \mathrm{E}[a\,X + b\,Y]^2 \\ &= \mathrm{E}[a^2\,X^2 + 2ab\,X\,Y + b^2\,Y^2] - (a\,m_X + b\,m_Y)^2 \\ &= a^2\,\mathrm{E}[X^2] + 2ab\,\mathrm{E}[XY] + b^2\,\mathrm{E}[Y^2] - a^2\,m_X^2 - 2ab\,m_X\,m_Y - b^2\,m_Y^2\end{aligned}$$

Ob upoštevanju enačb (6.15) in (6.27) v zadnji enačbi dobimo izraz za varianco linearne kombinacije dveh slučajnih spremenljivk

$$\mathrm{var}[Z] = \mathrm{var}[a\,X + b\,Y] = a^2\,\mathrm{var}[X] + b^2\,\mathrm{var}[Y] + 2ab\,\mathrm{cov}[X, Y]. \tag{6.29}$$

Zadnjo enačbo pogosto napišemo tudi v naslednji obliki

$$\sigma_Z^2 = a^2\,\sigma_X^2 + b^2\,\sigma_Y^2 + 2ab\,\rho_{XY}\,\sigma_X\,\sigma_Y.$$

Poseben primer linearne kombinacije je vsota dveh slučajnih spremenljivk. Pričakovana vrednost in varianca vsote dveh slučajnih spremenljivk sta

$$m_{X+Y} = m_X + m_Y, \qquad \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\,\rho_{XY}\,\sigma_X\,\sigma_Y.$$

Zadnja enačba opisuje zanimivo in včasih zelo uporabno dejstvo, da je lahko varianca vsote dveh slučajnih spremenljivk manjša od variance obeh posameznih slučajnih spremenljivk. Pogoj za to je negativna koreliranost slučajnih spremenljivk $X$ in $Y$. Vzemimo primer, ko sta varianci slučajnih spremenljivk $X$ in $Y$ enaki ($\sigma_X^2 = \sigma_Y^2$), korelacijski koeficient pa je $-0.9$

$$\sigma_Z^2 = 2\sigma_X^2 - 2 \cdot 0.9 \cdot \sigma_X^2 = 0.2\sigma_X^2.$$

V tem primeru je bila varianca vsote slučajnih spremenljivk $X$ in $Y$ kar petkrat manjša od variance slučajne spremenljivke $X$.

Obravnavajmo še primer, ko je korelacijski koeficient enak $-1$. V tem primeru je

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\sigma_X\,\sigma_Y = (\sigma_X - \sigma_Y)^2.$$

V posebnem primeru, ko sta varianci dveh slučajnih spremenljivk enaki, korelacijski koeficient pa je enak $-1$, je varianca vsote teh dveh slučajnih spremenljivk enaka nič: nenavaden rezultat, ki pravi, da vsota dveh slučajnih spremenljivk ni slučajna spremenljivka, temveč deterministična konstanta $Z = m_X + m_Y$.

Če sta slučajni spremenljivki nekorelirani (kovarianca oziroma korelacijski koeficient je enak nič), se zadnja enačba (6.29) poenostavi

$$\mathrm{var}[Z] = \mathrm{var}[X+Y] = \mathrm{var}[X] + \mathrm{var}[Y] = \sigma_X^2 + \sigma_Y^2. \tag{6.30}$$

Ker velja, da je korelacijski koeficient dveh neodvisnih slučajnih spremenljivk enak nič, velja zadnja enačba tudi za vsoto dveh neodvisnih slučajnih spremenljivk.

Na enak način lahko določimo tudi pričakovano vrednost in varianco razlike dveh slučajnih spremenljivk. Rezultati, ki jih bralec lahko sam preveri, so:

$$\mathrm{E}[X-Y] = \mathrm{E}[X] - \mathrm{E}[Y]$$
$$\mathrm{var}[X-Y] = \mathrm{var}[X] + \mathrm{var}[Y] - 2\,\mathrm{cov}[X,Y]$$

in za nekorelirani oziroma neodvisni slučajni spremenljivki

$$\mathrm{var}[X-Y] = \mathrm{var}[X] + \mathrm{var}[Y]. \tag{6.31}$$

Vidimo, da je varianca razlike dveh neodvisnih slučajnih spremenljivk vsota varianc teh dveh slučajnih spremenljivk.

**Primer 6.6:** *Dokažimo, da vrednost korelacijskega koeficienta $\rho_{XY}$ vedno leži med $-1$ in $1$!*

**Rešitev:** Vzemimo, da je slučajna spremenljivka $Z$ enaka

$$Z = X + t\,Y,$$

kjer je $t$ poljubno realno število. Varianca slučajne spremenljivke $Z$ je po enačbi (6.29) enaka

$$\mathrm{var}[Z] = \mathrm{var}[X] + 2t\,\mathrm{cov}[X,Y] + t^2\,\mathrm{var}[Y] = f(t) \geq 0.$$

Ta enačba predstavlja kvadratno funkcijo $f(t)$, ki mora biti za poljubni $t$ nenegativna, kar tudi pomeni, da ne sme imeti dveh realnih korenov. Pogoj za to je, da je diskriminanta enačbe manjša ali enaka nič:

$$\begin{aligned}
& 4\,\mathrm{cov}[X,Y] - 4\,\mathrm{var}[X]\,\mathrm{var}[Y] \leq 0 \quad \rightarrow \quad \mathrm{cov}[X,Y]^2 \leq \mathrm{var}[X]\,\mathrm{var}[Y] \\
\rightarrow \quad & \sqrt{\mathrm{cov}[X,Y]^2} \leq \sqrt{\mathrm{var}[X]\,\mathrm{var}[Y]} \quad \rightarrow \quad |\mathrm{cov}[X,Y]| \leq \sigma_X\,\sigma_Y \\
\rightarrow \quad & -\sigma_X\,\sigma_Y \leq \mathrm{cov}[X,Y] \leq \sigma_X\,\sigma_Y \quad \rightarrow \quad -1 \leq \frac{\mathrm{cov}[X,Y]}{\sigma_X\,\sigma_Y} \leq 1 \\
\rightarrow \quad & -1 \leq \rho_{XY} \leq 1
\end{aligned}$$

**Primer 6.7:** *Določimo pričakovano vrednost in varianco slučajne spremenljivke $Z$, ki je produkt slučajnih spremenljivk $X$ in $Y$*

$$Z = X\,Y.$$

**Rešitev:** Pričakovano vrednost slučajne spremenljivke $\mathrm{E}[XY]$ določimo neposredno iz enačbe (6.27)

$$\mathrm{E}[Z] = \mathrm{E}[XY] = \mathrm{cov}[X,Y] + \mathrm{E}[X]\,\mathrm{E}[Y].$$

Pričakovana vrednost produkta dveh slučajnih spremenljivk je torej odvisna tudi od kovariance med tema dvema slučajnima spremenljivkama. Le v primeru, da je kovarianca enaka nič, velja preprosta zveza, ki pravi, da je pričakovana vrednost produkta dveh slučajnih spremenljivk produkt pričakovanih vrednosti

$$\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]. \tag{6.32}$$

Varianca slučajne spremenljivke $Z$ je

$$\mathrm{var}[Z] = \mathrm{E}[Z^2] - \mathrm{E}[Z]^2 = \mathrm{E}[X^2\,Y^2] - \mathrm{E}[XY]^2. \tag{6.33}$$

Zadnja enačba vključuje moment četrtega reda $(2+2)$ slučajnega vektorja $X, Y$, ki ga lahko izračunamo le v primeru, da poznamo porazdelitev tega slučajnega vektorja. Naloga se nekoliko poenostavi, če sta slučajni spremenljivki $X$ in $Y$ neodvisni. Tedaj sta neodvisni tudi slučajni spremenljivki $X^2$ in $Y^2$. Zato lahko moment četrtega reda po enačbi (6.32) zapišemo kot produkt dveh momentov drugega reda

$$\mathrm{E}[X^2\,Y^2] = \mathrm{E}[X^2]\,\mathrm{E}[Y^2]. \tag{6.34}$$

Iz enačbe (6.15) sledi zveza

$$\mathrm{E}[X^2] = \sigma_X^2 + m_X^2,$$

ki jo skupaj z enačbama (6.32) in (6.34) vstavimo v enačbo (6.33) in po preureditvi dobimo

$$\mathrm{var}[Z] = \sigma_X^2\,\sigma_Y^2 + \sigma_X^2\,m_Y^2 + m_X^2\,\sigma_Y^2.$$

Pri tem primeru se je ponovno pokazala razlika med neodvisnostjo in nekoreliranostjo. Zato, da smo lahko izpeljali enačbo za določitev variance produkta dveh slučajnih spremenljivk, nekoreliranost ni bil zadosten pogoj. Slučajni spremenljivki sta morali biti neodvisni!

# 7 Verjetnostne porazdelitve

V prejšnjih poglavjih smo vpeljali pojme porazdelitev, porazdelitvenih funkcij, momentov porazdelitev in podobno. Razen v računskih primerih pa posebnosti posameznih porazdelitev nismo obravnavali. V naslednjem poglavju bomo opisali različne porazdelitve. Najprej bomo opisali najpreprostejšo diskretno in zvezno porazdelitev – enakomerna porazdelitev. Nato bomo obravnavali skupino porazdelitev, ki izhajajo iz preprostega poskusa z dvema možnima izidoma – Bernoullijev poskus. Na osnovi Bernoullijevega poskusa bomo izpeljali binomsko, geometrično, Pascalovo, Poissonovo, eksponentno porazdelitev in porazdelitev gama ($\Gamma$). Nadaljevali bomo z opisom limitnih porazdelitev, katere najpomembnejša in najbolj znana je normalna oziroma Gaussova porazdelitev. Druge limitne porazdelitve so še logaritemsko normalna in porazdelitve ekstremnih vrednosti. Z normalno porazdelitvijo so povezane pogosto uporabljane porazdelitve: porazdelitev $\chi^2$, študentova porazdelitev $t$ in porazdelitev $F$.

## 7.1 Enakomerna porazdelitev

### 7.1.1 Enakomerna diskretna porazdelitev

To najpreprostejšo porazdelitev smo v računskih primerih obravnavali že v prejšnjih poglavjih (glej primere 4.2, 5.1, 5.2 in 6.1). Zalogo vrednosti predstavlja množica $n$ števil $x_i$, ki so enakomerno razporejene, tako da je

$$x_{i+1} - x_i = d$$

za vse vrednosti $i = 1, \ldots, n-1$. Verjetnostna funkcija ima v tem primeru enake vrednosti za vse $x_i, i = 1, \ldots, n$, ki tvorijo zalogo vrednosti

$$p_X(x_i) = p = \frac{1}{n}, \qquad i = 1, \ldots, n.$$

Pogosto vzamemo, da je $x_i = i$. V tem primeru je pričakovana vrednost enaka

$$\mathrm{E}[X] = \frac{n+1}{2},$$

variacija je

$$\mathrm{var}[X] = \frac{n^2-1}{12},$$

koeficienta asimetrije in sploščenosti pa sta

$$\gamma_{1X} = 0, \qquad \gamma_{2X} = \frac{3(3n^2 - 7)}{5(n^2 - 1)}.$$

Ta preprosta porazdelitev je zelo pomembna porazdelitev pri teoriji iger na srečo, kjer so bolj zapleteni dogodki (na primer: met več kock, igra s kartami) sestavljeni iz zelo preprostih dogodkov, kjer so vsi izidi enako verjetni (na primer: met kocke, dvig posamezne karte, rezultat rulete).

### 7.1.2 Enakomerna zvezna porazdelitev

Tudi to porazdelitev smo že obravnavali (glej primera 4.4 in 5.5). V primeru 4.4 smo zapisali gostoto verjetnosti (4.9) in porazdelitveno funkcijo (4.10):

$$f_X(x) = \begin{cases} 0, & x < a, \\ \dfrac{1}{b-a}, & a \le x < b, \\ 0, & b \le x. \end{cases}$$

in

$$F_X(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \le x < b, \\ 1, & b \le x. \end{cases}$$

Osnovne momente (pričakovano vrednost, standardno deviacijo, koeficient asimetrije in koeficient kurtosis) izračunamo po enačbah (6.3)–(6.9)

$$\mathrm{E}[X] = m_X = \frac{a+b}{2},$$

$$\mathrm{var}[X] = \sigma_X^2 = \frac{(b-a)^2}{12},$$

$$\gamma_{1X} = 0,$$

$$\gamma_{2X} = 1.8.$$

To porazdelitev lahko uporabimo, če je gostota verjetnosti, da slučajna spremenljivka zasede poljubno vrednost na območju od $a$ do $b$, enaka. Pri inženirskih problemih so slučajne spremenljivke običajno neenakomerno porazdeljene. Včasih enakomerno porazdeljeno slučajno spremenljivko uporabijo zaradi neznanja ali nepoznavanja problema, ki ga obravnavajo.

Kljub vsemu je to zelo pomembna porazdelitev, saj s pojmom *generiranje slučajnega števila* najpogosteje mislimo prav na generiranje enakomerno porazdeljene slučajne spremenljivke, kjer je $a = 0$ in $b = 1$. Če v računalniškem programu uporabimo ukaz `Random` ali temu ukazu ustrezen ukaz, generiramo slučajno spremenljivko, ki je enakomerno porazdeljena od nič do ena. Tako generirana slučajna spremenljivka je osnova za generiranje poljubno porazdeljene slučajne spremenljivke. Kako se generira poljubno porazdeljeno slučajno spremenljivko, bomo obravnavali v poglavju Simulacije.

## 7.2 Porazdelitve, ki izvirajo iz Bernoullijevega poskusa

Bernoullijev[1] poskus je preprost poskus, ki ima dva možna izida. Rezultat Bernoullijevega poskusa je slučajna spremenljivka $X$, porazdeljena po Bernoullijevi porazdelitvi. Zaloga vrednosti vključuje dve vrednosti, ki ju običajno označimo z $0$ in $1$. Očitno je torej, da moramo to slučajno spremenljivko razvrstiti med diskretne slučajne spremenljivke. Po pomenu lahko vrednosti take slučajne spremenljivke predstavljajo: belo – črno, neuspeh – uspeh, številka – grb pri metu kovanca, visoko – nizko... Verjetnost, da se zgodi uspeh oziroma, da ima slučajna spremenljivka $X$ vrednost $1$, označimo s $p$. Verjetnost neuspeha oziroma, da ima slučajna spremenljivka vrednost 0, pa označimo s $q = 1 - p$.

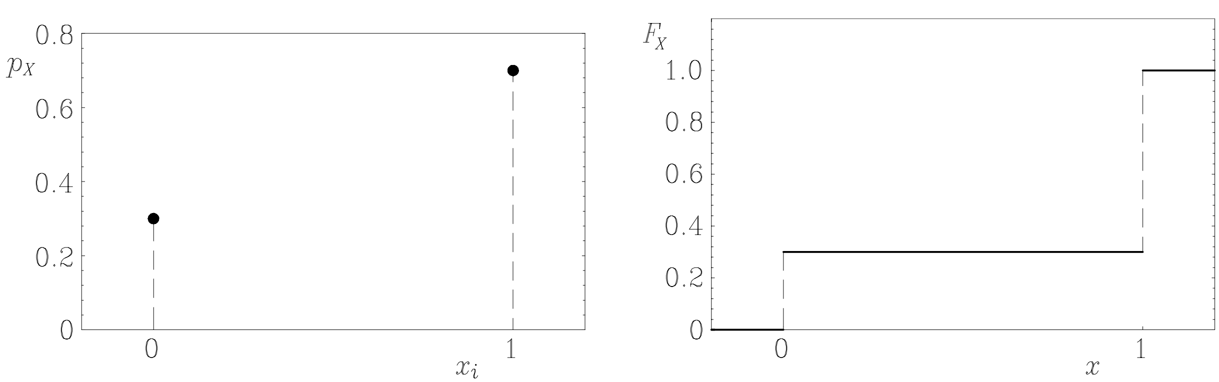
Verjetnostna funkcija slučajne spremenljivke $X$, ki je porazdeljena po Bernoullijevi porazdelitvi, je

$$p_X(x_i) = \begin{cases} (1-p) & X = x_1 = 0 \\ p & X = x_2 = 1. \end{cases}$$

Porazdelitvena funkcija slučajne spremenljivke, porazdeljene po Bernoullijevi porazdelitvi, je

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1-p & 0 < x \leq 1 \\ 1 & 1 < x. \end{cases}$$

Verjetnostno in porazdelitveno funkcijo slučajne spremenljivke, porazdeljene po Bernoullijevi porazdelitvi, prikazujemo na sliki (7.1).




Slika 7.1: Verjetnostna funkcija in porazdelitvena funkcija Bernoullijeve porazdelitve

Pričakovano vrednost in varianco slučajne spremenljivke, porazdeljene po Bernoullijevi porazdelitvi,

[1] Jacob (Jacques) Bernoulli, švicarski matematik in astronom (1654-1705).

izračunamo po enačbah (6.3) in (6.15)

$$\mathrm{E}[X] = \sum_{i=1}^{2} x_i\, p_X(x_i) = 0 \cdot (1-p) + 1 \cdot p = p$$
$$\mathrm{E}[X^2] = \sum_{i=1}^{2} x_i^2\, p_X(x_i) = 0^2 \cdot (1-p) + 1^2 \cdot p = p$$
$$\mathrm{var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = p - p^2 = p\,(1-p)$$

**Primer 7.1:** *Pri kolikšni vrednosti $p$ je varianca slučajne spremenljivke, porazdeljene po Bernoullijevi porazdelitvi, največja? Določimo pričakovano vrednosti in varianco za ta primer.*

**Rešitev**: Ekstremno vrednost variance določimo z odvajanjem $\mathrm{var}[X]$ po parametru $p$:

$$\frac{d\mathrm{var}[X]}{dp} = 1 - 2p = 0 \qquad \longrightarrow \qquad p_{max} = \frac{1}{2}.$$

Da je ta vrednost res maksimalna vrednost sledi iz dejstva, da je drugi odvod $d^2\mathrm{var}/dp^2$ manjši od nič. Vrednosti pričakovane vrednosti in variance sta v tem primeru

$$\mathrm{E}[X] = \frac{1}{2}, \qquad \mathrm{var}[X] = \frac{1}{4},$$

Poleg tega lahko opazimo, da je vrednost variance za $p = 0$ in $p = 1$ enaka nič. V teh dveh primerih pravzaprav nimamo opravka s slučajno spremenljivko, temveč z dvema določenima (determinističnima) konstantama, v prvem primeru je to vrednost 0, v drugem pa 1.

### 7.2.1 Binomska porazdelitev

Vzemimo, da opravimo oziroma opazujemo več zaporednih Bernoullijevih poskusov. Osnovni predpostavki, ki ju pri tem upoštevamo, sta: poskusi so medsebojno neodvisni, verjetnost uspeha se pri ponavljanju ne spreminja.

Če slučajna spremenljivka $Y$ predstavlja **število uspehov v $n$ ponovitvah Bernoullijevega poskusa**, pravimo, da je porazdeljena **binomsko**. Zaloga vrednosti te slučajne spremenljivke so vsa cela števila od 0 do $n$, saj se v $n$ ponovitvah poskusov ne more zgoditi manj kot nič uspehov in ne več kot $n$ uspehov. Pri določitvi verjetnostne funkcije te slučajne spremenljivke $Y$ nas torej zanima, kolikšna je verjetnost, da se je v $n$ ponovitvah Bernoullijevega poskusa zgodilo $y$ uspehov

$$p_Y(y_i) = P[Y = y_i], \qquad y_i = 0, 1, \ldots, n.$$

Opišimo nekaj primerov binomsko porazdeljene slučajne spremenljivke:

– Kovanec vržemo $n$ krat in opazujemo, ali pade grb ali številka. Število metov, ko je padel grb, označimo z $Y$, ki je slučajna spremenljivka, porazdeljena po binomski porazdelitvi. Pri poštenem kovancu je verjetnost, da pade grb, enaka $p = 0.5$.

– Glede na opazovanja v preteklosti, lahko ocenimo verjetnost $p$, da bo v enem letu na nekem območju prišlo do vsaj ene poplave. Slučajna spremenljivka $Y$, ki predstavlja število poplavnih let v obdobju naslednjih $n$ let, je binomsko porazdeljena slučajna spremenljivka.

– Verjetnost, da nek določen košarkar zadene prosti met je $p$. Ob predpostavki, da se ta verjetnost s časom ne spreminja in da je vsak met neodvisen od vseh drugih, lahko rečemo, da je slučajna spremenljivka $Y$ število zadetih prostih metov v $n$ poskusih, kolikor jih je imel naš košarkar na voljo v eni tekmi.

Preden izpeljemo verjetnostno funkcijo binomsko porazdeljene slučajne spremenljivke, rešimo preprosto nalogo.

**Primer 7.2:** *Vzemimo, da opravimo štiri enake in neodvisne Bernoullijeve poskuse, pri katerih je verjetnost uspeha enaka p=0.3. Označimo slučajne spremenljivke, ki predstavljajo rezultat teh poskusov z* $X_1$, $X_2$, $X_3$ *in* $X_4$. *Določimo verjetnosti, da je število uspehov* $Y$ *v štirih ponovitvah enako 0, 1, 2, 3 in 4, kar predstavlja zalogo vrednosti slučajne spremenljivke* $Y$*!*

**Rešitev**: Najprej določimo verjetnost, da je $Y = 0$. Ta dogodek je produkt štirih neodvisnih dogodkov in sicer, da je $X_1 = 0$, $X_2 = 0$, $X_3 = 0$ in $X_4 = 0$. Slučajna spremenljivka $Y$ dobi vrednost nič le v primeru, da imajo vse štiri slučajne spremenljivke $X_i$ vrednost nič. Zaradi neodvisnosti dogodkov $X_i = 0$ je verjetnost produkta teh dogodkov produkt verjetnosti. Zato dobimo:

$$\begin{aligned} P[Y=0] &= P[(X_1=0)\cap(X_2=0)\cap(X_3=0)\cap(X_4=0)] = \\ &= P[X_1=0]\,P[X_2=0]\,P[X_3=0]\,P[X_4=0] = (1-p)^4 = 0.2401 \end{aligned}$$

Dogodek $Y = 1$ je sestavljen dogodek iz dogodka, da je $X_1 = 1$, drugi pa so enaki nič, dogodka, da je $X_2 = 1$, ostali pa so enaki nič in tako dalje.

$$\begin{aligned} P[Y=1] &= P[((X_1=1)\cap(X_2=0)\cap(X_3=0)\cap(X_4=0))\cup \\ &\quad \cup((X_1=0)\cap(X_2=1)\cap(X_3=0)\cap(X_4=0))\cup \\ &\quad \cup((X_1=0)\cap(X_2=0)\cap(X_3=1)\cap(X_4=0))\cup \\ &\quad \cup((X_1=0)\cap(X_2=0)\cap(X_3=0)\cap(X_4=1))]. \end{aligned}$$

Ker so si sestavljeni dogodki, ki jih pri računu verjetnosti $P[Y = 1]$ seštejemo, medsebojno nezdružljivi, lahko verjetnost vsote zapišem kot vsoto verjetnosti

$$\begin{aligned} P[Y=1] &= P[(X_1=1)\cap(X_2=0)\cap(X_3=0)\cap(X_4=0)] + \\ &\quad + P[(X_1=0)\cap(X_2=1)\cap(X_3=0)\cap(X_4=0)] + \\ &\quad + P[(X_1=0)\cap(X_2=0)\cap(X_3=1)\cap(X_4=0)] + \\ &\quad + P[(X_1=0)\cap(X_2=0)\cap(X_3=0)\cap(X_4=1)]. \end{aligned}$$

Tako kot pri določitvi verjetnosti $P[Y = 0]$, so tudi sedaj posamezni dogodki v produktu neodvisni, zaradi česar lahko verjetnost produkta teh dogodkov zapišemo kot produkt verjetnosti

$$\begin{aligned} P[Y=1] &= P[X_1=1]\,P[X_2=0]\,P[X_3=0]\,P[X_4=0]\,+ \\ &+ P[X_1=0]\,P[X_2=1]\,P[X_3=0]\,P[X_4=0]\,+ \\ &+ P[X_1=0]\,P[X_2=0]\,P[X_3=1]\,P[X_4=0]\,+ \\ &+ P[X_1=0]\,P[X_2=0]\,P[X_3=0]\,P[X_4=1] = 4\cdot(1-p)^3\cdot p = 0.4116. \end{aligned} \tag{7.1}$$

Na podoben način določimo tudi verjetnosti, da je $Y = 2$, $Y = 3$ in $Y = 4$.

$$\begin{aligned} P[Y=2] &= P[((X_1=1)\cap(X_2=1)\cap(X_3=0)\cap(X_4=0))\,\cup \\ &\quad \cup((X_1=1)\cap(X_2=0)\cap(X_3=1)\cap(X_4=0))\,\cup \\ &\quad \cup((X_1=1)\cap(X_2=0)\cap(X_3=0)\cap(X_4=1))\,\cup \\ &\quad \cup((X_1=0)\cap(X_2=1)\cap(X_3=1)\cap(X_4=0))\,\cup \\ &\quad \cup((X_1=0)\cap(X_2=1)\cap(X_3=0)\cap(X_4=1))\,\cup \\ &\quad \cup((X_1=0)\cap(X_2=0)\cap(X_3=1)\cap(X_4=1))] = \\ &= P[X_1=1]\,P[X_2=1]\,P[X_3=0]\,P[X_4=0]\,+ \\ &+ P[X_1=1]\,P[X_2=0]\,P[X_3=1]\,P[X_4=0]\,+ \\ &+ P[X_1=1]\,P[X_2=0]\,P[X_3=0]\,P[X_4=1]\,+ \\ &+ P[X_1=0]\,P[X_2=1]\,P[X_3=1]\,P[X_4=0]\,+ \\ &+ P[X_1=0]\,P[X_2=1]\,P[X_3=0]\,P[X_4=1]\,+ \\ &+ P[X_1=0]\,P[X_2=0]\,P[X_3=1]\,P[X_4=1] = 6\cdot(1-p)^2\cdot p^2 = 0.2646, \end{aligned}$$

$$\begin{aligned} P[Y=3] &= P[((X_1=1)\cap(X_2=1)\cap(X_3=1)\cap(X_4=0))\,\cup \\ &\quad \cup((X_1=1)\cap(X_2=1)\cap(X_3=0)\cap(X_4=1))\,\cup \\ &\quad \cup((X_1=1)\cap(X_2=0)\cap(X_3=1)\cap(X_4=1))\,\cup \\ &\quad \cup((X_1=0)\cap(X_2=1)\cap(X_3=1)\cap(X_4=1))] = \\ &= P[X_1=1]\,P[X_2=1]\,P[X_3=1]\,P[X_4=0]\,+ \\ &+ P[X_1=1]\,P[X_2=1]\,P[X_3=0]\,P[X_4=1]\,+ \\ &+ P[X_1=1]\,P[X_2=0]\,P[X_3=1]\,P[X_4=1]\,+ \\ &+ P[X_1=0]\,P[X_2=1]\,P[X_3=1]\,P[X_4=1] = 4\cdot(1-p)\cdot p^3 = 0.0756, \end{aligned}$$

$$\begin{aligned} P[Y=4] &= P[(X_1=1)\cap(X_2=1)\cap(X_3=1)\cap(X_4=1)] = \\ &= P[X_1=1]\,P[X_2=1]\,P[X_3=1]\,P[X_4=1] = p^4 = 0.0081. \end{aligned}$$

Verjetnostno in porazdelitveno funkcijo slučajne spremenljivke, ki je porazdeljena binomsko, prikazujemo na sliki 7.2.

Iz izrazov, ki smo jih določili v zadnjem računskem primeru, lahko zapišemo splošni izraz za verjetnostno funkcijo binomsko porazdeljene slučajne spremenljivke $Y$

$$p_Y(y) = p^y (1-p)^{n-y} \binom{n}{y}, \qquad y = 0, 1, \ldots, n \tag{7.2}$$

Verjetnost, da se v $n$ ponovitvah Bernoullijevega poskusa zgodi $y$ uspehov, je torej produkt $p^y$, ki predstavlja verjetnost $y$ uspehov, $(1-p)^{n-y}$, ki predstavlja verjetnost $n-y$ neuspehov, ter $\binom{n}{y}$, ki predstavlja število razporeditev $y$ uspehov v $n$ poskusih. Na primer: v enačbi (7.1) smo zapisali, da se lahko en uspeh v štirih poskusih zgodi na štiri ($\binom{4}{1} = 4$) načine, v enačbi (7.1) pa, da se lahko dva uspeha v štirih poskusih zgodita na šest ($\binom{4}{2} = 6$) načinov. Število razporeditev določimo z binomskim koeficientom $\binom{n}{y}$, ki ga lahko izračunamo po naslednji enačbi

$$\binom{n}{y} = \frac{n!}{y!\,(n-y)!} = \frac{n\,(n-1)\cdots(n-y+1)}{y\,(y-1)\cdots 1}. \tag{7.3}$$

Binomske koeficiente za manjše vrednosti $n$ lahko določimo tudi z uporabo Pascalovega trikotnika,[2] ki ga tvorimo tako, da na robovih zapišemo enice, vmesne člene pa izračunamo kot vsoto sosednjih dveh členov v prejšnji vrstici:

$$\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array} \tag{7.4}$$

Iz Pascalovega trikotnika v (7.4) lahko na primer odčitamo, da je $\binom{4}{2} = 6$, $\binom{5}{2} = 10$ ali $\binom{6}{3} = 20$.
Število uspehov $Y$ v $n$ Bernoullijevih poskusih lahko smatramo tudi kot vsoto $n$ Bernoullijevo porazdeljenih slučajnih spremenljivk $X_i$.

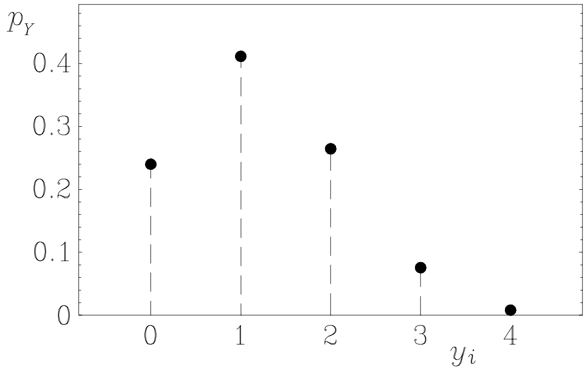
$$Y = \sum_{i=1}^{n} X_i.$$

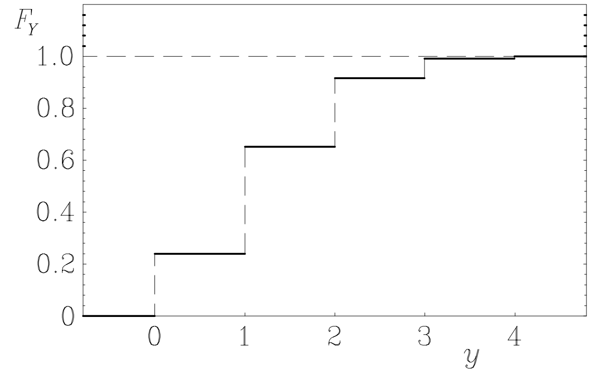
Pričakovano vrednost in varianco vsote slučajnih spremenljivk izračunamo po enačbah (6.28) in (6.31), pri tem pa upoštevamo, da so Bernoullijevi poskusi medsebojno neodvisni:

$$\mathrm{E}[Y] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} p = n\,p, \tag{7.5}$$

$$\mathrm{var}[Y] = \mathrm{var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{var}[X_i] = \sum_{i=1}^{n} p\,(1-p) = n\,p\,(1-p). \tag{7.6}$$

[2] Blaise Pascal francoski matematik (1623–1662).




Slika 7.2: Verjetnostna in porazdelitvena funkcija binomske porazdelitve za $n = 4$ in $p = 0.3$

**Primer 7.3:** *Izpeljimo enačbi za pričakovano vrednost in varianco binomsko porazdeljene slučajne spremenljivke po enačbah (6.3) in (6.6)!*

**Rešitev:** Enačbo (7.2) vstavimo v enačbo (6.3) in upoštevamo izraz za binomski koeficient

$$\mathrm{E}[Y] = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y} = \sum_{y=0}^{n} y \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}.$$

Vsoto od $y = 0$ do $n$ lahko spremenimo v vsoto od $y = 1$ do $n$, saj je sumand pri $y = 0$ enak nič. Iz enačbe izpostavimo $np$ okrajšamo $y$ in uredimo:

$$\mathrm{E}[Y] = \sum_{y=1}^{n} y \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} = np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!(n-y)!} p^{y-1} (1-p)^{n-y}.$$

Sedaj lahko vpeljemo novi oznaki $z = y - 1$ in $m = n - 1$ in zapišemo

$$\mathrm{E}[Y] = n\,p \sum_{z=0}^{m} \frac{m!}{z!(m-z)!} p^z (1-p)^{m-z}.$$

Vidimo lahko, da je v vsoti zapisana verjetnostna funkcija slučajne spremenljivke $Z$, ki je porazdeljena binomsko od 0 do $m$. Vsota vseh vrednosti verjetnostne funkcije slučajne spremenljivke $Z$ je enaka ena. Zato lahko zapišemo končni rezultat:

$$\mathrm{E}[Y] = np.$$

Za določitev $\mathrm{var}[Y]$ bomo najprej izračunali $\mathrm{E}[Y^2]$ na podoben način kot $\mathrm{E}[Y]$ z vpeljavo novih oznak $z = y - 1$ in $m = n - 1$

$$
\begin{aligned}
\mathrm{E}[Y^2] &= \sum_{y=0}^{n} y^2 \, \frac{n!}{y!(n-y)!} \, p^y (1-p)^{n-y} = \sum_{y=1}^{n} y^2 \, \frac{n!}{y!(n-y)!} \, p^y (1-p)^{n-y} = \\
&= np \sum_{y=1}^{n} y \, \frac{n-1!}{(y-1)!(n-y)!} \, p^{y-1} (1-p)^{n-y} = \\
&= np \sum_{z=0}^{m} (z+1) \, \frac{m!}{z!(m-z)!} \, p^z (1-p)^{m-z} = \\
&= np \left( \sum_{z=0}^{m} z \, \frac{m!}{z!(m-z)!} \, p^z (1-p)^{m-z} + \sum_{z=0}^{m} \frac{m!}{z!(m-z)!} \, p^z (1-p)^{m-z} \right) \\
&= np \, (mp+1) = np \Big( (n-1) \, p + 1 \Big) = n^2 p^2 - np^2 + np.
\end{aligned}
$$

Če v enačbi (6.15) upoštevamo izraza, ki smo jih izpljeli za $\mathrm{E}[Y]$ in $\mathrm{E}[Y^2]$, dobimo

$$
\mathrm{var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 = n^2 \, p^2 - n \, p^2 + n \, p - (n \, p)^2 = n \, p \, (1-p).
$$

### 7.2.2 Geometrična porazdelitev

Vzemimo, da opravimo oziroma opazujemo zaporedne Bernoullijeve poskuse. Osnovni predpostavki, ki ju pri tem upoštevamo, sta enaki kot pri binomski porazdelitvi: poskusi so medseboj neodvisni, verjetnost uspeha se pri ponavljanju ne spreminja. Definirajmo slučajno spremenljivko $N$, ki predstavlja **zaporedno številko** Bernoullijevega poskusa, pri katerem se je zgodil **prvi uspeh**.

Opišimo nekaj primerov geometrično porazdeljene slučajne spremenljivke:

- Vzemimo, da je verjetnost, da bo v enem letu na nekem območju prišlo do vsaj ene poplave, enaka $p$. Slučajna spremenljivka $N$ predstavlja zaporedno leto od začetka opazovanja, ko se poplava pojavi prvič.

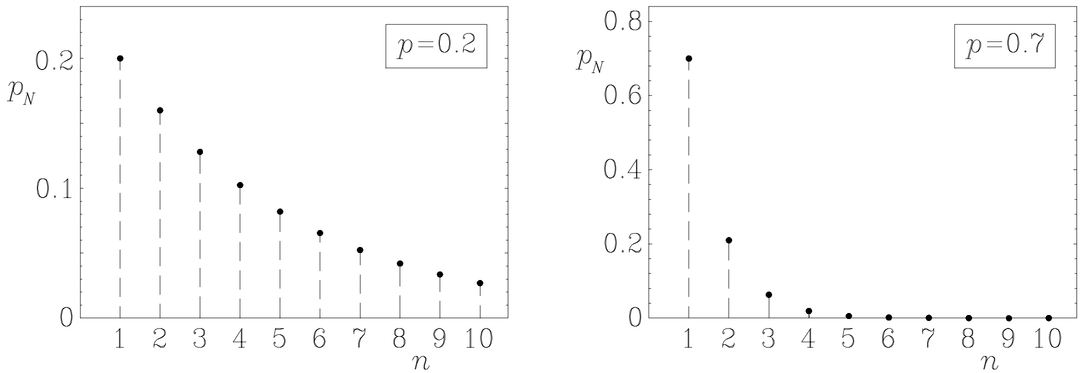
- Verjetnost, da nek določen košarkar zadene prosti met je $p$. Ob predpostavki, da se ta verjetnost s časom ne spreminja in da je vsak met neodvisen od vseh drugih, lahko rečemo, da je slučajna spremenljivka $N$ zaporedni prosti met, ko je naš košarkar prvič zadel.

Zaloga vrednosti geometrično porazdeljene slučajne spremenljivke so vsa naravna števila, saj se prvi poskus lahko zgodi najprej v prvem poskusu. Zaporedna številka poskusa, v katerem se zgodi prvi uspeh, navzgor ni omejena.

Izpeljava verjetnostne funkcije geometrično porazdeljene slučajne spremenjivke $p_N(n)$ je preprosta. Dogodek, da se prvi uspeh zgodi v $n$-tem poskusu je enak, dogodku, da se v prvih $(n-1)$ poskusih zgodijo sami neuspehi, v n-tem poskusu pa se zgodi uspeh. Ob upoštevanju neodvisnosti poskusov z enačbo to zapišemo takole:

$$
P[N=n] = (1-p)^{n-1} \, p, \qquad n = 1, 2 \ldots \tag{7.7}
$$

Verjetnostno funkcijo geometrično porazdeljene slučajne spremenjivke $N$ za dve različni vrednosti $p$ prikazujemo na sliki 7.3. Vidimo, da vrednost verjetnostne funkcije pada tako za nizke kot za visoke vrednosti $p$.




Slika 7.3: Verjetnostna funkcija geometrične porazdelitve za $p = 0.2$ in $p = 0.7$

Pričakovana vrednost in varianca geometrično porazdeljene slučajne spremenljivke $N$ sta

$$\mathrm{E}[N] = \frac{1}{p} \qquad\qquad \mathrm{var}[N] = \frac{1-p}{p^2}.$$

**Primer 7.4:** *Določimo pričakovano vrednost in varianco geometrično porazdeljene slučajne spremenljivke $N$ po enačbah (6.3) in (6.6)!*

**Rešitev:** Enačbo (7.7) vstavimo v enačbo (6.3)

$$\mathrm{E}[N] = \sum_{n=1}^{\infty} n\,(1-p)^{n-1}\,p = p\sum_{n=1}^{\infty} n\,(1-p)^{n-1}. \tag{7.8}$$

Ker varianco $\mathrm{var}[N]$ določamo po enačbi (6.15), zapišimo še pričakovano vrednost $\mathrm{E}[N^2]$:

$$\mathrm{E}[N^2] = \sum_{n=1}^{\infty} n^2\,(1-p)^{n-1}\,p = p\sum_{n=1}^{\infty} n^2\,(1-p)^{n-1}. \tag{7.9}$$

Vrednosti neskončnih vrst, ki ju vsebujeta zgornji dve enačbi, določimo iz preproste geometrične vrste (glej Bronštejn, Semendjajev, Musiol, Mühlig, Matematični priročnik, Tehniška založba Slovenije, 1997, str. 15.)

$$\sum_{n=0}^{\infty} q^n = \lim_{m\to\infty}\frac{1-q^{m+1}}{1-q} = \frac{1}{1-q}. \tag{7.10}$$

Če enačbo (7.10) odvajamo po $q$, dobimo izraz za vrsto, ki jo moramo določiti v enačbi (7.8)

$$\left(\sum_{n=0}^{\infty} q^n\right)' = \left(\frac{1}{1-q}\right)' \qquad\longrightarrow\qquad \sum_{n=1}^{\infty} n\,q^{n-1} = \frac{1}{(1-q)^2}. \tag{7.11}$$

Enačbo (7.11) pomnožimo s $q$ in ponovno odvajamo po $q$ ter dobimo izraz za vrsto, ki jo uporabimo v enačbi (7.9)

$$\left(\sum_{n=1}^{\infty} n\, q^n\right)' = \left(\frac{q}{(1-q)^2}\right)' \qquad \longrightarrow \qquad \sum_{n=1}^{\infty} n^2\, q^{n-1} = \frac{1+q}{(1-q)^3}. \tag{7.12}$$

Pričakovano vrednost določimo tako, da enačbo (7.11) vstavimo v (7.8) in upoštevamo, da je $q = 1-p$ oziroma $p = 1-q$

$$\mathrm{E}[N] = p\,\frac{1}{p^2} = \frac{1}{p}.$$

Varianco izračunamo po enačbi (6.15), v katero vstavimo izraza za vrsti (7.11) in (7.12)

$$\mathrm{var}[N] = \mathrm{E}[N^2] - \mathrm{E}[N]^2 = p\,\frac{1+(1-p)}{p^3} - \left(\frac{1}{p}\right)^2 = \frac{2-p-1}{p^2} = \frac{1-p}{p^2}.$$

### 7.2.3 Pascalova porazdelitev

Tudi pri obravnavanju Pascalove[3] oziroma negativne binomske porazdelitve opravimo oziroma opazujemo zaporedne Bernoullijeve poskuse. Osnovni predpostavki, ki ju pri tem upoštevamo, sta enaki kot pri binomski porazdelitvi. Slučajna spremenljivka $N_k$, ki predstavlja **zaporedno številko** Bernoullijevega poskusa, pri katerem se je zgodil ***k*-ti uspeh**, je porazdeljena po Pascalovi oziroma negativni binomski porazdelitvi.

Zaloga vrednosti te slučajne spremenljivke so vsa naravna števila večja ali enaka $k$, saj se $k$-ti uspeh ne more zgoditi prej kot v $k$-tem poskusu.

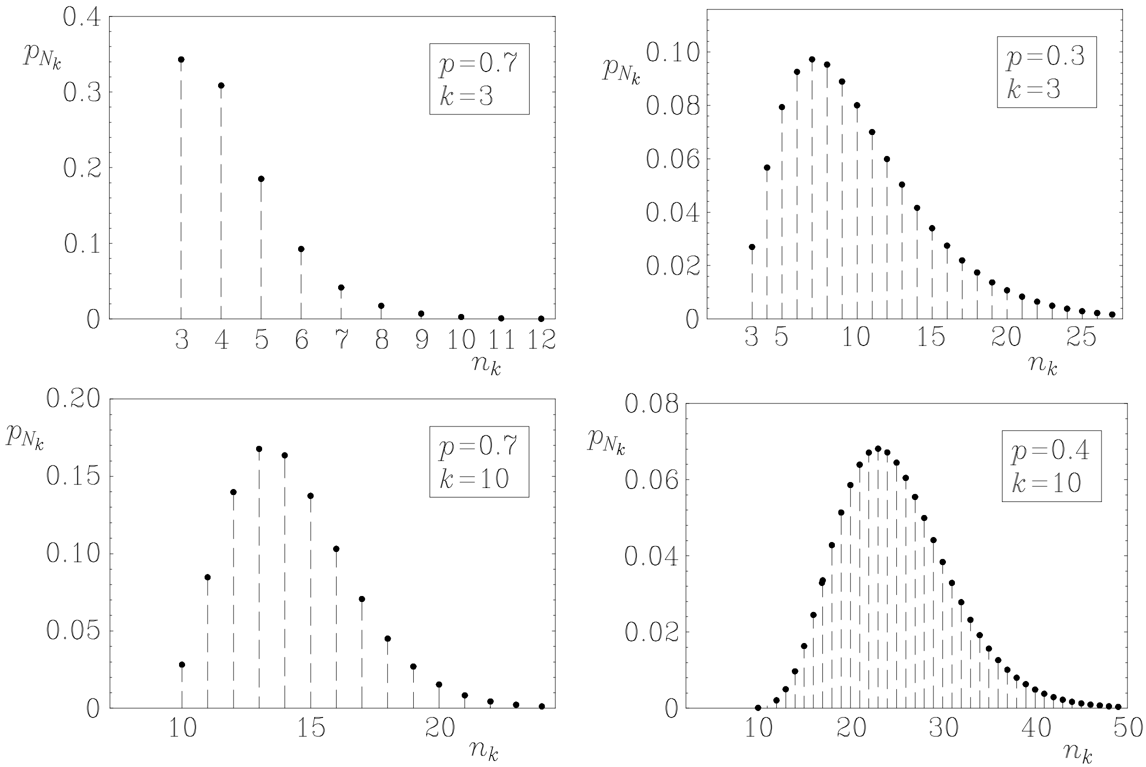
Izpeljavo verjetnostne funkcije slučajne spremenljivke $N_k$ naredimo podobno kot pri geometrični porazdelitvi. Dogodek, da se $k$-ti uspeh zgodi v $n_k$-tem poskusu je enak dogodku, da se v prvih $n_k - 1$ poskusih zgodi natanko $k-1$ uspehov in se v $n_k$-tem poskusu uspeh zgodi. Dogodek, da se v prvih $n_k - 1$ poskusih zgodi natanko $k-1$ uspehov opisuje binomsko porazdeljena slučajna spremenljivka. Ker so poskusi medsebojno neodvisni, je verjetnostna funkcija enaka

$$\begin{aligned} P[N_k = n_k] &= \binom{n_k-1}{k-1}(1-p)^{n_k-1-(k-1)}\,p^{k-1}\,p = \\ &= \binom{n_k-1}{k-1}(1-p)^{n_k-k}\,p^k, \qquad n_k = k, k+1, \ldots. \end{aligned} \tag{7.13}$$

Pascalova porazdelitev za $k=1$ je enaka geometrični porazdelitvi, kar lahko zaključimo iz same definicije obeh slučajnih spremenljivk, kot tudi s primerjavo verjetnostnih funkcij (7.7) in (7.13).

Verjetnostno funkcijo slučajne spremenjivke $N_k$, ki je porazdeljena po Pascalovi porazdelitvi, za štiri

[3] Blaise Pascal, francoski matematik (1623–1662).




Slika 7.4: Verjetnostna funkcija Pascalove porazdelitve za štiri različne vrednosti $k$ in $p$

različne vrednosti parametrov $p$ in $k$ prikazujemo na sliki 7.4. Očitno je lahko oblika za različne vrednosti parametrov zelo različna.

Zaporedna številka poskusa, v katerem se zgodi $k$-ti uspeh lahko obravnavamo kot vsoto $k$ geometrično porazdeljenih slučajnih spremenljivk

$$N_k = \sum_{i=1}^{k} N_i,$$

kjer z $N_i$ označimo geometrično porazdeljene slučajne spremenljivke. Pričakovano vrednost in varianco vsote slučajnih spremenljivk izračunamo po enačbah (6.28) in (6.31)

$$\mathrm{E}[N_k] = \mathrm{E}\left[\sum_{i=1}^{k} N_i\right] = \sum_{i=1}^{k} \mathrm{E}[N_i] = \sum_{i=1}^{k} \frac{1}{p} = \frac{k}{p}$$

in

$$\mathrm{var}[N_k] = \mathrm{var}\left[\sum_{i=1}^{k} N_i\right] = \sum_{i=1}^{k} \mathrm{var}[N_i] = \sum_{i=1}^{k} \frac{1-p}{p^2} = \frac{k\,(1-p)}{p^2}$$

### 7.2.4 Poissonova porazdelitev

Poissonova[4] porazdelitev je po pomenu enaka binomski porazdelitvi, saj predstavlja število uspehov. Razlika je le v tem, da pri Poissonovi porazdelitvi obravnavamo število uspehov v časovnem obdobju od 0 do $t$, medtem ko pri binomski porazdelitvi obravnavamo število uspehov v $n$ ponovitvah Bernoullijevega poskusa. Verjetnostno funkcijo te porazdelitve bomo izpeljali iz verjetnostne funkcije binomsko porazdeljene slučajne spremenljivke $Y$ (glej razdelek 7.2.1 Binomska porazdelitev). Vzemimo, da predstavlja en Bernoullijev poskus opazovanje v določenem časovnem intervalu $\Delta t$. Časovno obdobje $[0, t]$, ki ga analiziramo, razdelimo na $n$ intervalov dolžine $\Delta t$. Če časovni interval zmanjšujemo, se njihovo število seveda povečuje, medtem ko se pričakovano število uspehov v analiziranem obdobju ne spremeni, saj dolžina intervala, na katerega smo razdelili obravnavano obdobje, ne more vplivati na skupno število uspehov v tem obdobju. Opisani postopek zmanjševanja časovnega intervala z matematičnimi izrazi zapišemo takole:

$$\Delta t \to 0, \qquad n \to \infty, \qquad \mathrm{E}[Y] = n\,p = \nu = \text{konst.} \longrightarrow p = \frac{\nu}{n}$$

Če te izraze uporabimo v enačbi (7.2) in limitiramo z $n \to \infty$, dobimo

$$\begin{aligned}
p_Y(y) &= \lim_{n\to\infty} p^y (1-p)^{n-y} \binom{n}{y} = \lim_{n\to\infty} p^y (1-p)^{n-y} \frac{n!}{y!\,(n-y)!} = \\
&= \lim_{n\to\infty} \left(\frac{\nu}{n}\right)^y \left(1 - \frac{\nu}{n}\right)^{n-y} \frac{n\,(n-1)\,(n-2)\ldots(n-y+1)}{y!} = \\
&= \lim_{n\to\infty} \frac{\nu^y}{y!} \frac{\left(1-\frac{\nu}{n}\right)^n}{\left(1-\frac{\nu}{n}\right)^y} \frac{n\,(n-1)\,(n-2)\ldots(n-y+1)}{n^y}.
\end{aligned} \tag{7.14}$$

Določimo limite posameznih členov v zgornjem izrazu

$$\begin{aligned}
&\lim_{n\to\infty} \frac{n-y+i}{n} = 1, \qquad i = 1, \ldots, y, \\
&\lim_{n\to\infty} \left(1 - \frac{\nu}{n}\right)^y = 1, \\
&\lim_{n\to\infty} \left(1 - \frac{\nu}{n}\right)^n = \lim_{n\to\infty} \left(\frac{n-\nu}{n}\right)^n = \lim_{n\to\infty} \left(\frac{n}{n-\nu}\right)^{-n} = \lim_{n\to\infty} \left(1 + \frac{\nu}{n-\nu}\right)^{-n} = \\
&= \lim_{m\to\infty} \left(1 + \frac{1}{m}\right)^{-m\nu-\nu} = \lim_{m\to\infty} \left(\left(1+\frac{1}{m}\right)^m\right)^{-\nu} \left(1 + \frac{1}{m}\right)^{-\nu} = e^{-\nu},
\end{aligned} \tag{7.15}$$

kjer smo v zadnji vrstici vpeljali nove oznake ($m = (n-\nu)/\nu \leftrightarrow n = m\nu + \nu$) in upoštevali definicijo števila $e$

$$e = \lim_{m\to\infty} \left(1 + \frac{1}{m}\right)^m .$$

[4] Siméon Denis Poisson , francoski matematik (1781–1840).

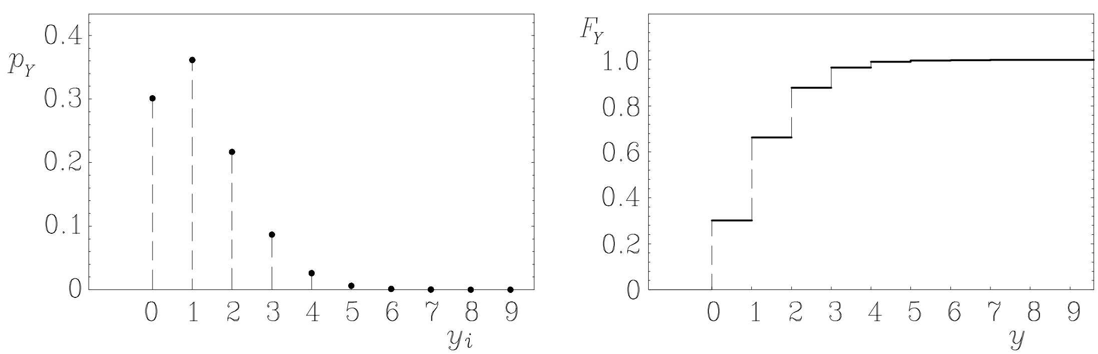
Ob upoštevanju izrazov v (7.15) v enačbi (7.14) lahko zapišemo verjetnostno funkcijo slučajne spremenljivke, porazdeljene po Poissonovi porazdelitvi

$$p_Y(y) = \frac{\nu^y}{y!}\, e^{-\nu}, \qquad y = 0, 1, \ldots, \infty. \tag{7.16}$$

Vidimo, da zaloga vrednosti te slučajne spremenljivke ni omejena, saj je verjetnost, da se v nekem časovnem obdobju zgodi mnogo uspehov različna od nič. To je bistvena razlika v primerjavi z binomsko porazdelitvijo, kjer število uspehov seveda ne more presegati števila Bernoullijevih poskusov $n$.

Tudi pričakovano vrednost in varianco za to porazdelitev lahko izpeljemo iz vrednosti za binomsko porazdelitev (enačbi (7.5) in (7.6))

$$\mathrm{E}[Y] = \lim_{n\to\infty} n\,\frac{\nu}{n} = \nu, \qquad \mathrm{var}[Y] = \lim_{n\to\infty} n\,\frac{\nu}{n}\left(1 - \frac{\nu}{n}\right) = \nu. \tag{7.17}$$




Slika 7.5: Verjetnostna in porazdelitvena funkcija Poissonove porazdelitve

Na sliki (7.5) prikazujemo verjetnostno in porazdelitveno funkcijo slučajne spremenljivke $Y$, porazdeljene po Poissonovi porazdelitvi z vrednostjo $\nu = 4 \cdot 0.3 = 1.2$. V primeru 7.2 obravnavamo podobno spremenljivko, ki je porazdeljena po binomski porazdelitvi.

Zanimiva in pomembna lastnost Poissonove porazdelitve je, da je vsota dveh Poissonovo porazdeljenih slučajnih spremenljivk $Y_1$ in $Y_2$ s parametroma $\nu_1$ in $\nu_2$ tudi Poissonovo porazdeljena slučajna spremenljivka s parametrom $\nu = \nu_1 + \nu_2$. Tej lastnosti pravimo **regenerativnost**.

Naštejmo nekaj primerov slučajnih spremenljivk, porazdeljenih po Poissonovi porazdelitvi:

- Na osnovi opazovanj v preteklosti, lahko ocenimo pričakovano število $\nu$ poplav v določenem časovnem obdobju (na primer 10 let). Slučajna spremenljivka, ki predstavlja število poplav v tem obdobju, je Poissonovo porazdeljena slučajna spremenljivka.
- Vzemimo, da vemo, da se na nekem cestnem odseku v nekem časovnem obdobju v povprečju zgodi $\nu$ nesreč. Število nesreč, ki se zgode na tem odseku in obdobju, je slučajna spremenljivka, ki je porazdeljena po Poissonu.

– Število pojavov kritične obremenitve na konstrukcijo je Poissonovo porazdeljena slučajna spremenljivka. Parameter porazdelitve $\nu$ predstavlja povprečno število pojava kritične obremenitve v nekem časovnem obdobju.

Poissonovo porazdelitev smo temeljili na opazovanju pojava v nekem časovnem obdobju. Seveda lahko tako porazdeljeno slučajno spremenljivko uporabimo tudi za primere, kjer opazujemo pojav na nekem območju.

– Obravnavajmo razporeditev sil na neki konstrukciji. Število delujočih sil na konstrukcijo je slučajna spremenljivka, porazdeljena po Poissonovi porazdelitvi.

V nadaljevanju bomo opisovali časovno odvisne Poissonove porazdelitve in ne bomo posebej poudarjali, da isti zaključki veljajo tudi za slučajno spremenljivko, ki opisuje razpored dogodkov oziroma elementov po nekem območju.

Pomanjkljivost Poissonove porazdelitve, ki jo definiramo z enačbo (7.16), je v tem, da je parameter porazdelitve $\nu$ odvisen od dolžine obravnavanega časovnega obdobja. Ob predpostavki, da se razmere s časom ne spreminjajo, lahko parameter Poissonove porazdelitve definiramo kot linearno funkcijo časa

$$\nu = \lambda\, t$$

V tem primeru tudi verjetnostno funkcijo zapišemo kot funkcijo časa

$$p_Y(y) = \frac{(\lambda\, t)^y}{y!}\, e^{-\lambda\, t}, \qquad y = 0, 1, \ldots, \infty.$$

Slučajnim spremenljivkam, ki so funkcije časa, pravimo **slučajni** oziroma **stohastični procesi**, pravkar opisanem slučajnem procesu pa **Poissonov proces**. V Poissonovem procesu je $Y(t)$ od časa odvisna slučajna spremenljivka in predstavlja število uspehov oziroma pojavov dogodka v časovnem obdobju od začetka obravnavanja do časa $t$. Poissonov proces ima tri pomembne značilnosti:

– **Stacionarnost**: Verjetnost, da se zgodi dogodek v časovnem intervalu od $t$ do $t+\Delta t$ je za poljubno vrednost $t$ enaka. Ta lastnost pomeni, da se razmere s časom ne spreminjajo.

– Verjetnost, da se v nekem zelo kratkem časovnem intervalu zgodita dva ali več uspehov, je zanemarljiva.

– **Neodvisnost**: Število uspehov v nekem časovnem obdobju je neodvisno od števila dogodkov v nekem drugem časovnem obdobju, ki se s prvim ne prekriva. Ta lastnost je značilna za slučajne procese, za katere pravimo, da so "brez spomina".

Različne pojave, ki jih srečujemo v inženirstvu, lahko opišemo s Poissonovim slučajnim procesom, zato je Poissonova porazdelitev za inženirstvo zelo pomembna porazdelitev:

– Pojav izjemnih dogodkov (neurja, poplave, izjemne obtežbe na konstrukcijo, grobe napake meritev in podobno) v določenem obdobju lahko modeliramo kot Poissonov proces.

– Slučajno prihajanje vozil ne določen cestni odsek je značilen primer Poissonovega procesa.

– Razpored začetnih napak oziroma mikrorazpok v materialu je primer Poissonovega procesa, kjer je Poissonova slučajna spremenljivka odvisna od velikosti obravnavanega območja in ne od časovnega obdobja.

Opišimo še pomen parametra $\lambda$ Poissonove porazdelitve oziroma procesa. Če za parameter $\nu$ velja, da je to pričakovano število uspehov oziroma dogodkov v nekem obravnavanem časovnem obdobju, potem parameter $\lambda = \nu/t$ pomeni pričakovano število uspehov oziroma dogodkov na enoto časa

$$\lambda = \frac{\mathrm{E}[Y]}{t}.$$

**Primer 7.5:** *Definirajmo* **50-letni dogodek***, kot dogodek, ki se v povprečju zgodi enkrat v 50 letih ali drugače: verjetnost, da se tak dogodek zgodi v posameznem letu je enaka $p = 1/50$. Primeri takih dogodkov so: 50-letni snežni metež, 50-letno neurje, 50-letna višina vode in podobno. Določimo verjetnost, da se v 50 letih zgodi vsaj en 50-letni dogodek! Določimo še verjetnost, da se zgodi eden, dva ali trije taki dogodki v 50 letih! To nalogo lahko rešimo z uporabo binomske ali Poissonove porazdelitve.*

**Rešitev**: Uporabimo najprej binomsko porazdelitev! Enoletno obdobje obravnavamo kot en Bernoullijev poskus. Verjetnost, da se v tem letu zgodi tak dogodek, je enaka $p = 1/50$. Ker obravnavamo petdesetletno obdobje, je število poskusov enako $n = 50$. Slučajna spremenljivka $Y$ predstavlja število takih dogodkov v obravnavanem obdobju. Verjetnost, da se je v petdesetih letih zgodil vsaj en tak dogodek ($P[Y > 0]$) izračunamo po naslednji enačbi

$$P[Y>0] = 1 - P[Y=0] = 1 - \binom{n}{0} p^0 (1-p)^n = 1 - 0.98^{50} = 1 - 0.3642 = 0.6358,$$

saj je dogodek, da se v tem obdobju ne zgodi noben tak dogodek, nasproten dogodku, da se zgodi vsaj eden. Na drugo vprašanje odgovorimo tako, da izračunamo verjetnosti, da se zgodi natanko eden, natanko dva in natanko trije dogodki. Nato te tri verjetnosti seštejemo, saj so ti dogodki medsebojno nezdružljivi:

$$P[Y=1] = \binom{n}{1} p^1 (1-p)^{n-1} = 50 \cdot 0.02 \cdot 0.98^{49} = 0.3716,$$
$$P[Y=2] = \binom{n}{2} p^2 (1-p)^{n-2} = 1225 \cdot 0.02^2 \cdot 0.98^{48} = 0.1858,$$
$$P[Y=3] = \binom{n}{3} p^3 (1-p)^{n-3} = 19600 \cdot 0.02^3 \cdot 0.98^{47} = 0.0607,$$
$$P[1 \le Y \le 3] = P[Y=1] + P[Y=2] + P[Y=3] = 0.3716 + 0.1858 + 0.0607 = 0.6181.$$

Podobne rezultate dobimo tudi, če uporabimo Poissonovo porazdelitev. Parameter Poissonove porazdelitve je $\nu = n\,p = 50 \cdot 0.02 = 1$. Verjetnost, da pride do vsaj enega dogodka določimo na podoben način, kot pri binomski porazdelitvi

$$P[Y>0] = 1 - P[Y=0] = 1 - \frac{\nu^0 e^{-\nu}}{0!} = 1 - e^{-1} = 1 - 0.3679 = 0.6321.$$

Izračunajmo še verjetnosti, da se zgodi en, dva ali trije dogodki v 50-letnem obdobju:

$$P[Y=1] = \frac{\nu^1 e^{-\nu}}{1!} = e^{-1} = 0.3679,$$
$$P[Y=2] = \frac{\nu^2 e^{-\nu}}{2!} = \frac{e^{-1}}{2} = 0.1839,$$
$$P[Y=3] = \frac{\nu^3 e^{-\nu}}{3!} = \frac{e^{-1}}{6} = 0.0613,$$
$$P[1 \le Y \le 3] = P[Y=1] + P[Y=2] + P[Y=3] = 0.3679 + 0.1839 + 0.0613 = 0.6131.$$

Vidimo, da so razlike med verjetnostmi, izračunanimi na osnovi binomske porazdelitve, in tistimi, izračunanimi na osnovi Poissonove, zelo majhne.

**Primer 7.6:** *Vzemimo, da imamo 30 sistemov, ki so vsi projektirani na 500 letni dogodek. Ugotovimo verjetnost, da v naslednjih 20 letih vsaj en sistem odpove vsaj enkrat zaradi pojava 500 letnega dogodka.*

**Rešitev:** Verjetnost, da se v nekem letu zgodi 500 letni dogodek, ki se v povprečju zgodi le enkrat v 500 letih, je $p = 1/500$. Določimo najprej verjetnost $p_1$, da katerikoli od obravnavanih 30 sistemov odpove vsaj enkrat v naslednjih 20 letih. To verjetnost bomo določili iz verjetnosti nasprotnega dogodka, da v naslednjih letih ne bo prišlo do takega dogodka

$$p_1 = 1 - P[\text{noben 500 letni dogodek}] = 1 - (1-p)^{20} = 0.03925.$$

Sedaj lahko izračunamo še verjetnost, da bo vsaj en sistem odpovedal v naslednjih 20 letih. Tudi to verjetnost izračunamo iz verjetnosti nasprotnega dogodka, da ne bo noben sistem odpovedal

$$P[\text{vsaj en sistem odpove}] = 1 - (1-p_1)^{30} = 0.699.$$

Vidimo lahko, da je verjetnost, da bo v naslednjih 20 letih vsaj eden izmed 30 neodvisnih sistemov odpovedal zaradi pojava 500 letnega dogodka, zelo visoka – skoraj 70%. To je na prvi pogled presenetljivo, saj se nam lahko zdi, da je projektiranje glede na 500 letni dogodek zelo zanesljivo. Razlog za ta rezultat je v večjem številu neodvisnih sistemov. Rezultat tega izračuna se žal potrjuje tudi v našem življenju: čeprav je vsaka izmed mnogih jedrskih central projektirana, izvedena in vzdrževana izredno zanesljivo in je verjetnost, da se bo kaj zgodilo v točno določeni elektrarni, zelo majhna, skoraj vsako leto pride do manjše ali večje nesreče v kakšni izmed njih. Razlog za to je poleg človeških napak in malomarnosti v kar velikem številu jedrskih central na naši Zemlji in verjetnost, da se v eni izmed njih zgodi nesreča, je vse prej kot zanemarljiva.

### 7.2.5 Eksponentna porazdelitev

Binomska, geometrična, Pascalova in Poissonova porazdelitve so diskretne, saj je njihova zaloga vrednosti diskretna. Prve tri temeljijo na končnem ali neskončnem številu Bernoullijevih poskusov, Poissonova pa izhaja iz opazovanja v nekem zveznem časovnem obdobju oziroma krajevnem območju. Eksponentna porazdelitev je analogna geometrični porazdelitvi, tako kot je Poissonova porazdelitev analogna binomski. Geometrično porazdeljena slučajna spremenljivka predstavlja število poskusov do prvega uspeha, slučajna spremenljivka $T$, ki predstavlja **čas do prvega uspeha**, pa je porazdeljena eksponentno.

Ker je čas zvezna količina, je tudi slučajna spremenljivka $T$ zvezna. Izpeljimo najprej porazdelitveno funkcijo eksponentne porazdelitve $F_T(t) = P[T \leq t]$. Izhajali bomo iz nasprotnega dogodka $T > t$, ki predstavlja dogodek, da se v obdobju od začetka opazovanja do časa $t$, ni zgodil noben uspeh. Verjetnost tega, lahko izrazimo z verjetnostno funkcijo Poissonovo porazdeljene slučajne spremenljivke

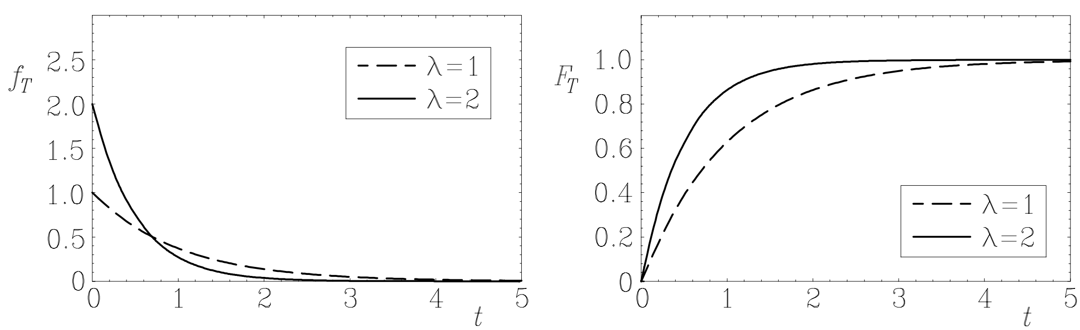
$$P[T > t] = P[Y = 0] = \frac{(\lambda\, t)^0\, e^{-\lambda\, t}}{0!} = e^{-\lambda\, t}, \qquad t \geq 0. \tag{7.18}$$

Iz enačbe (7.18) izhaja enačba porazdelitvene funkcije eksponentne porazdelitve

$$F_T(t) = 1 - e^{-\lambda\, t}, \qquad t \geq 0. \tag{7.19}$$

Gostoto verjenosti določimo z odvajanjem porazdelitvene funkcije (7.19)

$$f_T(t) = \frac{dF_T(t)}{dt} = \lambda\, e^{-\lambda\, t}, \qquad t > 0. \tag{7.20}$$




Slika 7.6: Gostota verjetnosti in porazdelitvena funkcija eksponentne porazdelitve

Pričakovano vrednost in varianco ter koeficienta simetričnosti in sploščenosti eksponentno porazdeljene slučajne spremenljivke smo izračunali že v računskem primeru 6.2:

$$\mathrm{E}[T] = \frac{1}{\lambda}, \qquad \mathrm{var}[T] = \frac{1}{\lambda^2}, \qquad \gamma_{1T} = 2, \qquad \gamma_{2T} = 9.$$

Eksponentna porazdelitev opisuje Poissonov stohastični proces podobno kot Poissonova porazdelitev, le da je pomen slučajne spremenljivke drugačen. Zaradi stacionarnosti Poissonovega procesa in neodvisnosti pojavljanja uspehov v tem procesu je vseeno, kdaj začnemo meriti čas do prvega uspeha, porazdelitev je vedno enaka. Zato se lahko odločimo in začnemo meriti čas po vsakem uspehu. V tem primeru slučajna spremenljivka $T$ predstavlja čas med dvema zaporednima uspehoma.

Parameter porazdelitve $\lambda$ ima enak pomen, kot pri Poissonovi porazdelitvi: število uspehov v časovni enoti. Zato je pričakovana vrednost slučajne spremenljivke $\mathrm{E}[T] = 1/\lambda$ in pomeni pričakovani čas do prvega oziroma med dvema zaporednima uspehoma.

Najbolj značilna uporaba eksponentne porazdelitve je pri simulaciji prometnih tokov, saj lahko čas med dvema zaporednima prihodoma vozil opišemo prav z eksponentno porazdeljeno slučajno spremenljivko.

### 7.2.6 Porazdelitev gama ($\Gamma$)

Podobno, kot smo pri obravnavanju zaporednih Bernoullijevih poskusih vpeljali slučajno spremenljivko, ki predstavlja število poskusov do $k$-tega uspeha (Pascalova porazdelitev), definiramo slučajno spremenljivko $T_k$, ki predstavlja **čas do $k$-tega uspeha**. Njeno porazdelitev imenujemo **porazdelitev gama**. Ime izhaja iz funkcije $\Gamma$, ki je vključena v funkciji gostote verjetnosti. Tako kot Poissonova in eksponentna porazdelitev, tudi porazdelitev gama opisuje Poissonov stohastični proces. Porazdelitev gama pogosto imenujejo tudi drugače: na primer Erlangova[5] porazdelitev ali Pearsonova[6] porazdelitev tretjega tipa.

Slučajno spremenljivko $T_k$, porazdeljeno po porazdelitvi gama, lahko obravnavamo kot vsoto $k$ eksponentno porazdeljenih slučajnih spremenljivk z istim parametrom $\lambda$. Zato lahko gostoto verjetnosti porazdelitve gama izpeljemo s konvolucijskim integralom gostot verjetnosti eksponentno porazdeljenih slučajnih spremenljivk (enačba (5.20)).

**Primer 7.7:** *Določimo gostoto verjetnosti slučajne spremenljivke $T_2$, ki je vsota dveh eksponentno porazdeljenih in neodvisnih slučajnih spremenljivk $T$.*

**Rešitev**: Gostota verjetnosti eksponentno porazdeljene slučajne spremenljivke je (7.20)

$$f_T(t) = \lambda\, e^{-\lambda t}, \qquad t > 0. \tag{7.21}$$

Konvolucijski integral izračunamo po enačbi (5.20)

$$f_{T_2}(t_2) = \int_{-\infty}^{\infty} f_T(t_2 - t)\, f_T(t)\, dt. \tag{7.22}$$

Ker je produkt $\lambda\, e^{-\lambda t}$ in $\lambda\, e^{-\lambda(t_2-t)}$ od nič različen le za $0 \le t \le t_2$, integral v enačbi (7.22) ob upoštevanju (7.21) zapišemo in izračunamo takole:

$$f_{T_2}(t_2) = \int_0^{t_2} \lambda\, e^{-\lambda t}\, \lambda\, e^{-\lambda(t_2-t)}\, dt = \lambda^2 t_2 e^{-\lambda t_2}, \qquad t_2 \ge 0.$$

[5] Agner Krarup Erlang, danski matematik (1878–1929).
[6] Karl Pearson, angleški matematik (1857–1936).

Gostote verjetnosti vsote dveh in več neodvisnih in eksponentno porazdeljenih slučajnih spremenljivk lahko določimo tudi s programom MATHEMATICA.[7] Rezultati integracije so:

$$
\begin{aligned}
f_{T_2}(t_2) &= \lambda^2\, t_2\, e^{-\lambda t_2}, \qquad && t_2 \geq 0,\\
f_{T_3}(t_3) &= \frac{\lambda^3\, t_3^2\, e^{-\lambda t_3}}{2}, \qquad && t_3 \geq 0,\\
f_{T_4}(t_4) &= \frac{\lambda^4\, t_4^3\, e^{-\lambda t_4}}{6}, \qquad && t_4 \geq 0,\\
f_{T_5}(t_5) &= \frac{\lambda^5\, t_5^4\, e^{-\lambda t_5}}{24}, \qquad && t_5 \geq 0
\end{aligned}
$$

oziroma splošno

$$
f_{T_k}(t_k) = \frac{\lambda\,(\lambda\, t_k)^{k-1} e^{-\lambda t_k}}{(k-1)!}, \qquad t_k \geq 0. \tag{7.23}
$$

Porazdelitev gama, ki jo opisuje gostota verjetnosti v (7.23) je odvisna od dveh parametrov: $\lambda$ predstavlja povprečno število uspehov v časovni enoti, $k$ pa število uspehov, do katerega merimo čas $T_k$. Gostoto verjetnosti (7.23) lahko posplošimo tako, da za parameter $k$ vzamemo poljubno pozitivno število. V tem primeru gostoto verjetnosti zapišemo z enačbo

$$
f_{T_k}(t_k) = \frac{\lambda\,(\lambda\, t_k)^{k-1} e^{-\lambda t_k}}{\Gamma(k)}, \qquad t_k \geq 0, \tag{7.24}
$$

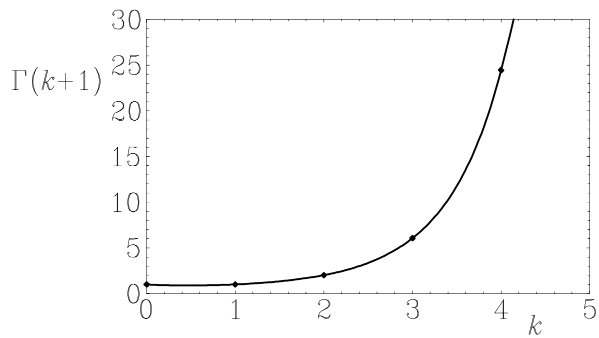
[7] Integracija, izvedena s programom MATHEMATICA:

```
la = 1;
fT1[t1_] = la E^(-la t1);
Print["fT2(t2) = ",fT2[t2_] = Integrate[fT1[t1] fT1[t2-t1],{t1,0,t2}]];
Print["fT3(t3) = ",fT3[t3_] = Integrate[fT2[t1] fT1[t3-t1],{t1,0,t3}]];
Print["fT4(t4) = ",fT4[t4_] = Integrate[fT3[t1] fT1[t4-t1],{t1,0,t4}]];
Print["fT5(t5) = ",fT5[t5_] = Integrate[fT4[t1] fT1[t5-t1],{t1,0,t5}]];

          t2
fT2(t2) = ---
           t2
          E
            2
           t3
fT3(t3) = -----
             t3
          2 E
            3
           t4
fT4(t4) = -----
             t4
          6 E
            4
           t5
fT5(t5) = ------
              t5
          24 E
```

kjer $\Gamma(k)$ predstavlja funkcijo gama, ki je definirana z naslednjo enačbo:

$$\Gamma(k) = \int_0^\infty e^{-x} x^{k-1} dx. \tag{7.25}$$

Vrednosti funkcije $\Gamma(x)$ so podane v preglednicah marsikatere knjige o statistike, vključena pa je tudi v marsikaterem programu (na primer: EXCEL (funkcija `gammaln`), MATHEMATICA (funkcija `Gamma[x]`), MATLAB (funkcija `gamma(x)`)...). Graf te funkcije prikazujemo na sliki (7.7).




Slika 7.7: Graf funkcije gama $\Gamma(k+1)$

Že pri definiciji porazdelitve gama smo zapisali, da je tako porazdeljena slučajna spremenljivka vsota $k$ neodvisnih eksponentno porazdeljenih slučajnih spremenljivk. Zato pričakovano vrednost in varianco izračunamo po enačbah (6.28) in (6.31)

$$\mathrm{E}[T_k] = \mathrm{E}\left[\sum_{i=1}^{k} T_i\right] = \sum_{i=1}^{k} \mathrm{E}[T_i] = \sum_{i=1}^{k} \frac{1}{\lambda} = \frac{k}{\lambda}$$

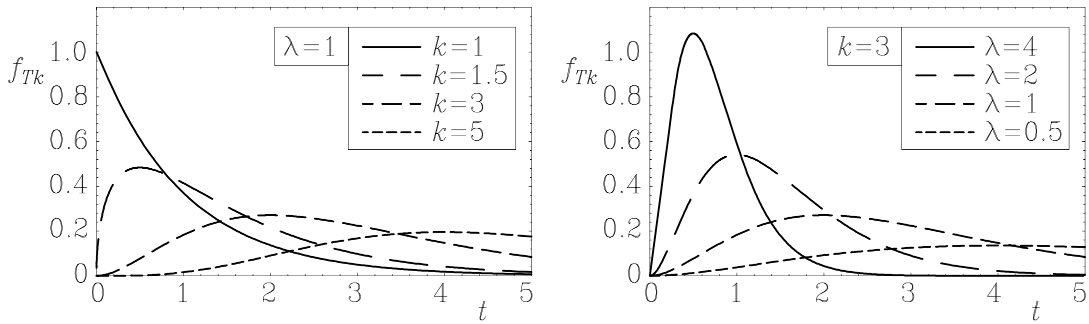
in

$$\mathrm{var}[T_k] = \mathrm{var}\left[\sum_{i=1}^{k} T_i\right] = \sum_{i=1}^{k} \mathrm{var}[T_i] = \sum_{i=1}^{k} \frac{1}{\lambda^2} = \frac{k}{\lambda^2}.$$

Koeficienta simetričnosti in kurtosis ste odvisna le od parametra $k$:

$$\gamma_{1T_k} = \frac{2}{\sqrt{k}},$$

$$\gamma_{2T_k} = 3 + \frac{6}{\sqrt{k}}.$$




Slika 7.8: Gostota verjetnosti porazdelitve gama za različne parametre $k$ in $\lambda$

## 7.3 Normalna porazdelitev

Pri obravnavanju porazdelitev, ki izhajajo iz Bernoullijevih poskusov (glej razdelek 7.2) smo pokazali, da vsota nekaj enako porazdeljenih slučajnih spremenljivk lahko da novo porazdelitev. Tako smo ugotovili, da je vsota nekaj Bernoullijevo porazdeljenih slučajnih spremenljivk slučajna spremenljivka, porazdeljena po binomski porazdelitvi, vsota nekaj geometrično porazdeljenih je slučajna spremenljivka, porazdeljena po Pascalovi porazdelitvi, vsota nekaj eksponentno porazdeljenih slučajnih spremenljivk pa je slučajna spremenljivka, porazdeljena po porazdelitvi gama. V tem razdelku bomo opisali porazdelitev, ki jo dobimo, če seštejemo neskončno mnogo enako porazdeljenih neodvisnih slučajnih spremenljivk. To porazdelitev imenujemo **normalna** ali **Gaussova**[8] **porazdelitev**.

Normalna porazdelitev je morda najpomembnejša oziroma najpogosteje uporabljena porazdelitev v statistiki, saj marsikatera količina predstavlja vsoto mnogih drugih in je zato vsaj približno normalno porazdeljena. Poglejmo nekaj primerov:

- Analizirajmo vzorec $n$ števil, ki pripadajo poljubno porazdeljeni slučajni spremenljivki. Povprečno vrednost vzorčenih števil izračunamo tako, da *vsoto* teh števil delimo z $n$.
- Rezultat izpita, ki je sestavljen iz večjega števila kratkih vprašanj, je *vsota* rezultatov posameznih vprašanj.
- Na eksperimentalno napako lahko vpliva mnogo dejavnikov, katerih vpliv na skupno napako običajno *seštejemo*. Gauss, ki je med prvimi[9] obravnaval normalno porazdelitev, jo je uporabil pri analizi eksperimentalnih napak. Zato so to porazdelitev imenovali tudi porazdelitev napake.
- Na določeno lastnost človeka (na primer višina) vpliva genska zasnova in okolje. Če vpliv okolja zanemarimo, ali pa ga upoštevamo posebej, lahko predpostavimo, da na določeno lastnost vpliva večje število genov. Določena lastnost je torej *vsota* vplivov večjega števila genov.

[8] Johann Carl Friedrich Gauss, nemški matematik (1777–1855).
[9] Približno sto let pred Gaussom je normalno porazdelitev izpeljal angleški matematik Abraham de Moivre (1667–1754).

### 7.3.1 Centralni limitni izrek

Vzemimo, da imamo $n$ neodvisnih in enako porazdeljenih slučajnih spremenljivk $X_i$, ki imajo enako pričakovano vrednost $m_{X_i} = m_X$ in standardno deviacijo $\sigma_{X_i} = \sigma_X > 0$. Naj bo slučajna spremenljivka $S_n$ vsota slučajnih spremenljivk $X_i$

$$S_n = \sum_{i=1}^{n} X_i.$$

Iz enačb (6.28) in (6.30) sledi, da lahko pričakovano vrednost $\mathrm{E}[S_n]$ in varianco $\mathrm{var}[S_n]$ zapišemo z naslednjima izrazoma

$$\mathrm{E}[S_n] = n\,\mathrm{E}[X] = n\,m_X, \qquad \mathrm{var}[S_n] = n\,\mathrm{var}[X] = n\,\sigma_X^2.$$

Centralni limitni izrek pravi, da če gre število slučajnih spremenljivk $n$ proti neskončnosti, gre porazdelitev vsote $S_n$ proti normalni porazdelitvi. Izrek lahko z enačbami zapišemo takole:

$$\lim_{n\longrightarrow\infty} P\left[\frac{S_n - n\,m_X}{\sqrt{n}\,\sigma_X} \le a\right] = \int\limits_{-\infty}^{a} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}\, du = F_U(a). \tag{7.26}$$

Iz te enačbe lahko razberemo, da je gostota verjetnosti te normalne porazdelitve enaka

$$f_U(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}, \qquad -\infty < u < \infty. \tag{7.27}$$

Dokaz tega pomembnega izreka vključuje med drugim tudi definicijo karakterističnih funkcij porazdelitev, kar presega okvir tega učbenika. Dokaz lahko najdete v različnih knjigah (na primer: R. Jamnik, Verjetnostni račun, D. C. Montgomery, G.C. Runger, Applied Statistics and Probabiliy for Engineers, C. M. Grinstead, J.L. Snell, Introduction to Probability), prvi pa ga je objavil že omenjeni Abraham de Moivre leta 1718. Tu bomo pokazali le nekaj primerov vsot enako porazdeljenih slučajnih spremenljivk.

**Primer 7.8:** *Določimo gostoto verjetnosti vsote dveh, treh in več enakomerno porazdeljenih slučajnih spremenljivk in jih primerjamo z ustrezno normalno porazdeljeno slučajno spremenljivko.*

**Rešitev**: Definirajmo gostoto verjetnosti slučajne spremenljivke $X$, ki je enakomerno porazdeljena od $-1$ do 1

$$f_X(x) = \begin{cases} 0 & x < -1, \\ 1/2 & -1 \le x < 1, \\ 0 & 1 \le x. \end{cases}$$

Vsoto dveh enakomerno porazdeljenih slučajnih spremenljivk izračunamo s konvolucijo (enačba (5.20))

$$f_{X_2}(x_2) = \int\limits_{-\infty}^{\infty} f_X(x_2 - x)\, f_X(x)\, dx = \begin{cases} \int\limits_{-1}^{x_2+1} 0.5^2 dx = 0.25(2 + x_2) & -2 \le x_2 < 0, \\ \int\limits_{x_2-1}^{1} 0.5^2 dx = 0.25(2 - x_2) & 0 \le x_2 \le 2, \\ 0 & \text{drugje.} \end{cases}$$

Gostota verjetnosti dveh enakomernih slučajnih spremenljivk je torej odsekoma linearna funkcija, ki jo prikazujemo na sliki 7.9. Vsoto treh enakomerno porazdeljenih slučajnih spremenljivk določimo s konvolucijo slučajnih spremenljivk $X_2$ in $X$

$$f_{X_3}(x_3) = \int_{-\infty}^{\infty} f_{X_2}(x_3 - x)\, f_X(x)\, dx =$$

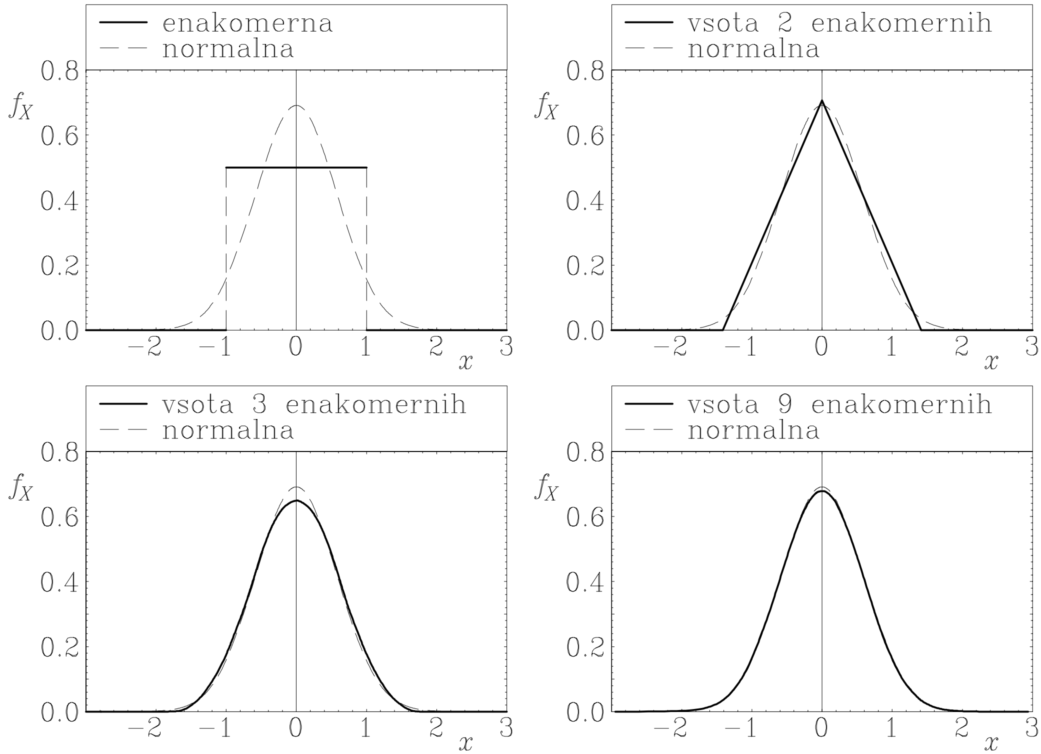
$$= \begin{cases} \int_{-1}^{x_3+2} 0.5 \cdot 0.25(2 + x_3 - x)dx & -3 \le x_3 < -1, \\ \int_{-1}^{x_3} 0.5 \cdot 0.25(2 - x_3 + x)dx + \int_{x_3}^{1} 0.5 \cdot 0.25(2 + x_3 - x)dx & -1 \le x_3 < 1, \\ \int_{x_3-2}^{1} 0.5 \cdot 0.25(2 - x_3 + x)dx & 1 \le x_3 \le 3, \\ 0 & \text{drugje.} \end{cases}$$

$$= \begin{cases} 0.5625 + 0.375\, x_3 + 0.0625\, x_3^2 & -3 \le x_3 < -1, \\ 0.375 - 0.125\, x_3^2 & -1 \le x_3 < 1, \\ 0.5625 - 0.375\, x_3 + 0.0625\, x_3^2 & 1 \le x_3 \le 3, \\ 0 & \text{drugje.} \end{cases}$$

Vsota treh enakomerno porazdeljenih slučajnih spremenljivk je odsekoma kvadratna funkcija. Po obliki je že zelo podobna normalni porazdelitvi. Tudi graf te gostote verjetnosti prikazujemo na sliki 7.9. Podobno izpeljemo tudi gostote verjetnosti za vsote štirih ($X_4 = X_3 + X$) in več enakomerno porazdeljenih slučajnih spremenljivk ($X_{n+1} = X_n + X$). Vsota devetih enakomernih slučajnih spremenljivk, katere gostoto verjetnosti prikazujemo na sliki 7.9, je odsekoma polinom osme stopnje, definiran na območju od $-9$ do 9. Na sliki vidimo, da se gostota verjetnosti vsote devetih enakomerno porazdeljenih slučajnih spremenljivk skoraj ne razlikuje od ustrezne normalno porazdeljene slučajne spremenljivke. Na sliki 7.9 prikazujemo tako transformirane slučajne spremenljivke, da imajo vse enako pričakovano vrednost $m_X = 0$ in standardno deviacijo $\sigma_X = \sqrt{1/3} = 0.5774$. Kako se gostota verjetnosti vsote enakomerno porazdeljenih slučajnih spremenljivk približuje normalni, lahko opazujemo tudi v kratkem animiranem filmu: HTTP://WWW.KM.FGG.UNI-LJ.SI/PREDMETI/SEI/FILM002.HTML

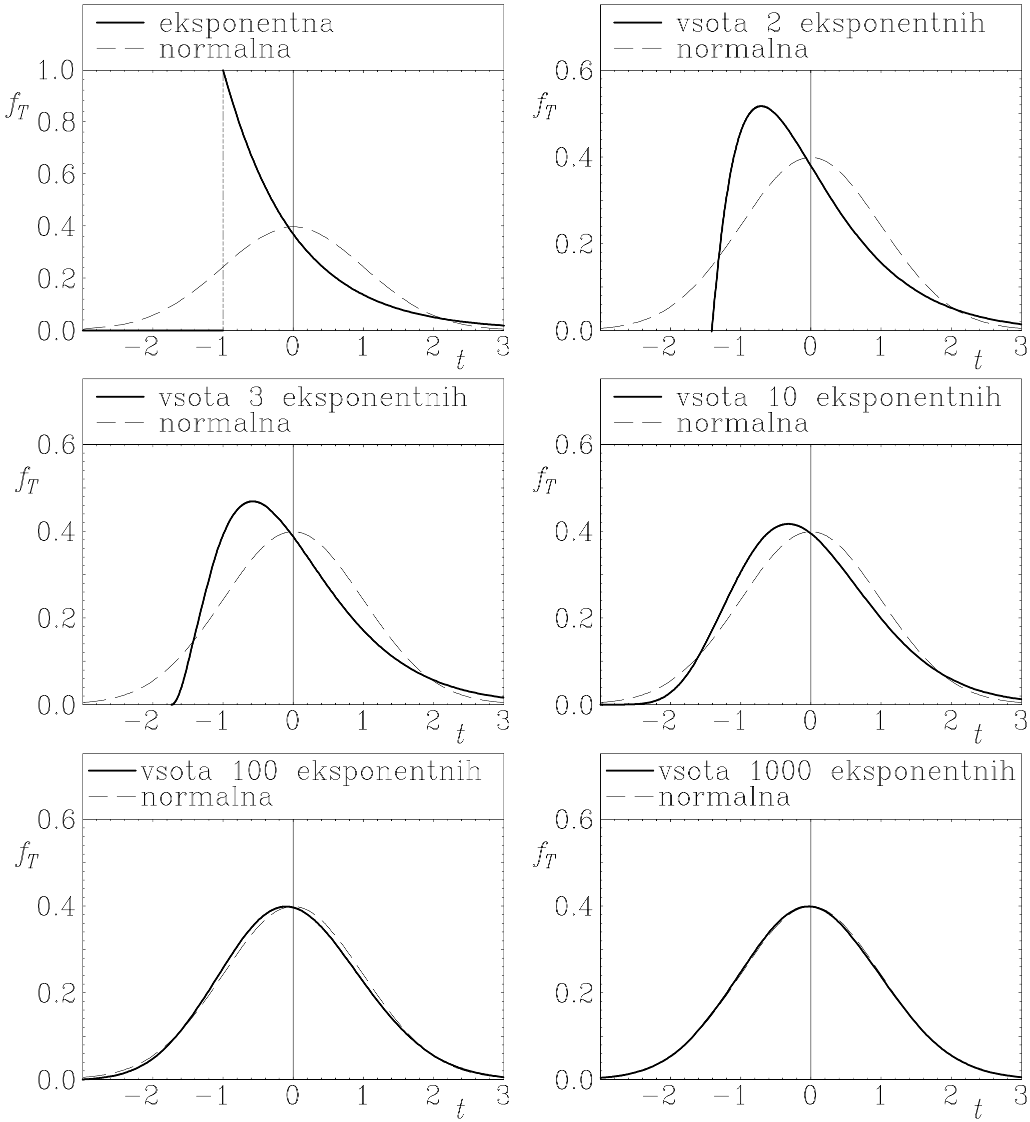
**Primer 7.9:** *Določimo gostoto verjetnosti vsote dveh, treh in več eksponentno porazdeljenih slučajnih spremenljivk in jih primerjamo z ustrezno normalno porazdeljeno slučajno spremenljivko.*

**Rešitev**: Vsote eksponentno porazdeljenih slučajnih spremenljivk smo izpeljali že pri obravnavanju porazdelitve gama, ki predstavlja vsoto več eksponentno porazdeljenih slučajnih spremenljivk (glej primer 7.7). Sedaj bomo le uporabili enačbo (7.23) in graf tako porazdeljenih slučajnih spremenljivk primerjali z normalno.




Slika 7.9: Gostota verjetnosti vsote dveh in več slučajnih spremenljivk

Vzemimo, da je $\lambda = 1$ in narišimo gostote verjetnosti eksponentne porazdelitve, vsote dveh, treh in več eksponentno porazdeljenih slučajnih spremenljivk. Porazdelitve transformiramo tako, da imajo vse enako pričakovano vrednost in standardno deviacijo (glej sliko 7.10). Vidimo, da se gostota verjetnosti vsote mnogih eksponentnih porazdelitev zelo lepo prilega normalni porazdelitvi. Kako se gostota verjetnosti vsote eksponentno porazdeljenih slučajnih spremenljivk približuje normalni, lahko opazujemo tudi v kratkem animiranem filmu: HTTP://WWW.KM.FGG.UNI-LJ.SI/PREDMETI/SEI/FILM001.HTML




Slika 7.10: Gostota verjetnosti vsote eksponentno porazdeljenih slučajnih spremenljivk

### 7.3.2 Standardizirana in splošna normalna porazdelitev

V enačbi (7.27) smo prvič zapisali gostoto verjetnosti normalne porazdelitve. Prva dva momenta te porazdelitve izračunamo po enačbah (6.3) in (6.6). Pričakovana vrednost je enaka nič, saj je gostota verjetnosti simetrična glede na ordinatno os, varianco slučajne spremenljivke pa lahko izračunamo s programom MATHEMATICA.[10]

$$\mathrm{E}[U] = m_U = \int\limits_{-\infty}^{\infty} u \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}\, du = 0,$$

$$\mathrm{var}[U] = \sigma_U^2 = \int\limits_{-\infty}^{\infty} u^2 \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2}\, du = 1.$$

Tako normalno porazdelitev, pri kateri je pričakovana vrednost enaka nič, varianca pa enaka ena, imenujemo **standardizirana normalna porazdelitev**. Normalno porazdelitev s poljubno pričakovano vrednostjo in varianco lahko definiramo z linearno transformacijo

$$X = g(U) = U\sigma_X + m_X. \tag{7.28}$$

Pokažimo, da sta pričakovana vrednost in varianca slučajne spremenljivke $X$ enaki $m_X$ in $\sigma_X^2$. Iz enačb (6.12) in (6.17) izračunamo pričakovano vrednost $\mathrm{E}[X]$ in varianco $\mathrm{var}[X]$

$$\mathrm{E}[X] = \mathrm{E}[U]\,\sigma_X + m_X = m_X, \qquad \mathrm{var}[X] = \mathrm{var}[U]\,\sigma_X^2 = \sigma_X^2.$$

Sedaj lahko izpeljemo še gostoto verjetnosti slučajne spremenljivke $X$. Ker je funkcija $g(U)$ v (7.28) monotona, lahko pri določitvi gostote verjenosti uporabimo enačbo (5.12) in upoštevamo, da je $g^{-1}(x) = (x - m_X)/\sigma_X$ in odvod $dg^{-1}(x)/dx = 1/\sigma_X$

$$f_X(x) = f_U(g^{-1}(x)) \left|\frac{dg^{-1}(x)}{dx}\right| = \frac{1}{\sqrt{2\pi}\,\sigma_X}\, e^{-\frac{1}{2}\left(\frac{x-m_X}{\sigma_X}\right)^2} \qquad -\infty < x < \infty. \tag{7.29}$$

Z enačbo (7.29) smo zapisali gostoto verjetnosti normalne porazdelitve s pričakovano vrednostjo $m_X$ in standardno deviacijo $\sigma_X$. Na sliki 7.11 prikazujemo gostote verjetnosti za nekaj normalno porazdeljenih slučajnih spremenljivk z različnimi parametri $m_X$ in $\sigma_X$.

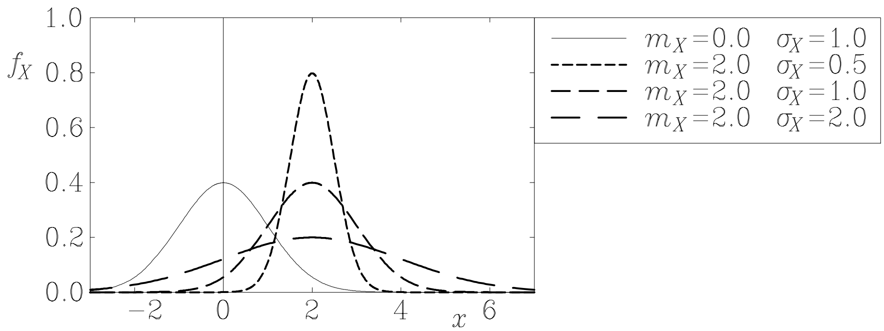
Zapišimo še višje centralne momente za normalno porazdelitev. Zaradi simetričnosti porazdelitve so vsi centralni momenti lihega reda enaki nič. Centralni moment reda $r \geq 2$ lahko zapišemo z naslednjo enačbo

$$\mu_X^{(r)} = \tfrac{1+(-1)^r}{2} \cdot \frac{r!}{2^{r/2}\,\left(\frac{r}{2}\right)!}\,\sigma_X^r.$$

Iz te enačbe sledi, da je koeficient simetričnosti $\gamma_{1X} = 0$, koeficient kurtosis, ki opisuje sploščenost porazdelitve, pa $\gamma_{2X} = 3$.

[10] Integracija, izvedena s programom MATHEMATICA:

```
Print["mu = ",Integrate[u 1/Sqrt[2 Pi] E^(-1/2 u^2),{u,-Infinity,Infinity}]];
Print["varu = ",Integrate[u^2 1/Sqrt[2 Pi] E^(-1/2 u^2),{u,-Infinity,Infinity}]];

mu = 0
varu = 1
```




Slika 7.11: Gostota verjetnosti normalno porazdeljenih slučajnih spremenljivk

### 7.3.3 Porazdelitvena funkcija normalno porazdeljene slučajne spremenljivke

Porazdelitveno funkcijo določimo z integriranjem funkcije gostote verjetnosti

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma_X}\, e^{-\frac{1}{2}\left(\frac{\tilde{x}-m_X}{\sigma_X}\right)^2} d\tilde{x}. \tag{7.30}$$

Integrala v enačbi (7.30) analitično ne moremo rešiti, zato lahko vrednosti porazdelitvene funkcije določimo le numerično. V nekaterih programih je porazdelitvena funkcija že vgrajena (na primer: EXCEL (funkciji *normdist* in *normsdist*)), v drugih pa je definirana sorodna funkcija, iz katere lahko preprosto določimo vrednosti porazdelitvene funkcije (na primer: MATHEMATICA (funkcija Erf[x])). Če želimo določiti vrednost porazdelitvene funkcije brez računalniških programov, moramo uporabiti preglednice. V takih preglednicah so podane vrednosti porazdelitvene funkcije standardizirane normalne porazdelitve – slučajna spremenljivka $U$. Za vrednost porazdelitvene funkcije slučajne spremenljivke $X$, porazdeljene po poljubni normalni porazdelitvi, moramo zapisati

$$F_X(x) = P[X \le x] = P\left[\frac{X - m_X}{\sigma_X} \le \frac{x - m_X}{\sigma_X}\right]. \tag{7.31}$$

Če v enačbi (7.31) upoštevamo (7.28), sledi

$$F_X(x) = P\left[U \le \frac{x - m_X}{\sigma_X}\right] = F_U\left(\frac{x - m_X}{\sigma_X}\right),$$

kar pomeni, da vrednost porazdelitvene funkcije $F_X(x)$ lahko določimo tako, da izračunamo vrednost porazdelitvene funkcije standardizirano normalno porazdeljene slučajne spremenljivke $U$ pri $(x-m_X)/\sigma_X$.

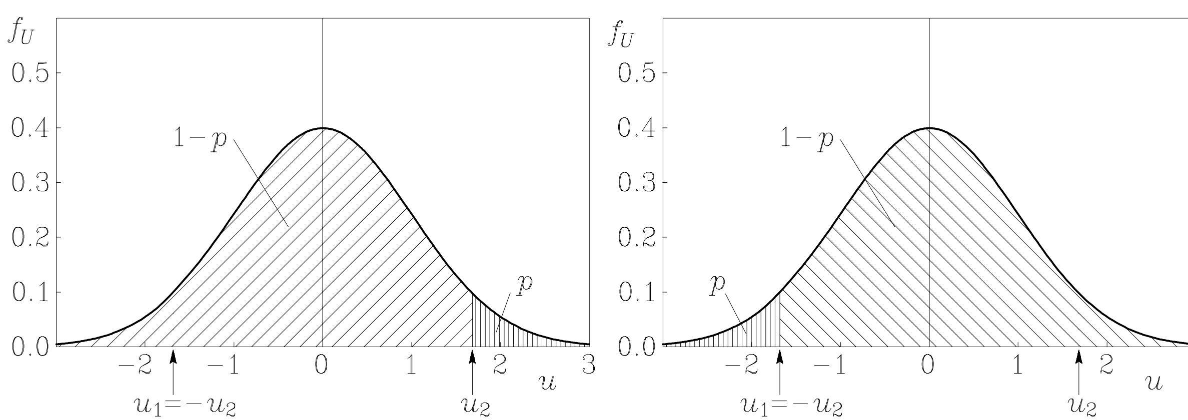
**Primer 7.10:** *Vzemimo, da je tlačna trdnost betona, ki ga uporabljamo pri gradnji konstrukcije, normalno porazdeljena slučajna spremenljivka $X$ s pričakovano vrednostjo $m_X = 27\,\mathrm{MPa}$ in standardno deviacijo $\sigma_X = 5\,\mathrm{MPa}$. Določimo verjetnost, da je trdnost betona nižja od $20\,\mathrm{MPa}$. Določimo tudi verjetnost, da je trdnost betona višja od $30\,\mathrm{MPa}$!*

**Rešitev:** Verjetnost, da je slučajna spremenljivka $X$ nižja od 20 MPa, določimo po naslednji enačbi:

$$P[X \leq 20] = F_X(20) = P\left[\frac{X - m_X}{\sigma_X} \leq \frac{20 - m_X}{\sigma_X}\right] = P\left[U \leq \frac{20 - 27}{5}\right] =$$
$$= P[U \leq -1.4] = F_U(-1.4) = 0.08076.$$

Vrednost porazdelitvene funkcije $F_X(20)$ oziroma $F_U(-1.4)$ lahko določimo z računalniškim programom EXCEL z ukazoma `NORMDIST(20;27;5;TRUE)` oziroma `NORMSDIST(-1.4)`. Če računalniških programov ne moremo ali ne želimo uporabiti, lahko rešitev odčitamo tudi iz preglednice, ki jo lahko najdemo v dodatku skoraj vsake knjige statistike in verjetnostnega računa in tudi na spletni strani: HTTP://WWW.KM.FGG.UNI-LJ.SI/PREDMETI/SEI/OTAB.PDF. Vidimo, da so v preglednici podane le vrednosti porazdelitvene funkcije za pozitivne vrednosti spremenljivke. Na primer: vrednost porazdelitvene funkcije standardizirane normalne porazdelitve pri $u = 1.23$ je $F_U(1.23) = 0.89065$. Za določitev vrednosti porazdelitvene funkcije pri negativnih vrednosti moramo upoštevati simetrijo normalne porazdeltive (glej sliko 7.12). Vzemimo, da poznamo vrednost porazdelitvene funkcije pri $u = u_2 > 0$. Če želimo določiti vrednost porazdelitvene funkcije pri $u = u_1 = -u_2$, moramo upoštevati simetrijo normalne porazdelitve. Iz slike vidimo, da sta verjetnosti $P[U \leq u_1]$ in $P[U > u_2]$ enaki in ju označimo s $p$. Ker je $P[U \leq u_2] + P[U > u_2] = 1$, lahko zapišemo $P[U \leq u_2] = 1 - p$. Sedaj lahko zapišemo enačbo

$$\left.\begin{array}{l} P[U \leq u_1] = F_U(u_1) = p \\ P[U \leq u_2] = F_U(u_2) = F_U(-u_1) = 1 - p \end{array}\right\} \longrightarrow \left\{\begin{array}{l} F_U(-u_1) = 1 - F_U(u_1) \\ F_U(u_1) = 1 - F_U(-u_1) \end{array}\right.$$




Slika 7.12: Simetrija standardizirane normalne porazdelitve

Verjetnost, da je trdnost betona manjša od 20 MPa, je enaka

$$P[X \leq 20] = P[U \leq -1.4] = F_U(-1.4) = 1 - F_U(1.4) = 1 - 0.91924 = 0.08076.$$

Verjetnost, da je trdnost betona višja od 30 MPa, določimo na podoben način

$$P[X > 30] = 1 - P[X \leq 30] = 1 - F_X(30) =$$
$$= 1 - F_U\left(\frac{30-27}{5}\right) = 1 - F_U(0.6) = 1 - 0.72575 = 0.27425,$$

kjer smo upoštevali, da za zvezno slučajno spremenljivko $X$ velja $P[X = 30] = 0$.

V zaključku tega primera bi se lahko vprašali, ali je pravilno, da trdnost materiala, za katero vemo, da je vedno večja od nič, modeliramo z normalno porazdelitvijo, za katero so zaloga vrednosti vsa realna števila, tudi negativna. Pomislek je seveda upravičen, a je v teh primerih verjetnost, da je slučajna spremenljivka negativna, tako nizka, da lahko to zanemarimo. V našem primeru je verjetnost, da je trdnost negativna, enaka

$$P[X < 0] = F_X(0) = F_U\left(\frac{0-27}{5}\right) = F_U(-5.4) = 3.33 \cdot 10^{-8} \approx 0.$$

**Primer 7.11:** *Predpostavimo, da se teža slovenskih prvošolčkov $X$ porazdeljuje normalno. Raziskave kažejo, da je 15% prvošolčkov lažjih od 25 kg, 10% pa težjih od 33 kg.*

*a) Določimo pričakovano vrednost $m_X$ in standardno deviacijo $\sigma_X$!*
*b) Tina tehta 22 kg. Ali je med 5% najlažjih?*

**Rešitev:** Zapišimo podatke z enačbami:

$$P[X \leq 25] = 0.15, \qquad P[X > 33] = 0.10 \;\rightarrow\; P[X \leq 33] = 0.90.$$

Neenačbe v izrazih za verjetnosti transformiramo tako, da dobimo standardizirano normalno porazdelitev

$$P\left[\frac{X-m_X}{\sigma_X} \leq \frac{25-m_X}{\sigma_X}\right] = P\left[U \leq \frac{25-m_X}{\sigma_X}\right] = F_U\left(\frac{25-m_X}{\sigma_X}\right) = 0.15,$$
$$P\left[\frac{X-m_X}{\sigma_X} \leq \frac{33-m_X}{\sigma_X}\right] = P\left[U \leq \frac{33-m_X}{\sigma_X}\right] = F_U\left(\frac{33-m_X}{\sigma_X}\right) = 0.90,$$

kjer je $F_U(u)$ porazdelitvena funkcija standardizirane normalne porazdelitve. Vrednosti inverzne funkcije $F_U^{-1}(0.15)$ in $F_U^{-1}(0.90)$ lahko odčitamo iz preglednic ali pa uporabimo računalniški program (na primer EXCEL: ukaza `NORMINV(0.15;0;1)` in `NORMINV(0.90;0;1)`) ali `NORMSINV(0.15)` in `NORMSINV(0.90)`. Ti vrednosti sta $F_U^{-1}(0.15) = -1.0364$ in $F_U^{-1}(0.90) = 1.2816$. Sedaj lahko zapišemo

$$\frac{25-m_X}{\sigma_X} = F_U^{-1}(0.15) = -1.0364, \qquad \frac{33-m_X}{\sigma_X} = F_U^{-1}(0.90) = 1.2816.$$

Če zgornji enačbi pomnožimo s $\sigma_X$, dobimo sistem dveh linearnih enačb z dvema neznankama ($m_X$ in $\sigma_X$). Rešitev tega sistema je

$$m_X = 28.5770\,\text{kg}, \qquad \sigma_X = 3.4513\,\text{kg}.$$

Določiti moramo še verjetnost $P[X \leq 22]$, kar lahko določimo tako, da neenačbo v izrazu za verjetnost transformiramo

$$P\left[\frac{X - 28.5770}{3.4513} \leq \frac{22 - 28.5770}{3.4513}\right] = P\left[U \leq -1.9057\right] = 0.0283,$$

kar lahko odčitamo iz preglednic za standardizirano normalno porazdelitev ali pa z uporabo računalniškega programa (na primer EXCEL: če računamo za standardizirano normalno porazdelitev, uporabimo ukaz `NORMSDIST(-1.9057)` ali `NORMDIST(-1.9057;0;1;TRUE)`, če pa računamo za poljubno normalno porazdelitev, uporabimo `NORMDIST(22;28.5770;3.4513;TRUE)`). Vidimo torej, da je Tina med 5% najlažjih prvošolčkov, saj je le 2.83% njenih sošolcev lažjih.

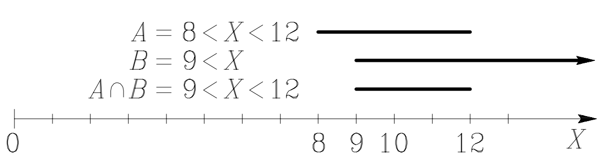
**Primer 7.12:** *Vzemimo, da je debelina asfaltne prevleke normalno porazdeljena slučajna spremenljivka $X$, s pričakovano vrednostjo $m_X = 10\,\mathrm{cm}$ in standardno deviacijo $\sigma_X = 1\,\mathrm{cm}$. Določimo verjetnost, da je odstopanje od pričakovane vrednosti manjše od 2 cm ob pogoju, da je debelina prevleke večja od 9 cm: $P[|X - m_X| \leq 2 | X > 9]$.*

**Rešitev:** Pogojno verjetnost izračunamo po enačbi (3.11)

$$P[|X - m_X| \leq 2 | X > 9] = \frac{P[|X - m_X| \leq 2 \cap X > 9]}{P[X > 9]}$$

Verjetnost, da je odstopanje od pričakovane vrednost manjše od 2 cm, lahko zapišemo tudi takole:

$$P[|X - m_X| \leq 2] = P[-2 \leq X - m_X \leq 2] = P[8 \leq X \leq 12].$$




Slika 7.13: Določitev produkta dveh dogodkov

Določiti moramo dogodek $(8 \leq X \leq 12) \cap (X > 9)$, kar je morda najlažje določiti s preproste slike (7.13), od koder vidimo, da je dogodek $(8 \leq X \leq 12) \cap (X > 9)$ enak dogodku $9 < X \leq 12$. Sedaj lahko ponovno zapišemo pogojno verjetnost

$$P[|X - m_X| \leq 2 | X > 9] = \frac{P[9 < X \leq 12]}{P[X > 9]} = \frac{F_X(12) - F_X(9)}{1 - F_X(9)}.$$

Z računalniškim programom EXCEL ali s pretvorbo na standardizirano normalno porazdelitveno funkcijo $F_U(u)$ določimo iskano verjetnost

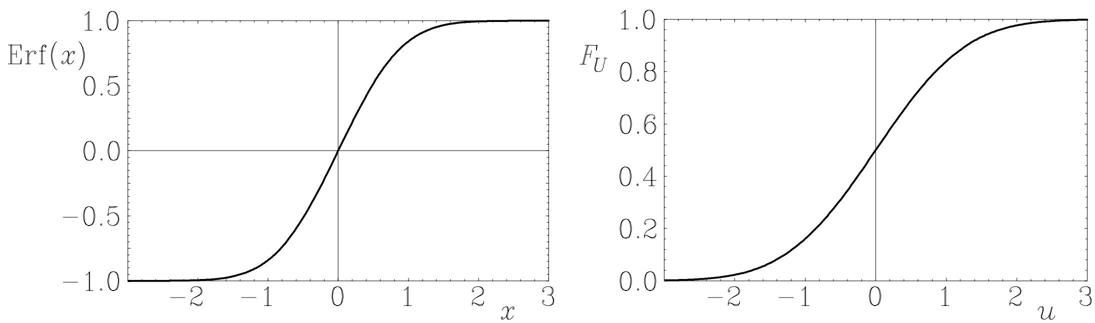
$$P[|X - m_X| \leq 2 | X > 9] = \frac{0.97725 - 0.15866}{1 - 0.15866} = \frac{0.81859}{0.84134} = 0.973$$

Verjetnost, da je odstopanje od pričakovane vrednosti manjše od 2 cm ob pogoju, da je debelina asfaltne prevleke večja od 9 cm, je 0.973.

**Primer 7.13:** *V računalniških programih* MATHEMATICA *in* MATLAB *je vgrajena funkcija* $\mathrm{Erf}(x)$*, ki je definirana z naslednjo enačbo*

$$\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}} \int\limits_0^x e^{-\tilde{x}^2}\, d\tilde{x}.$$

*Izrazimo porazdelitveno funkcijo* $F_U(u)$ *standardizirane normalne porazdelitve s funkcijo* $\mathrm{Erf}(x)$ *(glej sliko 7.14) in inverzno funkcijo* $F_u^{-1}(p)$*.*




Slika 7.14: Funkcija $\mathrm{Erf}(x)$ in porazdelitvena funkcija standardizirane normalne porazdelitve $F_U(u)$

**Rešitev:** Zaradi simetričnosti porazdelitve standardizirane normalne porazdelitve in zaradi enačbe (4.8) velja, da je

$$\int\limits_{-\infty}^0 f_U(u)\, du = 0.5.$$

Zato lahko napišemo porazdelitveno funkcijo $F_U(u)$ z naslednjo enačbo

$$F_U(u) = \frac{1}{2} + \int\limits_0^u \frac{1}{\sqrt{2\pi}} e^{-\tilde{u}^2/2} d\tilde{u}. \tag{7.32}$$

Z vpeljavo nove spremenljivke $\tilde{u}^2/2 = \tilde{x}^2$ bomo integral v (7.32) preoblikovali tako, da bomo porazdelitveno funkcijo $F_U(u)$ lahko izrazili s funkcjo $\mathrm{Erf}(x)$:

$$\tilde{u}^2/2 = \tilde{x}^2 \quad \longrightarrow \quad \tilde{u}/\sqrt{2} = \tilde{x} \quad \longrightarrow \quad d\tilde{u}/\sqrt{2} = d\tilde{x} \quad \longrightarrow \quad d\tilde{u} = \sqrt{2}\, d\tilde{x}. \tag{7.33}$$

Z vpeljavo nove spremenljivke se spremenijo tudi meje integriranja:

$$\tilde{u} \in [0, u] \quad \longrightarrow \quad \tilde{x} \in \left[0, \frac{u}{\sqrt{2}}\right]. \tag{7.34}$$

Če oznake v (7.33) in (7.34) vključimo v (7.32), dobimo

$$
\begin{aligned}
F_U(u) &= \frac{1}{2} + \int\limits_0^{u/\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-\tilde{x}^2} \sqrt{2}\, d\tilde{x} = \frac{1}{2} + \frac{1}{2}\frac{2}{\sqrt{\pi}} \int\limits_0^{u/\sqrt{2}} e^{-\tilde{x}^2}\, d\tilde{x} = \\
&= \frac{1}{2}\left(1 + \mathrm{Erf}\left(u/\sqrt{2}\right)\right).
\end{aligned}
\tag{7.35}
$$

Izraz v (7.35) smo izpeljali zato, da bomo lahko z računalniškim programom MATHEMATICA določili vrednost porazdelitvene funkcije standardizirane normalne porazdelitve.[11] Pri reševanju statističnih nalog pogosto uporabljamo porazdelitveni funkciji $F_U(u)$ inverzno funkcijo $F_U^{-1}(p)$, s katero lahko določimo tisto vrednost $u$, za katero velja

$$F_U(u) = P[U \leq u] = p \qquad \longrightarrow \qquad u = F_U^{-1}(p).$$

Ugotoviti moramo, kako lahko za določitev $F_U^{-1}(p)$ uporabimo inverzno funkcijo $\mathrm{Erf}^{-1}(x)$, ki je vgrajena v računalniški program MATLAB: funkcija `erfinv(x)`. Iz enačbe (7.35) sledi

$$\begin{aligned}
& p = F_U(u) = \frac{1}{2}\left(1 + \mathrm{Erf}\left(u/\sqrt{2}\right)\right) \quad \longrightarrow \\
& \longrightarrow \quad 2p - 1 = \mathrm{Erf}\left(u/\sqrt{2}\right) \quad \longrightarrow \\
& \longrightarrow \quad \mathrm{Erf}^{-1}(2p-1) = u/\sqrt{2} \quad \longrightarrow \\
& \longrightarrow \quad F_U^{-1}(p) = u = \sqrt{2}\,\mathrm{Erf}^{-1}(2p-1).
\end{aligned}$$

### 7.3.4 Vsota normalno porazdeljenih slučajnih spremenljivk

Glede na centralni limitni izrek, ki pravi, da je vsota mnogih neodvisnih slučajnih spremenljivk normalno porazdeljena slučajna spremenljivka, lahko pričakujemo, da je vsota dveh normalno porazdeljenih slučajnih spremenljivk tudi normalno porazdeljena. To lahko dokažemo tako, da določimo konvolucijski integral (enačba (5.20) za dve normalno porazdeljeni slučajni spremenljivki $X_1$ in $X_2$ s pričakovanima vrednostima $m_{X1}$ in $m_{X2}$ ter standardnima deviacijama $\sigma_{X1}$ in $\sigma_{X2}$

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{X1}}\, e^{-\frac{1}{2}\left(\frac{\tilde{x}-m_{X1}}{\sigma_{X1}}\right)^2} \frac{1}{\sqrt{2\pi}\sigma_{X2}}\, e^{-\frac{1}{2}\left(\frac{x-\tilde{x}-m_{X2}}{\sigma_{X2}}\right)^2} d\tilde{x}. \tag{7.36}$$

Ta integral lahko določimo s programom MATHEMATICA [12]. Rezultat, ki ga zaradi nepreglednosti ne prikazujemo tako, kot ga prikaže program MATHEMATICA, je

$$f_X(x) = \frac{1}{\sqrt{2\pi(\sigma_{X1}^2 + \sigma_{X2}^2)}}\, e^{-\frac{1}{2}\frac{(x-m_{X1}-m_{X2})^2}{\sigma_{X1}^2+\sigma_{X2}^2}}. \tag{7.37}$$

[11] Določitev vrednosti porazdelitvene funkcije $F_U(u)$ s programom MATHEMATICA:

```
FU[u_] = 1/2 ( 1 + Erf[u/Sqrt[2]]);
Print["FU[-1.4] = ",FU[-1.4]];
Print["FU[0.6] = ",FU[0.6]];

FU[-1.4] = 0.0807567
FU[0.6] = 0.725747
```

[12] Konvolucijski integral za dve normalno porazdeljeni slučajni spremenljivki s programom MATHEMATICA:

```
fx1[x_] = 1/(Sqrt[2 Pi] sx1) E^(-1/2 ((x-mx1)/sx1)^2);
fx2[x_] = 1/(Sqrt[2 Pi] sx2) E^(-1/2 ((x-mx2)/sx2)^2);
fy[y_] = Integrate[fx1[x] fx2[y-x],{x,-Infinity,Infinity}];
Print["fy(y) = ",fy[y]];
```

Iz enačbe (7.37) lahko zaključimo, da je slučajna spremenljivka $X$, ki predstavlja vsoto dveh normalno porazdeljenih slučajnih spremenljivk $X_1$ in $X_2$, tudi normalno porazdeljena slučajna spremenljivka s pričakovano vrednostjo $m_X$ in standardno deviacijo $\sigma_X$

$$m_X = m_{X1} + m_{X2} \qquad \text{in} \qquad \sigma_X^2 = \sigma_{X1}^2 + \sigma_{X2}^2. \tag{7.38}$$

Na podoben način lahko pokažemo, da je linearna kombinacija normalno porazdeljenih slučajnih spremenljivk $X_1$ in $X_2$

$$X = a_1\, X_1 + a_2\, X_2$$

tudi normalno porazdeljena slučajna spremenljivka s pričakovano vrednostjo $m_X$ in standardno deviacijo $\sigma_X$

$$m_X = a_1\, m_{X1} + a_2\, m_{X2} \qquad \text{in} \qquad \sigma_X^2 = a_1^2\, \sigma_{X1}^2 + a_2^2\, \sigma_{X2}^2. \tag{7.39}$$

**Primer 7.14:** *Če obnašanje konstrukcije obravnavamo dovolj poenostavljeno, lahko celoten vpliv obtežbe opišemo z eno slučajno spremenljivko $S$, celoten vpliv trdnosti oziroma odpornosti konstrukcije pa s slučajno spremenljivko $T$. Obe slučajni spremenljivki merimo z istimi enotami. Verjetnost porušitve konstrukcije $p_f$ torej lahko izračunamo z naslednjo enačbo*

$$p_f = P[S \geq T] = P[T - S \leq 0].$$

*Vzemimo, da sta slučajni spremenljivki $S$ in $T$ neodvisni in normalno porazdeljeni slučajni spremenljivki ($m_S = 30$, $\sigma_S = 4$, $m_T = 50$, $\sigma_T = 3$). Določimo verjetnost porušitve obravnavane konstrukcije!*

**Rešitev:** Definirajmo slučajno spremenljivko $M = T - S$ in jo imenujmo **varnost konstrukcije**. Rečemo lahko, da je konstrukcija varna, če je slučajna spremenljivka $M$ večja od nič. Ker sta $S$ in $T$ normalno porazdeljeni slučajni spremenljivki, je tudi njuna razlika porazdeljena normalno. Pričakovano vrednost $m_M$ in standardno deviacijo $\sigma_M$ izračunamo po enačbah (7.39)

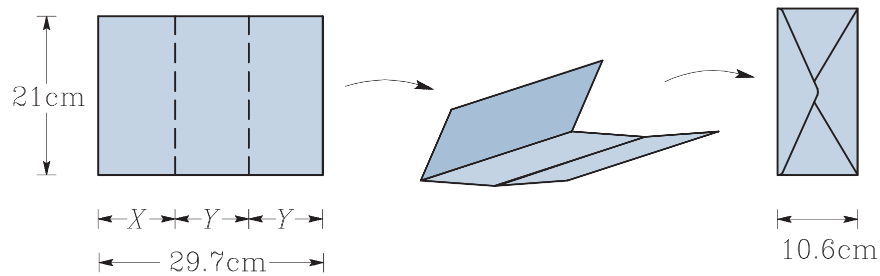
$$m_M = m_T - m_S = 20 \qquad \text{in} \qquad \sigma_M = \sqrt{\sigma_T^2 + \sigma_S^2} = 5.$$

Verjetnost porušitve lahko izračunamo po naslednji enačbi

$$P[M \leq 0] = P\left[\frac{M - m_M}{\sigma_M} \leq \frac{0 - 20}{5}\right] = P[U \leq -4] = F_U(-4) = 0.00003167,$$

kjer smo z $U$ označili standardizirano normalno porazdeljeno slučajno spremenljivko. V teoriji zanesljivosti konstrukcij vrednost $\frac{m_M}{\sigma_M}$ pogosto označimo z $\beta$ in imenujemo **indeks zanesljivosti**. Večji indeks zanesljivosti predstavlja bolj zanesljivo konstrukcijo. Verjetnost porušitve konstrukcije je v tem primeru manjša. Za običajne gradbeniške konstrukcije zahtevamo, da je $\beta$ večji od 3.8 za predvideno življenjsko dobo konstrukcije in 4.7 za enoletno obdobje.

**Primer 7.15:** *List papirja, na katerega smo napisali pismo, je formata A4* $(21 \times 29.7\,\mathrm{cm})$*. Preden pismo odpošljemo, moramo list prepogniti. Najprej ga prepognemo približno na tretjini (razdalja* $X$ *od enega roba), nato znamo preostanek prepogniti točno na polovici (glej sliko). Predpostavimo, da je* $X$ *normalno porazdeljena slučajna spremenljivka s pričakovano vrednostjo* $m_X = 9.9\,\mathrm{cm}$ *in standardno deviacijo* $\sigma_X = 1\,\mathrm{cm}$*. Določite porazdelitev* $Y$*. Kolikšna je verjetnost, da pisma ne bomo mogli spraviti v 10.6 cm široko kuverto.*




Slika 7.15: Prepogibanje lista papirja

**Rešitev:** Iz slike 7.15 lahko izpeljemo zvezo med slučajnima spremenljivkama $X$ in $Y$:

$$Y = \frac{29.7 - X}{2}.$$

Ker je slučajna spremenljivka $X$ porazdeljena normalno, je tako porazdeljena tudi slučajna spremenljivka $Y$. Pričakovano vrednost $m_Y$ in standardno deviacijo $\sigma_Y$ izračunamo po naslednjih enačbah

$$m_Y = \frac{29.7 - m_X}{2} = 9.9\,\mathrm{cm} \qquad \text{in} \qquad \sigma_Y^2 = \frac{\sigma_X^2}{4} \quad \rightarrow \quad \sigma_Y = \frac{\sigma_X}{2} = 0.5\,\mathrm{cm}.$$

Verjetnost, da pismo ne gre v kuverto, je enaka verjetnosti, da je $X$ ali $Y$ večja od 10.6 cm. Ker $X$ in $Y$ ne moreta biti istočasno večja od 10.6, sta dogodka $X > 10.6$ in $Y > 10.6$ nezdružljiva in lahko verjetnost vsote teh dogodkov napišem kot vsoto verjetnosti

$$\begin{aligned}
&P[X > 10.6 \cup Y > 10.6] = P[X > 10.6] + P[Y > 10.6] = \\
&= (1 - P[X \leq 10.6]) + (1 - P[Y \leq 10.6]) = 2 - F_X(10.6) - F_Y(10.6) = \\
&= 2 - F_U\left(\frac{10.6 - 9.9}{1}\right) - F_U\left(\frac{10.6 - 9.9}{0.5}\right) = 2 - F_U(0.7) - F_U(1.4) = \\
&= 2 - 0.91924 - 0.75804 = 0.3227
\end{aligned}$$

### 7.3.5 Normalna porazdelitev kot aproksimacija drugih porazdelitev

V prejšnjih razdelkih smo pokazali, da nekatere slučajne spremenljivke predstavljajo vsoto drugih slučajnih spremenljivk. Naštejmo take primere:

- Slučajna spremenljivka, porazdeljena po binomski porazdelitvi, je vsota slučajnih spremenljivk, porazdeljenih po Bernoullijevi porazdelitvi.

- Slučajna spremenljivka, porazdeljena po Pascalovi porazdelitvi, je vsota slučajnih spremenljivk, porazdeljenih po geometrični porazdelitvi.

- Slučajna spremenljivka, porazdeljena po porazdelitvi gama, je vsota slučajnih spremenljivk, porazdeljenih po eksponentni porazdelitvi.

Ker smo s centralnim limitnim izrekom pokazali, da je vsota mnogih neodvisnih in enako porazdeljenih slučajnih spremenljivk, normalno porazdeljena slučajna spremenljivka, lahko pričakujemo, da bo binomska porazdelitev za velike $n$ podobna normalni porazdelitvi. Enako velja za Pascalovo porazdelitev in porazdelitev gama za velike vrednosti $k$. Z naslednjim računskim primerom bomo pokazali, na kakšen način lahko aproksimiramo binomsko porazdelitev z normalno.

**Primer 7.16:** *Primerjajmo slučajni spremenljivki $Y$, ki je porazdeljena po binomski porazdelitvi, s slučajno spremenljivko $X$, porazdeljeno po normalni porazdelitvi z istimi pričakovanima vrednostima in standardnima deviacijama ($m_X = m_Y$ in $\sigma_X = \sigma_Y$). Parametra binomske porazdelitve sta $n = 10$ in $p = 0.5$. Določimo verjetnosti, da sta slučajni spremenljivki $X$ in $Y$ manjši ali enaki pet: $P[X \leq 5]$ in $P[Y \leq 5]$.*

**Rešitev:** Izračunajmo najprej verjetnost, da je slučajna spremenljivka $Y$, ki je porazdeljena po binomski porazdelitvi, manjša ali enaka 5. Ta verjetnost je vsota verjetnosti, da je $Y = 0$, $Y = 1$, $Y = 2$, $Y = 3$, $Y = 4$ in $Y = 5$:

$$
\begin{aligned}
P[Y = 0] &= \tbinom{10}{0} \cdot 0.5^0 \cdot 0.5^{10} = 0.5^{10} = 0.00098\\
P[Y = 1] &= \tbinom{10}{1} \cdot 0.5^1 \cdot 0.5^{9} = 10 \cdot 0.5^{10} = 0.00977\\
P[Y = 2] &= \tbinom{10}{2} \cdot 0.5^2 \cdot 0.5^{8} = 45 \cdot 0.5^{10} = 0.04395\\
P[Y = 3] &= \tbinom{10}{3} \cdot 0.5^3 \cdot 0.5^{7} = 120 \cdot 0.5^{10} = 0.11719\\
P[Y = 4] &= \tbinom{10}{4} \cdot 0.5^4 \cdot 0.5^{6} = 210 \cdot 0.5^{10} = 0.20508\\
P[Y = 5] &= \tbinom{10}{5} \cdot 0.5^5 \cdot 0.5^{5} = 252 \cdot 0.5^{10} = 0.24609
\end{aligned}
$$

$$
P[Y \leq 5] = \sum_{y=0}^{5} P[Y = y] = 0.632
$$

Verjetnost, da je slučajna spremenjivka $X$, ki je porazdeljena po normalni porazdelitvi, manjša ali enaka 5, izračunamo iz porazdelitvene funkcije. Najprej izračunajmo pričakovano vrednost $m_X$ in standardno deviacijo $\sigma_X$ te slučajne spremenljivke

$$
m_X = m_Y = n\,p = 5, \qquad \sigma_X = \sigma_Y = \sqrt{n\,p\,(1-p)} = \sqrt{2.5} = 1.581
$$

Sedaj lahko izračunamo verjetnost, da je $X$ manjši ali enak 5

$$
P[X \leq 5] = F_X(5) = F_U\left(\frac{5-5}{\sqrt{2.5}}\right) = F_U(0) = 0.5.
$$

Dobili smo rezultat, ki se zelo razlikuje od tistega, ki smo ga izračunali ob upoštevanju binomske porazdelitve. Hitro lahko ugotovimo, zakaj je do te velike razlike prišlo. Iz slike 7.16 vidimo, da verjetnosti, da slučajna spremenljivka $Y$ zavzame neko določeno vrednost $y$, ustreza verjetnosti, da slučajna spremenljivka $X$ zavzame vrednost v območju okoli vrednosti $y$. Tako lahko na primer napišemo

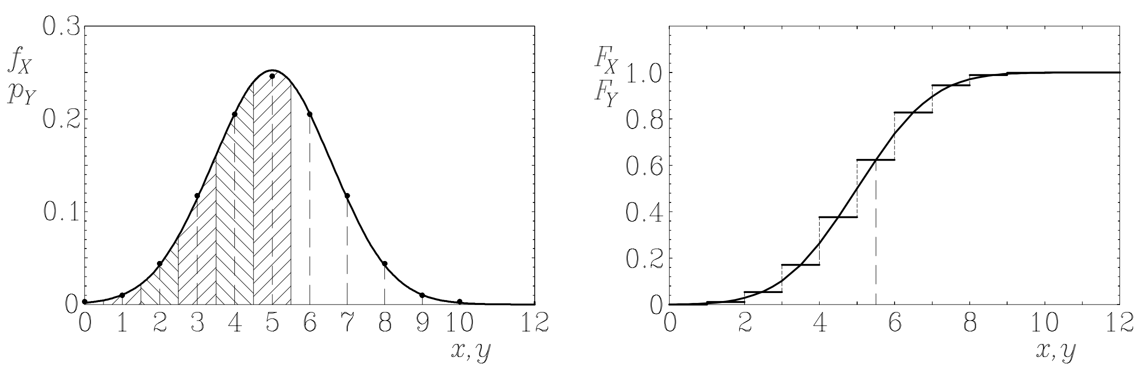
$$P[Y = 3] = 0.11719 \approx P[2.5 < X \leq 3.5] = F_X(3.5) - F_X(2.5) = 0.11447.$$

Zato sledi, moramo za določitev verjetnosti, da je slučajna spremenljivka manjša ali enaka 5, napisati naslednji izraz

$$P[Y \leq 5] \approx P[X \leq 5.5] = F_X(5.5) = F_U\left(\frac{5.5-5}{\sqrt{2.5}}\right) = F_U(0.316) = 0.624.$$

Očitno je ta približek bistveno boljši od prejšnjega. Ta rezultat lahko odčitamo tudi iz grafov porazdelitvenih funkcij za obe porazdelitvi, kjer lahko vidimo, da se porazdelitveni funkciji sekajo približno pri 5.5

$$F_Y(5.5) \approx F_X(5.5).$$




Slika 7.16: Primerjava binomske in normalne porazdelitve

Pri analizi rezultatov te naloge moramo poudariti, da so razlike med normalno in binomsko porazdelitvijo tako majhne tudi zato, ker nas je zanimala vrednost porazdelitvene funkcije na sredini porazdelitve. V primeru, da bi primerjali vrednosti porazdelitvenih funkcij na obeh repih porazdelitve, bi ugotovili, da so razlike bistveno večje. Primerjajmo na primer verjetnosti, da sta slučajni spremenljivki manjši od $-0.5$. Za binomsko porazdelitev seveda velja, da je $P[Y \leq -0.5] = 0$. Verjetnost, da je slučajna spremenljivka $Y$ manjša od $-0.5$ pa ni enaka nič, temveč je $P[X \leq -0.5] = F_X(-0.5) = 0.00025$, kar je sicer majhna a od nič različna verjetnost. V gradbeništvu pri analizi varnosti konstrukcije pogosto obravnavamo zelo majhne verjetnosti in je lahko razlika, med neverjetnim dogodkom in malo verjetnim dogodkom, zelo pomembna.

## 7.4 Logaritemsko normalna porazdelitev

V inženirstvu je marsikatera količina rezultat produkta mnogih drugih količin. Najbolj značilni primeri takih količin so: velikost delcev v sedimentih, velikost delcev agregata po drobljenju, trdnost nekaterih materialov, magnituda potresov in podobno. Zato je logaritemsko normalna porazdelitev, ki jo opisujemo v tem razdelku, za inženirstvo zelo pomembna.

Vzemimo, da imamo $n$ neodvisnih in enako porazdeljenih slučajnih spremenljivk $X_i$, ki imajo enako pričakovano vrednost $m_{X_i} = m_X$ in standardno deviacijo $\sigma_{X_i} = \sigma_X$. Definirajmo slučajno spremenljivko $Y$ kot produkt slučajnih spremenljivk $X_i$

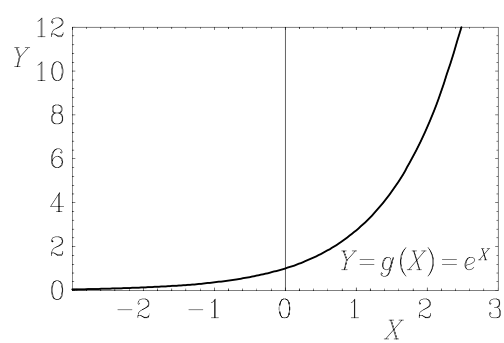
$$Y = \prod_{i=1}^{n} X_i. \tag{7.40}$$

Zanima nas, kako se porazdeljuje slučajna spremenljivka $Y$, ko gre $n$ proti neskončnosti. Nalogo rešimo tako, da enačbo (7.40) logaritmiramo in upoštevamo, da je logaritem produkta enak vsoti logaritmov

$$\ln Y = \ln\left(\prod_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \ln X_i = X. \tag{7.41}$$

Če so slučajne spremenljivke $X_i$ neodvisne in enako porazdeljene, so tudi slučajne spremenljivke $\ln X_i$ medsebojno neodvisne in enako porazdeljene. Zato po centralnem limitnem izreku velja, da je porazdelitev slučajne spremenljivke $X$ v enačbi (7.41) normalna, če gre število slučajnih spremenljivk $X_i$ proti neskončnosti. Zapišimo definicijo logaritemsko normalne porazdelitve: *Če je slučajna spremenljivka $X$ porazdeljena normalno in velja zveza $\ln Y = X$, potem je slučajna spremenljivka $Y$ porazdeljena* **logaritemsko normalno**. Izpeljimo gostoto verjetnosti $f_Y(y)$ logaritemsko normalne porazdelitve! Slučajna spremenljivka $Y$ je funkcija normalno porazdeljene slučajne spremenljivke $X$

$$Y = g(X) = e^X.$$




Slika 7.17: Zveza med normalno in logaritemsko normalno porazdeljeno slučajno spremenljivko

Zvezo med slučajnima spremenljivkama prikazujemo na sliki 7.17. Iz slike lahko tudi ugotovimo, da zaloga vrednosti slučajne spremenljivke $Y$ obsega vsa pozitivna realna števila. Očitno je, da je funkcija $g(X)$ monotona, zato za izpeljavo gostote verjetnosti uporabimo enačbo (5.12). Inverzna funkcija $g^{-1}(Y)$ in njen odvod sta

$$g^{-1}(Y) = \ln Y, \qquad \frac{dg^{-1}(Y)}{dY} = \frac{1}{Y}. \tag{7.42}$$

Gostoto verjetnosti slučajne spremenljivke $Y$ lahko sedaj zapišemo takole:

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right| = \frac{1}{\sqrt{2\pi}\,\sigma_X}\, e^{-\frac{1}{2}\left(\frac{\ln y - m_X}{\sigma_X}\right)^2}\frac{1}{y} \qquad 0 < y < \infty. \tag{7.43}$$

Enačba (7.43) predstavlja gostoto verjetnosti logaritemsko normalno porazdeljene slučajne spremenljivke $Y$, katere parametra sta pričakovana vrednost in standardna deviacija slučajne spremenljivke $X$. To je nekoliko nenavadno, zato bomo ta dva parametra raje zapisali tako, da bosta predstavljala slučajno spremenljivko $Y$. Namesto standardne deviacije $\sigma_X$ bomo pisali oznako $\sigma_{\ln Y}$. Namesto pričakovane vrednosti $m_X$ pa bomo vpeljali mediano $\tilde{m}_Y$ slučajne spremenljivke $Y$, za katero velja naslednja zveza

$$P[Y \leq \tilde{m}_Y] = 0.5.$$

Izraz znotraj oglatih oklepajev lahko logaritmiramo in upoštevamo zvezo med slučajnima spremenljivkama (7.42). Upoštevamo tudi, da za mediano slučajne spremenljivke $X$ velja enak izraz kot za mediano slučajne spremenljivke $Y$ in da zaradi simetričnosti normalne porazdelitve mediana $\tilde{m}_X$ sovpada s pričakovano vrednostjo $m_X$ ($\tilde{m}_X = m_X$)

$$\begin{aligned} P[Y \leq \tilde{m}_Y] = 0.5 &= P[\ln Y \leq \ln \tilde{m}_Y] = P[X \leq \ln \tilde{m}_Y] = \\ &= F_X(\ln \tilde{m}_Y) = F_X(\tilde{m}_X) = F_X(m_X). \end{aligned} \tag{7.44}$$

Iz enačbe (7.44) lahko sklepamo, da je

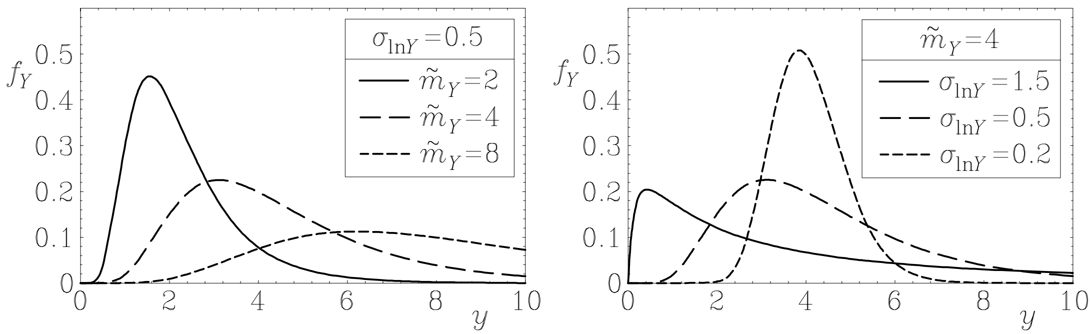
$$\ln \tilde{m}_Y = m_X, \tag{7.45}$$

zato bomo v enačbi (7.43) namesto pričakovane vrednosti $m_X$ pisali logaritem mediane $\ln \tilde{m}_Y$ slučajne spremenljivke $Y$. Gostoto verjetnosti $f_Y(y)$ logaritemsko normalno porazdeljene slučajne spremenljivke z novimi oznakami zapišemo takole:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma_{\ln Y}\, y}\, e^{-\frac{1}{2}\left(\frac{\ln y - \ln \tilde{m}_Y}{\sigma_{\ln Y}}\right)^2} \qquad 0 < y < \infty. \tag{7.46}$$

Na sliki 7.18 prikazujemo gostote verjenosti za nekaj različnih vrednosti parametrov $\tilde{m}_X$ in $\sigma_{\ln Y}$. Pričakovana vrednost $m_Y$ in varianca $\sigma_Y^2$ slučajne spremeljivke $Y$ sta

$$m_Y = \tilde{m}_Y\, e^{\frac{1}{2}\sigma_{\ln Y}^2}, \qquad \sigma_Y^2 = m_Y^2\left(e^{\sigma_{\ln Y}^2} - 1\right).$$




Slika 7.18: Gostota verjetnosti logaritemsko normalno porazdeljene slučajne spremenljivke

Pogosto moramo uporabiti obratne zveze, po katerih iz pričakovane vrednosti $m_Y$ in standardne deviacije $\sigma_Y$ izračunamo parametra $\tilde{m}_Y$ in $\sigma_{\ln Y}$

$$\begin{aligned}
\tilde{m}_Y &= m_Y\, e^{-\frac{1}{2}\sigma^2_{\ln Y}} = \frac{m_Y^2}{\sqrt{\sigma_Y^2 + m_Y^2}} = \frac{m_Y}{\sqrt{\frac{\sigma_Y^2}{m_Y^2} + 1}} = \frac{m_Y}{\sqrt{V_Y^2 + 1}}, \\
\sigma^2_{\ln Y} &= \ln\left(\frac{\sigma_Y^2}{m_Y^2} + 1\right) = \ln\left(V_Y^2 + 1\right),
\end{aligned} \tag{7.47}$$

kjer smo z $V_Y = \sigma_Y / m_Y = \sqrt{e^{\sigma^2_{\ln Y}} - 1}$ označili koeficient variacije. Zapišimo še koeficienta simetričnosti in sploščenosti, $\gamma_{1Y}$ in $\gamma_{2Y}$, za logaritemsko normalno slučajno spremenljivko $Y$:

$$\begin{aligned}
\gamma_{1Y} &= \sqrt{e^{\sigma^2_{\ln Y}} - 1}\,\left(e^{\sigma^2_{\ln Y}} + 2\right) = V_Y^3 - 3V_Y, \\
\gamma_{2Y} &= e^{4\sigma^2_{\ln Y}} + 2e^{3\sigma^2_{\ln Y}} + 3e^{2\sigma^2_{\ln Y}} - 3 = V_Y^8 + 6V_Y^6 + 15V_Y^4 + 16V_Y^2 + 3.
\end{aligned}$$

Preglednica 7.1: Momenti logaritemsko normalne porazdelitve

| $\tilde{m}_Y$ | $\sigma_{\ln Y}$ | $m_Y$ | $\sigma_Y$ | $\gamma_{1Y}$ | $\gamma_{2Y}$ |
|---|---|---|---|---|---|
| 2.0 | 0.50 | 2.266 | 1.208 | 1.750 | 8.898 |
| 4.0 | 0.50 | 4.533 | 2.416 | 1.750 | 8.898 |
| 8.0 | 0.50 | 9.065 | 4.831 | 1.750 | 8.898 |
| 4.0 | 1.50 | 12.321 | 35.895 | 33.468 | 10078.3 |
| 4.0 | 0.20 | 4.081 | 0.824 | 0.614 | 3.678 |
| 4.0 | 0.10 | 4.020 | 0.403 | 0.302 | 3.162 |
| 4.0 | 0.02 | 4.001 | 0.080 | 0.060 | 3.006 |

Vidimo, da sta koeficienta $\gamma_{1Y}$ in $\gamma_{2Y}$ neodvisna od mediane $\tilde{m}_Y$. V preglednici 7.1 prikazujemo nekaj vrednosti momentov lognormalno porazdeljene slučajne spremenljivke $Y$ za različne vrednosti parametrov porazdelitve $\tilde{m}_Y$ in $\sigma_{\ln Y}$. Za večje vrednosti $\sigma_{\ln Y}$ je porazdelitev bolj nesimetrična $\gamma_{1Y} \gg 0$ in bolj špičasta $\gamma_{2Y} \gg 3$, kar lahko vidimo tudi na sliki 7.18. Za zelo nizke vrednosti $\sigma_{\ln Y}$ pa je porazdelitev bolj podobna normalni porazdelitvi, saj je simetrična in srednje sploščena: $\gamma_{1Y} \approx 0$ in $\gamma_{2Y} \approx 3$.

Podobno kot pri normalni porazdelitvi tudi pri logaritemsko normalni porazdelitvi ne poznamo porazdelitvene funkcije $F_Y(y)$ v zaključeni analitični obliki. Njeno vrednost lahko izračunamo numerično z računalniškimi programi kot so MATHEMATICA, MATLAB ali EXCEL, ali pa porazdelitveno funkcijo $F_Y(y)$ preuredimo tako, da lahko uporabimo porazdelitveno funkcijo standardizirane normalne porazdelitev, katere vrednosti lahko najdemo v preglednicah. Če v izrazu za porazdelitveno funkcijo $F_Y(y)$ upoštevamo zvezi (7.41) in (7.45), lahko izpeljemo naslednjo uporabno enačbo

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[\ln Y \leq \ln y] = P[X \leq \ln y] = P\left[\frac{X - m_X}{\sigma_X} \leq \frac{\ln y - m_X}{\sigma_X}\right] = \\ &= P\left[U \leq \frac{\ln y - \ln \tilde{m}_Y}{\sigma_{\ln Y}}\right] = F_U\left(\frac{\ln y - \ln \tilde{m}_Y}{\sigma_{\ln Y}}\right) \end{aligned} \tag{7.48}$$

Na osnovi zadnje enačbe lahko z uporabo preglednice za porazdelitveno funkcijo normalne porazdelitve izračunamo vrednost porazdelitvene funkcije logaritemsko normalne porazdelitev. To enačbo lahko uporabimo tudi v računalniškem programu MATHEMATICA, pri čemer pa uporabimo tudi zvezo med porazdelitveno funkcijo standardizirane normalne porazdelitve $F_U(u)$ in vgrajeno funkcijo $\mathrm{Erf}(x)$ (enačba (7.35)).[13]

Ugotovimo, kako se porazdeljuje produkt dveh logaritemsko normalno porazdeljenih slučajnih spremenljivk $Y_1$ in $Y_2$

$$Y = Y_1 \cdot Y_2.$$

Zgornjo enačbo logaritmiramo in ugotovimo, da je slučajna spremenljivka $X = \ln Y$ porazdeljena normalno, saj sta tudi slučajni spremenljivki $X_1 = \ln Y_1$ in $X_2 = \ln Y_2$ porazdeljeni normalno. Iz te ugotovitve zaradi definicije logaritemsko normalne porazdelitve sledi sklep, da je tudi slučajna spremenljivka $Y = Y_1 \cdot Y_2$ porazdeljena logaritemsko normalno:

$$\ln Y = \ln(Y_1 \cdot Y_2) = \ln Y_1 + \ln Y_2 = X = X_1 + X_2. \tag{7.49}$$

Iz enačb (7.38), (7.46) in (7.49) lahko izpeljemo zveze med parametri slučajnih spremenljivk $Y_1$, $Y_2$ in $Y$:

$$m_X = m_{X_1} + m_{X_2} = \ln \tilde{m}_Y = \ln \tilde{m}_{Y_1} + \ln \tilde{m}_{Y_2} \quad \longrightarrow \quad \tilde{m}_Y = \tilde{m}_{Y_1} \tilde{m}_{Y_2},$$

$$\sigma_X^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2 \quad \longrightarrow \quad \sigma_{\ln Y}^2 = \sigma_{\ln Y_1}^2 + \sigma_{\ln Y_2}^2.$$

---

[13] Izračun porazdelitvene funkcije lognormalne porazdelitve s programom MATHEMATICA:

```
Fy[y_, myt_, Slny_] := 0.5 + 1/2 Erf[(Log[y]-Log[myt])/Slny/Sqrt[2]];
Print["Fy[3] = ", Fy[3, 4, 0.5]]

Fy[3] = 0.282523
```

**Primer 7.17:** *Naj bo največji letni pretok vode v neki rečici slučajna spremenljivka $Y$. Na podlagi meritev v zadnjih 40 letih lahko izračunamo oceni za pričakovano vrednost $m_Y$ in standardno deviacijo $\sigma_Y$ največjega letnega pretoka. Ti dve oceni sta:*

$$\hat{m}_Y = \bar{Y} = 115\,\mathrm{m}^3/\mathrm{s}, \qquad \hat{\sigma}_Y = S_Y^* = 45\,\mathrm{m}^3/\mathrm{s}.$$

*Ti dve oceni smo dobili tako, da smo v izračun povprečja in standardne deviacije vzorca vzeli največjo vrednost vodnega toka v vsakem letu. V vzorcu je bilo torej štirideset vrednosti. Vzemimo, da sta oceni pričakovane vrednosti in standardne deviacije pravi in izračunajmo verjetnost, da bo v naslednjem letu vodni pretok večji od $200\,\mathrm{m}^3/\mathrm{s}$. Določimo tudi stoletno vodo; verjetnost, da bo stoletna voda v naslednjem letu presežena, je enaka $1/100$. Pri tem upoštevajmo:*

*a) normalno porazdelitev slučajne spremenljivke $Y$,*

*b) logaritemsko normalno porazdelitev slučajne spremenljivke $Y$.*

**Rešitev:** Verjetnost, da je slučajna spremenljivka $Y$ večja od 200 določimo po naslednji enačbi

$$P[Y > 200] = 1 - P[Y \leq 200] = 1 - F_Y(200).$$

Stoletna voda $y_{100}$ je tista, za katero velja, da je verjetnost, da se v nekem letu zgodi taka voda ali večja, enaka 1/100:

$$P[Y > y_{100}] = \tfrac{1}{100} \qquad \longrightarrow \qquad P[Y \leq y_{100}] = 1 - \tfrac{1}{100} = 0.99 = F_Y(y_{100}).$$

a) V primeru normalne porazdelitve slučajne spremenljivke $Y$ verjetnost $P[Y \geq 200]$ določimo takole:

$$P[Y > 200] = 1 - F_Y(200) = 1 - F_U\left(\frac{200-115}{45}\right) =$$
$$= 1 - F_U(1.888889) = 1 - 0.97055 = 0.02945.$$

Stoletno vodo $y_{100}$ določimo z naslednjo enačbo

$$0.99 = F_Y(y_{100}) = F_U\left(\frac{y_{100}-115}{45}\right) \quad \longrightarrow \quad \frac{y_{100}-115}{45} = 2.32634 \quad \longrightarrow$$
$$\longrightarrow \quad y_{100} = 45 \cdot 2.32634 + 115 = 219.685\,\mathrm{m}^3/\mathrm{s}.$$

b) Pri logaritemsko normalni porazdelitvi moramo najprej določiti parametra porazdelitve $\tilde{m}_Y$ in $\sigma_{\ln Y}$ po enačbah (7.47)

$$\sigma_{\ln Y}^2 = \ln\left(\frac{\sigma_Y^2}{m_Y^2} + 1\right) = \ln\left(\frac{45^2}{115^2} + 1\right) = 0.14247,$$

$$\tilde{m}_Y = m_Y\, e^{-\frac{1}{2}\sigma_{\ln Y}^2} = 115\, e^{-\frac{0.14247}{2}} = 107.093.$$

Sedaj lahko izračunamo verjetnost, da je $Y$ večji od 200

$$\begin{aligned} P[Y > 200] &= 1 - F_Y(200) = 1 - F_U\left(\tfrac{\ln 200 - \ln 107.093}{0.37745}\right) = \\ &= 1 - F_U(1.65483) = 1 - 0.9510 = 0.0490. \end{aligned}$$

Stoletno vodo izračunamo z naslednjo enačbo

$$\begin{aligned} & 0.99 = F_Y(y_{100}) = F_U\left(\frac{\ln y_{100} - \ln 107.093}{0.37745}\right) \quad \longrightarrow \\ & \longrightarrow \quad \frac{\ln y_{100} - \ln 107.093}{0.37745} = 2.32634 \quad \longrightarrow \\ & \longrightarrow \quad y_{100} = 107.093 \cdot e^{0.37745 \cdot 2.32634} = 257.70\,\mathrm{m}^3/\mathrm{s}. \end{aligned}$$

Verjetnost $P[Y > 200]$ in vrednost stoletne vode $y_{100}$ lahko s programom EXCEL izračunamo z ukazoma za logaritemsko normalno porazdelitev
`1-LOGNORMDIST(200;LN(107.093);0.37745)` in
`LOGINV(0.99;LN(107.093);0.37745)`.

Ugotovimo lahko, da je razlika med uporabo normalne in logaritemsko normalne porazdelitve zelo velika. Razlike so tem večje, čimvečja je vrednost $\sigma_{\ln Y}$. Razlike so večje, če primerjamo rezultate na obeh repih porazdelitve, torej daleč od pričakovanih vrednosti oziroma median.

**Primer 7.18:** *Določimo verjetnost porušitve $p_f$ konstrukcije ob predpostavki, da sta slučajni spremenljivki $S$ in $T$, ki predstavljata obtežbo in odpornost konstrukcije, neodvisni in logaritemsko normalno porazdeljeni! Naloga je podobna tisti, ki smo jo rešili v primeru 7.14. Upoštevajmo isti pričakovani vrednosti in standardni deviaciji, kot v primeru 7.14 ($m_S = 30$, $\sigma_S = 4$, $m_T = 50$, $\sigma_T = 3$).*

**Rešitev:** Določimo najprej parametre logaritemsko normalno porazdeljenih slučajnih spremenljivk $S$ in $T$

$$\begin{aligned} \sigma^2_{\ln S} &= \ln\left(\frac{\sigma_S^2}{m_S^2} + 1\right) = \ln\left(\frac{4^2}{30^2} + 1\right) = 0.01762, \\ \tilde{m}_S &= m_S\, e^{-\frac{1}{2}\sigma^2_{\ln S}} = 30\, e^{-\frac{0.01762}{2}} = 29.737, \\ \sigma^2_{\ln T} &= \ln\left(\frac{\sigma_T^2}{m_T^2} + 1\right) = \ln\left(\frac{3^2}{50^2} + 1\right) = 0.003594, \\ \tilde{m}_T &= m_T\, e^{-\frac{1}{2}\sigma^2_{\ln T}} = 50\, e^{-\frac{0.003594}{2}} = 49.910. \end{aligned}$$

Verjetnost porušitve konstrukcije $p_f$ lahko izračunamo z naslednjo enačbo

$$p_f = P[S > T] = P\left[\frac{T}{S} \le 1\right].$$

Definirajmo slučajno spremenljivko $M = T/S$ in jo imenujemo **varnost konstrukcije**. Rečemo lahko, da je konstrukcija varna, če je slučajna spremenljivka $M$ večja od ena. Ker sta $S$ in $T$ logaritemsko normalno porazdeljeni slučajni spremenljivki, je tudi njun kvocient porazdeljen logaritemsko normalno. Parametra porazdelitve izračunamo z naslednjo enačbo

$$\tilde{m}_M = \frac{\tilde{m}_T}{\tilde{m}_S} = 1.6784 \qquad \text{in} \qquad \sigma_{\ln M} = \sqrt{\sigma^2_{\ln T} + \sigma^2_{\ln S}} = 0.14565$$

Verjetnost porušitve lahko izračunamo po naslednji enačbi

$$\begin{aligned} P[M \leq 1] &= P\left[\frac{\ln M - \ln \tilde{m}_M}{\sigma_{\ln M}} \leq \frac{\ln 1 - \ln 1.6784}{0.14565}\right] = \\ &= P[U \leq -3.555] = F_U(-3.555) = 0.0001888, \end{aligned}$$

kjer smo z $U$ označili standardizirano normalno porazdeljeno slučajno spremenljivko. Indeks zanesljivosti je v tem primeru $\beta = 3.555$ in je nižji od vrednosti, ki smo ga v podobnem primeru 7.14 izračunali za normalno porazdeljeni slučajni spremenljivki $T$ in $S$. Verjetnost porušitve $p_f$ je zato večja, kot v primeru 7.14.

## 7.5 Porazdelitve ekstremnih vrednosti

V inženirstvu nas pogosto zanimajo le ekstremne vrednosti določenih spremenljivk. Na primer, konstrukcija se poruši zaradi najmočnejšega vetra. Zato nas pravzaprav zanima porazdelitev hitrosti najmočnejšega vetra in ne porazdelitev hitrosti vetra v poljubnem trenutku. Podobno velja za trdnost materiala: zanima nas porazdelitev najnižje vrednosti trdnosti materiala.

### 7.5.1 Ekstremna vrednost končno mnogo slučajnih spremenljivk

Določimo porazdelitev slučajne spremenljivke $Y$, ki je definirana kot največja izmed $n$ **začetnih slučajnih spremenljivk** $X_i$, ki so medsebojno neodvisne in enako porazdeljene

$$Y = \max_i X_i.$$

Porazdelitvena funkcija $F_Y(y)$ je

$$F_Y(y) = P[Y \leq y] = P\left[\max_i X_i \leq y\right].$$

Če je največji izmed $X_i$ manjši od $y$, potem velja, da so vsi $X_i$ manjši od $y$

$$\begin{aligned} F_Y(y) &= P[X_1 \leq y \cap X_2 \leq y \cap \ldots \cap X_n \leq y] = \\ &= P[X_1 \leq y] \cdot P[X_2 \leq y] \cdots P[X_n \leq y] = \\ &= F_{X_1}(y) \cdot F_{X_2}(y) \cdots F_{X_n}(y) = \prod_{i=1}^{n} F_{X_i}(y) = \prod_{i=1}^{n} F_X(y) = F_X^n(y), \end{aligned} \tag{7.50}$$

kjer smo upoštevali, da so slučajne spremenljivke $X_i$ neodvisne in enako porazdeljene. Na podoben način lahko izpeljemo tudi porazdelitveno funkcijo $F_Z(z)$ slučajne spremenljivke $Z$, ki predstavlja najmanjšo izmed začetnih slučajnih spremenljivk $X_i$

$$Z = \min_i X_i.$$

Porazdelitvena funkcija $F_Z(z)$ je

$$F_Z(z) = P[Z \leq z] = P\left[\min_i X_i \leq z\right] = 1 - P\left[\min_i X_i \geq z\right].$$

Če je najmanjši izmed $X_i$ večji od $z$, so vsi $X_i$ večji od $z$. Zato lahko zapišemo naslednjo enačbo

$$\begin{aligned} F_Z(z) &= 1 - P[X_1 \geq z \cap X_2 \geq z \cap \ldots \cap X_n \geq z] = \\ &= 1 - P[X_1 \geq z] \cdot P[X_2 \geq z] \cdots P[X_n \geq z] = \\ &= 1 - (1 - F_{X_1}(z)) \cdot (1 - F_{X_2}(z)) \cdots (1 - F_{X_n}(z)) = \\ &= 1 - \prod_{i=1}^{n} (1 - F_{X_i}(z)) = 1 - \prod_{i=1}^{n} (1 - F_X(z)) = 1 - (1 - F_X(z))^n, \end{aligned} \tag{7.51}$$

kjer smo upoštevali, da so $X_i$ medsebojno neodvisne in enako porazdeljene slučajne spremenljivke.

Gostoti verjetnosti slučajnih spremenljivk $Y$ in $Z$ določimo z odvajanjem porazdelitvenih funkcij v (7.50) in (7.51).

$$f_Y(y) = \frac{dF_Y(y)}{dy} = n\, F_X^{n-1}(y)\, f_X(y) \tag{7.52}$$

in

$$f_Z(z) = \frac{dF_Z(z)}{dz} = n\,(1 - F_X(z))^{n-1}\, f_X(z). \tag{7.53}$$

**Primer 7.19:** *Določimo gostoti verjetnosti slučajnih spremenljivk $Y$ in $Z$, ki predstavljata največjo oziroma najmanjšo izmed treh začetnih slučajnih spremenljivk $X_i$. Vzemimo, da se začetne slučajne spremenljivke $X_i$ porazdeljujejo enakomerno od 0 do 1.*

**Rešitev:** Porazdelitvena funkcija začetne slučajne spremenljivke $X_i$ je

$$F_{X_i}(x) = F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & 1 < x \end{cases}$$

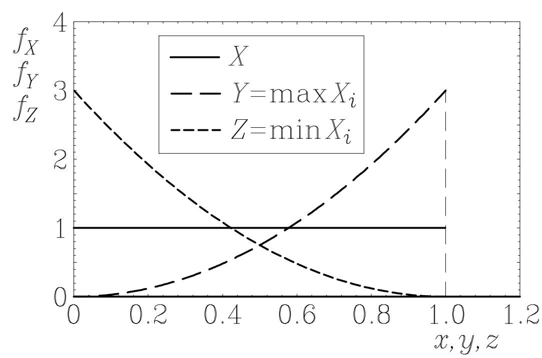
Gostoti verjetnosti slučajnih spremenljivk $Y$ in $Z$ določimo po enačbah (7.52) in (7.53)

$$f_Y(y) = n\, F_X^{n-1}(y)\, f_X(y) = \begin{cases} 0 & y < 0 \\ 3y^2 & 0 \leq y \leq 1 \\ 0 & 1 < y \end{cases}$$

oziroma

$$f_Z(z) = n\,(1 - F_X(z))^{n-1}\,f_X(z) = \begin{cases} 0 & z < 0 \\ 3(1-z)^2 & 0 \le z \le 1 \\ 0 & 1 < z \end{cases}$$

Na sliki 7.19 prikazujemo gostoto verjetnosti enakomerne slučajne spremenljivke ter gostoti verjetnosti slučajnih spremenljivk $Y$ in $Z$, ki predstavljata največjo oziroma najmanjšo vrednost izmed treh enakomerno porazdeljenih slučajnih spremenljivk $X_i$. Kot vidimo, porazdelitvi ekstremnih vrednosti nista simetrični, četudi je porazdelitev začetnih slučajnih spremenljivk $X_i$ simetrična. Vidimo tudi, da lahko pri porazdelitvi maksimuma pričakujemo višje vrednosti slučajne spremenljivke kot pri začetni porazdelitvi. Obratno velja, da lahko pri porazdelitvi minimuma pričakujemo nižje vrednosti.




Slika 7.19: Gostota verjetnosti porazdelitve minimuma in maksimuma

### 7.5.2 Asimptotične porazdelitve ekstremnih vrednosti

V prejšnjem razdelku smo opisali, kako določimo porazdelitve ekstremnih vrednosti končno mnogo slučajnih spremenljivk. Sedaj pa nas zanimajo še asimptotične porazdelitve ekstremnih vrednosti, ko gre število spremenljivk $X_i$ proti neskončnosti.

Centralni limitni izrek pove, da je vsota mnogih poljubnih enako porazdeljnih neodvisnih slučajnih spremenljivk porazdeljena normalno. Podobno velja, da je produkt mnogih poljubnih enako porazdeljenih neodvisnih slučajnih spremenljivk porazdeljen logaritemsko normalno. Pri porazdelitvah ekstremnih vrednosti je nekoliko drugače. Izkaže se, da so porazdelitve ekstremnih vrednosti odvisne od porazdelitve začetnih slučajnih spremenjivk. Na porazdelitev ekstremnih vrednosti vpliva predvsem oblika porazdelitve začetnih slučajnih spremenljivk na ustreznem repu porazdelitve. Na porazdelitev maksimuma vpliva porazdelitev začetnih slučajnih spremenljivk na **zgornjem repu**, to je pri velikih $x$, na porazdelitev minimuma pa porazdelitev začetnih slučajnih spremenljivk na **spodnjem repu** porazdelitve, to je pri majhnih $x$.

Ker oblika porazdelitve začetnih slučajnih spremenljivk $X_i$ vpliva na porazdelitev ekstremnih vrednosti, ločimo tri uporabljene tipe porazdelitev ekstremnih vrednosti. Ne glede na porazdelitev začetnih slučajnih spremenljivk je asimptotična porazdelitev ekstremnih vrednosti natanko ena izmed treh tipov porazdelitev ekstremnih vrednosti. Podrobnosti izpeljave vseh treh tipov porazdelitve lahko najdete v: D. Zupan, G. Turk, Porazdelitve ekstremnih vrednosti.

### 7.5.3 Porazdelitev ekstremnih vrednosti tipa I

Porazdelitev ekstremnih vrednosti tipa I imenujemo tudi **Gumbelova porazdelitev**.[14] Obravnavajmo najprej porazdelitev maksimuma. Če ima porazdelitvena funkcija $F_X(x)$ na zgornjem repu porazdelitve obliko

$$F_X(x) = 1 - e^{-g(x)},$$

kjer je $g(x)$ monotono naraščajoča funkcija, je porazdelitvena funkcija maksimuma slučajnih spremenljivk $X_i$, ko gre njihovo število proti neskončnosti, enaka

$$F_Y(y) = e^{-e^{-\alpha\,(y-u)}}, \qquad -\infty < y < \infty. \tag{7.54}$$

Konstanti $\alpha$ in $u$ sta parametra porazdelitve, ki sta odvisna od pričakovane vrednosti in variance slučajne spremenljivke $Y$. Gostoto verjetnosti te slučajne spremenljivke dobimo z odvajanjem porazdelitvene funkcije $F_Y(y)$ v enačbi (7.54)

$$f_Y(y) = \alpha\, e^{-\alpha\,(y-u) - e^{-\alpha\,(y-u)}}. \tag{7.55}$$

Pri reševanju nalog običajno potrebujemo zvezi med momentoma slučajne spremenljivke $m_Y$ in $\sigma_Y$ ter njenima parametroma $u$ in $\alpha$

$$m_Y = u + \frac{\gamma}{\alpha}, \qquad \sigma_Y^2 = \frac{\pi^2}{6\,\alpha^2},$$

kjer je $\gamma$ Eulerjeva[15] konstanta ($\gamma \approx 0.577216$). Brez težav lahko določimo tudi obratni zvezi:

$$\alpha = \frac{\pi}{\sqrt{6}\,\sigma_Y}, \qquad u = m_Y - \frac{\gamma}{\alpha}. \tag{7.56}$$

Na sliki 7.20 prikazujemo gostote verjetnosti za Gumbelovo porazdelitev maksimuma za nekaj različnih vrednosti parametrov $\alpha$ in $u$. Vidimo, da je razpršenost odvisna le od parametra $\alpha$, medtem ko je pričakovana vrednost odvisna od obeh parametrov $u$ in $\alpha$. Koeficienta simetričnosti in sploščenosti, $\gamma_{1Y}$ in $\gamma_{2Y}$, sta neodvisna od parametrov porazdelitve:

$$\gamma_{1Y} = 1.13955$$

in

$$\gamma_{2Y} = 5.4.$$

[14] Emil Julius Gumbel, nemško-ameriški matematik (1891-1966).
[15] Leonhard Euler, švicarski matematik (1707-1783).

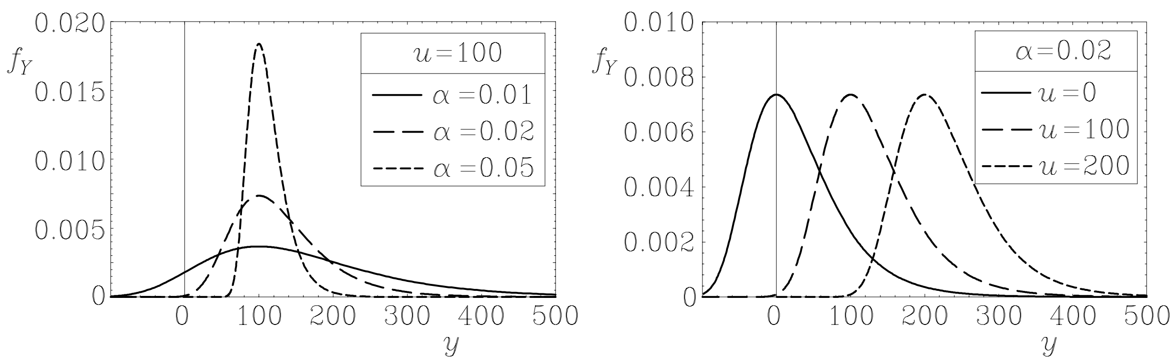
Porazdelitev je torej ne glede na parametre nesimetrična in bolj špičasta od normalne porazdelitve.

Če ima porazdelitvena funkcija na spodnjem repu porazdelitve obliko eksponentne funkcije

$$F_X(x) = e^{-g(x)},$$

je porazdelitvena funkcija minimuma slučajnih spremenljivk $X_i$, ko gre njihovo število proti neskončnosti, enaka

$$F_Z(z) = 1 - e^{-e^{\alpha\,(z-u)}}, \qquad -\infty < z < \infty. \tag{7.57}$$




Slika 7.20: Gostota verjetnosti Gumbelove porazdelitve maksimuma

Gostoto verjetnosti določimo z odvajanjem porazdelitvene funkcije

$$f_Z(z) = \alpha\, e^{\alpha\,(z-u) - e^{\alpha\,(z-u)}}.$$

Zveze med parametroma porazdelitve $\alpha$ in $u$ so podobne kot pri porazdelitvi maksimuma

$$m_Z = u - \frac{\gamma}{\alpha}, \qquad \sigma_Z^2 = \frac{\pi^2}{6\,\alpha^2}$$

in obratno

$$\alpha = \frac{\pi}{\sqrt{6}\,\sigma_Z}, \qquad u = m_Z + \frac{\gamma}{\alpha}. \tag{7.58}$$

Porazdelitev ekstremnih vrednosti tipa I oziroma Gumbelovo porazdelitev uporabljamo v hidrologiji in gradbeni konstruktivi. Za največji letni pretok nekega vodotoka pogosto predpostavimo, da se porazdeljuje po Gumbelovi porazdelitvi. To je pomembna predpostavka, saj na osnovi predpostavke o porazdelitvi in statističnih podatkov za neko relativno kratko obdobje določimo vrednost stoletnega pretoka, torej tistega pretoka, ki je v povprečju prekoračen le enkrat v 100 letih. Po Gumbelovi predpostavki se porazdeljujejo tudi nekatere obtežbe na konstrukcije ter trdnosti nekaterih materialov, kar moramo upoštevati pri določitvi zanesljivosti konstrukcij. Tudi pri tem je predpostavka o porazdelitvi zelo pomembna, saj močno vpliva na izračunano verjetnost porušitve konstrukcije.

**Primer 7.20:** *Vzemimo, da je največji letni pretok vode v neki rečici slučajna spremenljivka $Y$, porazdeljena po Gumbelovi porazdelitvi. Upoštevajmo enake podatke, kot v primeru logaritemsko normalne porazdelitve (glej primer 7.17). Izračunajmo verjetnost, da bo v naslednjem letu vodni pretok večji od* $200\,\mathrm{m}^3/\mathrm{s}$*. Določimo tudi stoletno vodo.*

**Rešitev:** Če sta oceni pričakovane vrednosti in standardne deviacije pravi, lahko iz enačb (7.56) izračunamo parametra porazdelitve $u$ in $\alpha$

$$\alpha = \frac{\pi}{\sqrt{6}\,\sigma_Y} = \frac{\pi}{\sqrt{6}\cdot 45} = 0.0285, \qquad u = m_Y - \frac{\gamma}{\alpha} = 115 - \frac{0.57722}{0.028501} = 94.75.$$

Verjetnost, da bo največji vodni pretok v naslednjem letu večji od 200, izračunamo po enačbi (7.54):

$$P[Y > 200] = 1 - P[Y \leq 200] = 1 - F_Y(200) = 1 - e^{-e^{-\alpha\,(200-u)}} = 1 - 0.9514 = 0.0486.$$

Tudi vrednost stoletne vode $y_{100}$ določimo iz enačbe (7.54), ki jo moramo v tem primeru invertirati:

$$P[Y > y_{100}] = 1 - P[Y \leq y_{100}] = 1 - F_Y(y_{100}) = 1 - e^{-e^{-\alpha\,(y_{100}-u)}} = 0.01 \qquad \longrightarrow$$
$$\longrightarrow y_{100} = u - \frac{1}{\alpha}\ln\left(-\ln(0.99)\right) = 256.15\,\mathrm{m}^3/\mathrm{s}$$

### 7.5.4 Porazdelitev ekstremnih vrednosti tipa II

Porazdelitev ekstremnih vrednosti tipa II imenujemo tudi **Fréchetova porazdelitev**.[16] Obravnavajmo porazdelitev maksimuma. Če je porazdelitvena funkcija $F_X(x)$ enaka

$$F_X(x) = 1 - \beta\left(\frac{1}{x}\right)^k, \qquad x > \beta^{\frac{1}{k}},$$

potem je porazdelitvena funkcija maksimuma slučajnih spremenljivk $X_i$, ko gre njihovo število proti neskončnosti, enaka

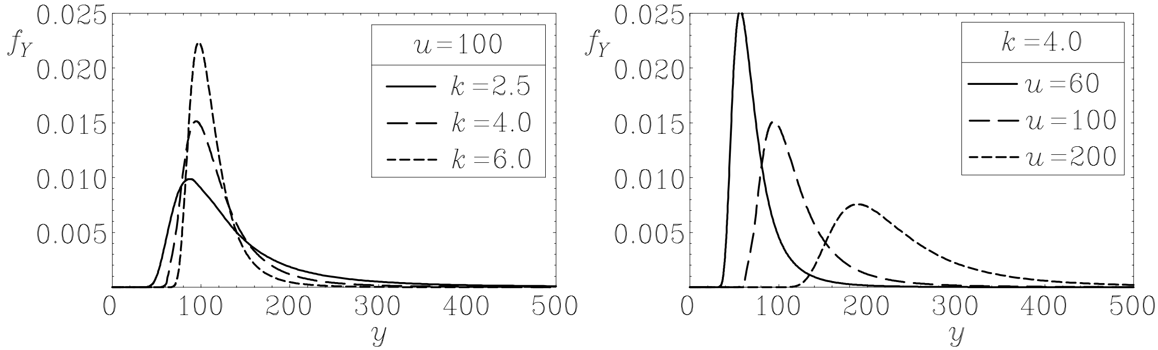
$$F_Y(y) = e^{-\left(\frac{u}{y}\right)^k}, \qquad y > 0. \tag{7.59}$$

Konstanti $k$ in $u$ sta parametra porazdelitve, ki sta odvisna od pričakovane vrednosti in variance slučajne spremenljivke $Y$. Gostoto verjetnosti te slučajne spremenljivke dobimo z odvajanjem porazdelitvene funkcije $F_Y(y)$ po enačbi (7.59)

$$f_Y(y) = \frac{k}{u}\left(\frac{u}{y}\right)^{k+1} e^{-\left(\frac{u}{y}\right)^k}, \qquad y > 0. \tag{7.60}$$

Na sliki 7.21 prikazujemo gostote verjetnosti za nekaj različnih vrednosti parametrov $k$ in $u$. Iz slike vidimo, da pri tej porazdelitvi oba parametra vplivata na njeno pričakovano vrednost in razpršenost.

[16] Maurice René Fréchet, francoski matematik (1878-1973).




Slika 7.21: Gostota verjetnosti Fréchetove porazdelitve maksimuma

Pri reševanju nalog potrebujemo zvezi med momentoma slučajne spremenljivke $m_Y$ in $\sigma_Y$ ter njenima parametroma $u$ in $k$

$$m_Y = u\,\Gamma\left(1-\tfrac{1}{k}\right), \quad k>1, \qquad \sigma_Y^2 = u^2\left(\Gamma\left(1-\tfrac{2}{k}\right)-\Gamma^2\left(1-\tfrac{1}{k}\right)\right), \quad k>2, \tag{7.61}$$

kjer je funkcija $\Gamma$ taka, kot smo jo definirali v enačbi (7.25). Obratnih zvez ne moremo zapisati v zaključeni obliki, zato parametre te porazdelitve določimo numerično, kakor bomo pokazali tudi v računskem primeru. Pogoja, da je $k>1$ oziroma $k>2$ izvirata iz definicijskega območja funkcije $\Gamma$. Zaradi tega, varianca slučajne spremenljivke $Y$ za $k \leq 2$ ne obstaja. Ker momenti višjega reda vključujejo člene, kot so $\Gamma(1-3/k)$, $\Gamma(1-4/k)$ ..., je pogoj za obstoj momenta $r$-tega reda, da je $k>r$.

Če ima začetna porazdelitev naslednjo obliko

$$F_X(x) = \beta\left(\frac{1}{x}\right)^k, \qquad x < \beta^{\frac{1}{k}},$$

potem je porazdelitvena funkcija minimuma slučajnih spremenljivk $X_i$, ko gre njihovo število proti neskončnosti, enaka

$$F_Z(z) = 1 - e^{-\left(\frac{u}{z}\right)^k}, \qquad z<0. \tag{7.62}$$

Gostoto verjetnosti določimo z odvajanjem porazdelitvene funkcije

$$f_Z(z) = k\left(\frac{u}{z}\right)^{k+1} e^{-\left(\frac{u}{z}\right)^k}, \qquad z<0.$$

Zvezi med momemtoma $m_Z$ in $\sigma_Z$ ter njenima parametroma $u$ in $k$ sta enaki, kot pri Fréchetovi porazdelitvi maksimuma (enačba (7.61))

$$m_Z = u\,\Gamma\left(1-\tfrac{1}{k}\right), \quad k>1, \qquad \sigma_Z^2 = u^2\left(\Gamma\left(1-\tfrac{2}{k}\right)-\Gamma^2\left(1-\tfrac{1}{k}\right)\right), \quad k>2, \tag{7.63}$$

Uporaba Fréchetove porazdelitve je podobna kot pri Gumbelovi porazdelitvi. Zato jo uporabimo pri obravnavanju hidroloških pojavov, nekaterih obtežb na konstrukcije in določenih karakteristik materialov oziroma konstrukcij. Značilnost Fréchetove porazdelitve v primerjavi z Gumbelovo je, da gostota verjetnosti bistveno počasneje pada, zato je pogostnost pojava zelo velikih vrednosti glede na pričakovano vrednost, večja.

**Primer 7.21:** *Rešimo isti problem, kot smo ga opisali v primeru 7.20, le da predpostavimo, da se največji letni pretok vode v neki rečici porazdeljuje po Fréchetovi porazdelitvi. Pričakovana vrednost in standardna deviacija oziroma njuni oceni naj bosta enaka kot v primeru 7.20. Izračunajmo verjetnost, da bo v naslednjem letu vodni pretok večji od* $200\,\mathrm{m}^3/\mathrm{s}$. *Določimo tudi 100-letno vodo; verjetnost, da bo 100-letna voda v naslednjem letu presežena, je enaka* $1/100$.

**Rešitev:** Iz enačb (7.61) lahko izpeljemo naslednjo enačbo:

$$\left(\frac{\sigma_Y}{m_Y}\right)^2 = \frac{\Gamma\left(1-\frac{2}{k}\right)}{\Gamma^2\left(1-\frac{1}{k}\right)} - 1, \tag{7.64}$$

ki predstavlja nelinearno enačbo spremenljivke $k$. Enačbo rešimo numerično s programom MATHEMATICA,[17] lahko pa tudi s programom Excel, pri čemer moramo funkcijo $\Gamma(x)$ zapisati kot `exp(gammaln(x))` in uporabiti orodje GOALSEEK. V našem primeru je rešitev enačbe enaka $k = 4.2389$. Iz enačbe (7.61) izračunamo tudi drugi parameter

$$u = \frac{m_Y}{\Gamma\left(1-\frac{1}{k}\right)} = 95.269.$$

Verjetnost, da bo v naslednjem letu največja voda večja od $200\,\mathrm{m}^3/\mathrm{s}$, izračunamo po enačbi (7.59)

$$P[Y > 200] = 1 - P[Y \le 200] = 1 - F_Y(200) = 1 - e^{-\left(\frac{u}{200}\right)^k} = 1 - 0.9578 = 0.0422.$$

Vrednost stoletne vode $y_{100}$ določimo iz enačbe (7.59) z invertiranjem

$$P[Y > y_{100}] = 1 - P[Y \le y_{100}] = 1 - F_Y(y_{100}) = 1 - e^{-\left(\frac{u}{y_{100}}\right)^k} = 0.01 \quad \longrightarrow$$
$$\longrightarrow y_{100} = u\left(-\ln(0.99)\right)^{-\frac{1}{k}} = 282.00\,\mathrm{m}^3/\mathrm{s}$$

**Primer 7.22:** *Določite pričakovano vrednost, standardno deviacijo ter koeficient simetričnosti* $\gamma_{1Y}$ *in sploščenosti* $\gamma_{2Y}$ *za slučajno spremenljivko* $Y$ *iz primera 7.21.*

---

[17] Iskanje rešitve enačbe (7.64), izvedeno s programom MATHEMATICA:

```
my = 115; Sy = 45; Vy = Sy/my;
k = kk /. FindRoot[Gamma[1-2/kk]/Gamma[1-1/kk]^2-1==Vy^2,{kk,3}];
Print["k = ", k]
Print["u = ", u = my/Gamma[1-1/k]];
Print["P[Y>200] = ", P200 = 1-E^(-(u/200)^k) ];
Print["y100 = ", y100 = u (-Log[0.99])^(-1/k)];

k = 4.2389
u = 95.2687
P[Y>200] = 0.0422089
y100 = 282.004
```

**Rešitev:** Pričakovano vrednost $m_Y$ in standardno deviacijo $\sigma_Y$ izračunamo z enačbo (7.61). S tem pravzaprav le preverimo, ali smo v primeru 7.21 pravilno izračunali parametra $k$ in $u$.

$$m_Y = u\,\Gamma\left(1-\tfrac{1}{k}\right) = 95.2687\,\Gamma\left(1-\tfrac{1}{4.2389}\right) = 115,$$
$$\sigma_Y = u\sqrt{\Gamma\left(1-\tfrac{2}{k}\right)-\Gamma^2\left(1-\tfrac{1}{k}\right)} = 95.2687\sqrt{\Gamma\left(1-\tfrac{2}{4.2389}\right)-\Gamma^2\left(1-\tfrac{1}{4.2389}\right)} = 45.$$

Vidimo, da smo izračunali enake vrednosti, kot so podane v besedilu primera 7.21. Koeficienta simetričnosti in sploščenosti lahko izračunamo numerično s programom MATHEMATICA.[18] Vidimo torej, da je koeficient simetričnosti enak $\gamma_{1Y} = 4.81979$, koeficient sploščenosti oziroma koeficient kurtosis pa je $\gamma_{2Y} = 214.803$, kar kaže na to, da je porazdelitev precej nesimetrična in zelo špičasta.

Izraze za koeficienta $\gamma_{1Y}$ in $\gamma_{2Y}$ lahko zapišemo tudi v analitični obliki. Pokažemo lahko (glej D. Zupan, G. Turk, Porazdelitve ekstremnih vrednosti), da je

$$\mathrm{E}[Y^r] = u^r\,\Gamma\left(1-\tfrac{r}{k}\right),$$

od koder lahko izračunamo centralne momente višjih redov po naslednjih enačbah:

$$\begin{aligned}
\mu_Y^{(3)} &= \mathrm{E}[(Y-m_Y)^3] = \mathrm{E}[Y^3 - 3\,Y^2\,m_Y + 3\,Y\,m_Y^2 - m_Y^3] = \\
&= \mathrm{E}[Y^3] - 3\,\mathrm{E}[Y^2]\,m_Y + 2\,m_Y^3 = \\
&= u^3\,\left(\Gamma\left(1-\tfrac{3}{k}\right) - 3\,\Gamma\left(1-\tfrac{2}{k}\right)\Gamma\left(1-\tfrac{1}{k}\right) + 2\,\Gamma^3\left(1-\tfrac{1}{k}\right)\right), \\
\mu_Y^{(4)} &= \mathrm{E}[(Y-m_Y)^4] = \mathrm{E}[Y^4 - 4\,Y^3\,m_Y + 6\,Y^2\,m_Y^2 - 4\,Y\,m_Y^3 + m_Y^4] = \\
&= \mathrm{E}[Y^4] - 4\,\mathrm{E}[Y^3]\,m_Y + 6\,\mathrm{E}[Y^2]\,m_Y^2 - 3\,m_Y^4] = \\
&= u^4\left(\Gamma\left(1-\tfrac{4}{k}\right) - 4\,\Gamma\left(1-\tfrac{3}{k}\right)\Gamma\left(1-\tfrac{1}{k}\right) + 6\,\Gamma\left(1-\tfrac{2}{k}\right)\Gamma^2\left(1-\tfrac{1}{k}\right) - 3\Gamma^4\left(1-\tfrac{1}{k}\right)\right).
\end{aligned}$$

Koeficienta simetričnosti in sploščenosti, $\gamma_{1Y}$ in $\gamma_{2Y}$ izračunamo po enačbah (6.8) in (6.9)

$$\gamma_{1Y} = \frac{\mu_Y^{(3)}}{\sigma_Y^3} = \frac{\Gamma\left(1-\tfrac{3}{k}\right) - 3\,\Gamma\left(1-\tfrac{2}{k}\right)\Gamma\left(1-\tfrac{1}{k}\right) + 2\,\Gamma^3\left(1-\tfrac{1}{k}\right)}{\left(\Gamma\left(1-\tfrac{2}{k}\right) - \Gamma^2\left(1-\tfrac{1}{k}\right)\right)^{\frac{3}{2}}}$$
$$\gamma_{2Y} = \frac{\mu_Y^{(4)}}{\sigma_Y^4} = \frac{\Gamma\left(1-\tfrac{4}{k}\right) - 4\,\Gamma\left(1-\tfrac{3}{k}\right)\Gamma\left(1-\tfrac{1}{k}\right) + 6\,\Gamma\left(1-\tfrac{2}{k}\right)\Gamma^2\left(1-\tfrac{1}{k}\right) - 3\Gamma^4\left(1-\tfrac{1}{k}\right)}{\left(\Gamma\left(1-\tfrac{2}{k}\right) - \Gamma^2\left(1-\tfrac{1}{k}\right)\right)^2}$$

[18] Določitev momentov slučajne spremenljivke $Y$ s programom MATHEMATICA:

```
FY[k_,u_,y_]=E^(-(u/y)^k);  fY[k_,u_,y_]=D[FY[k,u,y],y];
k = 4.2389;  u = 95.7476;
Print["my = ",my=N[Integrate[y fY[k,u,y],{y,0,Infinity}]]]
Print["Sy = ",sy=N[Sqrt[Integrate[(y-my)^2 fY[k,u,y],{y,0,Infinity}]]]]
Print["gama1 = ",N[Integrate[(y-my)^3 fY[k,u,y],{y,0,Infinity}]/sy^3]]
Print["gama2 = ",N[Integrate[(y-my)^4 fY[k,u,y],{y,0,Infinity}]/sy^4]]

my = 115.
Sy = 45.
gama1 = 4.81979
gama2 = 214.803
```

### 7.5.5 Porazdelitev ekstremnih vrednosti tipa III

Porazdelitev ekstremnih vrednosti tipa III imenujemo tudi **Weibullova porazdelitev**.[19] Posebnost te porazdelitve je, da so tako porazdelitve začetnih slučajnih spremenljivk $X_i$, kot tudi asimptotične porazdelitve ekstremnih vrednosti omejene na tisti strani porazdelitve, ki jo obravnavamo. Porazdelitev začetne slučajne spremenljivke za porazdelitev maksimuma je zato navzgor omejena s pogojem $x < \omega$ in ima v bližini $\omega$ naslednjo obliko

$$F_X(x) = 1 - c(\omega - x)^k, \quad k > 0, \ x < \omega.$$

Slučajna spremenljivka $Y$, ki predstavlja maksimum slučajnih spremenljivk $X_i$, v bližini $\omega$ porazdeljenih po zadnji enačbi, se v primeru, ko gre število slučajnih spremenljivk proti neskončnosti porazdeljuje po naslednji enačbi:

$$F_Y(y) = e^{-\left(\frac{\omega-y}{\omega-u}\right)^k}, \quad y < \omega. \tag{7.65}$$

Gostoto verjetnosti $f_Y(y)$ določimo z odvajanjem porazdelitvene funkcije (7.65)

$$f_Y(y) = \frac{k}{\omega - u}\left(\frac{\omega - y}{\omega - u}\right)^{k-1} e^{-\left(\frac{\omega-y}{\omega-u}\right)^k}, \quad y < \omega.$$

Na sliki 7.22 prikazujemo gostote verjetnosti za nekaj različnih vrednosti parametrov $k$, $u$ in $\omega$. Vidimo, da je vpliv parametrov $k$ in $\omega$ na prvi pogled zelo podoben, medtem ko parameter $u$ bolj kot druga dva vpliva na pričakovano vrednost slučajne spremenljivke $Y$.

Parametre porazdelitve $u$ in $k$ običajno določimo iz vzorca, medtem ko parameter $\omega$ določimo vnaprej. Včasih okoliščine natančno opredeljujejo parameter $\omega$, s katerim je slučajna spremenljivka navzgor omejena. Na primer: če slučajna spremenljivka $Y$ predstavlja nadmorsko višino vseh naših gora, bi morali kot zgornjo mejo vzeti nadmorsko višino Triglava, torej $\omega = 2864\,\mathrm{m}$. Drugič pa zgornje meje ne moremo določiti natančno in si jo izberemo. Za določitev parametrov $u$ in $k$ je dovolj, da poznamo dva momenta slučajne spremenljivke $Y$, običajno sta to pričakovana vrednost $m_Y$ in standardna deviacija $\sigma_Y$. Zvezi med $m_Y$ in $\sigma_Y$ ter parametroma $k$ in $u$ sta

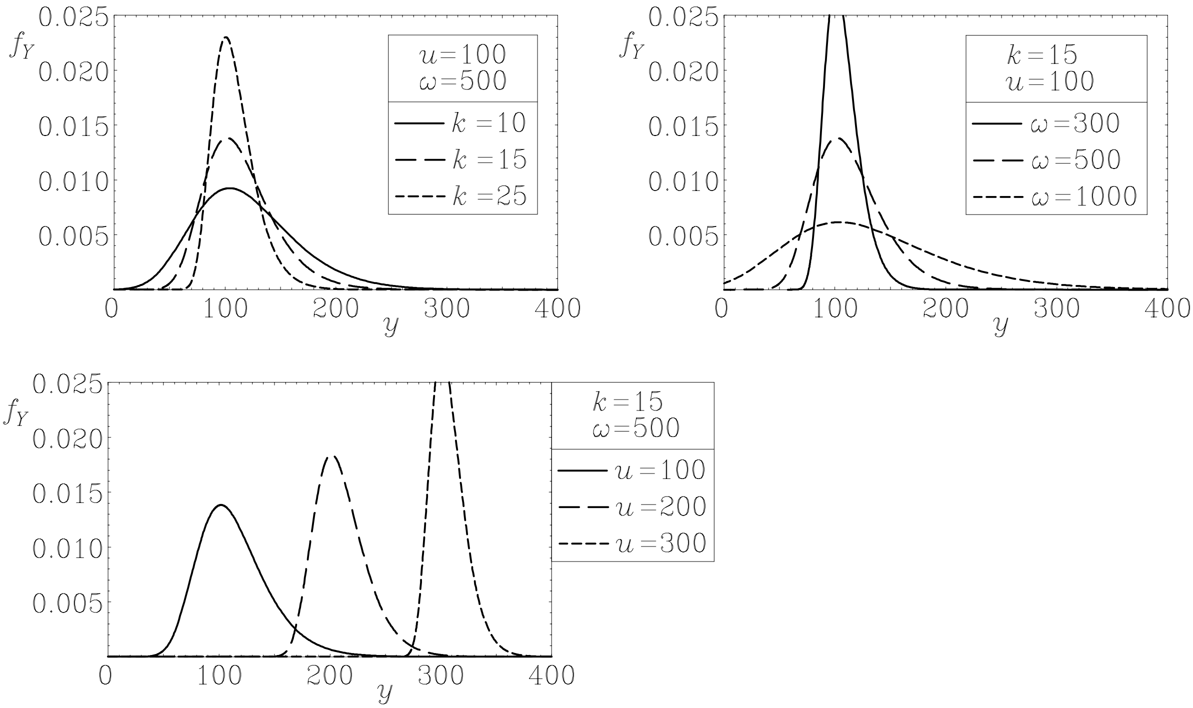
$$m_Y = \omega - (\omega - u)\,\Gamma\left(1 + \tfrac{1}{k}\right), \qquad \sigma_Y^2 = (\omega - u)^2\left(\Gamma\left(1 + \tfrac{2}{k}\right) - \Gamma^2\left(1 + \tfrac{1}{k}\right)\right), \tag{7.66}$$

obratne zveze v analitični obliki ne moremo zapisati, rešitev poiščemo z numeričnimi metodami, kakor bomo pokazali v računskem primeru.

**Primer 7.23:** *Rešimo isti problem, kot smo ga opisali v primerih 7.20 in 7.21, le da predpostavimo, da se največji letni pretok vode v neki rečici porazdeljuje po Weibullovi porazdelitvi. Pričakovana vrednost in standardna deviacija oziroma njuni oceni naj bosta enaka kot v primeru 7.20 in 7.21, za parameter $\omega$ pa izberimo vrednost $500\,\mathrm{m}^3/\mathrm{s}$. Izračunajmo verjetnost, da bo v naslednjem letu vodni pretok večji od $200\,\mathrm{m}^3/\mathrm{s}$. Določimo tudi 100-letno vodo.*

*Zberimo rezultate računskih primerov 7.17, 7.20, 7.21 in tega primera ter primerjajmo rezultate. Narišimo še gostote verjetnosti za vse uporabljene porazdelitve v teh računskih primerih.*

[19] Ernest Hjalmar Wallodi Weibull, švedski fizik in matematik (1887–1979).




Slika 7.22: Gostota verjetnosti Weibullove porazdelitve maksimuma

**Rešitev:** Iz enačb (7.66) lahko izpeljemo naslednjo enačbo:

$$\left(\frac{\sigma_Y}{\omega - m_Y}\right)^2 = \frac{\Gamma\left(1+\frac{2}{k}\right)}{\Gamma^2\left(1+\frac{1}{k}\right)} - 1, \tag{7.67}$$

ki predstavlja nelinearno enačbo spremenljivke $k$. Enačbo rešimo numerično s programom MATHEMATICA.[20] Rešitev enačbe je $k = 10.311$. Iz enačbe (7.66) izračunamo tudi drugi parameter

$$u = \omega - \frac{\omega - m_Y}{\Gamma\left(1+\frac{1}{k}\right)} = 95.831.$$

[20] Iskanje rešitve enačbe (7.67), izvedeno s programom MATHEMATICA:

```
my = 115; Sy = 45; w = 500;  Vym = Sy/(w-my);
k = kk /. FindRoot[Gamma[1+2/kk]/Gamma[1+1/kk]^2-1==Vym^2,{kk,5}];
Print["k = ", k]
Print["u = ", u = w-(w-my)/Gamma[1+1/k]];
Print["P[Y>200] = ", P200 = 1-E^(-((w-200)/(w-u))^k) ];
Print["y100 = ", y100 = w-(w-u) (-Log[0.99])^(1/k)];

k = 10.3106
u = 95.8312
P[Y>200] = 0.045224
y100 = 241.298
```

Verjetnost, da bo v naslednjem letu največja voda večja od $200\,\mathrm{m}^3/\mathrm{s}$, izračunamo po enačbi (7.59)

$$P[Y > 200] = 1 - P[Y \leq 200] = 1 - F_Y(200) = 1 - e^{-\left(\frac{\omega-200}{\omega-u}\right)^k} = 1 - 0.9548 = 0.0452.$$

Vrednost stoletne vode $y_{100}$ določimo iz enačbe (7.65) z invertiranjem

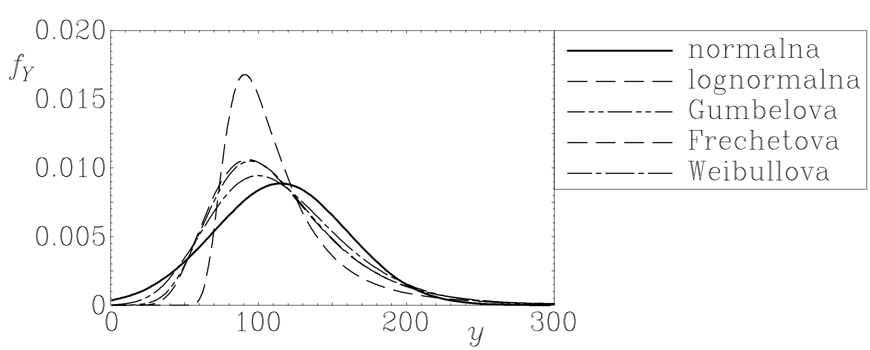
$$P[Y > y_{100}] = 1 - P[Y \leq y_{100}] = 1 - F_Y(y_{100}) = 1 - e^{-\left(\frac{\omega-y_{100}}{\omega-u}\right)^k} = 0.01 \quad \longrightarrow$$
$$\longrightarrow y_{100} = \omega - (\omega - u)\,(-\ln(0.99))^{\frac{1}{k}} = 241.30\,\mathrm{m}^3/\mathrm{s}$$

V preglednici 7.2 prikazujemo rezultate računskih primerov 7.17, 7.20, 7.21 in 7.23. Vidimo lahko, da najbolj izstopata rezultata za normalno porazdelitev, ki je bistveno nižja od drugih, in rezultati za Fréchetovo porazdelitev, ki je bistveno višja od drugih. Vrednosti za druge tri porazdelitve so zelo podobne. Omenimo lahko še, da na vrednosti, dobljene za Weibullovo porazdelitev, močno vpliva izbira parametra $\omega$. Če bi izbrali bistveno višjo vrednost, se rezultati približajo rezultatom Gumbelove porazdelitve (na primer, za $\omega = 10000$ dobimo $P[Y > 200] = 0.0485$ in $y_{100} = 255.61$).

Preglednica 7.2: Izračun verjetnosti $P[Y \geq 200]$ in stoletne vode $y_{100}$

| **porazdelitev** | $P[Y \geq 200]$ | $y_{100}$ |
|---|---|---|
| normalna | 0.0295 | 219.69 |
| logaritemsko normalna | 0.0490 | 257.70 |
| Gumbelova | 0.0486 | 256.15 |
| Fréchetova | 0.0422 | 282.00 |
| Weibullova | 0.0452 | 241.30 |

Na sliki 7.23 prikazujemo gostote verjetnosti za vse obravnavane porazdelitve. Vidimo lahko, da ponovno najbolj izstopata normalna in Fréchetova porazdelitev, druge tri pa so si zelo podobne.




Slika 7.23: Gostota verjetnosti različnih porazdelitev

Obravnavajmo sedaj še Weibullovo porazdelitev minimuma. V tem primeru so tako začetne porazdelitve $X_i$ kot tudi asimptotična porazdelitev ekstremnih vrednosti navzdol omejene s pogojem $\varepsilon < x$. V bližini

$\varepsilon$ ima porazdelitev začetnih slučajnih spremenljivk naslednjo obliko

$$F_X(x) = c(x-\varepsilon)^k \qquad k>0, \;\; x>\varepsilon.$$

Slučajna spremenljivka $Z$, ki predstavlja minimum slučajnih spremenljivk $X_i$, ko gre njihovo število proti neskončnosti in so v okolici $\varepsilon$ porazdeljeni po zgornji enačbi, se porazdeljuje takole:

$$F_Z(z) = 1 - e^{-\left(\frac{z-\varepsilon}{u-\varepsilon}\right)^k}, \qquad z>\varepsilon, \tag{7.68}$$

njena gostota verjetnosti pa je

$$f_Z(z) = \frac{k}{u-\varepsilon}\left(\frac{z-\varepsilon}{u-\varepsilon}\right)^{k-1} e^{-\left(\frac{z-\varepsilon}{u-\varepsilon}\right)^k}, \qquad z>\varepsilon,$$

Zvezi med momentoma $m_Z$ in $\sigma_Z$ slučajne spremenljivke $Z$ in njenima parametroma $u$ in $k$ sta

$$m_Z = \varepsilon + (u-\varepsilon)\,\Gamma\left(1+\tfrac{1}{k}\right), \qquad \sigma_Z^2 = (u-\varepsilon)^2\,\left(\Gamma\left(1+\tfrac{2}{k}\right) - \Gamma^2\left(1+\tfrac{1}{k}\right)\right). \tag{7.69}$$

Obratnih zvez ne moremo zapisati v zaključeni analitični obliki, vendar $k$ in $u$ za znane $m_Z$ in $\sigma_Z$ lahko izračunamo numerično na podoben način, kot smo to naredili v primerih 7.21 in 7.23. Vrednost $\varepsilon$ običajno določimo vnaprej, glede na lastnosti slučajne spremenljivke, ki jo obravnavamo. Zelo značilno je, da predpostavimo $\varepsilon = 0$ za količine, ki ne morejo biti negativne: trdnost materiala, količina padavin...

## 7.6 Druge porazdelitve

Poleg porazdelitev, ki smo jih spoznali v tem poglavju, obstaja še cela vrsta porazdelitev, uporabljenih pri različnih aplikacijah. Omenimo samo nekaj najpogstejše uporabljenih:

- **Porazdelitev** $\chi^2$ uporabljamo pri statistiki (območja zaupanja in preverjanje domnev);
- **Studentovo porazdelitev** ali **porazdelitev** $t$ uporabljamo pri statistiki (območja zaupanja in preverjanje domnev);
- **Porazdelitev** $F$ uporabljamo pri statistiki (območja zaupanja in preverjanje domnev);
- **Hipergeometrična porazdelitev** je pogosto uporabljena diskretna porazdelitev s podobnim pomenom kot binomska porazdelitev;
- **Cauchyjeva porazdelitev**, **logistična porazdelitev**, **dvojna eksponentna porazdelitev**, **porazdelitev** $\beta$, **von Misesova porazdelitev** in mnoge druge.

# 8 Vzorčenje

Do sedaj smo obravnavali slučajne spremenljivke in njihove porazdelitve in ob tem predpostavili, da lastnosti slučajne spremenljivke oziroma parametre njihovih porazdelitev povsem poznamo. Razen v nekaterih preprostih primerih parametrov porazdelitev slučajnih spremenljivk ne poznamo. Njihovo oceno lahko določimo z analizo vzorca. Podajmo nekaj osnovnih pojmov statističnih metod:

- **Populacija** zajema vse rezultate poskusov oziroma opazovanj pri obravnavanemu problemu. Populacija v statistiki ima torej precej širši pojem od tistega, ki smo ga sicer navajeni: pomeni lahko skupino ljudi, lahko pa tudi izid meta kocke, rezultat testiranja betonske kocke, število vozil, ki v nekem časovnem intervalu prevozijo opazovano križišče, izmera dolžine med dvema točkama, letna količina padavin v nekem kraju... Velikost populacije, ki je lahko končna ali neskončna, označimo z $N$. Populacija Slovencev je primer končne populacije, rezultati testiranja betonske kocke pa predstavljajo neskončno populacijo.

- **Vzorec** je tisti del populacije, iz katerega sklepamo na lastnosti populacije. Velikost vzorca označimo z $n$.

- S **parametrom** populacije opišemo njeno lastnost. Parametri so lahko pričakovana vrednost, varianca, določen parameter porazdelitve, po kateri se obravnavana populacija porazdeljuje. Pomembno je, da se zavedamo, da parameter populacije ni slučajna spremenljivka, ampak neznana deterministična količina.

- **Element vzorca ali populacije** $X_i$ predstavlja posamezni rezultat poskusa oziroma opazovanja. Običajno predpostavimo, da so elementi vzorca enako porazdeljene in medsebojno neodvisne slučajne spremenljivke. To pomeni, da so tudi pričakovane vrednosti in variance vseh slučajnih spremenljivk $X_i$ enake

  $$m_{X_i} = m_X \quad \text{in} \quad \sigma^2_{X_i} = \sigma^2_X. \tag{8.1}$$

- **Statistika** je količina, ki jo določimo iz vzorca. Na osnovi statistik določamo **točkovne** ali **intervalne ocene** parametrov populacije in sklepamo v primeru **preizkušanja domnev (testiranja hipotez)**. Osnovne statistike so: povprečje vzorca $\bar{X}$, varianca vzorca $S^2_X$, najmanjše in največje vrednosti vzorca ($X_{min}$ in $X_{max}$)... Statistika je slučajna spremenljivka, saj iz različnih vzorcev iste populacije dobimo različne vrednosti statistik.

**Vzorčenje** je proces, s katerim pripravimo vzorec. Vprašamo se lahko, zakaj moramo vzorčiti, saj bi bilo načeloma vedno bolje parametre populacije določiti tako, da bi pregledali celotno populacijo. Razlogov je več:

- Populacija je lahko neskončno velika, v tem primeru je v celoti ne moremo pregledati. Primer: meritev natezne trdnosti jeklenih palic.
- Populacija je zelo velika, pregled celotne populacije bi lahko bil zelo dolgotrajen ali predrag. Primer: populacija ljudi na Zemlji.
- Poskus oziroma opazovanje lahko vpliva ali pa spremeni element vzorca. Primer: v tovarni vozil za testiranje vpliva trkov izberejo manjše število vozil, ki jih med poskusom bolj ali manj poškodujejo.
- Nekateri elementi populacije so lahko povsem nedosegljivi. Primer: količina padavin na vrhu Mount Everesta v letu 1910. Drugi primer: količina padavin v Ljubljani leta 2111.
- V primeru, da se odločimo za pregled celotne populacije, je to pregledovanje lahko zelo dolgotrajno ali pa moramo v to vključiti veliko število izvajalcev pregledovanja. Zaradi tega lahko pride do relativno velikih napak, ki jih izvajalci pregledovanja storijo. Zato je pogosto bolje, da natančno opravimo pregled na relativno majhnem vzorcu, kot pa da površno pregledamo celotno populacijo.

Ločimo dva načina vzorčenja glede na to, ali elemente populacije po vsakem poskusu vrnemo v populacijo ali ne. Pri vzorčenju z vračanjem mora veljati, da poskus ne spremeni elementa. V tem primeru upoštevamo, da je populacija neskončno velika. Pri vzorčenju brez vračanja moramo pri končno velikih populacijah velikost populacije upoštevati. V primeru vzorčenja brez vračanja, so poskusi lahko tudi taki, da element populacije spremenijo ali celo uničijo.

Z vzorčenjem želimo dobiti vzorec, ki bi bil nepristranski, čimbolj natančen in čim cenejši. Prvi dve lastnosti se izključujeta z zadnjo, saj je večji (in zato tudi dražji) vzorec običajno tudi bolj natančen in bolj nepristranski. Poznamo več načinov vzorčenja:

- Povsem slučajno vzorčenje je najpreprostejše in glede nepristranskosti tudi najboljše vzorčenje. Slaba stran takega vzorčenja je, da za določeno stopnjo natančnosti potrebujemo relativno velike vzorce.
- Pri vzorčenju po skupinah pazimo, da so v vzorcu določene skupine elementov zastopane v takih razmerjih, kot veljajo za populacijo. Primer: pri anketiranju običajno pazimo, da razmerje med moškimi in ženskami enako razmerju v populaciji, ki jo obravnavamo, pazimo lahko tudi na starostno strukturo in strukturo izobrazbe. Na ta način dobimo nepristranski vzorec že z relativno manjšo velikostjo vzorca.
- Pri vzorčenju po blokih si povsem slučajno izberemo manjša področja, ki predstavljajo dele vzorca, ki jih združimo v en vzorec. Primer: pri opazovanju velikega polja si ne moremo privoščiti, da bi ga v celoti natančno pregledali. Zato si slučajno izberemo manjše parcele, na katerih lahko izvršimo opazovanja. Običajno pazimo, da vse parcele niso na robu ali v sredini polja.

- Subjektivnega vzorčenja, kjer o izbiri elementov vzorca odloča “strokovnjak”, pri statističnih analizah ne smemo uporabiti. Primer pristranskega, nenatančnega vzorca: Pri anketiranju, s katerim želimo ugotoviti javno mnenje o delovanju slovenske vlade, anketirance izbere predstavnik Vlade Slovenije.

## 8.1 Lastnosti osnovnih statistik

Omenili smo že, da so statistike slučajne spremenljivke. V nadaljevanju določimo pričakovano vrednost in varianco povprečja ($\mathrm{E}\left[\bar{X}\right]$ in $\mathrm{var}\left[\bar{X}\right]$) ter pričakovano vrednost in varianco variance vzorca ($\mathrm{E}\left[S_X^2\right]$ in $\mathrm{var}\left[S_X^2\right]$).

### 8.1.1 Povprečje vzorca

Povprečje vzorca $\bar{X}$ definiramo z naslednjo enačbo

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{8.2}$$

Pričakovano vrednost povprečja vzorca določimo ob upoštevanju lastnosti pričakovane vrednosti vsote slučajnih spremenljivk (enačba (6.28)) in enačbe (8.1)

$$\mathrm{E}\left[\bar{X}\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X] = \frac{n}{n}\mathrm{E}[X] = \mathrm{E}[X]. \tag{8.3}$$

Dobljena enačba je pomembna, saj nam pove, da je povprečje vzorca **nepristranska ocena** pričakovane vrednosti populacije.

Določimo še varianco povprečja vzorca

$$\mathrm{var}\left[\bar{X}\right] = \mathrm{var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}[X] = \frac{n}{n^2}\mathrm{var}[X] = \frac{\mathrm{var}[X]}{n}, \tag{8.4}$$

ki nam pove, da se varianca povprečja, ki predstavlja oceno pričakovane vrednosti populacije zmanjšuje z večanjem vzorca. Ko gre $n$ proti neskončnosti, gre varianca povprečja proti nič, kar pomeni, da je pričakovana vrednost populacije $m_X$ točno določena. V enačbi (8.4) smo predpostavili, da je velikost populacije neskončna oziroma, da vzorčimo z vračanjem. V primeru, da je populacija končno velika in vzorčimo z vračanjem, moramo enačbo (8.4) nekoliko spremeniti

$$\mathrm{var}\left[\bar{X}\right] = \frac{\mathrm{var}[X]}{n}\,\frac{N-n}{N-1}. \tag{8.5}$$

Limita zadnje enačbe, ko gre $N \to \infty$, je enačba (8.4). Poleg tega iz enačbe (8.5) sledi, da je v primeru, da je $N = n$, varianca povprečja enaka nič. To je logičen rezultat, saj v primeru, da je $N = n$, v vzorec vključimo celotno populacijo in je torej tudi povprečje tega vzorca enako pričakovani vrednosti populacije.

Pri statističnih analizah je zelo pomembno poznati porazdelitve uporabljenih statistik. Iz enačbe (8.2) vidimo, da je povprečje vzorca vsota enako porazdeljenih in neodvisnih slučajnih spremenljivk $X_i$ deljeno z velikostjo vzorca $n$. Pogosto predpostavimo, da so slučajne spremenljivke $X_i$ porazdeljene normalno, v tem primeru je tudi porazdelitev povprečja $\bar{X}$ normalna. Ne glede na porazdelitev slučajnih spremenljivk $X_i$ po centralnem limitnem izreku velja, da ko gre $n$ proti neskončnosti, gre porazdelitev $\bar{X}$ proti normalni porazdelitvi. Pri praktični uporabi pogosto predpostavimo, da je povprečje $\bar{X}$ porazdeljeno normalno, tudi če slučajne spremenljivke $X_i$ niso porazdeljene normalno in je $n$ le zmerno velik, na primer $n > 30$. Napaka, ki jo pri tem storimo je odvisna predvsem od porazdelitve $X_i$, pri nesimetričnih porazdelitvah je napaka običajno večja kot pri simetričnih.

### 8.1.2 Varianca vzorca

Varianco vzorca definiramo z enačbo

$$S_X^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 . \tag{8.6}$$

Pričakovano vrednost te statistike določimo podobno kot pričakovano vrednost povprečja vzorca:

$$\mathrm{E}\left[S_X^2\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right] = \mathrm{E}\left[\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} 2\bar{X}X_i + \sum_{i=1}^{n}\bar{X}^2\right)\right] . \tag{8.7}$$

Sedaj pokažimo, da je vsota produkta $\bar{X}X_i$ enaka vsoti $\bar{X}^2$

$$\sum_{i=1}^{n}\bar{X}X_i = \bar{X}\sum_{i=1}^{n} X_i = \bar{X}n\bar{X} = \sum_{i=1}^{n}\bar{X}^2$$

in to upoštevamo v enačbi (8.7):

$$\mathrm{E}\left[S_X^2\right] = \mathrm{E}\left[\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n}\bar{X}^2 + \sum_{i=1}^{n}\bar{X}^2\right)\right] = \mathrm{E}\left[\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n}\bar{X}^2\right)\right] .$$

Ob upoštevanju enačb (6.28) in (8.1) lahko zapišemo

$$\mathrm{E}\left[S_X^2\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[X_i^2\right] - \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\bar{X}^2\right] = \frac{n}{n}\mathrm{E}\left[X^2\right] - \frac{n}{n}\mathrm{E}\left[\bar{X}^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}\left[\bar{X}^2\right] . \tag{8.8}$$

Na desni strani te enačbe lahko odštejemo $\mathrm{E}[X]^2$ in prištejemo $\mathrm{E}\left[\bar{X}\right]^2$, kar zaradi enakosti (8.3) enačbe (8.8) ne pokvari. Nazadnje uporabimo še enačbo (6.15), s katero $\mathrm{E}\left[S_X^2\right]$ izrazimo z variancami $\mathrm{var}[X]$ in $\mathrm{var}\left[\bar{X}\right]$, in enačbo (8.4)

$$\mathrm{E}\left[S_X^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 - \left(\mathrm{E}\left[\bar{X}^2\right] - \mathrm{E}\left[\bar{X}\right]^2\right) = \mathrm{var}[X] - \mathrm{var}\left[\bar{X}\right] = \mathrm{var}[X] - \frac{\mathrm{var}[X]}{n}.$$

Pričakovana vrednost variance vzorca je torej

$$\mathrm{E}\left[S_X^2\right] = \frac{n-1}{n}\,\mathrm{var}[X] = \frac{n-1}{n}\,\sigma_X^2. \tag{8.9}$$

Ta enačba nam pove, da je varianca vzorca $S_X^2$ pristranska ocena variance $\sigma_X^2$ slučajne spremenljivke $X$ oziroma populacije. Pričakovana vrednost variance vzorca je namreč nekoliko nižja od variance populacije. Razlog za to, da varianca vzorca $S_X^2$ podceni varianco populacije je v tem, da smo v enačbi (8.6) računali razlike med elementi vzorca in povprečjem vzorca, ki je bilo izračunano iz istega vzorca in predstavlja oceno pričakovane vrednosti $m_X$ populacije. Zaradi tega so razlike v povprečju manjše, kot bi bile, če bi računali razlike glede na pravo pričakovano vrednost $m_X$.

Iz enačbe (8.9) sledi, da je nepristranska ocena variance $\sigma_X^2$ slučajne spremenljivke $X$ enaka

$$S_X^{*2} = \frac{n}{n-1}\,S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2, \tag{8.10}$$

saj je

$$\mathrm{E}\left[S_X^{*2}\right] = \frac{n}{n-1}\mathrm{E}\left[S_X^2\right] = \mathrm{var}[X] = \sigma_X^2.$$

Enačba (8.9) velja za neskončno veliko populacijo oziroma za vzorčenje z vračanjem. V primeru končno velike populacije pa moramo pri računu pričakovane vrednosti variance vzorca upoštevati enačbo (8.5)

$$\mathrm{E}\left[S_X^2\right] = \mathrm{var}[X] - \mathrm{var}[\bar{X}] = \mathrm{var}[X] - \frac{\mathrm{var}[X]}{n}\,\frac{N-n}{N-1} = \mathrm{var}[X]\frac{n-1}{n}\,\frac{N}{N-1} \tag{8.11}$$

in

$$\mathrm{E}\left[S_X^{*2}\right] = \frac{n}{n-1}\mathrm{E}\left[S_X^2\right] = \mathrm{var}[X]\frac{N}{N-1}. \tag{8.12}$$

Zadnji dve enačbi povesta, da je v primeru končno velike populacije tudi statistika $S_X^{*2}$ pristranska. V tem primeru bi si morali izbrati novo statistiko $S_X^{**2}$, ki je nepristranska tudi v primeru končne populacije

$$S_X^{**2} = \frac{N-1}{N}S_X^{*2} = \frac{N-1}{N}\,\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

Zanimivo je tudi preveriti, kaj se zgodi v primeru, da je $n = N$, torej ko je v vzorec vključena celotna populacija. Vidimo lahko, da v tem primeru statistiki $S_X^2$ in $S_X^{**2}$ sovpadata

$$S_X^2 = S_X^{**2} = \frac{1}{N}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \mathrm{var}[X] = \sigma_X^2.$$

V primeru neskončno velikega vzorca, ko gre $n = N$ proti neskončnosti, statistiki $S_X^2$ in $S_X^{**2}$ limitirata proti $S_X^{*2}$ iz enačbe (8.10).

Brez izpeljave zapišimo varianci statistik $S_X^2$ in $S_X^{*2}$ za neskončno velike populacije:

$$\mathrm{var}[S_X^2] = \sigma_X^4 \frac{2(n-1)}{n^2} \quad \text{in} \quad \mathrm{var}[S_X^{*2}] = \sigma_X^4 \frac{2}{n-1}.$$

Kot zanimivost lahko še poudarimo, da standardna deviacija vzorca $S_X^*$ ni nepristranska ocena standardne deviacije populacije $\sigma_X$. Pričakovana vrednost standardne deviacije vzorce $S_X^*$ je odvisna od tipa porazdelitve slučajne spremenljivke $X$. V primeru, da je porazdelitev normalna, je pričakovana vrednost standardne deviacije $\mathrm{E}[S_X^*]$ enaka

$$\mathrm{E}[S_X^*] = \frac{4n-4}{4n-3}\,\sigma_X.$$

# 9 Ocenjevanje parametrov

Slučajno spremenljivko $X$ oziroma populacijo, ki jo ta predstavlja, v celoti opišemo z njeno porazdelitvijo. Vsaki porazdelitvi pripadajo določeni parametri, ki jih običajno ne poznamo. Na osnovi vzorca oziroma statistik, ki jih iz vzorca izračunamo, lahko parametre populacije ocenimo.

Ločimo dva tipa ocen: **točkovna ocena**, s katero oceno parametra podamo z eno samo vrednostjo, ter **intervalna ocena**, s katero podamo interval oziroma območje, za katerega lahko z določeno verjetnostjo trdimo, da vključuje parameter. Parameter označimo z $a$, njegovo točkovno oceno označimo z $\hat{a}$. Ocena $\hat{a}$ je **nepristranska**, če velja, da je

$$\mathrm{E}[\hat{a}] = a.$$

Intervalno oceno oziroma **interval zaupanja** podamo z naslednjo enačbo:

$$a \in [a_s, a_z] \quad \text{tako, da je} \quad P[a_s \leq a \leq a_z] = 1 - \alpha,$$

kjer $1 - \alpha$ predstavlja stopnjo zaupanja, s katerim izrečemo trditev, da parameter $a$ leži v intervalu med $a_s$ in $a_z$.

## 9.1 Točkovne ocene

Vzemimo, da želimo določiti točkovne ocene parametrov porazdelitve $a_i,\ i = 1, \ldots, n_a$. Pri porazdelitvah, ki jih običajno uporabljamo, je število parametrov $n_a$ enako ena (geometrijska, Poissonova, eksponentna...) ali dve (binomska, Pascalova, normalna...).

Tu bomo obravnavali dve metodi določanja točkovnih ocen parametrov porazdelitev: **metoda momentov** in **metoda največjega verjetja**. Pri prvi ocene parametrov določimo iz ocen prvih nekaj momentov slučajne spremenljivke: $\hat{m}_X = \bar{X}$, $\hat{\sigma}_X^2 = S_X^{*2}$. Pri drugi pa ocene parametrov določimo z uporabo verjetnostne funkcije pri diskretnih slučajnih spremenljivkah in gostoto verjetnosti pri zveznih slučajnih spremenljivkah.

### 9.1.1 Metoda momentov

Določiti želimo take ocene parametrov $a_i$, da bodo momenti slučajne spremenljivke izbrane porazdelitve enaki iz vzorca izračunanim ocenam teh momemtov.

Že pri obravnavanju slučajnih spremenljivk v poglavju 7 smo pri vsaki porazdelitvi zapisali zvezo med parametri porazdelitve in njenimi momenti. Na primer: pri slučajni spremenljivki $Y$, ki je binomsko porazdeljena s parametroma $n_b$ in $p$ sta pričakovana vrednost in varianca enaki: $m_Y = n_b\,p$ in $\sigma_Y^2 = n_b\,p\,(1-p)$.

S to metodo običajno ocenjujemo parametre porazdelitev z enim ali dvema parametroma. Če je parameter le eden, potem za oceno $\hat{a}$ parametra $a$ uporabimo oceno pričakovane vrednosti $\hat{m}_X = \bar{X}$:

$$m_X = f(a) \quad \rightarrow \quad a = f^{-1}(m_X) \quad \rightarrow \quad \hat{a} = f^{-1}(\hat{m}_X) = f^{-1}\left(\bar{X}\right),$$

kjer $f^{-1}(.)$ predstavlja inverzno zvezo funkcije $f(a)$. V splošnem je ta zveza nelinearna in moremo torej rešiti nelinearno enačbo.

Če sta parametra dva ($a_1$ in $a_2$), njuni oceni $\hat{a}_1$ in $\hat{a}_2$ določimo iz naslednjih enačb

$$\left.\begin{aligned} m_X &= f_1(a_1, a_2) \\ \sigma_X^2 &= f_2(a_1, a_2) \end{aligned}\right\} \quad \longrightarrow \quad \begin{aligned} f_1(\hat{a}_1, \hat{a}_2) &= \bar{X} \\ f_2(\hat{a}_1, \hat{a}_2) &= S_X^{*2} \end{aligned} \quad ,$$

kjer zadnji dve enačbi predstavljata sistem dveh, v splošnem nelinearnih, enačb za dve neznani oceni $\hat{a}_1$ in $\hat{a}_2$.

**Primer 9.1:** *Določimo oceno parametra slučajne spremenljivke $X$, če je porazdeljena:*

*a) binomsko s parametroma $n_b = 80$ in $p$ – v tem primeru določamo le oceno parametra $p$,*

*b) eksponentno s parametrom $\lambda$.*

*Iz vzorca s 30 elementi smo izračunali oceno pričakovane vrednosti slučajne spremenljivke: $\bar{X} = 20$.*

**Rešitev**: Pri obeh porazdelitvah opravimo enak postopek: določimo zvezo med pričakovano vrednostjo $m_X$ in parametrom in namesto nje vstavimo njeno oceno $\bar{X}$:

a) **Binomska porazdelitev s parametroma $n_b = 80$ in $p$:**

V tem primeru določimo le oceno parametra $\hat{p}$:

$$m_X = n_b\,p \quad \rightarrow \quad p = \frac{m_X}{n_b} \quad \rightarrow \quad \hat{p} = \frac{\bar{X}}{n_b} = 0.25.$$

a) **Eksponentna porazdelitev s parametrom $\lambda$:**

Določamo oceno parametra $\hat{\lambda}$:

$$m_X = \frac{1}{\lambda} \quad \rightarrow \quad \lambda = \frac{1}{m_X} \quad \rightarrow \quad \hat{\lambda} = \frac{1}{\bar{X}} = 0.05. \qquad (9.1)$$

**Primer 9.2:** *Določimo oceni parametrov slučajne spremenljivke $X$, če je porazdeljena:*

*a) enakomerno s parametroma $a$ in $b$,*
*b) binomsko s parametroma $n_b$ in $p$,*
*c) normalno s parametroma $m_X$ in $\sigma_X$,*
*d) lognormalno s parametroma $\tilde{m}_X$ in $\sigma_{\ln X}$,*
*e) po Gumbelovi porazdelitvi s parametroma $\alpha$ in $u$.*

*Iz vzorca s 30 elementi smo izračunali oceni pričakovane vrednosti in standardnega odklona slučajne spremenljivke: $\bar{X} = 30$, $S_X^* = 5$.*

**Rešitev**: Pri vseh porazdelitvah opravimo enak postopek: določimo zveze med momentoma in parametroma in namesto momentov vstavimo njuni oceni:

a) **Enakomerna porazdelitev:**

Iz sistema linearnih enačb

$$m_X = \frac{a+b}{2}$$
$$\sigma_X = \frac{b-a}{\sqrt{12}}$$

lahko izračunamo zveze

$$\left.\begin{aligned} b &= m_X + \sqrt{3}\,\sigma_X \\ a &= m_X - \sqrt{3}\,\sigma_X \end{aligned}\right\} \longrightarrow \quad \begin{aligned} \hat{b} &= \bar{X} + \sqrt{3}\,S_X^* = 38.66 \\ \hat{a} &= \bar{X} - \sqrt{3}\,S_X^* = 21.34 \end{aligned}$$

b) **Binomska porazdelitev s parametroma $n_b$ in $p$:**

Iz sistema enačb

$$m_X = n_b\,p$$
$$\sigma_X = \sqrt{n_b\,p\,(1-p)}$$

lahko izračunamo zveze

$$\left.\begin{aligned} p &= 1 - \frac{\sigma_X^2}{m_X} \\ n_b &= \frac{m_X}{p} \end{aligned}\right\} \longrightarrow \quad \begin{aligned} \hat{p} &= 1 - \frac{S_X^{*2}}{\bar{X}} = 0.16667 \\ \hat{n}_b &= \frac{\bar{X}}{\hat{p}} = 180 \end{aligned} \tag{9.2}$$

c) **Normalna porazdelitev s parametroma $m_X$ in $\sigma_X$:**

V tem primeru je določitev preprosta, saj velja, da je

$$\left.\begin{aligned} m_X &= m_X \\ \sigma_X &= \sigma_X \end{aligned}\right\} \longrightarrow \quad \begin{aligned} \hat{m_X} &= \bar{X} = 30 \\ \hat{\sigma_X} &= S_X^* = 5 \end{aligned} \tag{9.3}$$

d) **Lognormalna porazdelitev s parametroma $\tilde{m}_X$ in $\sigma_{\ln X}$:**

Zvezi med parametroma $\tilde{m}_X$ in $\sigma_{\ln X}$ in momentoma $m_X$ in $\sigma_X$ smo že izpeljali, ko smo obravnavali lognormalno porazdelitev (enačba (7.47))

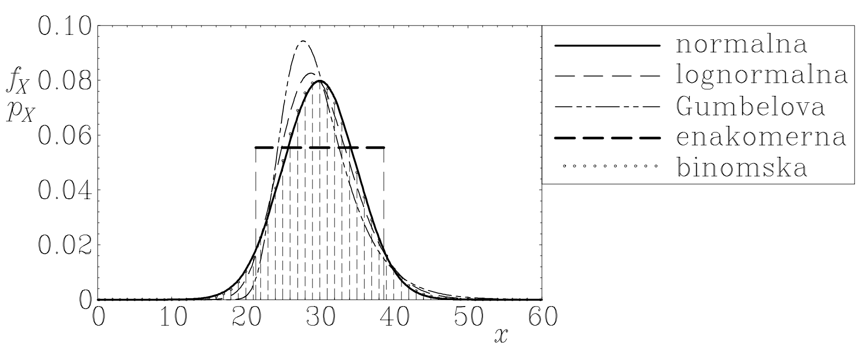
$$\left.\begin{aligned} \tilde{m}_X &= m_X\, e^{-\frac{1}{2}\sigma^2_{\ln X}}, \\ \sigma^2_{\ln X} &= \ln\left(\frac{\sigma_X^2}{m_X^2}+1\right) \end{aligned}\right\} \longrightarrow \quad \begin{aligned} \hat{\tilde{m}}_X &= \bar{X}\, e^{-\frac{1}{2}\hat{\sigma}^2_{\ln X}} = 29.59, \\ \hat{\sigma}^2_{\ln X} &= \ln\left(\frac{S_X^{*2}}{\bar{X}^2}+1\right) = 0.02740. \end{aligned} \tag{9.4}$$

e) **Gumbelova porazdelitev s parametroma $\alpha$ in $u$:**

Zvezi med parametroma $\alpha$ in $u$ in momentoma $m_X$ in $\sigma_X$ smo že izpeljali, ko smo obravnavali lognormalno porazdelitev (enačba (7.56))

$$\left.\begin{aligned} \alpha &= \frac{\pi}{\sqrt{6}\,\sigma_X}, \\ u &= m_X - \frac{\gamma}{\alpha}, \end{aligned}\right\} \longrightarrow \quad \begin{aligned} \hat{\alpha} &= \frac{\pi}{\sqrt{6}\,S_X^*} = 0.2565, \\ \hat{u} &= \bar{X} - \frac{\gamma}{\hat{\alpha}} = 27.75, \end{aligned}$$

kjer je $\gamma$ Eulerjeva konstanta ($\gamma \approx 0.577216$).




Slika 9.1: Različne porazdelitve z $m_X = 30$ in $\sigma_X = 5$

### 9.1.2 Metoda največjega verjetja

Vzorec $X_j,\ \ j = 1,\dots,n$ je slučajen. Verjetnost, da se zgodi nek vzorec, je odvisna od porazdelitve slučajne spremenljivke in njenih parametrov $a_i,\ \ i = 1,\dots,n_a$. Določiti želimo take ocene parametrov $\hat{a}_i$, da bo verjetnost, da se je zgodil vzorec, ki ga imamo na voljo, največja.

#### Diskretne slučajne spremenljivke

Verjetnost, da slučajna spremenljivka $X_j$ zavzame določeno vrednost, izračunamo z verjetnostno funkcijo

$$P[X_j = x_j] = p_X(x_j),$$

kjer smo upoštevali, da so vsi $X_j$ enako porazdeljeni. Verjetnost, da slučajne spremenljivke $X_j$ zavzamejo vrednosti $x_j$ pa lahko zapišemo kot produkt verjetnostnih funkcij, saj predpostavimo, da so $X_j$ medsebojno neodvisne – vrednost enega elementa vzorca ne vpliva na drugega:

$$
\begin{aligned}
&P[X_1 = x_1 \cap X_2 = x_2 \cap \ldots \cap X_n = x_n] = \\
&= P[X_1 = x_1]\, P[X_2 = x_2] \cdots P[X_n = x_n] = \\
&= p_X(x_1)\, p_X(x_2) \ldots p_X(x_n) = \prod_{j=1}^{n} p_X(x_j).
\end{aligned}
$$

Sedaj lahko definiramo **funkcijo verjetja** $L(a_i)$ kot verjetnost, da se zgodi nek vzorec

$$
L(a_i) = P[\text{vzorca}] = \prod_{j=1}^{n} p_X(x_j). \tag{9.5}
$$

Ocene parametrov $\hat{a}_i$ določimo z zahtevo, da ima funkcija $L(a_i)$ maksimum.

Podrobneje bomo to prikazali na preprostem primeru binomsko porazdeljene slučajne spremenljivke.

**Primer 9.3:** *Vzemimo, da imamo vzorec binomsko porazdeljene slučajne spremenljivke $Y$, za katero vemo, da je $n_b = 40$. Določiti moramo parameter $p$ tako, da bo verjetnost, da se je tak vzorec zgodil, največja. Vzorec predstavljajo tri slučajne spremenljivke $Y_j$, ki so zavzele naslednje vrednosti: 5, 4 in 7. Velikost vzorca je torej enaka $n = 3$.*

**Rešitev:** Verjetnost, da se je določen vzorec zgodil, zapišemo v obliki funkcije $L(p)$:

$$
L(p) = P[\text{vzorca}] = P[Y_1 = y_1] \cdot P[Y_2 = y_2] \cdot P[Y_3 = y_3],
$$

saj predpostavimo, da so posamezni elementi vzorca medsebojno neodvisni. Za vse tri slučajne spremenlivke $Y_i$ predpostavimo isto binomsko porazdelitev, zato lahko zapišemo:

$$
\begin{aligned}
L(p) &= \binom{n_b}{y_1} p^{y_1} (1-p)^{n_b - y_1} \binom{n_b}{y_3} p^{y_2} (1-p)^{n_b - y_2} \binom{n_b}{y_2} p^{y_3} (1-p)^{n_b - y_3} = \\
&= \binom{n_b}{y_1} \binom{n_b}{y_2} \binom{n_b}{y_3} p^{y_1 + y_2 + y_3} (1-p)^{3n_b - y_1 - y_2 - y_3}.
\end{aligned} \tag{9.6}
$$

Določiti želimo tak $p$, da je verjetnost vzorca $L(p) = P[\text{vzorca}]$ maksimalna, kar običajno naredimo tako, da postavimo pogoj, da je odvod $dL(p)/dp = 0$

$$
\frac{dL(p)}{dp} = \frac{d}{dp} \left( \binom{n_b}{y_1} \binom{n_b}{y_2} \binom{n_b}{y_3} \hat{p}^{y_1 + y_2 + y_3} (1-\hat{p})^{3n_b - y_1 - y_2 - y_3} \right) = 0.
$$

Če moramo določiti ekstrem funkcije, ki ima obliko produkta, to lažje opravimo tako, da enačbo logaritmiramo in nato odvajamo. V tem primeru logaritmiramo enačbo (9.6)

$$
\ln L(p) = \ln \left( \binom{n_b}{y_1} \binom{n_b}{y_2} \binom{n_b}{y_3} \right) + (y_1 + y_2 + y_3) \ln p + (3n_b - y_1 - y_2 - y_3) \ln(1-p)
$$

in po odvajanju dobimo

$$\frac{d\ln L(p)}{dp} = \frac{y_1+y_2+y_3}{\hat{p}} - \frac{3n_b - y_1 - y_2 - y_3}{1-\hat{p}} = 0 \quad \longrightarrow$$
$$\longrightarrow \quad \frac{(1-\hat{p})(y_1+y_2+y_3) - \hat{p}(3n_b - y_1 - y_2 - y_3)}{\hat{p}(1-\hat{p})} = 0 \quad \longrightarrow$$
$$\longrightarrow \quad (y_1+y_2+y_3) - 3\,\hat{p}\,n_b = 0 \quad \longrightarrow \quad \hat{p} = \frac{y_1+y_2+y_3}{3\,n_b} = \frac{\bar{Y}}{n_b} = \frac{5.333}{40} = 0.1333.$$

Vidimo, da smo dobili enako enačbo kot pri metodi momentov (enačba (9.2)). Kot bomo videli v enem izmed naslednjih primerov, metoda največjega verjetja ne da vedno iste enačbe za račun ocene parametrov kot metoda momentov.

### Zvezne slučajne spremenljivke

Pri zveznih slučajnih spremenljivkah je verjetnost, da se zgodi nek točno določen vzorec, enaka nič. Zato funkcijo verjetja $L(a_i)$ zapišemo v odvisnosti od gostote verjetnosti in rečemo, da moramo določiti take ocene $\hat{a}_i$, da bo gostota verjetnosti pri danih vrednostih vzorca največja

$$L(a_i) = \prod_{j=1}^{n} f_X(x_j). \tag{9.7}$$

Poiskati moramo take vrednosti $a_i$, da ima funkcija $L(a_i)$ maksimum.

Podrobneje bomo uporabo te metode prikazali na računskih primerih.

**Primer 9.4:** *Vzemimo, da imamo vzorec eksponentno porazdeljene slučajne spremenljivke $T$. Določimo enačbo za določitev ocene parametra $\hat{\lambda}$ po metodi največjega verjetja. Vzemimo, da je velikost vzorca enaka $n = 30$, povprečje vzorca pa je $\bar{T} = 20$.*

**Rešitev:** Funkcija verjetja je produkt

$$L(\lambda) = \prod_{j=1}^{n} \lambda e^{-\lambda\, t_j}.$$

Tudi v tem primeru bo izpeljava preprostejša če zgornjo enačbo logaritmiramo in nato odvajamo po parametru $\lambda$

$$\ln L(\lambda) = \sum_{j=1}^{n} (\ln\lambda - \lambda\, t_j) = n\,\ln\,\lambda - \lambda \sum_{j=1}^{n} t_j.$$

Če želimo določiti tako oceno parametra $\lambda$, da ima funkcija $L(\lambda)$ maksimum, zahtevamo, da je odvod $dL(\lambda)/d\lambda$ oziroma $d\ln(L(\lambda))/d\lambda$ enak nič

$$\frac{d\ln(L(\lambda))}{d\lambda} = 0 \quad \longrightarrow \quad \frac{n}{\hat{\lambda}} - \sum_{j=1}^{n} t_j = 0 \quad \longrightarrow \quad \hat{\lambda} = \frac{n}{\sum\limits_{j=1}^{n} t_j} = \frac{1}{\bar{T}} = \frac{1}{20} = 0.05.$$

Vidimo, da smo dobili enak rezultat kot pri metodi momentov (enačba (9.1)).

**Primer 9.5:** *Določimo oceni parametrov slučajne spremenljivke $X$, če je porazdeljena po:*

*a) normalno s parametroma $m_X$ in $\sigma_X$,*

*b) lognormalno s parametroma $\tilde{m}_X$ in $\sigma_{\ln X}$,*

*c) po Gumbelovi porazdelitvi s parametroma $\alpha$ in $u$.*

*Vzorce 30 elementov za različne porazdelitve prikazujemo v naslednji preglednici:*

Preglednica 9.1: Vzorec slučajnih spremenljivk

| **Normalna porazdelitev** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 26.56 | 23.14 | 36.84 | 24.53 | 31.13 | 31.46 | 31.95 | 36.85 | 28.38 | 31.97 |
| 25.11 | 24.08 | 32.03 | 33.43 | 22.71 | 30.66 | 30.40 | 31.93 | 32.27 | 27.46 |
| 26.91 | 28.86 | 31.66 | 27.93 | 32.41 | 38.90 | 29.70 | 30.48 | 24.25 | 34.28 |
| **Lognormalna porazdelitev** | | | | | | | | | |
| 31.43 | 27.31 | 29.37 | 31.39 | 22.49 | 23.65 | 37.99 | 32.39 | 27.58 | 28.98 |
| 23.64 | 34.16 | 30.00 | 35.05 | 29.51 | 24.01 | 25.48 | 22.06 | 25.96 | 36.30 |
| 28.94 | 36.82 | 20.09 | 25.32 | 45.12 | 36.70 | 31.27 | 35.88 | 41.25 | 26.72 |
| **Gumbelova porazdelitev** | | | | | | | | | |
| 30.88 | 23.54 | 31.76 | 34.18 | 29.80 | 30.90 | 27.07 | 34.31 | 27.42 | 25.09 |
| 29.56 | 25.46 | 31.26 | 37.49 | 29.73 | 22.22 | 29.23 | 23.33 | 34.55 | 28.20 |
| 32.21 | 29.11 | 24.90 | 31.68 | 27.40 | 31.27 | 29.35 | 24.87 | 26.88 | 51.89 |

**Rešitev:** Funkcija verjetja so v tem primeru funkcije dveh spremenljivk, saj imajo vse tri obravnavane slučajne spremenljivke po dva parametra.

a) Pri normalni porazdelitvi funkcijo verjetja $L(m_X, \sigma_X)$ zapišemo z naslednjo enačbo:

$$L(m_X, \sigma_X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_X}\, e^{-\frac{1}{2}\left(\frac{x_i - m_X}{\sigma_X}\right)^2}. \tag{9.8}$$

Podobno kot v primeru 9.4, kjer smo obravnavali eksponentno porazdelitev, tudi tokrat nalogo lažje rešimo, če enačbo (9.8) logaritmiramo in jo nato odvajamo po parametrih $m_X$ in $\sigma_X$

$$\ln L(m_X, \sigma_X) = \sum_{i=1}^{n} \left( -\ln(2\pi) - \ln\sigma_X - \frac{1}{2}\left(\frac{x_i - m_X}{\sigma_X}\right)^2 \right) =$$

$$= -n\,\ln(2\pi) - n\,\ln\sigma_X - \frac{1}{2\,\sigma_X^2} \sum_{i=1}^{n} (x_i - m_X)^2.$$

Z odvajanjem po neznanih parametrih $m_X$ in $\sigma_X$ dobimo sistem dveh enačb

$$\frac{2}{2\,\hat{\sigma}_X^2}\sum_{i=1}^{n}(x_i - \hat{m}_X) = 0$$

$$-\frac{n}{\hat{\sigma}_X} + \frac{2}{2\,\hat{\sigma}_X^3}\sum_{i=1}^{n}(x_i - m_X)^2 = 0.$$

Enačbi preuredimo in dobimo obrazca za določitev ocen parametrov normalne porazdelitve $\hat{m}_X$ in $\hat{\sigma}_X$ po metodi največjega verjetja

$$\sum_{i=1}^{n}(x_i - \hat{m}_X) = 0 \qquad \longrightarrow \qquad \hat{m}_X = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X},$$

$$-n + \frac{1}{\hat{\sigma}_X^2}\sum_{i=1}^{n}(x_i - \hat{m}_X)^2 = 0 \qquad \longrightarrow \qquad \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2 = S_X^2.$$

Vidimo lahko, da smo za račun ocene pričakovane vrednosti dobili enako oceno kot po metodi momentov (enačba (9.3)), oceno variance $\hat{\sigma}_X$ pa določimo s statistiko $S_X^2$ in ne s $S_X^{*2}$ kot pri metodi momentov. Ob upoštevanju podatkov v preglednici 9.1 dobimo po metodi momentov oceni $\hat{m}_X = 29.94$ in $\hat{\sigma}_X = 4.130$, medtem ko po metodi največjega verjetja dobimo $\hat{m}_X = 29.94$ in $\hat{\sigma}_X = 4.061$.

- Funkcijo $L(\tilde{m}_X, \sigma_{\ln X})$ za lognormalno porazdelitev zapišemo takole:

$$L(\tilde{m}_X, \sigma_{\ln X}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\, x\, \sigma_{\ln X}}\, e^{-\frac{1}{2}\left(\frac{\ln x_i - \ln \tilde{m}_X}{\sigma_{\ln X}}\right)^2}. \tag{9.9}$$

Enačbo (9.9) logaritmiramo in jo nato odvajamo po parametrih $\tilde{m}_X$ in $\sigma_{\ln X}$

$$\ln L(\tilde{m}_X, \sigma_{\ln X}) = \sum_{i=1}^{n}\left(-\ln(2\pi) - \ln x_i - \ln \sigma_{\ln X} - \frac{1}{2}\left(\frac{\ln x_i - \ln \tilde{m}_X}{\sigma_{\ln X}}\right)^2\right) =$$

$$= -n\, \ln(2\pi) - \sum_{i=1}^{n} \ln x_i - n\, \ln \sigma_{\ln X} - \frac{1}{2\,\sigma_{\ln X}^2}\sum_{i=1}^{n}(\ln x_i - \ln \tilde{m}_X)^2.$$

Z odvajanjem po neznanih parametrih $\tilde{m}_X$ in $\sigma_{\ln X}$ dobimo sistem dveh enačb

$$\frac{2}{2\,\hat{\sigma}_{\ln X}^2}\sum_{i=1}^{n}(\ln x_i - \ln \hat{\tilde{m}}_X)\frac{1}{\hat{\tilde{m}}_X} = 0$$

$$-\frac{n}{\hat{\sigma}_{\ln X}} + \frac{2}{2\,\hat{\sigma}_{\ln X}^3}\sum_{i=1}^{n}(\ln x_i - \ln \hat{\tilde{m}}_X)^2 = 0.$$

Enačbi preuredimo in dobimo obrazca za določitev ocen parametrov lognormalne porazdelitve $\hat{\tilde{m}}_X$ in $\hat{\sigma}_{\ln X}$ po metodi največjega verjetja

$$\sum_{i=1}^{n}(\ln x_i - \ln \hat{\tilde{m}}_X) = 0 \qquad \longrightarrow \qquad \hat{\tilde{m}}_X = \left(\prod_{i=1}^{n} x_i\right)^{1/n},$$

$$-n + \frac{1}{\hat{\sigma}_X^2}\sum_{i=1}^{n}(\ln x_i - \ln \hat{\tilde{m}}_X)^2 = 0 \qquad \longrightarrow \qquad \hat{\sigma}_{\ln X}^2 = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \ln \hat{\tilde{m}}_X)^2.$$

Vidimo, da smo tokrat dobili drugačne enačbe kot pri metodi momentov (enačbi (9.4). Po metodi momentov ob upoštevanju podatkov iz preglednice (9.1) dobimo oceni $\hat{\tilde{m}}_X = 29.643$ in $\hat{\sigma}_{\ln X} = 0.19775$, po metodi največjega verjetja pa dobimo nekoliko drugačni oceni $\hat{\tilde{m}}_X = 29.661$ in $\hat{\sigma}_{\ln X} = 0.19441$.

- Funkcijo $L(u, \alpha)$ za Gumbelovo porazdelitev zapišemo takole (glej enačbo (7.55)):

$$L(u, \alpha) = \prod_{i=1}^{n} \alpha\, e^{-\alpha\,(x_i - u) - e^{-\alpha\,(x_i - u)}} \tag{9.10}$$

Po logaritmiranju pa dobimo nekoliko drugačno obliko

$$\ln L(u, \alpha) = n \,\ln \alpha - \alpha \sum_{i=1}^{n} x_i + n\,\alpha\, u - \sum_{i=1}^{n} e^{-\alpha\,(x_i - u)} \tag{9.11}$$

Poiskati moramo maksimum funkcije $L(m_X, \sigma_X)$ po enačbi (9.10) oziroma $\ln L(u, \alpha)$ po enačbi (9.11). Če enačbo (9.11) odvajamo po parametrih $\alpha$ in $u$, dobimo dve nelinearni enačbi:

$$\frac{\partial \ln L}{\partial \alpha} = 0 \quad \rightarrow \quad \frac{n}{\hat{\alpha}} - \sum_{i=1}^{n} x_i + n\,\hat{u} + \sum_{i=1}^{n}(x_i - \hat{u})\, e^{-\hat{\alpha}\,(x_i - \hat{u})} = 0,$$

$$\frac{\partial \ln L}{\partial u} = 0 \quad \rightarrow \quad n\,\hat{\alpha} - \sum_{i=1}^{n} \hat{\alpha} e^{-\hat{\alpha}(x_i - \hat{u})} = 0.$$

Zadnji enačbi predstavljata sistem dveh nelinearnih enačb, ki jih analitično ne znamo rešiti. Lahko pa ju rešimo numerično, če enačbi nekoliko preuredimo:

$$\begin{aligned} \hat{\alpha} &= \frac{n}{\sum\limits_{i=1}^{n} x_i - n\,\hat{u} - \sum\limits_{i=1}^{n}(x_i - \hat{u})\, e^{-\hat{\alpha}\,(x_i - \hat{u})}}, \\ \hat{u} &= \frac{1}{\hat{\alpha}} \,\ln \left( \frac{n}{\sum\limits_{i=1}^{n} e^{-\hat{\alpha}\, x_i}} \right). \end{aligned} \tag{9.12}$$

Enačbi (9.12) uporabimo pri numerični rešitvi problema tako, da v njiju najprej vstavimo začetne vrednosti parametrov in nato ponavljamo izračun s popravljenimi vrednostmi, dokler se le-te ne razlikujejo za več kot izbrano odstopanje. V našem primeru vzamemo vrednosti, ki ju dobimo po metodi momentov:

$$\left.\begin{aligned} \alpha &= \frac{\pi}{\sqrt{6}\,\sigma_X}, \\ u &= m_X - \frac{\gamma}{\alpha}, \end{aligned}\right\} \quad \longrightarrow \quad \begin{aligned} \hat{\alpha} &= \frac{\pi}{\sqrt{6}\,S_X^*} = 0.2318, \\ \hat{u} &= \bar{X} - \frac{\gamma}{\hat{\alpha}} = 27.36, \end{aligned}$$

To preprosto iterativno metodo lahko izvedemo z različnimi programi (programski jeziki C++, FORTAN, BASIC..., ali računalniški programi kot so EXCEL, MATLAB ali MATHEMATICA).[1]

[1] Rešitev sistema dveh nelinearnih enačb s preprosto iteracijsko metodo v programu MATHEMATICA:

```
yy={30.88, 23.54, 31.76, 34.18, 29.80, 30.90, 27.07, 34.31, 27.42, 25.09,
    29.56, 25.46, 31.26, 37.49, 29.73, 22.22, 29.23, 23.33, 34.55, 28.20,
    32.21, 29.11, 24.90, 31.68, 27.40, 31.27, 29.35, 24.87, 26.88, 51.89};
ny=Length[yy]; my=Sum[yy[[i]],{i,ny}]/ny;
               sy=Sqrt[Sum[(yy[[i]]-my)^2,{i,ny}]/(ny-1)];
alfa = N[Pi/Sqrt[6]/sy];  u = N[my - EulerGamma/alfa];
Print["Ocena po metodi momentov"]
Print["my   = ",my,"       sy = ",sy]
Print["alfa = ",alfa,"      u  = ",u]

alfa1 = 0; it = 1;
While[Abs[alfa-alfa1]>0.000001,
  alfa1 = alfa;   u1 = u;   It++;
  alfa = ny/(Sum[yy[[i]],{i,ny}]- ny u - Sum[(yy[[i]]-u)
         E^(-alfa (yy[[i]]-u)),{i,ny}]);
  u = 1/alfa Log[ny/Sum[E^(-alfa yy[[i]]),{i,ny}]];]
Print["\nOcena po metodi najvecjega verjetja"]
Print["iter.    alfa      u"];
Print[" ",it,"     ",alfa,"  ", u];

Ocena po metodi momentov
my   = 29.8513       sy = 5.53308
alfa = 0.231797      u  = 27.3612

Ocena po metodi najvecjega verjetja
iter.    alfa      u
 28     0.26184  27.6023
```

Numerična rešitev tega sistema enačb je $\hat{\alpha} = 0.2618$ in $\hat{u} = 27.60$, kar je precejšnje odstopanje od ocen, ki smo jih dobili po metodi momentov.

Naloge se lahko lotimo tudi tako, da uporabimo pripravljena orodja za iskanje ekstremnih vrednosti. V računalniškem programu EXCEL lahko uporabimo orodje SOLVER, v programu MATHEMATICA pa lahko uporabimo ukaz FINDMINIMUM. Ker pravzaprav iščemo maksimum funkcije $\ln L$, moramo pri uporabi ukaza FINDMINIMUM namesto funkcije $\ln L$ zapisati $-\ln L$.[2] Na ta način smo dobili enaki oceni kot po enačbi (9.12).

**Primer 9.6:** *Vzorec enakomerno od* $0$ *do* $b$ *porazdeljene slučajne spremenljivke prikazujemo v preglednici (9.2). Velikost vzorca je* $n = 5$*. Določimo enačbo za določitev ocene parametra* $\hat{b}$ *po metodi momentov in metodi največjega verjetja.*

Preglednica 9.2: Vzorec enakomerno porazdeljene slučajne spremenljivke

| | | | | |
|---|---|---|---|---|
| 1.65 | 0.30 | 3.70 | 0.15 | 1.25 |

**Rešitev:** Za določitev ocene parametra $\hat{b}$ po metodi momentov uporabimo zvezo med parametrom $b$ in pričakovano vrednostjo te slučajne spremenlivke

$$m_X = \frac{b}{2} \quad \rightarrow \quad b = 2\,m_X \quad \rightarrow \quad \hat{b} = 2\,\bar{X} = 2.82.$$

Funkcija verjetja je

$$L(b) = \prod_{i=1}^{n} f_X(x_i) = \begin{cases} 0 & \text{za } \ b < \max x_i \\ \dfrac{1}{b^n} & \text{za } \ b \geq \max x_i. \end{cases}$$

Na sliki 9.2 je ta primer ilustriran. Če je $b = b_1 < \max x_i$, potem je vsaj ena izmed $f_X(x_i)$ enaka nič in je s tem tudi $L(b)$ enaka nič. Če je $b = b_2 > \max x_i$, so vse vrednosti $f_X(x_i) = 1/b$, zato je funkcija verjetja $L(b) = 1/b^n$. Graf funkcije verjetja (slika 9.2) je torej monotono podajoča funkcija za $b > \max x_i$, kar pomeni, da je njena največja vrednost prav v točki $\hat{b} = \max x_i$, kar je tudi ocena tega parametra po metodi največjega verjetja.

[2] Iskanje maksimuma funkcije verjetnosti $\ln L$ v programu MATHEMATICA:

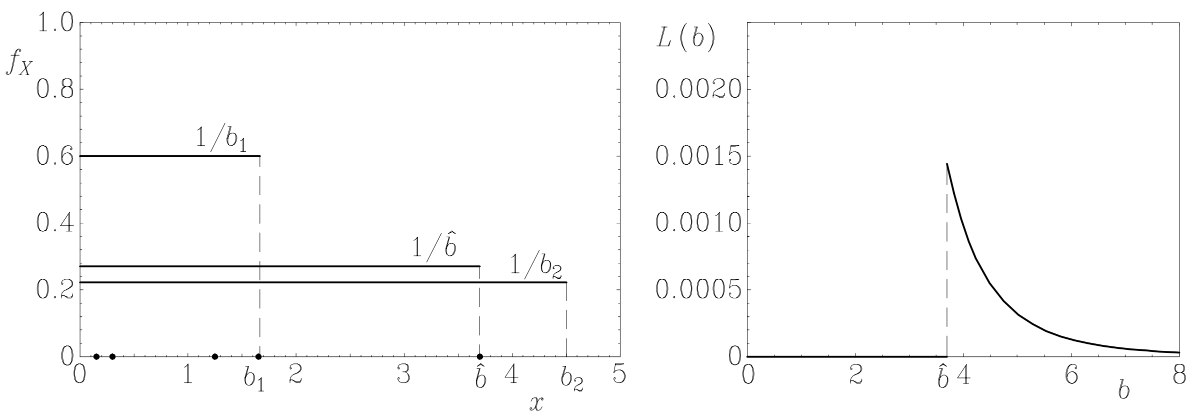
```
LL[alfa1_,u1_] = Log[Product[alfa1 E^(-alfa1 (yy[[i]]-u1) -
                 E^(-alfa1 (yy[[i]]-u1))),{i,ny}]];
resitev = FindMinimum[-LL[alfa1,u1],{alfa1,0.5},{u1,my}];
alfa = alfa1 /. resitev[[2,1]];
u = u1 /. resitev[[2,2]];
Print["\nOcena po metodi najvecjega verjetja"]
Print["alfa = ",alfa,"       u  = ",u]

Ocena po metodi najvecjega verjetja
alfa = 0.26184       u  = 27.6026
```

V našem primeru je ocena po metodi največjega verjetja enaka $\hat{b} = 3.7$, ki se precej razlikuje od vrednosti, ki smo jo dobili po metodi momentov ($\hat{\hat{b}} = 2.82$). Vprašamo se lahko, katera ocena je boljša. Na to lahko odgovorimo vsaj na dva načina:

- Ocena po metodi momentov ($\hat{\hat{b}} = 2.82$) je rezultat, za katerega lahko rečemo, da je neverjeten. Vzorec slučajne spremenljivke, ki je enakomerno porazdeljena od 0 do 2.82, seveda ne more vključevati vrednosti 3.7 (glej preglednico 9.2). Sklep: ocena po metodi največjega verjetja je boljša.
- Za oceno po metodi največjega verjetja ($\hat{b} = 3.7$) bi težko trdili, da je nepristranska, saj je odvisna le od enega elementa vzorca. Če je prišlo pri pripravi vzorca do napake, katere rezultat je previsoka vrednost enega elementa vzorca, bo to odločilno vplivalo na oceno. Sklep: boljša je ocena po metodi momentov.

Bralec naj se sam odloči, kateri odgovor mu je ljubši.

S tem primerom smo razložili tudi prvi primer 1.1 v tem učbeniku, kjer je bila naloga enaka (za $n = 1$), odgovor pa podan "po občutku".




Slika 9.2: Gostota verjetnosti slučajne spremenljivke $X$ in funkcija verjetja

## 9.2 Intervali zaupanja

Na osnovi slučajnega vzorca lahko ob izbrani stopnji zaupanja $1-\alpha$ določimo interval zaupanja za izbrani parameter. Ker se podatki vzorcev razlikujejo, se razlikujejo vzorčne ocene parametrov in zato tudi izračunane meje intervala zaupanja za ocenjevani parameter. Pri tem moramo posebej poudariti, da je izbrani parameter deterministična količina, meja intervala zaupanja pa slučajna spremenljivka. Z intervalom zaupanja torej s stopnjo zaupanja $1-\alpha$ določimo meje, za katere velja, da z verjetnostjo $1-\alpha$ vključujejo izbrani parameter.

V nadaljevanju bomo opisali določevanje intervalov zaupanja za nekatere parametre populacij: pričakovano vrednost, varianco, razliko v pričakovanih vrednostih in delež populacij.

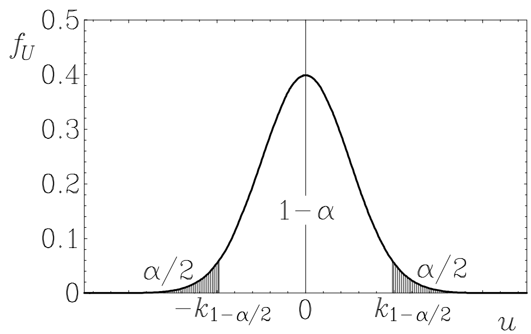
### 9.2.1 Interval zaupanja za pričakovano vrednost

V prejšnjem poglavju smo zapisali, da je povprečje vzorca $\bar{X}$ nepristranska ocena pričakovane vrednosti $m_X$. Povprečje vzorca $\bar{X}$ je slučajna spremenljivka s pričakovano vrednostjo $m_X$ in varianco $\sigma_X^2/n$ (za neskončno velike populacije). Če so elementi vzorca $X_i$ normalno porazdeljeni ali pa je vzorec razmeroma velik, lahko vzamemo, da je $\bar{X}$ porazdeljena normalno.

Iz vzorca lahko določimo statistiko

$$U = \frac{\bar{X} - m_X}{\sigma_X/\sqrt{n}}, \tag{9.13}$$

za katero vemo, da je porazdeljena normalno s pričakovano vrednostjo 0 in standardno deviacijo 1, oziroma standardizirano normalno.




Slika 9.3: Gostota verjetnosti statistike $U$

Verjetnost, da slučajna spremenljivka $U$ leži znotraj intervala $(-k_{1-\alpha/2}, k_{1-\alpha/2})$, je enaka $1-\alpha$ (glej tudi sliko 9.3), kar lahko napišemo z naslednjo enačbo

$$P[-k_{1-\alpha/2} \leq U \leq k_{1-\alpha/2}] = 1 - \alpha. \tag{9.14}$$

Iz slike 9.3 lahko vidimo, da vrednost $k_{1-\alpha/2}$ določimo s porazdelitveno funkcijo standardizirane normalne porazdelitve oziroma njeno inverzno funkcijo:

$$P[U \leq k_{1-\alpha/2}] = F_U(k_{1-\alpha/2}) = 1 - \alpha/2 \quad \rightarrow \quad k_{1-\alpha/2} = F_U^{-1}(1 - \alpha/2).$$

V enačbo (9.14) vstavimo enačbo (9.13) in dobimo:

$$\begin{aligned} &P\left[-k_{1-\alpha/2} \leq \frac{\bar{X} - m_X}{\sigma_X/\sqrt{n}} \leq k_{1-\alpha/2}\right] = 1 - \alpha \quad \rightarrow \\ &\rightarrow \quad P\left[\bar{X} - k_{1-\alpha/2}\,\sigma_X/\sqrt{n} \leq m_X \leq \bar{X} + k_{1-\alpha/2}\,\sigma_X/\sqrt{n}\right] = 1 - \alpha. \end{aligned} \tag{9.15}$$

Sedaj lahko zapišemo interval zaupanja za pričakovano vrednost $m_X$

$$m_X \in \left[\bar{X} - k_{1-\alpha/2}\,\sigma_X/\sqrt{n}, \bar{X} + k_{1-\alpha/2}\,\sigma_X/\sqrt{n}\right]. \tag{9.16}$$

Vidimo, da meji intervala zaupanja nista deterministični vrednosti, ampak slučajni. Meji intervala zaupanja sta za različne vzorce različni. Verjetnost, da iz vzorca določen interval zaupanja zajema pravo vrednost $m_X$, je enaka $1-\alpha$. Če bi ponavljali vzorčenje mnogokrat in iz vzorca vsakič določili interval zaupanja, lahko pričakujemo, da bo v $1-\alpha$ primerih interval zaupanja zajemal pravo vrednost $m_X$, v $\alpha$ primerih pa bo interval zaupanja določen napačno – prava vrednost $m_X$ bo ležala zunaj intervala zaupanja.

V enačbah (9.13) in (9.15) smo predpostavili, da je standardna deviacija $\sigma_X$ znana. Običajno seveda prave vrednosti standardne deviacije ne poznamo, temveč jo iz vzorca lahko ocenimo s statistiko $S^*_X$. V tem primeru s podobno enačbo kot (9.13) določimo statistiko $T$

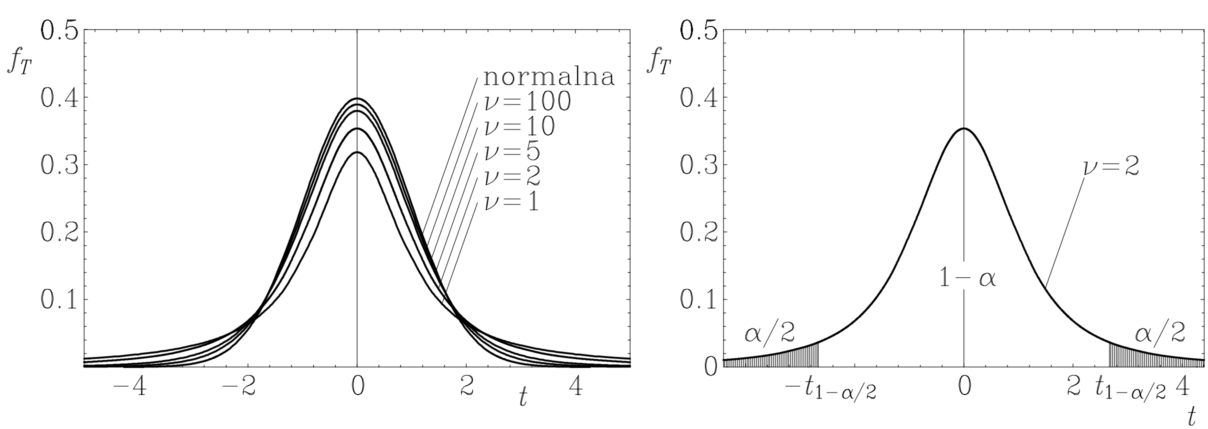
$$T = \frac{\bar{X} - m_X}{S^*_X/\sqrt{n}}, \tag{9.17}$$

ki je ob predpostavki, da je slučajna spremenljivka $X$ porazdeljena normalno, porazdeljena po Studentovi porazdelitvi $t$ z $n-1$ prostostnimi stopnjami. Lastnosti in oblika Studentove porazdelitve so zelo podobni standardizirani normalni porazdelitvi. Zato tudi verjetnost, da je vrednost slučajne spremenljivke $T$ znotraj intervala $(-t_{1-\alpha/2}, t_{1-\alpha/2})$, določimo s podobno enačbo kot (9.14) (glej tudi sliko 9.4)

$$P[-t_{1-\alpha/2} \le T \le t_{1-\alpha/2}] = 1-\alpha. \tag{9.18}$$

V zadnjo enačbo vstavimo enačbo (9.17) in izpeljemo interval zaupanja za pričakovano vrednost $m_X$

$$\begin{aligned}
&P\left[-t_{1-\alpha/2} \le \frac{\bar{X} - m_X}{S^*_X/\sqrt{n}} \le t_{1-\alpha/2}\right] = 1-\alpha \quad \rightarrow \\
&\rightarrow \quad P\left[\bar{X} - t_{1-\alpha/2}\, S^*_X/\sqrt{n} \le m_X \le \bar{X} + t_{1-\alpha/2}\, S^*_X/\sqrt{n}\right] = 1-\alpha \quad \rightarrow \\
&\rightarrow \quad m_X \in \left[\bar{X} - t_{1-\alpha/2}\, S^*_X/\sqrt{n}, \bar{X} + t_{1-\alpha/2}\, S^*_X/\sqrt{n}\right].
\end{aligned} \tag{9.19}$$




Slika 9.4: Gostota verjetnosti statistike $T$

Z enačbama (9.15) in (9.19) smo zapisali dvostranski interval zaupanja. Včasih pa nas zanima enostranski interval zaupanja, ki ga lahko opišemo z naslednjo enačbo (glej tudi sliko 9.5)

$$P[-t_{1-\alpha} \leq T] = 1 - \alpha. \tag{9.20}$$

Če v enačbo (9.20) vstavimo enačbo (9.17), lahko določimo enostranski interval zaupanja

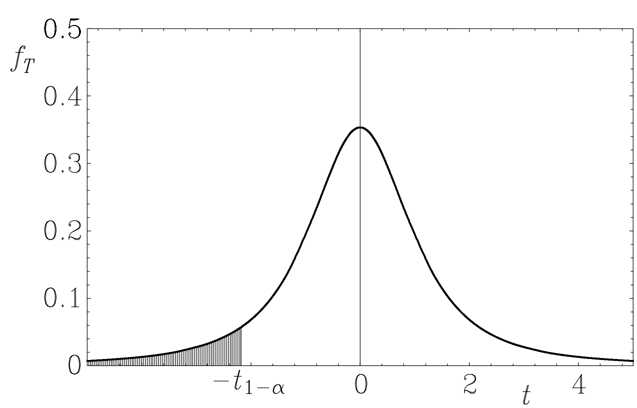
$$\begin{aligned} &P\left[-t_{1-\alpha} \leq \frac{\bar{X} - m_X}{S_X^*/\sqrt{n}}\right] = 1 - \alpha \quad \to \quad P\left[m_X \leq \bar{X} + t_{1-\alpha}\, S_X^*/\sqrt{n}\right] = 1 - \alpha \quad \to \\ &\to \quad m_X \in \left(-\infty, \bar{X} + t_{1-\alpha}\, S_X^*/\sqrt{n}\right]. \end{aligned} \tag{9.21}$$

Na podoben način iz enačbe

$$P[T \leq t_{1-\alpha}] = 1 - \alpha. \tag{9.22}$$

določimo tudi enostranski interval zaupanja na drugi strani porazdelitve

$$\begin{aligned} &P\left[\frac{\bar{X} - m_X}{S_X^*/\sqrt{n}} \leq t_{1-\alpha}\right] = 1 - \alpha \quad \to \quad P\left[\bar{X} - t_{1-\alpha}\, S_X^*/\sqrt{n} \leq m_X\right] = 1 - \alpha. \quad \to \\ &\to \quad m_X \in \left[\bar{X} - t_{1-\alpha}\, S_X^*/\sqrt{n}, \infty\right). \end{aligned} \tag{9.23}$$




Slika 9.5: Gostota verjetnosti statistike $T$ in enostranski interval zaupanja

**Primer 9.7:** *V pekarni pečejo hlebce. Ker vseh hlebcev nimajo časa tehtati, stehtajo le* $n$ *slučajno izbranih hlebcev. Na osnovi vzorca ugotovijo, da je povprečna masa hlebcev enaka* 2.1 kg*, standardni odklon vzorca pa je* 50 g*. Določimo mejo enostranskega intervala zaupanja tako, da bomo z določeno zanesljivostjo lahko trdili, da je pričakovana vrednost mase hlebcev večja od določene vrednosti. Tako vprašanje bi si zastavil prodajalec, ki bi rad oglašal, da imajo hlebci, ki jih prodaja, v povprečju večjo maso od določene vrednosti. Velikost vzorca* $n = 20$*, stopnja zaupanja je* 99*-odstotna.*

**Rešitev**: Ker je vzorec relativno majhen, moramo uporabiti porazdelitev $T$. Interval zaupanja izračunamo po enačbi (9.23). Vrednost $t_{1-\alpha,\nu}$ poiščemo v preglednici ali pa izračunamo z enem od računalniških programov (na primer EXCEL): $t_{1-\alpha,\nu} = t_{0.99,19} = 2.539$.

$$P\left[\bar{X} - t_{1-\alpha}\, S_X^*/\sqrt{n} \le m_X\right] = 1 - \alpha$$
$$P\left[2.1 - 2.539 \cdot 0.05/\sqrt{20} \le m_X\right] = P\left[2.07 \le m_X\right] = 0.99 \quad \rightarrow \quad m_X \in [2.07, \infty).$$

Z gotovostjo 0.99 lahko trdimo, da je pričakovana vrednost mase hlebcev večja od 2.07 kg.

**Primer 9.8:** *V laboratoriju preizkušajo trdnost betona. Rezultate meritev (v* $\mathrm{MPa}$*) prikazujemo v naslednji preglednici 9.3. Določite obojestranski interval zaupanja za pričakovano vrednost pri* $92$*-odstotni stopnji zaupanja. Določite tudi mejno vrednost* $x_k$*, za katero lahko trdite, da je pričakovana vrednost trdnosti betona z verjetnostjo* $85\ \%$ *večja od nje.*

Preglednica 9.3: Rezultati merjenja trdnosti betona

| | | | | | | |
|---|---|---|---|---|---|---|
| 28.329 | 28.386 | 31.803 | 29.173 | 31.973 | 23.323 | 33.612 |
| 26.409 | 30.306 | 24.643 | 23.788 | 21.926 | 36.715 | 37.433 |

**Rešitev**: Izračunajmo najprej povprečje in varianco oziroma standardno deviacijo vzorca:

$$\bar{X} = \frac{\sum_{i=1}^{14} x_i}{14} = \frac{28.329 + 28.386 + 31.803 + \cdots}{14} = 29.130,$$
$$S_X^{*2} = \frac{\sum_{i=1}^{14} (x_i - \bar{X})^2}{13} = \frac{(28.329 - 29.130)^2 + (28.386 - 29.130)^2 + \cdots}{13} = 23.543,$$
$$S_X^* = \sqrt{S_X^{*2}} = 4.852.$$

V vzorcu je 14 elementov, zato je število prostostnih stopenj $\nu = 14 - 1 = 13$. Obojestranski interval zaupanja določimo po enačbi (9.19), vrednost $t_{1-\alpha/2,13}$ pa izračunamo z računalniškim programom EXCEL: $t_{0.96,13} = 1.899$.

$$P\left[29.130 - 1.899 \cdot 4.852/\sqrt{14} \le m_X \le 29.130 + 1.899 \cdot 4.852/\sqrt{14}\right] = 0.92,$$
$$P\left[26.667 \le m_X \le 31.592\right] = 0.92 \quad \rightarrow \quad m_X \in [26.667, 31.592]\,.$$

Enostranski interval zaupanja določimo po enačbi (9.23), vrednost $t_{1-\alpha,13}$ pa izračunamo z računalniškim programom EXCEL: $t_{0.85,13} = 1.079$.

$$P\left[29.130 - 1.079 \cdot 4.852/\sqrt{14} \le m_X\right] = P\left[27.730 \le m_X\right] = 0.85 \quad \rightarrow \quad m_X \in [27.730, \infty)\,.$$

Z zanesljivostjo 0.92 lahko trdimo, da leži pričakovana vrednost trdnosti betona med 26.667 MPa in 31.592 MPa. Da je pričakovana vrednost trdnosti betona večja od 27.730 MPa, lahko trdimo s 85-odstotnim zaupanjem.

Preverimo še, kakšne rezultate bi dobili, če bi namesto Studentove porazdelitve vzeli kar normalno porazdelitev:

$$P\,[26.860 \le m_X \le 31.400] = 0.92 \quad \rightarrow \quad m_X \in [26.860, 31.400]\,,$$
$$P\,[27.786 \le m_X] = 0.85 \quad \rightarrow \quad m_X \in [27.786, \infty)\,.$$

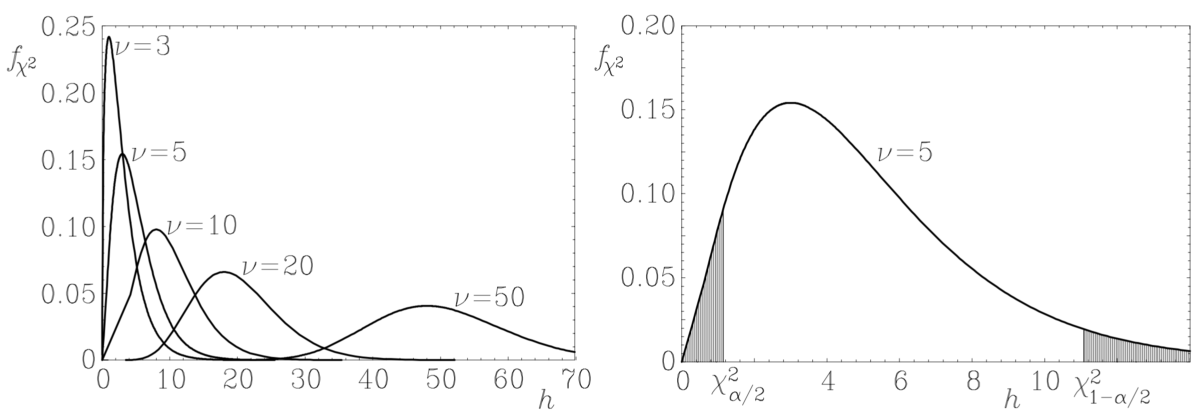
Vidimo, da razlike kljub majhnemu vzorcu $n = 14$ niso zelo velike. Razlog za to je predvsem v relativno nizki stopnji zaupanja, saj se vrednosti $k_{1-\alpha}$ od vrednosti $t_{1-\alpha}$ bolj razlikujejo pri nizkih vrednostih $\alpha$ .

### 9.2.2 Interval zaupanja za varianco

Pri določanju intervala zaupanja za varianco uporabimo oceno variance $\hat{\sigma}_X^2 = S_X^{*2}$, oziroma normirano statistiko $H$

$$H = \frac{(n-1)S_X^{*2}}{\sigma_X^2} \sim \chi^2_{\nu=n-1}, \tag{9.24}$$

ki je porazdeljena po **porazdelitvi $\boldsymbol{\chi^2}$** z $n-1$ prostostnimi stopnjami. Porazdelitev $\chi^2$ je definirana le za pozitivne vrednosti slučajne spremenljivke, je unimodalna in tem bolj asimetrična v desno, čim manjše je število prostostnih stopenj (slika 9.6). Ob večanju števila enot v vzorcu se porazdelitev $\chi^2$ vedno bolj približuje normalni porazdelitvi.




Slika 9.6: Gostota verjetnosti statistike $H$ in dvostranski interval zaupanja

Statistiko $H$ lahko uporabimo za določitev intervala zaupanja pri ocenjevanju populacijske variance $\sigma_X^2$. Verjetnost, da vrednost statistike $H$ leži znotraj intervala $(\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2})$ je

$$P\left[\chi^2_{\alpha/2} \le H \le \chi^2_{1-\alpha/2}\right] = 1-\alpha.$$

Če v zadnjo enačbo vstavimo enačbo (9.24), dobimo izraz za verjetnost, da varianca $\sigma_X^2$ leži znotraj intervala zaupanja

$$P\left[\frac{(n-1)S_X^{*2}}{\chi^2_{1-\alpha/2}} \le \sigma_X^2 \le \frac{(n-1)S_X^{*2}}{\chi^2_{\alpha/2}}\right] = 1-\alpha. \tag{9.25}$$

Interval zaupanja za varianco $\sigma_X^2$ je

$$\sigma_X^2 \in \left[\frac{(n-1)S_X^{*2}}{\chi^2_{1-\alpha/2}}, \frac{(n-1)S_X^{*2}}{\chi^2_{\alpha/2}}\right]. \tag{9.26}$$

Ker porazdelitev $\chi^2$ ni simetrična glede na koordinatno izhodišče, moramo v preglednici poiskati vrednosti za vsako stran porazdelitve posebej: $\chi^2_{1-\alpha/2}$ za desno stran in $\chi^2_{\alpha/2}$ za levo stran porazdelitve $\chi^2$.

Na podoben način kot pri določitvi intervala zaupanja za pričakovano vrednost lahko tudi pri intervalih zaupanja za varianco določimo enostranski interval zaupanja:

$$\begin{aligned} P\left[\chi^2_\alpha \leq H\right] = 1-\alpha \quad &\rightarrow \quad P\left[\sigma_X^2 \leq \frac{(n-1)S_X^{*2}}{\chi^2_\alpha}\right] = 1-\alpha \quad \rightarrow \\ &\rightarrow \quad \sigma_X^2 \in \left[0, \frac{(n-1)S_X^{*2}}{\chi^2_\alpha}\right]. \end{aligned} \tag{9.27}$$

in

$$\begin{aligned} P\left[H \leq \chi^2_{1-\alpha}\right] = 1-\alpha. \quad &\rightarrow \quad P\left[\frac{(n-1)S_X^{*2}}{\chi^2_{1-\alpha}} \leq \sigma_X^2\right] = 1-\alpha \quad \rightarrow \\ &\rightarrow \quad \sigma_X^2 \in \left[\frac{(n-1)S_X^{*2}}{\chi^2_{1-\alpha}}, \infty\right). \end{aligned} \tag{9.28}$$

**Primer 9.9:** *Vzemimo podatke iz primera 9.8 in določimo interval zaupanja za varianco $\sigma_X^2$. Stopnja zaupanja naj bo 95 %.*

**Rešitev**: Že pri reševanju primera 9.8 smo izračunali povprečje $\bar{X}$ in varianco vzorca $S_X^{*2}$

$$\bar{X} = 29.13, \qquad s_X^{*2} = 23.543.$$

Vrednosti $\chi^2_{1-\alpha}$ in $\chi^2_\alpha$ odčitamo iz preglednic oziroma z enim izmed računalniških programov EXCEL, MATHEMATICA..., in sta za 95-odstotono stopnjo zaupanja in število prostorstnih stopenj $\nu = n-1 = 13$ enaki

$$\chi^2_{1-\alpha,\nu} = \chi^2_{0.975,13} = 24.736, \qquad \chi^2_{\alpha,\nu} = \chi^2_{0.025,13} = 5.009.$$

Obojestranski interval zaupnja izračunamo po enačbi (9.26)

$$\sigma_X^2 \in [12.373, 61.105].$$

### 9.2.3 Interval zaupanja za razliko med pričakovanima vrednostima dveh populacij

Denimo, da imamo dve populaciji, ki ju predstavljata dve slučajni spremenljivki $X$ in $Y$. Predpostavimo najprej, da imata obe slučajni spremenljivki enako varianco $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Na voljo imamo vzorca obeh populacij $X_i, i = 1, \ldots, n_X$ in $Y_j, j = 1, \ldots, n_Y$. Povprečji $\bar{X}$ in $\bar{Y}$ sta ob predpostavki, da sta $X$

in $Y$ porazdeljeni normalno, ali je velikost obeh vzorcev zelo velika $n_X >> 1$ in $n_Y >> 1$, porazdeljeni normalno s pričakovanima vrednostima $m_X$ in $m_Y$ ter variancama $\sigma^2/n_X$ in $\sigma^2/n_Y$.

Ker variance $\sigma^2$ običajno ne poznamo, iz obeh vzorcev določimo statistiko $T$ po naslednji enačbi:

$$T = \frac{\bar{X} - \bar{Y} - (m_X - m_Y)}{S^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}, \tag{9.29}$$

ki je porazdeljena po Studentovi porazdelitvi $t$ z $\nu = n_X + n_Y - 2$ prostostnimi stopnjami. Oceno variance $\hat{\sigma}^2 = S^{*2}$ izračunamo takole:

$$S^{*2} = \frac{(n_X - 1)S_X^{*2} + (n_Y - 1)S_Y^{*2}}{n_X + n_Y - 2}.$$

Verjetnost, da statistika $T$ zavzame vrednost v območju $(-t_{1-\alpha/2}, t_{1-\alpha/2})$, je

$$P[-t_{1-\alpha/2} \leq T \leq t_{1-\alpha/2}] = 1 - \alpha. \tag{9.30}$$

Če v zadnjo enačbo vstavimo izraz (9.29), dobimo obojestranski interval zaupanja za razliko v pričakovanih vrednostih

$$\begin{aligned} & P\left[\bar{X} - \bar{Y} - t_{1-\alpha/2} S^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \leq m_X - m_Y \leq \bar{X} - \bar{Y} + t_{1-\alpha/2} S^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}\right] = 1 - \alpha \quad \rightarrow \\ & \rightarrow \quad m_X - m_Y \in \left[\bar{X} - \bar{Y} - t_{1-\alpha/2} S^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}, \bar{X} - \bar{Y} + t_{1-\alpha/2} S^* \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}\right] \end{aligned} \tag{9.31}$$

V primeru, da varianci $\sigma_X^2$ in $\sigma_Y^2$ nista enaki, moramo določiti statistiko

$$T = \frac{\bar{X} - \bar{Y} - (m_X - m_Y)}{\sqrt{\frac{S_X^{*2}}{n_X} + \frac{S_Y^{*2}}{n_Y}}}, \tag{9.32}$$

za katero pa porazdelitve ne poznamo v analitični obliki. Približno velja, da je statistika v enačbi (9.32) porazdeljena po Studentovi porazdelitvi $t$ z $\nu$ prostostnimi stopnjami, kjer $\nu$ izračunamo z naslednjo enačbo:

$$\nu = \frac{\left(\frac{S_X^{*2}}{n_X} + \frac{S_Y^{*2}}{n_Y}\right)^2}{\frac{1}{n_X - 1}\left(\frac{S_X^{*2}}{n_X}\right)^2 + \frac{1}{n_Y - 1}\left(\frac{S_Y^{*2}}{n_Y}\right)^2}. \tag{9.33}$$

V tem primeru interval zaupanja določimo tako, da v enačbo (9.30) vstavimo izraz (9.32)

$$\begin{aligned} & P\left[\bar{X} - \bar{Y} - t_{1-\alpha/2} \sqrt{\frac{S_X^{*2}}{n_X} + \frac{S_Y^{*2}}{n_Y}} \leq m_X - m_Y \leq \bar{X} - \bar{Y} + t_{1-\alpha/2} \sqrt{\frac{S_X^{*2}}{n_X} + \frac{S_Y^{*2}}{n_Y}}\right] = 1 - \alpha \quad \rightarrow \\ & \rightarrow \quad m_X - m_Y \in \left[\bar{X} - \bar{Y} - t_{1-\alpha/2} \sqrt{\frac{S_X^{*2}}{n_X} + \frac{S_Y^{*2}}{n_Y}}, \bar{X} - \bar{Y} + t_{1-\alpha/2} \sqrt{\frac{S_X^{*2}}{n_X} + \frac{S_Y^{*2}}{n_Y}}\right] \end{aligned}$$

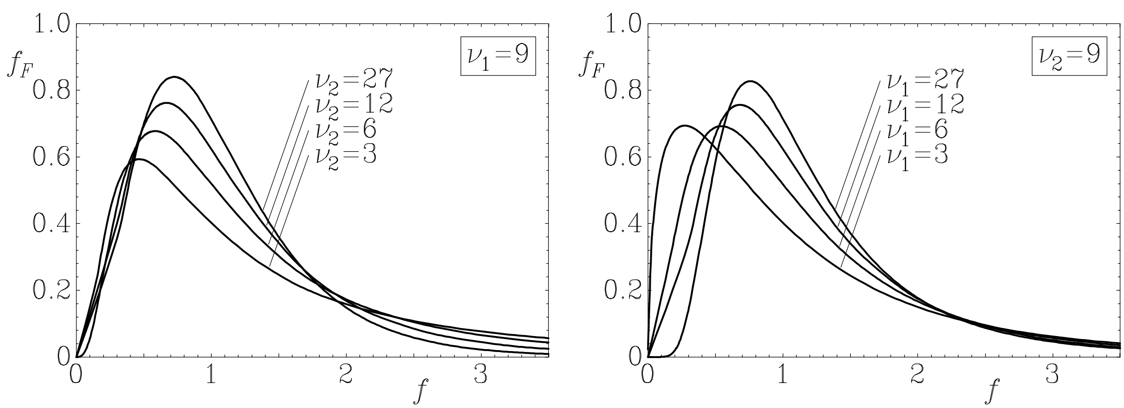
(9.34)

### 9.2.4 Interval zaupanja za razmerje varianc dveh populacij

Ponovno obravnavamo dve populaciji, ki ju predstavljata slučajni spremenljivki $X$ in $Y$, za kateri predpostavimo, da sta porazdeljeni normalno. Statistika $F$, ki jo določimo po naslednji enačbi

$$F = \frac{\frac{S_X^{*2}}{\sigma_X^2}}{\frac{S_Y^{*2}}{\sigma_Y^2}} \tag{9.35}$$

je porazdeljena po porazdelitvi $F$ z $\nu_X = n_X - 1$ in $\nu_Y = n_Y - 1$ prostostnimi stopnjami (slika 9.7). Porazdelitev $F$ po obliki spominja na porazdelitev $\chi^2$ in ima tudi isto zalogo vrednosti, ki vključuje vsa pozitivna realna števila.




Slika 9.7: Gostota verjetnosti statistike $F$ za različne vrednosti prostostnih stopenj $\nu_1$ in $\nu_2$

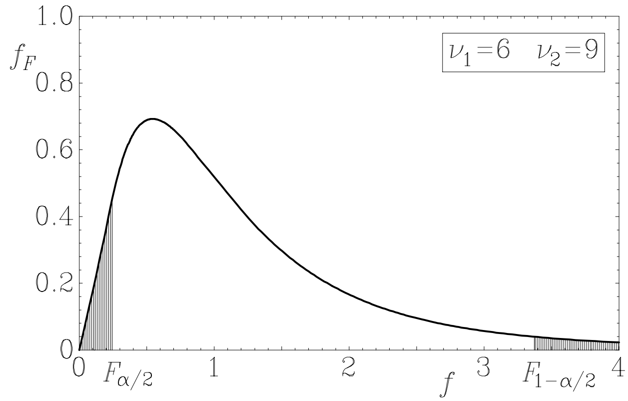
Verjetnost, da je statistika $F$ znotraj intervala $(F_{\alpha/2}, F_{1-\alpha/2})$, je enaka (slika 9.8)

$$P\left[F_{\alpha/2} \leq F \leq F_{1-\alpha/2}\right] = 1 - \alpha.$$

Če v zadnjo enačbo vstavimo enačbo (9.35), lahko določimo interval zaupanja za razmerje med variancama dveh populacij

$$P\left[\frac{1}{F_{1-\alpha/2}}\frac{S_X^{*2}}{S_Y^{*2}} \leq \frac{\sigma_X^2}{\sigma_Y^2} \leq \frac{1}{F_{\alpha/2}}\frac{S_X^{*2}}{S_Y^{*2}}\right] = 1 - \alpha \quad \rightarrow \quad \frac{\sigma_X^2}{\sigma_Y^2} \in \left[\frac{1}{F_{1-\alpha/2}}\frac{S_X^{*2}}{S_Y^{*2}}, \frac{1}{F_{\alpha/2}}\frac{S_X^{*2}}{S_Y^{*2}}\right]. \tag{9.36}$$

Vrednosti $F_{\alpha/2,\nu_1,\nu_2}$ in $F_{1-\alpha/2,\nu_1,\nu_2}$ lahko odčitamo iz preglednic ali pa izračunamo z računalniškim programom (na primer EXCEL). V programu Excel je vgrajena funkcija `finv(p;nu1;nu2)`, na primer za $\alpha = 0.05$, $\nu_1 = 6$ in $\nu_2 = 9$ je vrednost $F_{\alpha/2,\nu_1,\nu_2} = F_{0.025,6,9} = \text{finv}(0.975; 6; 9) = 0.1810$ in $F_{1-\alpha/2,\nu_1,\nu_2} = F_{0.975,6,9} = \text{finv}(0.025; 6; 9) = 4.320$. Bodite pozorni na pomen parametra $p$ v funkciji




Slika 9.8: Dvostranski intervali zaupanja za porazdelitev $F$

`finv(p;nu1;nu2)`! V preglednicah so običajno podane le vrednosti inverzne porazdelitvene funkcije na zgornjem repu, za vrednosti $1-\alpha$ oziroma $1-\alpha/2$. Če želimo določiti vrednosti še na spodnjem repu porazdelitve, moramo upoštevati zvezo

$$F_{\alpha,\nu_1,\nu_2} = \frac{1}{F_{1-\alpha,\nu_2,\nu_1}}. \tag{9.37}$$

Če želimo iz preglednic določiti vrednost $F_{0.025,6,9}$, najprej odčitamo vrednost $F_{0.975,9,6} = 5.523$ in nato izračunamo

$$F_{0.025,6,9} = \frac{1}{5.523} = 0.1810.$$

**Primer 9.10:** *Dva proizvajalca črpalk sta prodala po 20 črpalk istemu uporabniku. V naslednji preglednici so podatki o trajanju delovanja črpalk pred prvo okvaro. Določimo interval zaupanja za razliko pričakovanih vrednosti in razmerje v variancah dolžine delovanja črpalk dveh uporabnikov. Stopnja zaupanja je 95 %.*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 510 | 450 | 478 | 512 | 506 | 485 | 501 | 481 | 452 | 494 |
| | 514 | 507 | 487 | 467 | 502 | 508 | 503 | 492 | 502 | 499 |
| B | 510 | 513 | 497 | 506 | 493 | 501 | 547 | 514 | 487 | 490 |
| | 495 | 497 | 508 | 493 | 522 | 502 | 527 | 486 | 531 | 497 |

**Rešitev:** Najprej izračunamo povprečje in varianco obeh vzorcev:

$$\bar{X}_A = 492.5, \qquad S_A^{*2} = 353.4211,$$
$$\bar{X}_B = 505.8, \qquad S_B^{*2} = 257.6421.$$

Pri določitvi intervala zaupanja za razliko pričakovanih vrednosti moramo upoštevati, da sta varianci populacij različni in neznani. Interval zaupanja zato določimo z izrazom (9.34), pri čemer število prostostnih stopenj $\nu$ določimo po enačbi (9.33)

$$\nu = \frac{\left(\frac{S_A^{*2}}{n_A} + \frac{S_B^{*2}}{n_B}\right)^2}{\frac{1}{n_A-1}\left(\frac{S_A^{*2}}{n_A}\right)^2 + \frac{1}{n_B-1}\left(\frac{S_B^{*2}}{n_B}\right)^2} = 37.1,$$

kjer smo upoštevali, da je velikost obeh vzorcev enaka $n_A = n_B = 20$. Iz preglednic ali z računalnikom določimo vrednost $t_{1-\alpha/2,\nu} = t_{0.975,37.1} \approx t_{0.975,37} = 2.0262$. Interval zaupanja določimo z izrazom (9.34)

$$m_A - m_B \in [-24.4997, -2.1003]\,.$$

Interval zaupanja za razmerje med variancama populacij $\sigma_A^2/\sigma_B^2$ določimo z izrazom (9.36), pred tem pa iz preglednic ali z računalnikom določimo vrednosti $F_{\alpha/2,\nu_A,\nu_B} = F_{0.025,19,19} = 0.3958$ in $F_{1-\alpha/2,\nu_A,\nu_B} = F_{0.975,19,19} = 2.5264$. Interval zaupanja za $\sigma_A^2/\sigma_B^2$ je

$$\frac{\sigma_A^2}{\sigma_B^2} \in [0.5430, 3.4657]\,.$$

### 9.2.5 Interval zaupanja za delež v populaciji

Včasih želimo iz vzorčnih podatkov oceniti **delež elementov populacije**, ki ustrezajo nekemu kriteriju. Verjetnost, da slučajen element vzorca oziroma populacije ustreza kriteriju, je $p$. Točkovna ocena parametra $p$ je

$$\hat{p} = \frac{N_u}{n}, \tag{9.38}$$

kjer je $N_u$ število elementov vzorca, ki ustrezajo kriteriju, in $n$ velikost vzorca. Število $N_u$ je slučajna spremenljivka, saj je število elementov, ki ustrezajo določenemu kriteriju odvisno tudi od izbire vzorca. Slučajna spremenljivka $N_u$ je porazdeljena binomsko s parametroma $n$ in $p$ ter pričakovano vrednostjo $n\,p$ in varianco $n\,p\,(1-p)$. Pri obravnavanju normalne porazdelitve smo povedali, da je pri velikih $n$ normalna porazdelitev dobra aproksimacija binomske porazdelitve. Zato lahko tvorimo statistiko $U$

$$U = \frac{N_u - n\,p}{\sqrt{n\,p\,(1-p)}} = \frac{\hat{p} - p}{\sqrt{\frac{p\,(1-p)}{n}}}, \tag{9.39}$$

ki je porazdeljena približno standardizirano normalno. Na podoben način, kot smo določili interval zaupanja za pričakovano vrednost (glej enačbo (9.14) in sliko 9.3), zapišemo enačbo

$$P[-k_{1-\alpha/2} \le U \le k_{1-\alpha/2}] = 1 - \alpha,$$

v katero vstavimo enačbo (9.39)

$$P\left[-k_{1-\alpha/2} \le \frac{\hat{p}-p}{\sqrt{\dfrac{p\,(1-p)}{n}}} \le k_{1-\alpha/2}\right] = 1-\alpha. \tag{9.40}$$

Iz enačbe (9.40) lahko izpeljemo izraz za določitev intervala zaupanja za parameter $p$ tako, da izraz v oglatem oklepaju kvadriramo in zapišemo kvadratno neenačbo

$$\begin{aligned} &\frac{(\hat{p}-p)^2}{\dfrac{p\,(1-p)}{n}} \le k^2_{1-\alpha/2} \quad \rightarrow \\ &\rightarrow \quad \hat{p}^2 - 2\,\hat{p}\,p + p^2 - \frac{k^2_{1-\alpha/2}}{n}(p-p^2) \le 0 \quad \rightarrow \\ &\rightarrow \quad p^2\left(1+\frac{k^2_{1-\alpha/2}}{n}\right) - p\left(2\,\hat{p} + \frac{k^2_{1-\alpha/2}}{n}\right) + \hat{p}^2 \le 0 \end{aligned}$$

Ničli kvadratne funkcije v zadnji neenačbi sta:

$$\begin{aligned} p_{1,2} &= \frac{\left(2\,\hat{p} + \dfrac{k^2_{1-\alpha/2}}{n}\right) \pm \sqrt{\left(2\,\hat{p} + \dfrac{k^2_{1-\alpha/2}}{n}\right)^2 - 4\left(1+\dfrac{k^2_{1-\alpha/2}}{n}\right)\hat{p}^2}}{2\left(1+\dfrac{k^2_{1-\alpha/2}}{n}\right)} = \\ &= \frac{\hat{p} + \dfrac{k^2_{1-\alpha/2}}{2n} \pm k_{1-\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n} + \dfrac{k^2_{1-\alpha/2}}{4n^2}}}{1+\dfrac{k^2_{1-\alpha/2}}{n}}. \end{aligned}$$

Sedaj lahko zapišemo verjetnost in interval zaupanja za delež v populaciji:

$$\begin{aligned} P\Bigg[&\frac{\hat{p} + \dfrac{k^2_{1-\alpha/2}}{2n} - k_{1-\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n} + \dfrac{k^2_{1-\alpha/2}}{4n^2}}}{1+\dfrac{k^2_{1-\alpha/2}}{n}} \le p \le \\ &\le \frac{\hat{p} + \dfrac{k^2_{1-\alpha/2}}{2n} + k_{1-\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n} + \dfrac{k^2_{1-\alpha/2}}{4n^2}}}{1+\dfrac{k^2_{1-\alpha/2}}{n}}\Bigg] = 1-\alpha \end{aligned}$$

$$p \in \left[ \frac{\hat{p} + \frac{k_{1-\alpha/2}^2}{2n} - k_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{k_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{k_{1-\alpha/2}^2}{n}}, \frac{\hat{p} + \frac{k_{1-\alpha/2}^2}{2n} + k_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{k_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{k_{1-\alpha/2}^2}{n}} \right] \tag{9.41}$$

Če imamo **velik vzorec** opazovanj, se izraz (9.41) poenostavi. Upoštevamo, da je ocena $\hat{p}$ približno enaka $p$, zato pod korenom v enačbi (9.40) namesto $p$ zapišemo kar $\hat{p}$

$$P\left[-k_{1-\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}} \leq k_{1-\alpha/2}\right] =$$

$$=P\left[\hat{p} - k_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \leq p \leq \hat{p} + k_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right] = 1-\alpha,$$

Interval zaupanja za delež je tedaj

$$p \in \left[\hat{p} - k_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + k_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{9.42}$$

Ta poenostavljeni izraz lahko uporabimo namesto (9.41), če je $n$ dovolj velik in vrednost $p$ ni blizu nič ali ena.

**Primer 9.11: Volitve 2002.** *Vsakih nekaj let se v vsaki demokratični državi zgodijo volitve. Večinoma so uradni rezultati znani nekaj dni po volitvah, prvi neuradni delni rezultati volitev pa nekaj ur po zaprtju volišč. Nestrpni državljani pa bi radi izvedeli za rezultate takoj, ko se volišča zaprejo. Zato medijske hiše (Televizija Slovenija, Pop TV...) naročijo izdelavo vzporednih volitev, katerih rezultati so znani takoj po zaprtju volišč. Rezultati so pravzaprav znani že prej, vendar jih pred tem zaradi volilnega molka nihče ne sme objaviti.*

*Obravnavajmo rezultate vzporednih volitev drugega kroga predsedniških volitev leta 2002 v Sloveniji. Vzporedne volitve sta za obe največji televizijski hiši v Sloveniji delali dve podjetji:*
*CATI (http://www.cati.si/) in*
*Gral iteo (http://www.graliteo.si/).*
*Rezultati so povzeti v naslednji preglednici:*

| **Podjetje** | **Velikost vzorca** $n$ | **Rezultat vzporednih volitev** $\hat{p}$ (delež volilcev dr. Janeza Drnovška) |
|---|---|---|
| CATI | 9615 | 57.8 % |
| Gral iteo | 9311 | 57.7 % |

*Določimo interval zaupanja za delež populacije slovenskih volilcev, ki so volili za dr. Janeza Drnovška. Upoštevajmo 99-odstotno stopnjo zaupanja. Končni uradni rezultat, ki ga je objavila republiška volilna komisija je 56.54 %. Primerjajte končni rezultat z izračunanim intervalom zaupanja.*
**Rešitev:** Uporabimo izraza (9.41) in (9.42) za določitev intervala zaupanja. Najprej iz preglednic ali z računalnikom določimo vrednost $k_{1-\alpha/2} = k_{0.995} = 2.5758$. V naslednji preglednici zapišemo rezultate za interval zaupanja za obe podjetji in oba izraza:

| **Podjetje** | **Interval zaupanja (9.41)** | **Interval zaupanja (9.42)** |
|---|---|---|
| CATI | [0.56498,0.59092] | [0.56503,0.59097] |
| Gral iteo | [0.56376,0.59013] | [0.56381,0.59019] |

Vidimo, da so razlike v intervalu zaupanja pri uporabi izrazov (9.41) ali (9.42) neznatne, kar bi lahko pričakovali, saj je vzorec relativno velik, vrednost $p$ pa ni zelo blizu nič ali ena. Opazimo lahko tudi, da sta obe podjetji dobili zelo podobna rezultata, kar ni presenetljivo, saj sta uporabili podobno metodologijo vzorčenja in podobno velikost vzorca.

Končni uradni rezultat volitev v vseh primerih leži znotraj izračunanega intervala zaupanja, čeprav je dober odstotek nižji od rezultatov vzporednih volitev.

**Primer 9.12:** *V neki tovarni izdelujejo izdelke. Verjetnost, da je poljubno izbran izdelek pokvarjen, je enaka $p$. Vzemimo, da smo preizkusili sto izdelkov, od katerih so bili trije pokvarjeni. Določimo interval zaupanja za delež pokvarjenih izdelkov. Uporabimo izraza (9.41) in (9.42) ter primerjajmo rezultate. Stopnja zaupanja naj bo 95-odstotna.*
**Rešitev:** Vrednost $k_{1-\alpha/2} = k_{0.975} = 1.960$ odčitamo iz preglednice ali določimo z računalnikom. Točkovna ocena deleža pokvarjenih izdelkov je

$$\hat{p} = \frac{3}{100} = 0.03.$$

Interval zaupanja po točnem izrazu (9.41) je

$$p \in [0.01025, 0.08452].$$

Interval zaupanja po približnem izrazu (9.42) pa je

$$p \in [-0.00343, 0.06343].$$

Tokrat so razlike občutnejše. Spodnja meja intervala zaupanja po približnem izrazu je celo negativna, kar je seveda napačno, saj vrednost $p$ mora ležati med 0 in 1. V tem primeru vidimo, da je bolje, da uporabimo točni izraz (9.41).

### 9.2.6 Določitev velikosti vzorca

Velikokrat nastopi problem, ko moramo iz populacije izbrati vzorec elementov. Koliko elementov bomo vključili v vzorec, je odvisno od:

- variance populacije (večja kot je varianca, večji vzorec potrebujemo za isto natančnost rezultatov),
- vrste statističnega izračuna (kateri parameter populacije obravnavamo).

V primeru, da bomo z vzorcem ocenjevali pričakovano vrednost populacije, potem izračunamo velikost vzorca po enačbi:

$$n > \left(\frac{k_{1-\alpha/2}\,\sigma_X}{E}\right)^2,$$

kjer je $\sigma_X$ znana oziroma predpostavljena standardna deviacija populacije, $E$ pa polovica največje dovoljene velikosti intervala zaupanja.

V primeru, da želimo z vzorcem oceniti delež populacije, pa velikost vzorca izračunamo takole:

$$n > \frac{k_{1-\alpha/2}^2\,p(1-p)}{E^2},$$

kjer je $p$ znan oziroma predpostavljen delež populacije, $E$ pa polovica največje dovoljene velikosti intervala zaupanja. Pri zadnji enačbi neenačbi smo izhajali iz približnega obrazca (9.42).

**Primer 9.13:** *V neki tovarni izdelujejo izdelke. Verjetnost, da je poljubno izbran izdelek pokvarjen, je enaka $p$. Vodstvo tovarne želi določiti 95-odstotni enostranski interval zaupanja za delež slabih izdelkov*

$$p \in [p_s, 1] \textit{ tako, da velja } P[p_s \leq p] = 1 - \alpha = 0.95.$$

*Pri tem so postavili zahtevo, da meja območja zaupanja ne sme odstopati od ocenjenega deleža $\hat{p}$ za več kot 1 %. Določimo najmanjše velikosti vzorcev, pri katerih bo izpolnjen zahtevani pogoj za vrednosti $\hat{p} = 0.10$, $0.05$ in $0.02$.*

**Rešitev:** Rešitev za uporabo približnega izraza (9.42) je preprosta:

$$P[p_s \leq p] = 1 - \alpha \quad \rightarrow \quad P\left[\hat{p} - k_{1-\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p\right] = 1 - \alpha.$$

Vrednost $k_{1-\alpha} = k_{0.95} = 1.645$ določimo iz preglednica ali računalnikom. Pri postavljenem pogoju o največji dovoljeni širini intervala zaupanja,

$$\hat{p} - \left(\hat{p} - k_{1-\alpha}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right) = k_{1-\alpha}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} \leq E = 0.01,$$

lahko zapišemo pogoj za zahtevano velikost vzorca

$$n \geq \left(\frac{k_{1-\alpha}}{E}\right)^2 \hat{p}\,(1-\hat{p}). \tag{9.43}$$

Če želimo uporabiti točni izraz (9.41), je problem bistveno zahtevnejši. Verjetnost, da $p$ leži znotraj intervala zapišemo takole:

$$P[p_s \leq p] = P\left[\frac{\hat{p} + \dfrac{k_{1-\alpha}^2}{2n} - k_{1-\alpha}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n} + \dfrac{k_{1-\alpha}^2}{4n^2}}}{1 + \dfrac{k_{1-\alpha}^2}{n}} \leq p\right] = 1 - \alpha.$$

Največja dovoljena razlika med mejo intervala in točkovno oceno je

$$\hat{p} - \frac{\hat{p} + \dfrac{k_{1-\alpha}^2}{2n} - k_{1-\alpha}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n} + \dfrac{k_{1-\alpha}^2}{4n^2}}}{1 + \dfrac{k_{1-\alpha}^2}{n}} \leq E = 0.01. \tag{9.44}$$

Iskano vrednost velikosti vzorca lahko določimo numerično (v programu EXCEL z ukazom `Goal Seek`, v programu MATHEMATICA pa z ukazom `FindRoot`. Rezultate za velikost vzorca za različne vrednosti $\hat{p}$ po obeh izrazih za interval zaupanja podajamo v naslednji preglednici:

| $\hat{p}$ | 0.02 | 0.05 | 0.10 |
|---|---|---|---|
| $n$ (9.43) | $\geq 531$ | $\geq 1286$ | $\geq 2435$ |
| $n$ (9.44) | $\geq 268$ | $\geq 1039$ | $\geq 2216$ |

Razlike med obema načinoma določanja velikosti vzorca so precejšnje. To pomeni, da bi v tem primeru morali uporabiti točen izraz (9.41).

# 10 Preizkušanje domnev

**Statistična domneva** oziroma **hipoteza** je vsaka domneva o slučajni spremenljivki $X$. Razlikujemo parametrične in neparametrične domneve. **Parametrična domneva** je domneva o vrednosti nekega parametra porazdelitve. **Neparametrična domneva** je domneva o neki neparametrični lastnosti (tip porazdelitve, neodvisnost ...) porazdelitve slučajne spremenljivke $X$ oziroma slučajnega vektorja $\boldsymbol{X}$ $\{x_1, x_2, \ldots\}$.

**Ničelna domneva** $H_0$ je domneva, ki jo v danih okoliščinah želimo preizkusiti. **Alternativne domneve** $H_i (i = 1, 2, ..., k)$ so domneve, ki so nezdružljive z ničelno.

Primeri domnev so:

- parametrični domnevi dvostranskega testa:
  $H_0$: $m_X = 15$
  $H_1$: $m_X \neq 15$;

- parametrični domnevi enostranskega testa:
  $H_0$: $\sigma_X = 16$
  $H_1$: $\sigma_X > 16$;

- neparametrični domnevi:
  $H_0$: porazdelitev je normalna
  $H_1$: porazdelitev ni normalna;

- neparametrični domnevi:
  $H_0$: kocka je poštena
  $H_1$: kocka ni poštena.

**Preizkus domneve** ali **test** je vsak postopek, po katerem lahko na temelju vzorca slučajne spremenljivke $X(x_1, x_2, ..., x_n)$ ničelno domnevo zavrnemo ali ne.

Domnevo preizkusimo po naslednjem postopku:

1. Postavimo ničelno in alternativno domnevo o parametru:
   $H_0$: ničelna domneva,
   $H_1$: alternativna domneva.

2. Izberemo statistiko, ki ustreza ničelni domnevi, in določimo njeno porazdelitev.

3. Izberemo tveganje oziroma stopnjo značilnosti $\alpha$. Na osnovi tveganja in porazdelitve statistike določimo meje kritičnega območja.

4. Na vzorčnih podatkih izračunamo vrednost statistike.

5. Sklep:
   - Če vrednost statistike pade v kritično območje (območje zavrnitve ničelne domneve), ničelno domnevo zavrnemo in sprejmemo alternativno domnevo ob tveganju $\alpha$.
   - Če vrednost statistike ne pade v kritično območje, ničelne domneve ne moremo zavrniti ob tveganju $\alpha$.

## 10.1 Preizkušanje domneve o pričakovani vrednosti

Na voljo imamo vzorec $X_i, i = 1, \ldots, n$ z $n$ elementi. Iz vzorca lahko izračunamo povprečje $\bar{X}$ in varianco vzorca $S_X^{*2}$ po enačbah (8.2) in (8.10).

Postavimo ničelno in alternativno domnevo

$H_0$: $m_X = m_0$,
$H_1$: $m_X \neq m_0$,

Če je spremenljivka $X$ porazdeljena normalno $N(m_X, \sigma_X)$ z **znano** standardno deviacijo $\sigma_X$ in neznano pričakovano vrednostjo $m_X$ in velja ničelna domneva $H_0$: $m_X = m_0$, potem je statistika

$$U = \frac{\bar{X} - m_0}{\sigma_X / \sqrt{n}} \tag{10.1}$$

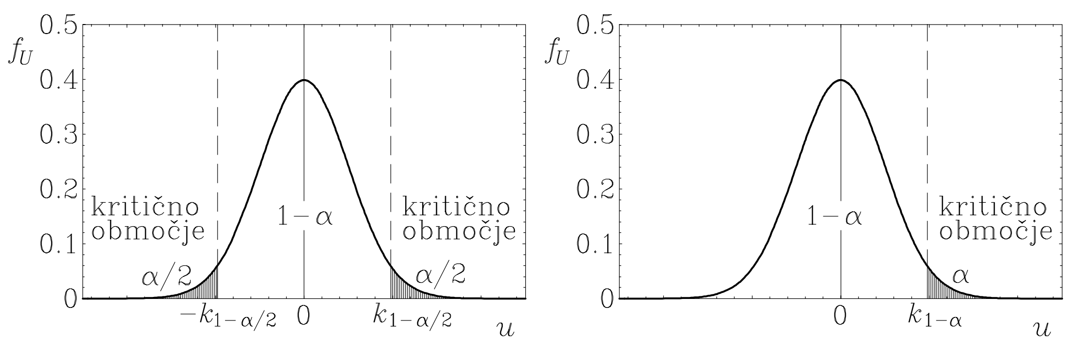
porazdeljena standardizirano normalno ($U \sim N(0, 1)$), kjer je $\bar{X}$ povprečje vzorca in $n$ velikost vzorca (enačbi (8.3) in (8.4) v poglavju o lastnostih osnovnih statistik).

Ob predpostavki, da velja ničelna domneva, je statistika $U$ torej slučajna spremenljivka, ki je porazdeljena standardizirano normalno (slika 10.1). Pri ponavljanju vzorčenja enako velikih vzorcev bi dobili različne vrednosti $U$. Pri običajnem testiranju imamo na voljo le en vzorec, na osnovi katerega moramo odločiti, ali ničelno domnevo zavrnemo ali ne. Kritično območje oziroma območje zavrnitve je: $(-\infty, -k_{1-\alpha/2}] \cup [k_{1-\alpha/2}, \infty)$ (slika 10.1). Če se vrednost statistike zelo razlikuje od 0 in pade v kritično območje, sta možna dva dogodka:

- ničelna domneva velja, vzorec je bil slučajno tako izbran, da je vrednost statistike padla v kritično območje,
- ničelna domneva ne velja.

Pri statističnem preizkušanju se v tem primeru vedno odločimo za drugo možnost in zaključimo, da ničelna domneva ne velja, temveč velja alternativna. Pri tem lahko storimo napako $I$. vrste: zavrnitev ničelne domneve, čeprav ta velja. Tveganje, da storimo napako $I$. vrste, je enako tveganju $\alpha$.

Če vrednost statistike $U$ ne pade v kritično območje, ničelne domneve ne bomo zavrnili. Pomembno je, da v tem primeru **ničelne domneve ne sprejmemo**. Če bi ničelno domnevo sprejeli, bi lahko storili napako $II$. vrste: sprejem ničelne domneve, ki ne velja. Verjetnost, da storimo napake $II$. vrste, označimo z $\beta$. Ta verjetnost je odvisna od dejanske vrednosti $m_X$, ki je seveda ne poznamo. Poleg tega lahko vedno izberemo tako majhno tveganje $\alpha$, da statistika ne pade v kritično območje. Zato bomo pri vseh naših preizkušanjih domnev v primerih, ko statistika ne pade v kritično območje zaključili le, da **ničelne domneve ne moremo zavrniti**. Razlika med izjavama *"Ničelno domnevo sprejmemo"* in *"Ničelne domneve ne moremo zavrniti"* se morda zdi majhna, a je izredno pomembna. Morda razliko med tema dvema izjavama še bolje ilustriramo s podobnima izjavama iz pravniškega življenja: *"Obtoženec je kriv"* in *"Za obtoženca ne moremo trditi, da je nedolžen"*. Bi obtoženca v obeh primerih obsodili?




Slika 10.1: Porazdelitev statistike $U$ in kritična območja za dvostranski in enostranski test

Če je spremenljivka $X$ porazdeljena normalno $N(m_X, \sigma_X)$ z neznano standardno deviacijo $\sigma_X$ in **neznano** pričakovano vrednostjo $m_X$, pri tem preizkusu domneve uporabimo statistiko

$$T = \frac{\bar{X} - m_0}{S_X^*/\sqrt{n}}, \tag{10.2}$$

ki je porazdeljena po porazdelitvi $t$ z $\nu = n - 1$ prostostnimi stopnjami.

Kritično območje oziroma območje zavrnitve je v tem primeru: $(-\infty, -t_{1-\alpha/2}] \cup [t_{1-\alpha/2}, \infty)$. Če statistika $T$ zavzame vrednost manjšo od $-t_{1-\alpha/2}$ ali pa večjo od $t_{1-\alpha/2}$, ničelno domnevo zavrnemo in s stopnjo značilnosti $\alpha$ trdimo, da velja alternativna domneva.

**Primer 10.1:** *V laboratoriju preizkušajo trdnost betona. Rezultate meritev (v* MPa*) prikazujemo v naslednji preglednici 10.1.*

*Preizkusimo domnevo, da je pričakovana vrednost trdnosti betona enaka* $30$. *Ugotoviti moramo, ali je pričakovana vrednost statistično značilno različna od* $30$. *Tveganje naj bo enako* $5\,\%$.

Preglednica 10.1: Rezultati merjenja trdnosti betona

| | | | | | | |
|---|---|---|---|---|---|---|
| 28.329 | 28.386 | 31.803 | 29.173 | 31.973 | 23.323 | 33.612 |
| 26.409 | 30.306 | 24.643 | 23.788 | 21.926 | 36.715 | 37.433 |

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: $m_X = 30\,\mathrm{MPa}$,
$H_1$: $m_X \neq 30\,\mathrm{MPa}$.

Statistika, ki ustreza ničelni domnevi

$$T = \frac{\bar{X} - 30}{S_X^*/\sqrt{n}},$$

je porazdeljena po Studentovi porazdelitvi z $\nu = n - 1 = 13$ prostostnimi stopnjami.

Kritično območje oziroma območje zavrnitve ničelne domneve je $(-\infty, -t_{1-\alpha/2}] \cup [t_{1-\alpha/2}, \infty)$. Vrednost $t_{1-\alpha/2,\nu} = t_{0.975,13} = 2.160$. Če bo statistika $T$ večja od $2.160$ ali manjša od $-2.160$, bomo ničelno domnevo zavrnili.

Za izračun statistike moramo najprej izračunati povprečje in varianco oziroma standardni odklon vzorca:

$$\bar{X} = \frac{\sum_{i=1}^{14} x_i}{14} = \frac{28.329 + 28.386 + 31.803 + \cdots}{14} = 29.130,$$
$$S_X^{*2} = \frac{\sum_{i=1}^{14} (x_i - \bar{X})^2}{13} = \frac{(28.329 - 29.130)^2 + (28.386 - 29.130)^2 + \cdots}{13} = 23.543,$$
$$S_X^* = \sqrt{S_X^{*2}} = 4.852.$$

Vrednost statistike $T$ je

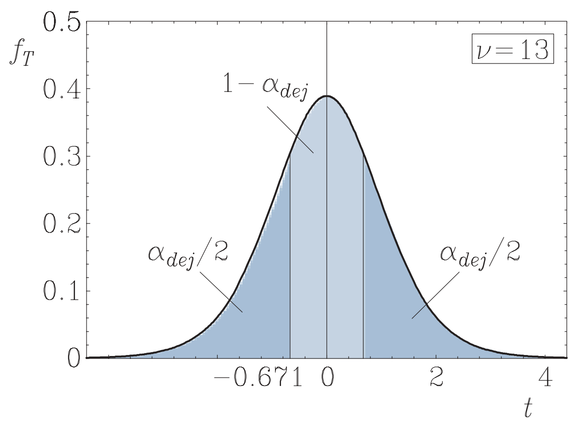
$$T = \frac{29.130 - 30}{4.852/\sqrt{14}} = -0.671.$$

Ker statistika $T$ ni v kritičnem območju, ničelne domneve ne moremo zavrniti. Pričakovana vrednost torej ni statistično značilno različna od 30 MPa. Seveda pa s tem ne želimo trditi, da je pričakovana vrednost enaka 30 MPa.

Včasih želimo določiti dejansko tveganje za zavrnitev ničelne domneve, to je verjetnost $\alpha_{dej}$, da smo ob zavrnitvi ničelne domneve naredili napako I. vrste. Iz slike 10.2 vidimo, da je $\alpha_{dej}$ enaka

$$\begin{aligned}\alpha_{dej} &= P[T \leq -0.671 \cup T \geq 0.671] = 2\,P[T \leq -0.671] = 2\,P[T \geq 0.671] = \\ &= 2\,(1 - P[T \leq 0.671] = 2\,(1 - F_T(0.671)) = 2 \cdot 0.257 = 0.514,\end{aligned}$$

pri čemer smo upoštevali simetričnost Studentove porazdelitve. Dejansko tveganje, da storimo napako I. vrste in zavrnemo ničelno domnevo, je v tem primeru enako 51.4 %, kar je seveda več od predpisanega tveganja 5 %. Zato ničelne domneve ne zavrnemo. Pričakovana vrednost trdnosti betona ni statistično značilno različna od 30.




Slika 10.2: Določitev dejanskega tveganja $\alpha_{dej}$

**Primer 10.2:** *Uporabimo iste podatke kot v prejšnjem primeru 10.1. Tokrat preizkusimo ničelno domnevo, da je pričakovana vrednost trdnosti betona enaka* 25*. Alternativna domneva naj tokrat pravi, da je pričakovana vrednost večja od* 25*. Tveganje naj bo enako* $1\,\%$*.*

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: $m_X = 25\,\text{MPa}$,
$H_1$: $m_X > 25\,\text{MPa}$.

Statistika, ki ustreza ničelni domnevi

$$T = \frac{\bar{X} - 25}{S_X^*/\sqrt{n}},$$

je porazdeljena po Studentovi porazdelitvi z $\nu = n - 1 = 13$ prostostnimi stopnjami.

Kritično območje oziroma območje zavrnitve ničelne domneve je $[t_{1-\alpha}, \infty)$. Vrednost $t_{1-\alpha,\nu} = t_{0.99,13} =$ 2.650. Če bo statistika $T$ večja od 2.650, bomo ničelno domnevo zavrnili.

Povprečje in standardni odklon vzorca (glej prejšnji primer) sta

$$\bar{X} = 29.130, \qquad S_X^* = 4.852.$$

Vrednost statistike $T$ je

$$T = \frac{29.130 - 25}{4.852/\sqrt{14}} = 3.1849.$$

Ker statistika $T$ leži v kritičnem območju, ničelno domnevo zavrnemo. Pričakovana vrednost je statistično značilno večja od 25 MPa.

## 10.2 Preizkušanje domneve o razliki med pričakovanima vrednostima dveh populacij

Naši podatki sta dva vzorca: vzorec slučajne spremenljivke $X$, $X_i, i = 1, \ldots, n_X$, in vzorec slučajne spremenljivke $Y$, $Y_j, j = 1, \ldots, n_Y$. Velikosti vzorcev $n_X$ in $n_Y$ sta lahko različni. Iz vzorcev lahko izračunamo povprečja $\bar{X}$ in $\bar{Y}$ ter varianci $S_X^{*2}$ in $S_Y^{*2}$ po enačbah (8.2) in (8.10).

Če se slučajni spremenljivki $X$ in $Y$ porazdeljujeta normalno $\mathrm{N}(m_X, \sigma_X)$ in $\mathrm{N}(m_Y, \sigma_Y)$ z **znanima** standardnima deviacijama $\sigma_X$ in $\sigma_Y$ ter neznanima pričakovanima vrednostma $m_X$ in $m_Y$, lahko izvedemo preizkus domnev o razliki pričakovanih vrednosti. Postavimo ničelno in alternativno domnevo:

$H_0$: $m_X - m_Y = 0$ oziroma $m_X = m_Y$,
$H_1$: $m_X - m_Y \neq 0$ oziroma $m_X \neq m_Y$.

V tem primeru tvorimo statistiko $U$

$$U = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}, \tag{10.3}$$

ki je porazdeljena standardizirano normalno. Kritično območje oziroma območje zavrnitve je: $(-\infty, -k_{1-\alpha/2}] \cup [k_{1-\alpha/2}, \infty)$. Če statistika $U$ zavzame vrednost manjšo od $-k_{1-\alpha/2}$ ali pa večjo od $k_{1-\alpha/2}$, ničelno domnevo zavrnemo in s stopnjo značilnosti $\alpha$ trdimo, da velja alternativna domneva.

V primeru, da se slučajni spremenljivki $X$ in $Y$ porazdeljujeta normalno $\mathrm{N}(m_X, \sigma_X)$ in $\mathrm{N}(m_Y, \sigma_Y)$ z **neznano**, vendar **enako**, standardno deviacijo $\sigma_X = \sigma_Y = \sigma$ ter neznanima pričakovanima vrednostima $m_X$ in $m_Y$, moramo določiti drugo statistiko

$$T = \frac{\bar{X} - \bar{Y}}{S^* \sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}}, \tag{10.4}$$

ki je porazdeljena po porazdelitvi $t$ z $\nu = n_X + n_Y - 2$ prostostnimi stopnjami. Oceno variance $\sigma^2$ izračunamo po naslednji enačbi

$$S^{*2} = \frac{(n_X - 1)S_X^{*2} + (n_Y - 1)S_Y^{*2}}{n_X + n_Y - 2}. \tag{10.5}$$

V primeru, ko neznani standardni deviaciji nista enaki, pa se statistika

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\dfrac{S_X^{*2}}{n_X} + \dfrac{S_Y^{*2}}{n_Y}}} \tag{10.6}$$

porazdeljuje približno po porazdelitvi $t$ z $\nu$ prostostnimi stopnjami. Število prostostnih stopenj v tem primeru izračunamo z enačbo

$$\nu = \frac{\left(\dfrac{S_X^{*2}}{n_X} + \dfrac{S_Y^{*2}}{n_Y}\right)^2}{\dfrac{(S_X^{*2}/n_X)^2}{n_X - 1} + \dfrac{(S_Y^{*2}/n_Y)^2}{n_Y - 1}}. \tag{10.7}$$

Kritično območje oziroma območje zavrnitve je v tem primeru: $(-\infty, -t_{1-\alpha/2}] \cup [t_{1-\alpha/2}, \infty)$. Če statistika $T$ zavzame vrednost manjšo od $-t_{1-\alpha/2}$ ali pa večjo od $t_{1-\alpha/2}$, ničelno domnevo zavrnemo in s stopnjo značilnosti $\alpha$ trdimo, da velja alternativna domneva.

**Primer 10.3:** *Dva proizvajalca črpalk sta prodala po 20 črpalk istemu uporabniku. V naslednji preglednici so podatki o trajanju delovanja črpalk pred prvo okvaro.*

Preglednica 10.2: Trajanje delovanja črpalk [dni]

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 510 | 450 | 478 | 512 | 506 | 485 | 501 | 481 | 452 | 494 |
| | 514 | 507 | 487 | 467 | 502 | 508 | 503 | 492 | 502 | 499 |
| B | 510 | 513 | 497 | 506 | 493 | 501 | 547 | 514 | 487 | 490 |
| | 495 | 497 | 508 | 493 | 522 | 502 | 527 | 486 | 531 | 497 |

*Ugotovite, ali lahko s 5 % tveganjem trdite, da je pričakovana vrednost trajanja delovanja črpalk prvega proizvajalca nižja od pričakovane vrednosti trajanja delovanja črpalk drugega.*

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: $m_{X_1} = m_{X_2}$,
$H_1$: $m_{X_1} < m_{X_2}$.

Statistika, ki ustreza ničelni domnevi

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{S_{X_1}^{*2}}{n_1} + \dfrac{S_{X_2}^{*2}}{n_2}}},$$

je porazdeljena po Studentovi porazdelitvi z

$$\nu = \frac{\left(\dfrac{S_{X_1}^{*2}}{n_1} + \dfrac{S_{X_2}^{*2}}{n_2}\right)^2}{\dfrac{(S_{X_1}^{*2}/n_1)^2}{n_1 - 1} + \dfrac{(S_{X_2}^{*2}/n_2)^2}{n_2 - 1}}.$$

prostostnimi stopnjami, kjer sta velikosti vzorca za oba proizvajalca enaki $n_1 = n_2 = 20$, povprečji in standardni deviaciji vzorcev pa sta:

$$\bar{X}_1 = 492.5\,\text{dni}, \qquad S_{X_1}^{*} = 18.80\,\text{dni},$$
$$\bar{X}_2 = 505.8\,\text{dni}, \qquad S_{X_1}^{*} = 16.05\,\text{dni}.$$

Kritično območje oziroma območje zavrnitve ničelne domneve je $(-\infty, -t_{1-\alpha,\nu}]$. Število prostostnih stopenj $\nu = 37.09$, meja kritičnega območja pa je $t_{0.95,37.09} = 1.687$.

Vrednost statistike $T$ je

$$T = \frac{492.5 - 505.8}{\sqrt{\dfrac{18.80^2}{20} + \dfrac{16.05^2}{20}}} = -2.406.$$

Ker vrednost statistike $T = -2.406$ leži v kritičnem območju $(-\infty, -1.687]$, moramo ničelno domnevo zavrniti in trdimo, da je pričakovana vrednost trajanja delovanja črpalk prvega proizvajalca statistično značilno krajša od pričakovane vrednosti trajanja delovanja črpalk drugega proizvajalca.

V prikazani rešitvi smo predpostavili, da sta varianci obeh populacij različni. Če bi predpostavili, da sta varianci enaki $\sigma_{X_1} = \sigma_{X_2}$, bi statistiko namesto po enačbah (10.6–10.7) izračunali po enačbah (10.4–10.5).

Število prostostnih stopenj za račun meje kritičnega območja je $\nu = n_1 + n_2 - 2 = 38$, kritično območje je $(-\infty, -1.686]$. Varianco obeh vzorcev $S^{*2}$ izračunamo po enačbi (10.5)

$$S^{*2} = \frac{19 \cdot 18.80^2 + 19 \cdot 16.05^2}{20 + 20 - 2} = 305.53 \qquad \rightarrow \qquad S^{*} = 17.48.$$

Statistika $T$ je sedaj enaka

$$T = \frac{492.5 - 505.8}{17.48\sqrt{\frac{1}{20} + \frac{1}{20}}} = -2.345.$$

Tudi v tem primeru vrednost statistike leži v kritičnem območju, zato ničelno domnevo zavrnemo. Trdimo lahko, da je pričakovana vrednost trajanja delovanja črpalk prvega proizvajalca statistično značilno krajša od pričakovane vrednosti trajanja delovanja črpalk drugega proizvajalca.

Vidimo, da smo v obeh primerih prišli do istega zaključka. Razliko opazimo le, če določimo dejansko tveganje ob zavrnitvi ničelne domneve. Ob predpostavki o enakih variancah, je dejansko tveganje $\alpha_{dej} = 0.0122$, če pa predpostavimo, da sta varianci različni, je tveganje nekoliko nižje $\alpha_{dej} = 0.0106$.

## 10.3 Preizkušanje domneve o varianci

Za preizkus domneve o varianci $\sigma_X^2$ potrebujemo vzorec $X_i, i = 1, \ldots, n$ slučajne spremenljivke $X$, iz katerega lahko izračunamo varianco vzorca $S_X^{*2}$ po enačbi (8.10).

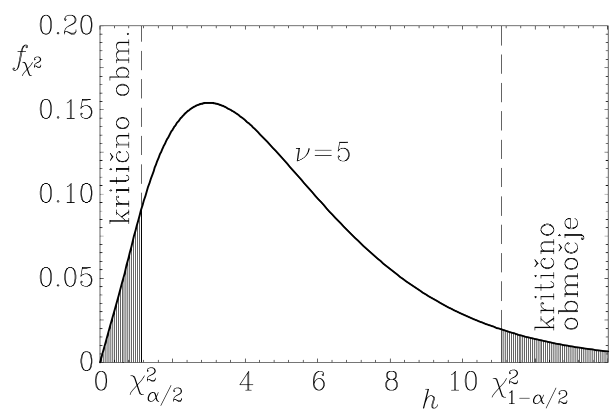
Denimo, da se spremenljivka $X$ porazdeljuje normalno N$(m_X, \sigma_X)$. Postavimo ničelno in alternativno domnevo

$H_0$: $\sigma_X = \sigma_0$,
$H_1$: $\sigma_X \neq \sigma_0$.

Statistika

$$H = \frac{(n-1)S_X^{*2}}{\sigma_0^2} \tag{10.8}$$

je porazdeljena po porazdelitvi $\chi^2$ z $\nu = n - 1$ prostostnimi stopnjami.




Slika 10.3: Porazdelitev statistike $H$ in kritični območji za dvostranski test

Na sliki 10.3 lahko vidimo, da je kritično območje $[0, \chi^2_{\alpha/2}] \cup [\chi^2_{1-\alpha/2}, \infty)$. Če je torej statistika $H$ manjša od $\chi^2_{\alpha/2}$ ali večja od $\chi^2_{1-\alpha/2}$, ničelno domnevo zavrnemo in s stopnjo značilnosti $\alpha$ trdimo, da velja alternativna.

**Primer 10.4:** *Uporabimo iste podatke kot v primeru 10.1. Preizkusimo domnevo, da je varianca trdnosti betona enaka* $25$. *Ugotoviti moramo, ali je pričakovana vrednost statistično značilno različna od* $25$. *Tveganje naj bo enako* $5\,\%$.

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: $\sigma_X^2 = 25$,
$H_1$: $\sigma_X^2 \neq 25$.

Statistika, ki ustreza ničelni domnevi

$$H = \frac{(n-1)S_X^{*2}}{\sigma_X^2},$$

je porazdeljena po porazdelitvi $\chi^2$ z $\nu = n - 1 = 13$ prostostnimi stopnjami.

Kritično območje oziroma območje zavrnitve ničelne domneve je $[0, \chi^2_{\alpha/2}] \cup [\chi^2_{1-\alpha/2}, \infty)$. Meji kritičnega območja sta $\chi^2_{0.025,13} = 5.009$ in $\chi^2_{0.975,13} = 24.736$. V primeru 10.1 smo že izračunali varianco vzorca:

$$S_X^{*2} = 23.543,$$

zato lahko izračunamo statistiko $H$

$$H = \frac{13 \cdot 23.543}{25} = 12.24.$$

Ker vrednost statistike ne leži v kritičnem območju, ničelne domneve ne moremo zavrniti. Varianca ni statistično značilno različna od 25.

## 10.4 Preizkušanje domneve o razmerju med variancama dveh populacij

Podobno kot pri preizkusu domneve o razliki med pričakovanima vrednostima, imamo tudi tukaj na voljo dva vzorca $X_i, i = 1, \ldots, n_X$ in $Y_j, j = 1, \ldots, n_Y$ dveh slučajnih spremenljivk $X$ in $Y$. Iz vzorcev lahko izračunamo vzorčni varianci $S_X^{*2}$ in $S_Y^{*2}$ po enačbi (8.10).

Pri preizkušanju domnev o razliki v srednjih vrednostih dveh populacij včasih predpostavimo, da sta varianci obeh populacij enaki. To domnevo lahko preizkusimo s **testom homogenosti**, ki ga pogosto imenujemo tudi **test *F***. Denimo, da se dve slučajni spremenljivki $X$ in $Y$ porazdeljujeta normalno $\mathrm{N}(m_X, \sigma_X)$ in $\mathrm{N}(m_Y, \sigma_Y)$ z neznanima standardnima deviacijama $\sigma_X$ in $\sigma_Y$, potem se v **porazdelitvi *F*** (tudi **Fisher-Snedecorjevi porazdelitvi**) z $\nu_X = n_X - 1$ in $\nu_Y = n_Y - 1$ prostostnimi stopnjami porazdeljuje izraz

$$F = \frac{S_X^{*2}/\sigma_X^2}{S_Y^{*2}/\sigma_Y^2} \qquad (10.9)$$

kjer sta $S_X^{*2}$ in $S_Y^{*2}$ varianci vzorcev.

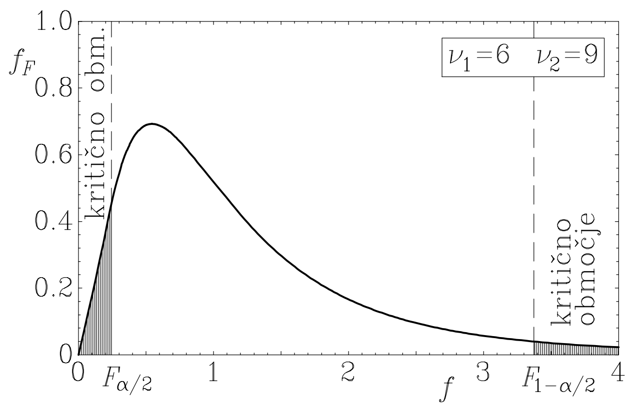
Običajno postavimo ničelno in alternativno domnevo takole:

$H_0$: $\sigma_X^2 = \sigma_Y^2$,
$H_1$: $\sigma_X^2 \neq \sigma_Y^2$.

V tem primeru je testna statistika $F$

$$F = \frac{S_X^{*2}}{S_Y^{*2}} \tag{10.10}$$

porazdeljena po porazdelitvi $F$ z $\nu_X = n_X - 1$ in $\nu_Y = n_Y - 1$.




Slika 10.4: Porazdelitev statistike $F$ in kritični območji za dvostranski test

Na sliki 10.4 lahko vidimo, da je kritično območje $[0, F_{\alpha/2}] \cup [F_{1-\alpha/2}, \infty)$. Če je torej statistika $F$ manjša od $F_{\alpha/2}$ ali večja od $F_{1-\alpha/2}$, ničelno domnevo zavrnemo in s stopnjo značilnosti $\alpha$ trdimo, da velja alternativna.

**Primer 10.5:** *Uporabimo iste podatke kot v primeru 10.3. Preizkusimo domnevo, da sta varianci trajanja delovanja črpalk dveh proizvajalcev enaki. Ugotoviti moramo, ali je varianca trajanja delovanja črpalk prvega proizvajalca statistično značilno večja od trajanja delovanja črpalk drugega proizvajalca. Tveganje naj bo enako $5\,\%$.*

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: $\sigma_{X_1}^2 = \sigma_{X_2}^2$,
$H_1$: $\sigma_{X_1}^2 > \sigma_{X_2}^2$.

Statistika, ki ustreza ničelni domnevi

$$F = \frac{S_{X_1}^{*2}}{S_{X_2}^{*2}},$$

je porazdeljena po porazdelitvi $F$ z $\nu_1 = n_1 - 1 = 19$ in $\nu_2 = n_2 - 1 = 19$ prostostnimi stopnjami.

Kritično območje oziroma območje zavrnitve ničelne domneve je $[F_{1-\alpha}, \infty)$. Meja kritičnega območja je $F_{0.95,19,19} = 2.168$. V primeru 10.3 smo že izračunali varianci vzorca:

$$S^{*2}_{X_1} = 353.421, \qquad S^{*2}_{X_2} = 257.642,$$

zato lahko izračunamo statistiko $F$

$$F = \frac{353.421}{257.642} = 1.372.$$

Ker vrednost statistike ne leži v kritičnem območju, ničelne domneve ne moremo zavrniti. Varianca trajanja delovanja črpalk prvega proizvajalca ni statistično značilno večja od variance trajanja delovanja črpalk drugega.

## 10.5 Preizkušanje domneve o deležu v populaciji

Tu preizkušamo domnevo o verjetnosti $p$, da populacija ustreza določenemu pogoju. Z ničelno domnevo prepostavimo, da je verjetnost $p$, da populacija ustreza določenemu pogoju, enaka neki vrednosti $p_0$. V tem primeru je vzorec predstavljen z rezultati $n$ Bernoullijevih poskusov, s katerimi ugotavljamo, ali elementi vzorca oziroma populacije ustrezajo izbranemu pogoju. Število *uspešnih poskusov*, pri katerih je bil izpolnjen pogoj, označimo z $N_u$. Točkovno oceno $\hat{p}$ parametra $p$ izračunamo po enačbi (9.38).

Postavimo ničelno in alternativno domnevo:

$H_0$: $p = p_0$,
$H_1$: $p \neq p_0$.

Ob predpostavki, da ničelna domneva drži, lahko zapišemo statistiko $U$ (enačba (9.39))

$$U = \frac{\hat{p} - p}{\sqrt{\frac{p\,(1-p)}{n}}},$$

ki se porazdeljuje približno po standardizirani normalni porazdelitvi. Kritično območje oziroma območje zavrnitve ničelne domneve je $(-\infty, -k_{1-\alpha/2}]$ ali $[k_{1-\alpha/2}, \infty)$. Če je torej statistika $U < -k_{1-\alpha/2}$ ali $U > k_{1-\alpha/2}$, ničelno domnevo zavrnemo in trdimo, da je $p$ statistično značilno različen od $p_0$.

**Primer 10.6: Volitve 2002.** *Obravnavajmo predsedniške volitve iz leta 2002, ki smo jih obravnavali že v primeru 9.11. Tokrat preizkusimo ničelno domnevo, da je rezultat volitev $p = 0.5$. Seveda je v tem primeru bolj smiselno, da opravimo enostranski test, saj nas zanima, ali je $p > 0.5$, torej ali kandidat, ki je na vzporednih volitvah dobil več glasov lahko zagotovo upa tudi na končno zmago. Tokrat vzemimo le rezultate enega podjetja, ki je izvajalo vzporedne volitve (na primer CATI: velikost vzorca: $n = 9615$ in $\hat{p} = 0.578$. Ugotovimo, ali lahko ob tveganju $\alpha = 0.001$ zavrnemo ničelno domnevo, da je $p = 0.5$.*

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: $p = 0.5$,

$H_1$: $p > 0.5$.

Statistika je v tem primeru enaka

$$U = \frac{\hat{p} - p}{\sqrt{\frac{p\,(1-p)}{n}}} = \frac{0.578 - 0.5}{\sqrt{\frac{0.5\,(1-0.5)}{9615}}} = 15.297.$$

Kritično območje za ta enostranski test pa je $[k_{1-\alpha}, \infty)$. Meja kritičnega območja je $k_{1-\alpha} = 3.090$. Ker je vrednost statistike $U = 15.297$ bistveno večja od kritične vrednosti, lahko ničelno domnevo zavrnemo in zaključimo, da je $p$ statistično značilno večji od $0.5$, torej da lahko z zelo veliko zanesljivostjo trdimo, da je zmagal kandidat, ki je na vzporednih volitvah dobil več glasov (na teh volitvah je bil to dr. Drnovšek). Dejansko tveganje je v tem primeru izredno majhno: $\alpha_{dej} = 4.0 \cdot 10^{-53}$.

## 10.6 Preizkušanje skladnosti – test $\chi^2$

Ta test lahko uporabimo pri preizkušanju zelo različnih domnev. Rezultate meritev oziroma vzorec $(X_j, j = 1, \ldots, n)$ razvrstimo v $k$ razredov. Tako dobimo opazovana števila elementov v posameznem razredu $\hat{n}_i$, $i = 1, \ldots, k$. Ob predpostavki, da velja ničelna domneva, lahko določimo teoretično število elementov v posameznem razredu $n_i$, $i = 1, \ldots, k$.

$H_0$: Vzorec je skladen s predpostavko,
$H_1$: Vzorec ni skladen s predpostavko.

Statistika

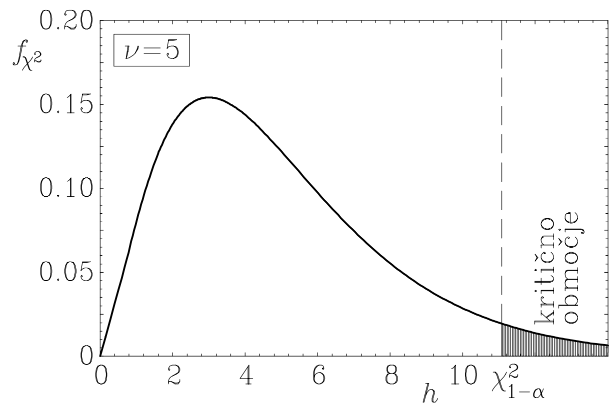
$$H = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{n_i} \tag{10.11}$$

se porazdeljuje po porazdelitvi $\chi^2$ z $\nu = k - p - 1$ prostostnimi stopnjami, kjer je $p$ število parametrov, ki smo jih ocenili iz vzorca. Test $\chi^2$ najpogosteje uporabljamo pri preizkušanju, ali vzorec ustreza predpostavljeni porazdelitvi, pa tudi pri preizkušanju statistične odvisnosti.

Območje zavrnitve ničelne domneve je $[\chi^2_{1-\alpha}, \infty)$ (slika 10.5). Če je statistika $H$ večja od $\chi^2_{1-\alpha}$, ničelno domnevo zavrnemo in trdimo, da vzorec ni skladen s predpostavko.
Primeri ničelnih domnev pri preizkušanju domnev o porazdelitvi populacije:

- $H_0$: Slučajna spremenljivka $X$ je porazdeljena normalno,
  $H_1$: Slučajna spremenljivka $X$ ni porazdeljena normalno.

  Vzorec $X_j$, $j = 1, \ldots, n$ razvrstimo v $k$ razredov. Iz vzorca izračunamo oceni za pričakovano vrednost $\hat{m}_X = \bar{X}$ in standardno deviacijo $\hat{\sigma}_X = S^*_X$, torej je število parametrov, ki smo jih




Slika 10.5: Porazdelitev statistike $H$ pri testu skladnosti

ocenili iz vzorca enako $p = 2$. Glede na izbrane meje razredov določimo teoretične velikosti razredov po naslednji enačbi:

$$n_i = n\,P[x_{i,min} < X < x_{i,max}]. \tag{10.12}$$

Statistiko $H$ izračunamo po enačbi (10.11) in jo primerjamo z mejno vrednostjo $\chi^2_{1-\alpha,\nu}$, ki jo določimo iz preglednic ali z računalnikom.

- $H_0$: Slučajna spremenljivka $X$ je porazdeljena normalno ($m_X = 12$ in $\sigma_X = 3$),
  $H_1$: Slučajna spremenljivka $X$ ni porazdeljena normalno ($m_X = 12$ in $\sigma_X = 3$).

  Ker sta parametra normalne porazdelitve že podana v ničelni domnevi, ju iz vzorca ni potrebno oceniti. Zato je število parametrov, ki jih ocenimo iz vzorca, enako nič $p = 0$. Sicer je postopek preizkušanja enak, kot v prejšnem primeru.

- $H_0$: Igralna kocka je poštena (slučajna spremenljivka $X$ je porazdeljena enakomerno od 1 do 6),
  $H_1$: Igralna kocka ni poštena (slučajna spremenljivka $X$ ni porazdeljena enakomerno od 1 do 6).

  Slučajna spremenljivka $X$ predstavlja rezultat meta kocke. V tem primeru rezultatov ne razvrstimo v razrede, ampak vsi možni rezultati predstavljajo po en razred. Opazovane velikosti razredov $\hat{n}_i$ so števila posameznih zadetkov (na primer: $\hat{n}_1$ pove, kolikokrat je padla enica). Teoretična števila elementov v razredih so enaka:

  $$n_i = \frac{n}{6}, \qquad i = 1, \ldots, 6.$$

  kjer je $n$ število metov kocke.

**Primer 10.7:** *Proizvajalec tovornjakov trdi, da je življenska doba njegovih tovornjakov normalno porazdeljena s pričakovano vrednostjo* $m_X = 350\,000\,\mathrm{km}$ *in standardno deviacijo* $\sigma_X = 70\,000\,\mathrm{km}$. *Podatke o življenski dobi 40 tovornjakov tega proizvajalca prikazujemo v preglednici 10.3.*

Preglednica 10.3: Življenska doba tovornjakov v 1000 km

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 329 | 261 | 367 | 439 | 434 | 471 | 197 | 334 | 427 | 274 |
| 302 | 232 | 221 | 282 | 296 | 202 | 310 | 322 | 359 | 324 |
| 327 | 324 | 444 | 344 | 337 | 314 | 488 | 411 | 516 | 304 |
| 466 | 237 | 388 | 413 | 484 | 344 | 313 | 397 | 323 | 403 |

*Preizkusite ničelno domnevo, ki pravi, da je življenska doba tovornjakov porazdeljena normalno s pričakovano vrednostjo* $m_X = 350\,000\,\mathrm{km}$ *in standardno deviacijo* $\sigma_X = 70\,000\,\mathrm{km}$. *Tveganje naj bo enako* $5\,\%$.

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: Slučajna spremenljivka $X$ je porazdeljena normalno $(m_X = 350, \sigma_X = 70)$,
$H_1$: Slučajna spremenljivka $X$ ni porazdeljena tako.

---

Najprej moramo določiti število razredov, v katere razporedimo elemente vzorca iz preglednice 10.3. Za to pogosto uporabimo **Sturgesovo formulo**[1]

$$k = 1 + 3.32 \log n,$$

kjer je $n$ velikost vzorca, $k$ pa je število razredov. V tem primeru, ko je velikost vzorca $n = 40$, dobimo

$$k = 1 + 3.32 \log 40 = 6.3 \approx 6.$$

Meje razredov lahko določimo na različne načine. V tem primeru se odločimo za štiri enako široke razrede, najnižji in najvišji razred pa sta ob predpostavki o normalni porazdelitvi neskončno široka. Meje med razredi in razporeditev podatkov iz preglednice 10.3 prikazujemo v naslednji preglednici 10.4.

Preglednica 10.4: Meje razredov in razporeditev podatkov v razrede

| Razred $i$ | Meje razredov | Opazovane velikosti razredov $\hat{n}_i$ | Teoretične velikosti razredov $n_i$ |
|---|---|---|---|
| 1 | $(-\infty, 250]$ | 5 | 3.06 |
| 2 | $(250, 300]$ | 4 | 6.44 |
| 3 | $(300, 350]$ | 15 | 10.50 |
| 4 | $(350, 400]$ | 4 | 10.50 |
| 5 | $(400, 450]$ | 7 | 6.44 |
| 6 | $(450, \infty)$ | 5 | 3.06 |

Teoretične velikosti razredov določimo po enačbi (10.12). Pokažimo izračun teoretične velikosti razre-

[1] H. A. Sturges, The Choise of a Class Interval, J. Amer. Statist. Assoc., Vol. 21, 65–66, 1926.

dov za prva dva razreda.

$$n_1 = n\,P[X \le 250] = n\,F_X(250) = n\,F_U\left(\frac{250-350}{70}\right) = n\,F_U(-1.4286) = 3.06,$$
$$n_2 = n\,P[250 < X \le 300] = n\,(F_X(300) - F_X(250)) =$$
$$= n\,(F_U(-0.7143) - F_U(-1.4286)) = 6.44.$$

Teoretične velikosti razredov za vse razrede prikazujemo v preglednici 10.4. Statistiko $H$ izračunamo po enačbi (10.11)

$$H = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{n_i} = \frac{(3.06-5)^2}{3.06} + \frac{(6.44-4)^2}{6.44} + \cdots = 9.376.$$

Območje zavrnitve ničelne domneve je $[\chi^2_{1-\alpha,\nu}, \infty)$. V našem primeru je število prostostnih stopenj $\nu = k - 1 = 5$, saj iz vzorca nismo določali ocen parametrov populacije. Meja kritičnega območja je $\chi^2_{0.95,5} = 11.07$.

Ker statistika ne pade v kritično območje, ničelne hipoteze ne moremo zavrniti. Ne moremo trditi, da porazdelitev ni normalna s pričakovano vrednostjo $m_X = 350\,000\,\mathrm{km}$ in standardno deviacijo $\sigma_X = 70\,000\,\mathrm{km}$.

Dejansko tveganje je $\alpha_{dej} = 0.095$.

**Primer 10.8:** *Preizkusite ničelno domnevo, da je življenska doba tovornjakov porazdeljena normalno. Uporabimo podatke prejšnjega primera 10.7. Tveganje naj bo enako* $5\,\%$.

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: Slučajna spremenljivka $X$ je porazdeljena normalno,
$H_1$: Slučajna spremenljivka $X$ ni porazdeljena normalno.

Tokrat v ničelni domnevi ni trditve o pričakovani vrednosti in standardni deviaciji. Zato moramo njuni vrednosti oceniti iz vzorca. Povprečje in standardna deviacija vzorca sta

$$\hat{m}_X = \bar{X} = 349.02, \qquad \hat{\sigma}_X = S_X^* = 81.71.$$

Teoretične velikosti razredov moramo izračunati glede na iz vzorca ocenjene vrednosti pričakovane vrednosti in standardne deviacije (glej preglednico 10.5)
Statistiko $H$ izračunamo po enačbi (10.11)

$$H = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{n_i} = \frac{(4.51-5)^2}{4.51} + \frac{(6.46-4)^2}{6.46} + \cdots = 7.691.$$

Območje zavrnitve ničelne domneve je $[\chi^2_{1-\alpha,\nu}, \infty)$. V tem primeru je število prostostnih stopenj $\nu = k - p - 1 = 6 - 2 - 1 = 3$, saj smo tokrat iz vzorca ocenili dva parametra porazdelitve ($m_X$ in $\sigma_X$), Preglednica 10.5: Meje razredov in razporeditev podatkov v razrede

| Razred $i$ | Meje razredov | Opazovane velikosti razredov $\hat{n}_i$ | Teoretične velikosti razredov $n_i$ |
|---|---|---|---|
| 1 | $(-\infty, 250]$ | 5 | 4.51 |
| 2 | $(250, 300]$ | 4 | 6.46 |
| 3 | $(300, 350]$ | 15 | 9.22 |
| 4 | $(350, 400]$ | 4 | 9.16 |
| 5 | $(400, 450]$ | 7 | 6.32 |
| 6 | $(450, \infty)$ | 5 | 4.33 |

zato moramo vzeti, da je $p = 2$. Meja kritičnega območja je $\chi^2_{0.95,3} = 7.815$.

Ker statistika ne pade v kritično območje, ničelne hipoteze ne moremo zavrniti. Ne moremo trditi, da porazdelitev ni normalna. Dejansko tveganje ob zavrnitvi ničelne domneve je $\alpha_{dej} = 0.0528$, torej le nekoliko višje od izbranega tveganja $\alpha = 0.05$.

Zanimiva je primerjava dejanskega tveganja pri zadnjih dveh primerih. Vidimo, da je dejansko tveganje v primeru, kjer v ničelni domnevi predpostavimo tudi pričakovano vrednost in standardno deviacijo, bistveno višje. Razlog za tako razliko je v bistvenem zmanjšanju števila prostostnih stopenj v primeru, ko smo morali iz vzorca določiti oceni pričakovane vrednosti in standardne deviacije. Rešitev problema bi bilo povečanje števila razredov, ki pa bi ga v primeru tako majhnega vzorca težko izvedli. Ena izmed rešitev je tudi uporaba testa Kolmogorova in Smirnova, ki ga bomo opisali v nadaljevanju.

**Primer 10.9:** *Ugotoviti želimo, ali je igralna kocka goljufiva.*

*Igralno kocko vržemo 240-krat in beležimo rezultate, ki jih prikazujemo v preglednici 10.6. Tveganje naj bo enako $5\,\%$.*

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: Igralna kocka je poštena (slučajna spremenljivka je porazdeljena enakomerno),
$H_1$: Igralna kocka ni poštena.

---

V tem primeru imamo šest razredov, vsak različen rezultat pri metu kocke razvrstimo v svoj razred. Ob predpostavki, da ničelna domneva velja, so teoretične velikosti vseh razredov enake

$$n_i = n\,P[X = x_i] = n\,\frac{1}{6} = 40.$$

Rezultate metov kocke, podane v preglednici 10.6, razvrstimo v šest razredov. Opazovane in teoretične velikosti razredov prikazujemo v preglednici 10.7.

Preglednica 10.6: Rezultati 240 metov sumljive igralne kocke

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 6 | 1 | 4 | 4 | 6 | 3 | 4 | 1 | 6 | 2 | 3 | 6 | 5 | 6 | 3 | 2 | 3 | 5 | 6 | 1 | 2 | 6 | 3 |
| 6 | 3 | 3 | 6 | 1 | 5 | 2 | 6 | 6 | 2 | 6 | 4 | 6 | 3 | 2 | 6 | 1 | 2 | 2 | 2 | 6 | 2 | 4 | 3 |
| 1 | 4 | 3 | 1 | 4 | 2 | 4 | 5 | 4 | 1 | 5 | 5 | 2 | 3 | 6 | 1 | 4 | 5 | 4 | 1 | 3 | 3 | 3 | 2 |
| 5 | 5 | 4 | 1 | 2 | 6 | 4 | 3 | 5 | 1 | 2 | 3 | 1 | 2 | 1 | 2 | 4 | 1 | 5 | 3 | 2 | 3 | 2 | 5 |
| 2 | 6 | 6 | 6 | 3 | 6 | 2 | 6 | 6 | 4 | 1 | 6 | 6 | 2 | 1 | 3 | 5 | 3 | 5 | 6 | 5 | 6 | 3 | 4 |
| 5 | 6 | 3 | 6 | 3 | 5 | 1 | 1 | 3 | 4 | 1 | 5 | 6 | 3 | 2 | 2 | 6 | 2 | 5 | 6 | 4 | 6 | 6 | 6 |
| 6 | 2 | 4 | 4 | 1 | 4 | 6 | 6 | 1 | 5 | 5 | 4 | 6 | 2 | 3 | 3 | 2 | 6 | 2 | 2 | 6 | 3 | 2 | 5 |
| 1 | 2 | 5 | 4 | 3 | 6 | 5 | 5 | 2 | 1 | 3 | 3 | 2 | 6 | 2 | 5 | 3 | 5 | 2 | 4 | 3 | 4 | 1 | 1 |
| 6 | 5 | 1 | 2 | 2 | 2 | 6 | 6 | 4 | 6 | 3 | 3 | 3 | 5 | 2 | 1 | 2 | 1 | 3 | 3 | 3 | 6 | 6 | 5 |
| 3 | 5 | 6 | 3 | 6 | 1 | 2 | 4 | 4 | 6 | 5 | 4 | 6 | 5 | 3 | 4 | 6 | 6 | 2 | 1 | 6 | 6 | 5 | 4 |

Statistiko $H$ izračunamo po enačbi (10.11)

$$H = \sum_{i=1}^{k} \frac{(n_i - \hat{n}_i)^2}{n_i} = \frac{(40-31)^2}{40} + \frac{(40-43)^2}{40} + \cdots = 13.5.$$

Območje zavrnitve ničelne domneve je $[\chi^2_{1-\alpha,\nu}, \infty)$. V našem primeru je število prostostnih stopenj $\nu = k - 1 = 5$. Meja kritičnega območja je $\chi^2_{0.95,5} = 11.07$.

Preglednica 10.7: Velikosti razredov pri preizkušanju igralne kocke

| Rezultat $i$ | Opazovane velikosti razredov $\hat{n}_i$ | Teoretične velikosti razredov $n_i$ |
|---|---|---|
| 1 | 31 | 40 |
| 2 | 43 | 40 |
| 3 | 43 | 40 |
| 4 | 31 | 40 |
| 5 | 34 | 40 |
| 6 | 58 | 40 |

Ker vrednost statistike $H = 13.5$ pade v kritično območje $[11.07, \infty)$, moramo ničelno domnevo zavrniti. S petodstotnim tveganjem lahko trdimo, da kocka ni poštena.

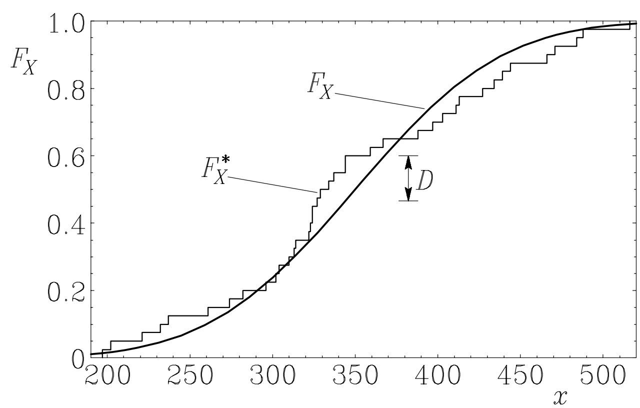
(Dejansko so bili rezultati meta kocke, prikazani v preglednici 10.6, generirani z računalnikom. Predpostavljeno je bilo, da so verjetnosti, da pade 1, 2, 3, 4 ali 5 enake 0.15, medtem ko je verjetnost, da pade 6 enaka 0.25. Ta kocka je torej zares goljufiva.)

## 10.7 Preizkus Kolmogorova in Smirnova

Cilj tega preizkusa je enak kot pri preizkusu $\chi^2$: Ugotoviti želimo, ali lahko na podlagi vzorca zavrnemo ničelno domnevo o porazdelitvi populacije. Zato je ničelna domneva enaka, kot pri testu $\chi^2$:

$H_0$: Slučajna spremenljivka $X$ je porazdeljena po določeni porazdelitvi,
$H_1$: Slučajna spremenljivka $X$ ni porazdeljena tako.

Test Kolmogorova in Smirnova temelji na porazdelitveni funkciji $F_X(x)$ in empirični porazdelitveni funkciji $F_X^*(x_j)$.




Slika 10.6: Test Kolmogorova in Smirnova

Vrednosti elementov vzorca $x_j$, $j = 1, \ldots, n$ razvrstimo po velikosti od najmanjše vrednosti proti največji in tvorimo empirično porazdelitveno funkcijo

$$F_X^*(x_j) = j/n, \tag{10.13}$$

kjer $j$ predstavlja zaporedno številko elementa v razvrščenem vzorcu. Graf empirične porazdelitvene funkcije je stopničast, kot kaže tudi slika (10.6).

Statistiko $D$ določimo tako, da poiščemo največjo absolutno razliko med porazdelitveno funkcijo $F_X(x_j)$ in $F_X^*(x_j)$

$$D_n = \max_{j=1}^{n} |F_X(x_j) - F_X^*(x_j)| \,. \tag{10.14}$$

Če je statistika $D_n$ večja od mejne vrednosti pri določenem tveganju $\alpha$, ki jo odčitamo iz preglednic ali pa jo izračunamo z računalnikom, moramo ničelno domnevo zavrniti in lahko s tveganjem $\alpha$ trdimo, da porazdelitev ni taka, kot pravi ničelna domneva. Vrednost $n$ je enaka velikosti vzorca, torej številu elementov vzorca.

V primerjavi s testom $\chi^2$ ima test Kolmogorova in Smirnova bistveno prednost, da elemente vzorca ni potrebno razvrščati v razrede. Število in širine razredov lahko namreč bistveno vplivajo na rezultat testa $\chi^2$. Zato je tudi potrebna velikost vzorca $n$ za učinkovito uporabo tega testa manjša kot v primeru testa $\chi^2$.

Pomanjkljivost tega testa je, da ga lahko zanesljivo uporabimo le za zvezne porazdelitve, katerih parametre poznamo. Če moramo parametre predpostavljene porazdelitve oceniti iz vzorca, test ne da točne ocene tveganja pri zavrnitvi ničelne domneve. Kljub temu ta test uporabljamo tudi za diskretne porazdelitve in tudi v primeru, da moramo parametre porazdelitve izračunati iz vzorca. Nekatere pomanjkljivosti testa Kolmogorova in Smirnova odpravlja test Anderson-Darling, ki pa obstaja le za določene porazdelitve.

**Primer 10.10:** *S preizkusom Kolmogorova in Smirnova preizkusite ničelno domnevo, ki pravi, da je življenska doba tovornjakov porazdeljena normalno s pričakovano vrednostjo* $m_X = 350\,000\,\mathrm{km}$ *in standardno deviacijo* $\sigma_X = 70\,000\,\mathrm{km}$.

*Uporabimo podatke primera 10.7. Tveganje naj bo enako* $5\,\%$.

**Rešitev**: Postavimo ničelno in alternativno domnevo:

$H_0$: Slučajna spremenljivka $X$ je porazdeljena normalno $(m_X = 350, \sigma_X = 70)$,
$H_1$: Slučajna spremenljivka $X$ ni porazdeljena tako.

---

Elemente vzorca, prikazanega v preglednici 10.3, razvrstimo po velikosti. Sedaj lahko določimo vrednosti empirične porazdelitvene funkcije $F_X^*$ po enačbi (10.13) ter vrednost statistike $D$ po enačbi (10.14).

Preglednica 10.8: Določitev statistike $D$ v preizkusu Kolmogorova in Smirnova

| $i$ | $x_i$ | $F_X^*(x_i)$ | $F_X(x_i)$ | $\Delta F_X$ | $i$ | $x_i$ | $F_X^*(x_i)$ | $F_X(x_i)$ | $\Delta F_X$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 197 | 0.025 | 0.0144 | 0.0106 | 21 | 334 | 0.525 | 0.4096 | 0.1154 |
| 2 | 202 | 0.050 | 0.0172 | 0.0328 | 22 | 337 | 0.550 | 0.4263 | 0.1237 |
| 3 | 221 | 0.075 | 0.0327 | 0.0423 | 23 | 344 | 0.575 | 0.4658 | 0.1092 |
| 4 | 232 | 0.100 | 0.0459 | 0.0541 | 24 | 344 | 0.600 | 0.4658 | 0.1342 |
| 5 | 237 | 0.125 | 0.0532 | 0.0718 | 25 | 359 | 0.625 | 0.5512 | 0.0738 |
| 6 | 261 | 0.150 | 0.1018 | 0.0482 | 26 | 367 | 0.650 | 0.5959 | 0.0541 |
| 7 | 274 | 0.175 | 0.1388 | 0.0362 | 27 | 388 | 0.675 | 0.7064 | 0.0314 |
| 8 | 282 | 0.200 | 0.1657 | 0.0343 | 28 | 397 | 0.700 | 0.7490 | 0.0490 |
| 9 | 296 | 0.225 | 0.2202 | 0.0048 | 29 | 403 | 0.725 | 0.7755 | 0.0505 |
| 10 | 302 | 0.250 | 0.2464 | 0.0036 | 30 | 411 | 0.750 | 0.8082 | 0.0582 |
| 11 | 304 | 0.275 | 0.2555 | 0.0195 | 31 | 413 | 0.775 | 0.8159 | 0.0409 |
| 12 | 310 | 0.300 | 0.2839 | 0.0161 | 32 | 427 | 0.800 | 0.8643 | 0.0643 |
| 13 | 313 | 0.325 | 0.2986 | 0.0264 | 33 | 434 | 0.825 | 0.8849 | 0.0599 |
| 14 | 314 | 0.350 | 0.3035 | 0.0465 | 34 | 439 | 0.850 | 0.8982 | 0.0482 |
| 15 | 322 | 0.375 | 0.3446 | 0.0304 | 35 | 444 | 0.875 | 0.9103 | 0.0353 |
| 16 | 323 | 0.400 | 0.3499 | 0.0501 | 36 | 466 | 0.900 | 0.9513 | 0.0513 |
| 17 | 324 | 0.425 | 0.3552 | 0.0698 | 37 | 471 | 0.925 | 0.9581 | 0.0331 |
| 18 | 324 | 0.450 | 0.3552 | 0.0948 | 38 | 484 | 0.950 | 0.9722 | 0.0222 |
| 19 | 327 | 0.475 | 0.3712 | 0.1038 | 39 | 488 | 0.975 | 0.9757 | 0.0007 |
| 20 | 329 | 0.500 | 0.3821 | 0.1179 | 40 | 516 | 1.000 | 0.9911 | 0.0089 |

Z $\Delta F_X$ smo označili $|F_X^*(x_i) - F_X(x_i)|$.

Iz preglednice 10.8 lahko odčitamo, da je vrednost statistike $D = 0.1342$. Kritično območje oziroma območje zavrnitve ničelne domneve je $[0.2102, 1]$, kar lahko odčitamo iz preglednice Kolmogorova in Smirnova.

Ker vrednost statistike ne leži v območju zavrnitve ničelne domneve, je ne moremo zavrniti. Ne moremo trditi, da slučajna spremenljivka $X$ (življenska doba tovornjakov) ni porazdeljena normalno s pričakovano vrednostjo $m_X = 350\,000\,\mathrm{km}$ in standardno deviacijo $\sigma_X = 70\,000\,\mathrm{km}$.

Dejansko tveganje ob zavrnitvi ničelne domneve je enako $\alpha_{dej} = 0.415$, kar bi bilo izredno visoko tveganje. V tem primeru se je pokazalo, da je preizkus Kolmogorova in Smirnova boljši od preizkusa skladnosti s porazdelitvijo $\chi^2$, kjer je bilo dejansko tveganje precej blizu izbranega tveganja $\alpha = 0.05$. Prednost preizkusa Kolmogorova in Smirnova je posebej očitna pri relativno majhnih vzorcih, pri katerih je razporeditev elementov v razrede lahko težavna.

(Vzorec vrednosti življenskih dob tovornjakov, prikazan v preglednici 10.3, je bil dejansko generiran z računalnikom. Predpostavljeno je bilo, da je porazdelitev normalna s pričakovano vrednostjo $m_X = 350$ tisoč km in standardno deviacijo $\sigma_X = 70$ tisoč km.)

# 11 Bivariatna analiza

V tem poglavju obravnavamo statistično analizo slučajnega vektorja dveh slučajnih spremenljivk. Iz vzorca in z uporabo ustreznih statističnih metod lahko ugotovimo, ali sta dve slučajni spremenljivki statistično značilno medsebojno odvisni. Drugi del tega poglavja opisuje analizo linearne povezanosti dveh slučajnih spremenljivk.

Vzorec običajno sestavljajo pari vrednosti slučajnih spremenljivk: $X_i, Y_i, i = 1, \ldots, n$, kjer je $n$ velikost vzorca.

## 11.1 Preizkušanje statistične odvisnosti

Postavimo ničelno in alternativno domnevo:

$H_0$: spremenljivki $X$ in $Y$ sta neodvisni,
$H_1$: spremenljivki $X$ in $Y$ sta odvisni.

Za preizkušanje domneve o statistični povezanosti med dvema slučajnima spremenljivkama $X$ in $Y$ na osnovi vzorčnih podatkov uporabimo test $\chi^2$. Ta test temelji na primerjavi empiričnih (dejanskih) frekvenc s teoretičnimi frekvencami. Vzorec slučajnega vektorja $X_l, Y_l, l = 1, \ldots, n$ razporedimo v razrede ($k_X$ razredov za spremenljivko $X$ in $k_Y$ razredov za spremenljivko $Y$). Števila elementov vzorca v razredih, to so empirične oziroma dejanske frekvence $\hat{n}_{ij}, i = 1, \ldots, k_X, j = 1, \ldots, k_Y$, prikažemo v **kontingenčni preglednici** (preglednica 11.1).

Teoretične frekvence oziroma teoretične velikosti razredov $n_{ij}$ v kontingenčni preglednici izračunamo po naslednji enačbi:

$$n_{ij} = n\,P[(X = x_i) \cap (Y = y_j)]. \tag{11.1}$$

Z izrazom $(X = x_i) \cap (Y = y_j)$ opišemo dogodek, da slučajna spremenljivka $X$ zavzame vrednost v $i$-tem razredu, slučajna spremenljivka $Y$ pa v $j$-tem razredu.

Ob predpostavki, da velja ničelna hipoteza, da sta slučajni spremenljivki $X$ in $Y$ neodvisni, lahko verjetnost produkta dogodkov $(X = x_i) \cap (Y = y_j)$ zapišemo kot produkt verjetnosti

$$P[X = x_i \cap Y = y_j] = P[X = x_i]\, P[Y = y_j].$$

Preglednica 11.1: Dejanske velikosti razredov

| Spremenljivka $Y$ | Spremenljivka $X$ | | | | Vsota za vse razrede Y |
|---|---|---|---|---|---|
| | 1 | 2 | $\cdots$ | $k_X$ | |
| 1 | $\hat{n}_{11}$ | $\hat{n}_{21}$ | $\cdots$ | $\hat{n}_{k_X 1}$ | $\hat{n}_{Y1}$ |
| 2 | $\hat{n}_{12}$ | $\hat{n}_{22}$ | $\cdots$ | $\hat{n}_{k_X 2}$ | $\hat{n}_{Y2}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $k_Y$ | $\hat{n}_{1\,k_Y}$ | $\hat{n}_{2\,k_Y}$ | $\cdots$ | $\hat{n}_{k_X k_Y}$ | $\hat{n}_{Y\,k_Y}$ |
| Vsota za vse razrede $X$ | $\hat{n}_{X1}$ | $\hat{n}_{X2}$ | $\cdots$ | $\hat{n}_{X\,k_X}$ | $n$ |

Verjetnosti, da je $X = x_i$ oziroma $Y = y_j$ lahko ocenimo iz vzorca:

$$P[X = x_i] = \frac{\hat{n}_{Xi}}{n} \quad \text{in} \quad P[Y = y_j] = \frac{\hat{n}_{Yj}}{n}. \tag{11.2}$$

Če enačbi (11.2) upoštevamo v enačbi (11.1), lahko zapišemo končni izraz za določitev teoretičnih velikosti razredov $n_{ij}$

$$n_{ij} = \frac{\hat{n}_{Xi}\,\hat{n}_{Yj}}{n}. \tag{11.3}$$

Sestavimo kontingenčno preglednico teoretičnih frekvenc $n_{ij}$ ter jih s statistiko $H$ primerjamo z dejanskimi:

$$H = \sum_{i=1}^{k_X} \sum_{j=1}^{k_Y} \frac{(n_{ij} - \hat{n}_{ij})^2}{n_{ij}}. \tag{11.4}$$

Statistika $H$ se porazdeljuje po porazdelitvi $\chi^2$ z $\nu = (k_X - 1)(k_Y - 1)$ prostostnimi stopnjami.

Kritično območje za zavrnitev ničelne domneve je $[\chi^2_{1-\alpha,\nu}, \infty)$. Če je statistika $H > \chi^2_{1-\alpha,\nu}$, ničelno domnevo zavrnemo in trdimo, da sta slučajni spremenljivki statistično značilno medsebojno odvisni.

**Primer 11.1:** *Anketirance, ki jih razporedimo po starosti v tri skupine (mlajši, srednji, starejši), vprašamo, kaj si mislijo o nekem ukrepu našega župana. Možna sta dva odgovora: “za” ali “proti”. Rezultate podajamo v naslednji kontingenčni preglednici.*

Preglednica 11.2: Dejanske velikosti razredov

| Mnenje o ukrepu | Starost | | | Vsota |
|---|---|---|---|---|
| | mlajši | srednji | starejši | |
| za | 182 | 213 | 203 | 598 |
| proti | 154 | 138 | 110 | 402 |
| Vsota | 336 | 351 | 313 | 1000 |

*Ugotoviti želimo ali starost meščanov vpliva na mnenje o ukrepu župana. Tveganje naj bo enako* $5\ \%$.

**Rešitev:** Postavimo ničelno in alternativno domnevo:

$H_0$ : Mnenje o ukrepu je neodvisna od starosti meščanov,
$H_1$ : Mnenje o ukrepu je odvisna od starosti meščanov.

Statistiko $H$ bomo določili po enačbi (11.4). Zato moramo najprej izračunati teoretične velikosti razredov $n_{ij}$ po enačbi (11.3), ki jih prikazujemo v naslednji preglednici.

Preglednica 11.3: Teoretične velikosti razredov

| Mnenje o ukrepu | Starost | | | Vsota |
|---|---|---|---|---|
| | mlajši | srednji | starejši | |
| za | 200.9 | 209.9 | 187.2 | 598 |
| proti | 135.1 | 141.1 | 125.8 | 402 |
| Vsota | 336 | 351 | 313 | 1000 |

Glede na tveganje $\alpha = 0.05$ lahko zapišemo kritično območje: $[\chi^2_{1-\alpha,\nu}, \infty)$, kjer je število prostostnih stopenj enako

$$\nu = (3-1)(2-1) = 2.$$

Če bo statistika $H$ večja od $\chi^2_{1-\alpha,\nu}$, ničelno hiptezo zavrnemo in trdimo, da sta mnenje o ukrepu in starost anketirancev odvisni slučajni spremenljivki. Mejo kritičnega območja odčitamo iz preglednic ali izračunamo z računalniškim programom (na primer EXCEL) $\chi^2_{1-\alpha,\nu} = \chi^2_{0.95,2} = 5.991$.

Statistika $H$ je

$$H = \sum_{i=1}^{k_X} \sum_{j=1}^{k_Y} \frac{(n_{ij} - \hat{n}_{ij})^2}{n_{ij}} = \frac{(200.9-182)^2}{200.9} + \frac{(209.9-213)^2}{209.9} + \cdots + \frac{(125.8-110)^2}{125.8} = 7.878.$$

Ker je $H = 7.878 > \chi^2_{0.95,2} = 5.991$, ničelno hipotezo zavrnemo in trdimo, da je mnenje o ukrepu statistično značilno odvisno od starosti meščanov.

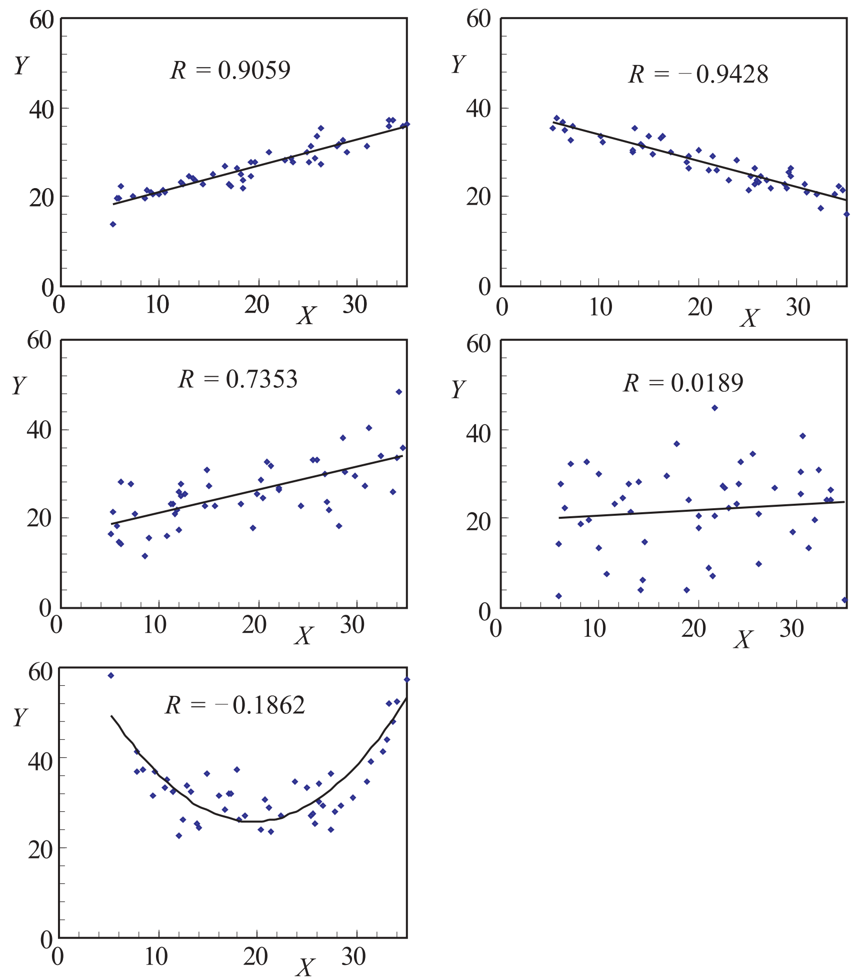
Ugotovimo še dejansko tveganje ob zavrnitvi ničelne hipoteze! Iz porazdelitvene funkcije porazdelitve $F_{\chi^2}$ (glej sliko 10.5) lahko izračunamo

$$1-\alpha = F_{\chi^2}(7.878) \quad \rightarrow \quad \alpha = 1 - F_{\chi^2}(7.878) = 1 - 0.9805 = 0.0195.$$

Tveganje je torej nekoliko pod 2 %. Ker je to tveganje nižje od predpisanega tveganja $\alpha = 5\,\%$, smo ničelno hipotezo zavrnili.

## 11.2 Preizkušanje linearne povezanosti

Povezanost med dvema številskima spremenljivkama grafično ponazorimo z razsevnim grafom (slika 11.1).




Slika 11.1: Vzorci $X$ in $Y$ z različno linearno povezanostjo

Linearno povezanost med dvema spremenljivkama merimo s **kovarianco** (glej tudi primera 6.3 in 6.4):

$$\sigma_{XY} = \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} (x_i - m_X)(y_j - m_Y)\, p_{XY}(x_i, y_j)$$

oziroma

$$\sigma_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_X)(y - m_Y)\, f_{XY}(x, y)\, dx\, dy.$$

Brezdimenzijska mera linerne povezanosti je Pearsonov **koeficient korelacije**:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Iz vzorčnih podatkov $X_i, Y_i, i = 1, \ldots, n$ lahko ocenimo kovarianco po naslednji enačbi:

$$\hat{\sigma}_{XY} = S_{XY} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}),$$

kjer je $n$ število opazovanj v vzorcu, $\bar{X}$ povprečje vzorca $X_i$ in $\bar{Y}$ povprečje vzorca $Y_i$.

Ocena koeficienta korelacije pa je

$$\hat{\rho}_{XY} = R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

Pomen parametra $\rho_{XY}$ oziroma $R_{XY}$ za različne vzorce prikazujemo na sliki 11.1. V primerih, kjer je parameter $R$ blizu nič, lahko govorimo o zelo slabi linearni povezanosti. Če se $R$ približuje vrednosti ena, sta spremenljivki močno pozitivno linearno povezani, če pa se približuje vrednosti $-1$, sta spremenljivki močno negativno linearno povezani.

**Statistično sklepanje o linearni povezanosti**. Postavimo ničelno in alternativno domnevo:

$H_0$: $\rho_{XY} = 0$ (spremenljivki nista linearno povezani)
$H_1$: $\rho_{XY} \neq 0$ (spremenljivki sta linearno povezani)

Statistika $T$

$$T = \frac{R_{XY}\sqrt{n-2}}{\sqrt{1 - R_{XY}^2}} \tag{11.5}$$

se v tem primeru porazdeljuje po Studentovi porazdelitvi $t$ z $\nu = n - 2$ prostostnimi stopnjami. Kritično območje oziroma območje zavrnitve je $(-\infty, -t_{1-\alpha/2}], [t_{1-\alpha/2}, \infty)$. Če je torej vrednost statistike $T$ manjša od $-t_{1-\alpha/2}$ ali večja od $t_{1-\alpha/2}$, lahko s tveganjem $\alpha$ zaključimo, da sta spremenljivki statistično značilno linearno povezani.

## 11.3 Regresija

Regresijska funkcija $\hat{Y} = f(X)$ opisuje, kakšen je vpliv spremenljivke $X$ na $Y$ brez drugih vplivov, ki so lahko posledica vpliva drugih spremenljivk ali slučajnega odstopanja. Slučajno spremenljivko $Y$ lahko zapišemo kot vsoto dveh spremenljivk

$$Y = \hat{Y} + \varepsilon = f(X) + \varepsilon,$$

kjer spremenljivko $X$ imenujemo neodvisna spremenljivka, slučajno spremenljivko $Y$ pa odvisna spremenljivka ter $\varepsilon$ napaka (ali slučajno odstopanje). Neodvisna spremenljivka $X$ je deterministična ali slučajna. Poglejmo dva primera:

1. Ugotavljamo zvezo med trdnostjo zemljine $Y$ in globino od površja $X$. Glede na to, da si globino lahko sami izberemo, lahko privzamemo, da je neodvisna slučajna spremenljivka deterministična.
2. Analiziramo, kako sta povezana elastični modul $X$ in trdnost $Y$ betona. V tem primeru moramo narediti preizkus, kjer na istem preizkušancu najprej izmerimo elastični modul, nato pa še trdnost. V tem primeru sta obe spremenljivki slučajni.

Običajno predpostavimo, da se $\varepsilon$ porazdeljuje normalno s pričakovano vrednostjo $0$ in standardno deviacijo $\sigma$

$$\mathrm{E}[\varepsilon] = 0 \qquad \qquad \mathrm{var}[\varepsilon] = \sigma^2.$$

### 11.3.1 Linearna regresija

V primeru, da je regresijska funkcija linearna

$$\hat{Y} = f(X) = a + b\,X,$$

zapišemo regresijsko enačbo takole:

$$Y = \hat{Y} + \varepsilon = f(X) + \varepsilon = a + b\,X + \varepsilon.$$

Za posamezni element vzorca $X_i$ in $Y_i$ zapišemo naslednjo regresijsko enačbo:

$$Y_i = \hat{Y}_i + \varepsilon_i = a + b\,X_i + \varepsilon_i.$$

Z regresijo določimo tiste vrednosti ocen $\hat{a}$ in $\hat{b}$, da je prileganje regresijske premice elementom vzorca čimboljše. Če za določitev ocene parametrov $\hat{a}$ in $\hat{b}$ uporabimo **metodo najmanjših kvadratov**, moramo poiskati minimum funkcije $S(a, b)$, ki predstavlja vsoto kvadratov odstopanj $\varepsilon_i$

$$S(a, b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - (a + b\,X_i))^2.$$

Funkcijo $S(a,b)$ odvajamo po $a$ in $b$ in zahtevamo, da so ti odvodi enaki nič

$$\frac{\partial S}{\partial a} = \sum_{i=1}^{n} 2(Y_i - \hat{a} - \hat{b}\,X_i)(-1) = 0,$$
$$\frac{\partial S}{\partial b} = \sum_{i=1}^{n} 2(Y_i - \hat{a} - \hat{b}\,X_i)(-X_i) = 0.$$

Po preureditvi zgornjih izrazov dobimo sistem dveh linearnih enačb z dvema neznakama $\hat{a}$ in $\hat{b}$. Temu sistemu pravimo tudi sistem **normalnih enačb**:

$$\begin{aligned} n\,\hat{a} \quad &+ \left(\sum_{i=1}^{n} X_i\right)\hat{b} &= \sum_{i=1}^{n} Y_i, \\ \left(\sum_{i=1}^{n} X_i\right)\hat{a} \quad &+ \left(\sum_{i=1}^{n} X_i^2\right)\hat{b} &= \sum_{i=1}^{n} Y_i\,X_i, \end{aligned}$$

ki ga lahko zapišemo tudi v matrični obliki

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} Y_i\,X_i \end{bmatrix}.$$

Sistem linearnih enačb lahko rešimo na različne načine. Morda najbolj običajen način je reševanje z Gaussovo eliminacijo, tako da sistem preoblikujemo tako, da je matrika sistema zgornja trikotna. Prvo enačbo pomnožimo z $-\sum_{i=1}^{n} X_i/n$ in prištejemo drugi enačbi:

$$\begin{aligned} n\,\hat{a} \;+ &\qquad\qquad \left(\sum_{i=1}^{n} X_i\right)\hat{b} &= \sum_{i=1}^{n} Y_i, \\ 0\,\hat{a} \;+ &\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right)\hat{b} &= \sum_{i=1}^{n} Y_i\,X_i - \left(\sum_{i=1}^{n} Y_i\right)\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right). \end{aligned} \tag{11.6}$$

Iz druge enačbe v sistemu (11.6) lahko določimo oceno $\hat{b}$

$$\hat{b} = \frac{\sum_{i=1}^{n} Y_i\,X_i - \left(\sum_{i=1}^{n} Y_i\right)\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)}{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2}.$$

Če imenovalec in števec delimo z $n$ in upoštevamo enačbi (2.7) in (2.16), dobimo

$$\hat{b} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)\frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \bar{Y}\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2} = \frac{S_{XY}}{S_X^2}. \tag{11.7}$$

Uporabimo še prvo enačbo iz sistema (11.6) in določimo oceno parametra $\hat{a}$

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n} X_i \hat{b} = \bar{Y} - \bar{X}\,\hat{b} = \bar{Y} - \frac{S_{XY}}{S_X^2}\bar{X} \tag{11.8}$$

Oceni parametrov $\hat{a}$ in $\hat{b}$ sta slučajni spremenljivki, za katere lahko zapišemo srednji vrednosti in varianci:

$$\begin{aligned} \mathrm{E}[\hat{a}] &= a & \mathrm{var}[\hat{a}] &= \frac{\sigma^2}{n}\left(1 + \frac{\bar{X}^2}{S_X^2}\right) \\ \mathrm{E}[\hat{b}] &= b & \mathrm{var}[\hat{b}] &= \frac{\sigma^2}{S_X^2 n} \end{aligned} \tag{11.9}$$

Izraza za srednji vrednosti pričata, da sta obe oceni nepristranski. Iz izrazov za varianco pa vidimo, da velikost odstopanj $\varepsilon$, ki se odraža z varianco $\sigma^2$ vpliva na povečanje variance obeh ocen, medtem ko ta varianca pada z velikostjo vzorca $n$.

Varianci obeh ocen parametrov smo izrazili z varianco $\sigma^2$ slučajne spremenljivke $\varepsilon$, ki predstavlja odstopanja elementov vzorca od modela oziroma regresijske premice. Te variance običajno ne poznamo, zato jo moramo oceniti iz vzorca. Nepristransko oceno $\hat{\sigma}^2$ določimo po naslednji enačbi

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}\,X_i)^2 = \frac{n}{n-2}\,S_Y^2\,(1 - R_{XY}^2), \tag{11.10}$$

kjer smo pri računu variance delili z $n-2$, saj se je število prostostnih stopenj pri določitvi dveh ocen parametrov $\hat{a}$ in $\hat{b}$ zmanjšalo za dve.

**Preizkušanje domneve o vrednosti koeficienta** $b$. Postavimo ničelno in alternativno domnevo:

$H_0$: $b = b_0$,
$H_1$: $b \neq b_0$.

---

Testna statistika $T$ je normirana ocena parametra $\hat{b}$ tako, da ima srednjo vrednost enako nič in standardno deviacijo ena. Če namesto variance $\sigma^2$ zapišemo njeno oceno $\hat{\sigma}^2$ in uporabimo enačbi (11.7) in (11.10), dobimo naslednji izraz

$$T = \frac{\hat{b} - \mathrm{E}[\hat{b}]}{\sqrt{\mathrm{var}[\hat{b}]}} = \frac{\hat{b} - b_0}{\frac{\sigma}{S_X\sqrt{n}}} = \frac{\frac{S_{XY}}{S_X^2} - b_0}{\frac{\sqrt{n}}{\sqrt{n-2}}\, S_Y \sqrt{1 - R_{XY}^2}\,\frac{1}{S_X\sqrt{n}}}.$$

Ta statistika se porazdeljuje po porazdelitvi $t$ z $\nu = n - 2$ prostostnimi stopnjami. Z ničelno domnevo najpogosteje predpostavimo, da je $b = 0$, kar ustreza predpostavki, da sta slučajni spremenljivki $X$ in $Y$ linearno neodvisni. V tem primeru lahko iz zgornjega izraza izpeljemo enačbo za določitev statistike $T$

$$T = \frac{S_{XY}}{S_X\, S_Y}\,\frac{\sqrt{n-2}}{\sqrt{1 - R_{XY}^2}} = \frac{R_{XY}\sqrt{n-2}}{\sqrt{1 - R_{XY}^2}}, \tag{11.11}$$

ki smo jo že uporabili pri preizkušanju domnev o linearni neodvisnosti med dvema slučajnima spremenljivkama.

**Primer 11.2:** *V naslednji preglednici podajamo letne povprečne koncentracije žveplovega dioksida* $\mathrm{SO_2}$ *v Mariboru za obdobje od 1992 do 2002. Glavni vir onesnaženja z žveplovim dioksidom so termocentrale, ki za gorivo uporabljajo premog (pri nas Šoštanj in Trbovlje), manjši izvori pa so termocentrale oziroma toplarne, ki za gorivo uporabljajo nafto. Mejna vrednost povprečne letne koncentracije* $\mathrm{SO_2}$ *po Direktivi Sveta EU (99/30/EEC) je enaka* $20\,\mu\mathrm{g/m^3}$.

Preglednica 11.4: Letne povprečne koncentracije $\mathrm{SO_2}$ v Mariboru v $\mu\mathrm{g/m^3}$

| Leto | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{SO_2}$ | 47 | 42 | 30 | 28 | 24 | 23 | 18 | 17 | 13 | 10 | 8 |

Vir: Onesnaženost zraka v Sloveniji v letu 2002, Poročilo Agencije Republike Slovenije za okolje (ARSO), Ministrstvo za okolje in prostor

*Določimo oceni za parametra* $\hat{a}$ *in* $\hat{b}$ *linearne regresije. Če bi veljal linearni model, katerega leta bi bila koncentracija enaka nič? Določimo oceno standardne deviacije odstopanj od modela* $\widehat{sigma}$ *in preizkusimo domnevo o parametru* $b$ *oziroma domnevo o linearni neodvisnosti med spremenljivkama. Tveganje naj bo enaodstotno.*

**Rešitev:** Preden začnemo z reševanjem naloge, bomo vrednosti za leta spremenili tako, da bodo leta (spremenljivka $X$) tekla od 0 do 10. To naredimo zato, da vrednosti $S_X^2$, $S_{XY}$ in ocene parametra $\hat{a}$ ne bodo prevelike številke. Povprečno letno koncentracijo $\mathrm{SO_2}$ označimo z $Y$. Nove podatke podajamo v preglednici 11.5, skupaj z rezultati za vrednost modela in odstopanj od modela.

Preglednica 11.5: Letne povprečne koncentracije $SO_2$, linearni model in odstopanja

| $X$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 47 | 42 | 30 | 28 | 24 | 23 | 18 | 17 | 13 | 10 | 8 |
| Model | 41.91 | 38.26 | 34.60 | 30.95 | 27.29 | 23.64 | 19.98 | 16.33 | 12.67 | 9.02 | 5.36 |
| Odstopanja | −5.09 | −3.75 | 4.60 | 2.95 | 3.29 | 0.64 | 1.98 | −0.67 | −0.33 | −0.98 | −2.64 |

Oceni parametrov $\hat{a}$ in $\hat{b}$ ter oceno standardne deviacije $\hat{\sigma}$ določimo iz enačb (11.7), (11.8) in (11.10). Iz vrednosti osnovnih statistik za dane podatke

$$
\begin{aligned}
\bar{X} &= 5, & \bar{Y} &= 23.636,\\
S_X^2 &= 10, & S_Y^2 &= 142.050,\\
S_{XY} &= -36.545 & R_{XY} &= -0.9696
\end{aligned}
$$

lahko izračunamo ocene

$$
\begin{aligned}
\hat{b} &= \frac{S_{XY}}{S_X^2} = \frac{-36.545}{10} = -3.6545,\\
\hat{a} &= \bar{Y} - \hat{b}\,\bar{X} = 23.636 - (-3.655 \cdot 5) = 41.909,\\
\hat{\sigma}^2 &= \frac{n}{n-2}\, S_Y^2\,(1 - R_{XY}^2) = \frac{11}{9} \cdot 142.050 \cdot \bigl(1 - (-0.9696)^2\bigr) = 10.3798.
\end{aligned}
$$

Zapišemo lahko linearni model, ki se po metodi najmanjših kvadratov najbolje prilega podatkom:

$$
\hat{Y} = f(X) = 41.909 - 3.6545\,X \tag{11.12}
$$

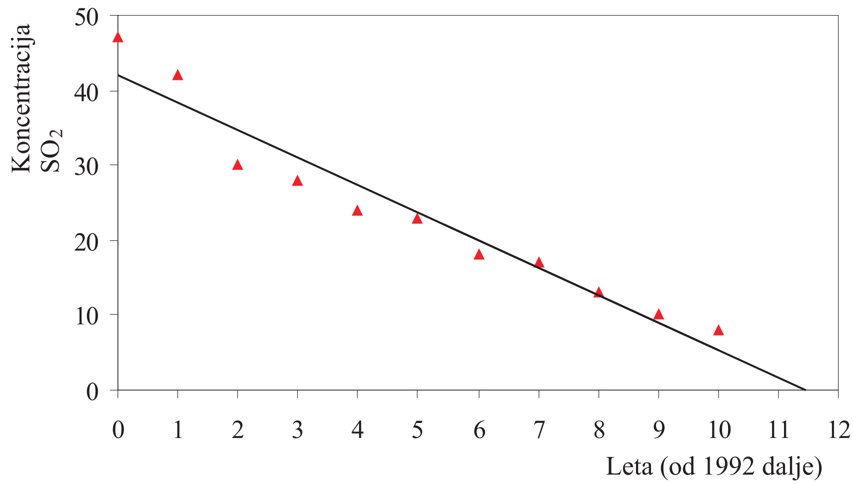
Z enačbo (11.12) lahko izračunamo vrednosti modela v točkah, kjer imamo podatke in odstopanje modela od pravih vrednosti (preglednica 11.5).

Iz enačbe (11.12) lahko izračunamo tudi leto, ko bi morala biti povprečna letna koncentracija $SO_2$ enaka nič, če bi bil linearni model ustrezen.

$$
41.909 - 3.6545\,X = 0 \qquad \to \qquad X = \frac{41.909}{3.6545} = 11.47.
$$

To bi pomenilo, da bi morala biti povprečna letna koncentracija $SO_2$ enaka nič najpozneje v letu 2004. To se seveda ni zgodilo, kar pomeni, da je linearni model sicer ustrezen za obdobje od 1992 do 2002, ekstrapolacija z linearno funkcijo pa v tem primeru ni bila ustrezna.

Grafični prikaz podatkov in linearnega modela prikazujemo na sliki 11.2.




Slika 11.2: Povprečna letna koncentracija $\mathrm{SO_2}$ v obdobju od 1992 do 2002

Nazadnje preizkusimo še domnevo o linearni neodvisnosti spremenljivk $X$ in $Y$. Postavimo ničelno in alternativno domnevo:

$H_0$: $b = 0$,
$H_1$: $b \neq 0$.

Statistiko $T$ izračunamo po enačbi (11.5)

$$T = \frac{R_{XY}\sqrt{n-2}}{\sqrt{1-R_{XY}^2}} = \frac{-0.9696 \cdot \sqrt{9}}{\sqrt{1-(-0.9696)^2}} = -11.897.$$

Meja kritičnega območja je $t_{1-\alpha/2} = 3.250$. Ker je vrednost statistike $T = 11.897$ manjša od $-t_{1-\alpha/2} = -3.250$, moramo ničelno domnevo zavrniti in trdimo, da je parameter $b$ statistično značilno različen od 0. Zaključimo lahko tudi z izjavo, da sta spremenljivki $X$ in $Y$ statistično značilno linearno odvisni.

### 11.3.2 Nelinearna regresija

Rezultati pogosto kažejo, da zveza med dvema spremenljivkama ni linearna. Obravnavajmo najprej funkcijo z dvema parametroma $a$ in $b$

$$Y = \hat{Y} + \varepsilon = f(X, a, b) + \varepsilon.$$

Oceni parametrov $\hat{a}$ in $\hat{b}$ določimo po metodi najmanjših kvadratov tako, da poiščemo minimum funkcije

$$S(a,b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n}(Y_i - f(X,a,b))^2.$$

Ta funkcija je v splošnem nelinearna glede na parametra $a$ in $b$, zato je iskanje minimuma lahko zelo zahtevna. Nelinearna funkcija ima lahko mnogo lokalnih ekstremov, iskanje globalnega minimuma pa je še vedno problem, ki ga poskušajo rešiti mnogi raziskovalci. Metod in njihovih različic je mnogo, v grobem jih lahko delimo na gradientne metode (na primer Newtonova metoda) in genetske algoritme.

Problem se zelo poenostavi, če je funkcija $f(X, a, b)$ linearna glede na parametra $a$ in $b$, sicer pa je lahko poljubna nelinearna funkcija

$$f(X, a, b) = a\, f_1(X) + b\, f_2(X).$$

Taki primeri so:

$$\begin{aligned} f(X, a, b) &= a + b\, X^2, \\ f(X, a, b) &= a \sin X + b\, \cos X, \\ f(X, a, b) &= a \log X + b\, \cos X^2. \end{aligned}$$

Oceni parametrov $\hat{a}$ in $\hat{b}$ določimo po metodi najmanjših kvadratov

$$S(a, b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - a\, f_1(X) + b\, f_2(X))^2.$$

Funkcijo $S(a, b)$ odvajamo po parametrih $a$ in $b$ ter zahtevamo, da sta odvoda enaka nič

$$\begin{aligned} \frac{\partial S}{\partial a} &= \sum_{i=1}^{n} 2(Y_i - \hat{a}\, f_1(X_i) - \hat{b}\, f_2(X_i))(-f_1(X_i)) = 0, \\ \frac{\partial S}{\partial b} &= \sum_{i=1}^{n} 2(Y_i - \hat{a}\, f_1(X_i) - \hat{b}\, f_2(X_i))(-f_2(X_i)) = 0. \end{aligned}$$

Podobno kot pri linearni regresiji dobimo tudi v tem primeru sistem linearnih enačb

$$\begin{aligned} \left(\sum_{i=1}^{n} f_1^2(X_i)\right) \hat{a} &+ \left(\sum_{i=1}^{n} f_1(X_i)\, f_2(X_i)\right) \hat{b} &= \sum_{i=1}^{n} f_1(X_i)\, Y_i, \\ \left(\sum_{i=1}^{n} f_1(X_i)\, f_2(X_i)\right) \hat{a} &+ \left(\sum_{i=1}^{n} f_2^2(X_i)\right) \hat{b} &= \sum_{i=1}^{n} f_2(X_i)\, Y_i, \end{aligned}$$

ki ima enolično rešitev le v primeru, da je determinanta sistema različna od nič. Ta pogoj je izpolnjen v primeru, da sta funkciji $f_1(x)$ in $f_2(x)$ linearno neodvisni. Iz zadnjega sistema enačb lahko hitro izpeljemo sistem enačb za linearno regresijo, če zapišemo, da je $f_1(x) = 1$ in $f_2(x) = x$.

Včasih je funkcija $f(X, a, b)$ taka, da jo s preprosto transformacijo prevedemo v linearno funkcijo. Tudi v tem primeru je določitev ocen parametrov $\hat{a}$ in $\hat{b}$ relativno preprosto. Nekaj takih funkcij z ustrezno transformacijo opisujemo z naslednjimi izrazi:

$$
\begin{array}{llllll}
Y = \dfrac{1}{a+b\,X} & \xrightarrow{\frac{1}{Y}} & \dfrac{1}{Y} = a + b\,X & \rightarrow & Z = a + b\,X \\
Y = a\,b^X & \xrightarrow{\ln Y} & \ln Y = \ln a + X\,\ln b & \rightarrow & Z = A + B\,X \\
Y = a\,e^{bX} & \xrightarrow{\ln Y} & \ln Y = \ln a + X\,b & \rightarrow & Z = A + b\,X \\
Y = a\,X^b & \xrightarrow{\ln Y} & \ln Y = \ln a + b\,\ln X & \rightarrow & Z = A + b\,W \\
Y = \ln(a+b\,X) & \xrightarrow{e^Y} & e^Y = a + b\,X & \rightarrow & Z = a + b\,X
\end{array}
\tag{11.13}
$$

Druga in tretja enačba predstavljata isti model, drugačen je le zapis enačbe.

**Primer 11.3:** *Vrnimo se k problemu, ki smo ga obravnavali v primeru 11.2. Namesto linearne regresije poskusimo, ali je morda bolje uporabiti nelinearno (eksponentno) regresijo. S primerjavo ocen variance odstopanj $\hat{\sigma}^2$ lahko ugotovimo, kateri model se bolje prilega podatkom.*

**Rešitev:** Regresijska funkcija je v tem primeru

$$\hat{Y} = f(X) = a\,e^{bX}.$$

Enačbo logaritmiramo in dobimo

$$\ln \hat{Y} = \ln a + b\,X \quad \rightarrow \quad Z = A + b\,X.$$

Podatke iz preglednice 11.5 preuredimo tako, da izračunamo še vrednosti spremenljivke $Z = \ln Y$. V preglednici 11.6 prikazujemo vrednosti spremenljivke $Z$ ter vrednosti, ki jih določimo z eksponentnim modelom, in odstopanja eksponentnega modela od dejanskih podatkov.

Preglednica 11.6: Vrednosti spremenljivk $X$, $Y$ in $Z$, eksponentni model in odstopanja

| $X$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $Y$ | 47 | 42 | 30 | 28 | 24 | 23 | 18 | 17 | 13 | 10 | 8 |
| $Z = \ln Y$ | 3.850 | 3.738 | 3.401 | 3.332 | 3.178 | 3.135 | 2.890 | 2.833 | 2.565 | 2.303 | 2.079 |
| Model | 47.64 | 40.30 | 34.10 | 28.85 | 24.41 | 20.65 | 17.47 | 14.78 | 12.51 | 10.58 | 8.95 |
| Odstopanja | 0.64 | -1.70 | 4.10 | 0.85 | 0.41 | -2.35 | -0.53 | -2.22 | -0.49 | 0.58 | 0.95 |

Izračunati moramo še osnovne statistike za slučajno spremenljivko $Z$:

$$
\begin{aligned}
\bar{Z} &= 3.028, \\
S_Z^2 &= 0.2866, \\
S_{XZ} &= -1.6717
\end{aligned}
$$

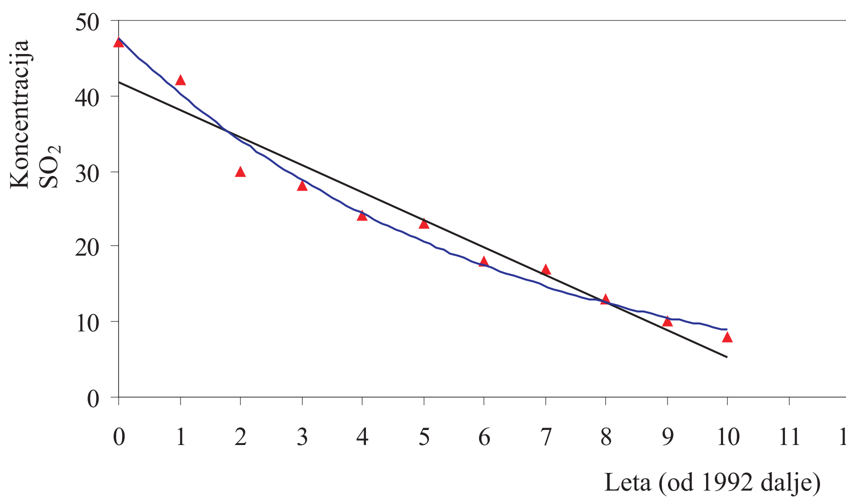
in oceni parametrov $\hat{A}$ in $\hat{b}$

$$\hat{b} = \frac{S_{XZ}}{S_X^2} = \frac{-1.6717}{10} = -0.16717,$$
$$\hat{A} = \bar{Z} - \hat{b}\,\bar{X} = 3.028 - (-0.16717) \cdot 5 = 3.8636,$$
$$\hat{a} = e^{\hat{A}} = 47.636.$$

Če primerjamo vrednosti modela v preglednicah 11.5 in 11.6, vidimo, da so odstopanja bistveno manjša pri eksponentni regresiji. Oceno variance odstopanj $\hat{\sigma}^2$ določimo po enačbi (11.10)

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \varepsilon_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(Y_i - \hat{a}e^{\hat{b}\,X_i}\right)^2 = \frac{1}{9}\left(0.64^2 + (-1.70)^2 + \cdots\right) = 3.686.$$

Ocena variance odstopanj $\hat{\sigma}^2$ je torej bistveno manjša kot v primeru linearne regresije. Primerjavo med linearno in eksponentno regresijo prikazujemo tudi na sliki 11.3. Tudi iz slike je očitno, da je v tem primeru eksponentna regresija boljša od linearne.




Slika 11.3: Povprečna letna koncentracija $SO_2$ v obdobju od 1992 do 2002

### 11.3.3 Linearna regresija več spremenljivk

Linearna regresija več spremenljivk oziroma multipla linearna regresija je posplošitev linearne regresije ene same neodvisne spremenljivke. Ta problem pravzaprav ne sodi v poglavje Bivariatna analiza, saj tu analiziramo več kot dve spremenljivki. Vseeno bomo to snov podali na koncu tega poglavja, saj se neposredno navezuje na linearno regresijo.

V tem primeru obravnavamo model odvisnosti med spremenljivkami $X_j, j = 1, \ldots, k$ in slučajno spremenljivko $Y$.

Vzorec je v tem primeru običajno podan z vrednostmi $X_{ij}, i = 1, \ldots, n, j = 1, \ldots, k$ neodvisnih spremenljivk in ustrezne vrednosti $Y_i, i = 1, \ldots, n$ spremenljivke $Y$. Linearni model zapišemo z enačbo

$$Y = \hat{Y} + \varepsilon = a + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k + \varepsilon,$$

kjer $\varepsilon$, podobno kot pri linearni regresiji ene spremenljivke, predstavlja odstopanje od modela. Običajno predpostavimo, da je $\varepsilon$ porazdeljen normalno s pričakovano vrednostjo nič in standardno deviacijo $\sigma$.

Za posamezni element vzorca $X_{ij}, Y_i$ zapišemo regresijsko enačbo takole:

$$Y_i = a + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_k X_{ik} + \varepsilon_i = a + \sum_{j=1}^{k} b_j X_{ij} + \varepsilon_i.$$

Z regresijo želimo določiti ocene neznanih regresijskih parametrov $\hat{a}$ in $\hat{b}_j$ tako, da bodo odstopanja dejanskih vrednosti $Y_i$ od modela čimmanjša. Uporabimo metodo najmanjših kvadratov, s katero iščemo minimum funkcije

$$S(a, b_j) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left( Y_i - a - \sum_{j=1}^{k} b_j X_{ij} \right)^2 = \sum_{i=1}^{n} (Y_i - a - b_1 X_{i1} - b_2 X_{i2} - \cdots - b_k X_{ik})^2$$

Minimum te funkcije določimo tako, da jo odvajamo po $a$ in $b_j$ ter postavimo pogoj, da so ti odvodi enaki nič:

$$\begin{aligned}
\frac{\partial S}{\partial a} &= \sum_{i=1}^{n} 2 \left( Y_i - \hat{a} - \hat{b}_1 X_{i1} - \hat{b}_2 X_{i2} - \cdots - \hat{b}_k X_{ik} \right) (-1) = 0 \\
\frac{\partial S}{\partial b_1} &= \sum_{i=1}^{n} 2 \left( Y_i - \hat{a} - \hat{b}_1 X_{i1} - \hat{b}_2 X_{i2} - \cdots - \hat{b}_k X_{ik} \right) (-X_{i1}) = 0 \\
\frac{\partial S}{\partial b_2} &= \sum_{i=1}^{n} 2 \left( Y_i - \hat{a} - \hat{b}_1 X_{i1} - \hat{b}_2 X_{i2} - \cdots - \hat{b}_k X_{ik} \right) (-X_{i2}) = 0 \\
&\vdots \\
\frac{\partial S}{\partial b_k} &= \sum_{i=1}^{n} 2 \left( Y_i - \hat{a} - \hat{b}_1 X_{i1} - \hat{b}_2 X_{i2} - \cdots - \hat{b}_k X_{ik} \right) (-X_{ik}) = 0
\end{aligned}$$

Po preureditvi zgornjih enačb dobimo sistem $k+1$ linearnih enačb za $k+1$ neznano oceno parametrov $\hat{a}$ in $\hat{b}_j$

$$\begin{array}{rcrcrcrcl}
n\,\hat{a} & + & \sum X_{i1}\,\hat{b}_1 & + & \sum X_{i2}\,\hat{b}_2 & + \cdots + & \sum X_{ik}\,\hat{b}_k & = & \sum Y_i, \\
\sum X_{i1}\,\hat{a} & + & \sum X_{i1}^2\,\hat{b}_1 & + & \sum X_{i2} X_{i1}\,\hat{b}_2 & + \cdots + & \sum X_{ik} X_{i1}\,\hat{b}_k & = & \sum Y_i X_{i1}, \\
\sum X_{i2}\,\hat{a} & + & \sum X_{i1} X_{i2}\,\hat{b}_1 & + & \sum X_{i2}^2\,\hat{b}_2 & + \cdots + & \sum X_{ik} X_{i2}\,\hat{b}_k & = & \sum Y_i X_{i2}, \\
\vdots & & \vdots & & \vdots & \ddots & \vdots & & \vdots \\
\sum X_{ik}\,\hat{a} & + & \sum X_{i1} X_{ik}\,\hat{b}_1 & + & \sum X_{i2} X_{ik}\,\hat{b}_2 & + \cdots + & \sum X_{ik}^2\,\hat{b}_k & = & \sum Y_i X_{ik},
\end{array}$$

kjer smo z znakom za vsoto $\sum$ zapisali vsoto $\sum_{i=1}^{n}$. Ta sistem normalnih enačb lahko zapišemo tudi v matrični obliki

$$\begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ik} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i2}\, X_{i1} & \cdots & \sum X_{ik}\, X_{i1} \\ \sum X_{i2} & \sum X_{i1}\, X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{ik}\, X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{ik} & \sum X_{i1}\, X_{ik} & \sum X_{i2}\, X_{ik} & \cdots & \sum X_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_k \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum Y_i\, X_{i1} \\ \sum Y_i\, X_{i2} \\ \vdots \\ \sum Y_i\, X_{ik} \end{bmatrix} \tag{11.14}$$

**Primer 11.4:** *V okviru projekta za določitev trdnosti lesa smo izmerili različne lastnosti na istih lesenih deskah. Na spodnji preglednici prikazujemo rezultate za 20 desk. Predpostavimo, da je upogibna trdnost $f$ lesenih desk odvisna od elastičnega modula $E$ in gostote $\rho$. Določimo ocene za koeficiente regresijske funkcije*

$$y = a + b_1\, x_1 + b_2\, x_2.$$

| $x_1 = E\,[\mathrm{MPa}]$ | $x_2 = \rho\,[\mathrm{kg/m^3}]$ | $y = f\,[\mathrm{MPa}]$ |
|---|---|---|
| 9150 | 538 | 31.03 |
| 10487 | 500 | 22.57 |
| 11357 | 423 | 32.29 |
| 4683 | 394 | 13.22 |
| 8728 | 413 | 36.39 |
| 14633 | 519 | 58.64 |
| 10615 | 509 | 31.32 |
| 8204 | 384 | 33.92 |
| 12235 | 509 | 22.00 |
| 6711 | 500 | 17.31 |
| 11647 | 471 | 41.82 |
| 11673 | 423 | 33.00 |
| 13785 | 442 | 45.66 |
| 5787 | 423 | 13.38 |
| 10369 | 471 | 28.02 |
| 13530 | 432 | 46.29 |
| 11131 | 500 | 29.98 |
| 10268 | 413 | 25.38 |
| 12753 | 452 | 30.71 |
| 10658 | 423 | 33.16 |

**Rešitev:** Za določitev ocen $\hat{a}$, $\hat{b}_1$ in $\hat{b}_2$ uporabimo prve tri vrstice v enačbi (11.14). Zato moramo izračunati naslednje vsote:

$$\sum_{i=1}^{n} X_{i1} = 9150 + 10487 + \cdots = 208404,$$
$$\sum_{i=1}^{n} X_{i2} = 538 + 500 + \cdots = 9139,$$
$$\sum_{i=1}^{n} X_{i1}^2 = 9150^2 + 10487^2 + \cdots = 2301198018,$$
$$\sum_{i=1}^{n} X_{i2}^2 = 538^2 + 500^2 + \cdots = 4216487,$$
$$\sum_{i=1}^{n} X_{i1}\, X_{i2} = 9150 \cdot 538 + 10487 \cdot 500 + \cdots = 95922910,$$
$$\sum_{i=1}^{n} Y_i = 31.03 + 22.57 + \cdots = 626.09,$$
$$\sum_{i=1}^{n} X_{i1}\, Y_i = 9150 \cdot 31.03 + 10487 \cdot 22.57 + \cdots = 6956369,$$
$$\sum_{i=1}^{n} X_{i2}\, Y_i = 538 \cdot 31.03 + 500 \cdot 22.57 + \cdots = 287201.8.$$

Iz linearnega sistema enačb

$$\begin{bmatrix} 20 & 208404 & 9139 \\ 208404 & 2301198018 & 95922910 \\ 9139 & 95922910 & 4216487 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 626.09 \\ 6956369 \\ 287201.8 \end{bmatrix}$$

izračunamo ocene parametrov linearne regresije

$$\begin{bmatrix} \hat{a} \\ \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} 9.664 \\ 0.003512 \\ -0.03272 \end{bmatrix}.$$

# 12 Analiza variance

Z **analizo variance** ugotavljamo, kako ena ali več neodvisnih spremenljivk vpliva na slučajno spremenljivko $Y$. Vzorec je urejen tako, da so vrednosti neodvisnih spremenljivk razvrščene v razrede, ali pa neodvisna spremenljivka predstavlja opisni znak. Neodvisni spremenljivki pogosto rečemo tudi **faktor**.

Prvi primer: Ugotoviti želimo, ali je telesna višina $Y$ prvošolčkov odvisna od spola $X$. V tem primeru neodvisna spremenljivka $X$ predstavlja opisni znak, ki lahko zavzame le dve vrednosti: ženski in moški spol.

Drugi primer: Ugotoviti želimo, ali oddaljenost učenčevega bivališča $X$ od šole vpliva na njegov učni uspeh $Y$. V tem primeru oddaljenosti od šole razvrstimo v primerno število razredov, tako da neodvisna spremenljivka lahko zavzame le toliko vrednosti, kolikor je razredov.

Tretji primer: Ugotoviti želimo, ali stopnja izobrazbe $X_1$ in način zaposlitve $X_2$ vplivata na dohodek prebivalca Slovenije. V tem primeru sta obe neodvisni spremenljivki opisna znaka.

## 12.1 Analiza variance za en faktor oziroma eno neodvisno spremenljivko

Vzemimo, da vrednosti neodvisne spremenljivke $X$ lahko razvrstimo v $a$ razredov in da je vzorec pripravljen tako, da je v vsakem razredu enako število elementov $n$. Vzorec torej vsebuje $an$ elementov. Osnovni model takih podatkov lahko zapišemo s preprosto enačbo

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad i = 1, \ldots, a, \;\; j = 1, \ldots, n, \tag{12.1}$$

kjer $Y_{ij}$ predstavljajo vzorčne vrednosti slučajne spremenljivke $Y$, $\mu$ je pričakovana vrednost, $\alpha_i$ predstavlja vplive posameznih razredov faktorja, $\varepsilon_{ij}$ pa predstavlja vzorčna odstopanja od modela. Običajno predpostavimo, da so odstopanja $\varepsilon_{ij}$ porazdeljena normalno s pričakovano vrednostjo nič in standardno deviacijo $\sigma$.

Primer podatkov za analizo variance prikazujemo v preglednici 12.1.

Preglednica 12.1: Podatki za analizo variance – vzorec

| Razred | Izmerjeni podatki $Y_{ij}$ | | | | $\bar{Y}_i$ |
|---|---|---|---|---|---|
| 1 | $Y_{11}$ | $Y_{12}$ | $\cdots$ | $Y_{1n}$ | $\bar{Y}_1$ |
| 2 | $Y_{21}$ | $Y_{22}$ | $\cdots$ | $Y_{2n}$ | $\bar{Y}_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $a$ | $Y_{a1}$ | $Y_{a2}$ | $\cdots$ | $Y_{an}$ | $\bar{Y}_a$ |

Pri analizi variance preizkusimo domnevo, ali faktor ne vpliva na spremenljivko $Y$. Glede na model, ki smo ga zapisali z enačbo (12.1), to ustreza pogoju, da so vsi $\alpha_i$ enaki nič. Ničelno in alternativno domnevo zapišemo takole:

$H_0$: $\alpha_i = 0$, za vse $i = 1, \ldots, a$, kar pomeni, da faktor ne vpliva,
$H_1$: $\alpha_i \neq 0$, za vsaj en $i = 1, \ldots, a$, kar pomeni, da faktor vpliva.

Določiti moramo statistiko, na osnovi katere bomo preizkusili ničelno domnevo. Za to moramo izračunati povprečja za posamezne razrede ter skupno povprečje:

$$\bar{Y}_i = \frac{1}{n}\sum_{j=1}^{n} Y_{ij}, \qquad i = 1, \ldots, a, \qquad\qquad \bar{Y} = \frac{1}{n\,a}\sum_{i=1}^{a}\sum_{j=1}^{n} Y_{ij}. \tag{12.2}$$

Nato tvorimo vsote kvadratov razlik, ki jih označimo s $SS$ (angl. “summed squares”):

$$\begin{aligned} SS_T &= \sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \bar{Y})^2, \\ SS_A &= n\sum_{i=1}^{a}(\bar{Y}_i - \bar{Y})^2, \\ SS_E &= \sum_{i=1}^{a}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2, \end{aligned} \tag{12.3}$$

S $SS_T$ označimo celotno vsoto kvadratov. Vsota kvadratov zaradi vpliva faktorja $SS_A$ je vsota kvadratov zaradi razlik med povprečnimi vrednostmi $\bar{Y}_i$ v razredih, zato tej vsoti na kratko rečemo tudi vsota kvadratov med razredi. Vsota kvadratov zaradi vpliva napak oziroma nepojasnjenih odstopanj $SS_E$ je vsota kvadratov razlik med posameznimi elementi vzorca $Y_{ij}$ in povprečji $\bar{Y}_i$ posameznih razredov, zato to na kratko opišemo kot vsoto kvadratov znotraj razredov. Računanje vsot kvadratov je dobro preveriti, ali izpolnjuje preprosto zvezo $SS_T = SS_A + SS_E$.

Na osnovi vsot kvadratov pripravimo **preglednico analize variance**, ki jo krajše imenujemo kar **preglednica ANOVA**:

| Vir odstopanj | Vsota kvadratov | Prostostne stopnje | Povprečni kvadrati | Statistika $F$ |
|---|---|---|---|---|
| Faktor | $SS_A$ | $n_{ps_A}$ | $MS_A$ | $F$ |
| Napaka | $SS_E$ | $n_{ps_E}$ | $MS_E$ | |
| Skupaj | $SS_T$ | $n_{ps_T}$ | | |

Števila prostostnih stopnj so:

$$n_{ps_A} = a - 1, \quad n_{ps_E} = a(n-1), \quad n_{ps_T} = an - 1,$$

kjer velja tudi $n_{ps_T} = n_{ps_A} + n_{ps_E}$. Povprečne kvadrate izračunamo po preprostih enačbah:

$$MS_A = \frac{SS_A}{n_{ps_A}}, \quad MS_E = \frac{SS_E}{n_{ps_E}}.$$

S prikazano preglednico izračunamo statistiko

$$F = \frac{MS_A}{MS_E},$$

ki se porazdeljuje po porazdelitvi $F$ s prostostnimi stopnjami $\nu_1 = n_{ps_A}$ in $\nu_2 = n_{ps_E}$. Kritično območje oziroma območje zavrnitve ničelne domneve je $[F_{1-\alpha}, \infty)$. Če je statistika $F$ večja od kritične vrednosti $F_{krit} = F_{1-\alpha,\nu_1,\nu_2}$ oziroma presega mejo območja zavrnitve ničelne hipoteze, ničelno hipotezo zavrnemo in trdimo, da je vpliv faktorja na spremenljivko $Y$ statistično značilen.

**Primer 12.1:** *Na avtomatskih merilnih postajah, ki so postavljene v več slovenskih krajih, merijo najvišje urne in povprečne letne koncentracije $SO_2$, $NO_2$, ozona, delcev $PM_{10}$ (T. Bolte, s sod., Kakovost zraka v Sloveniji v letu 2006, Agencija RS za okolje, Ljubljana, 2007.) Na preglednici 12.2 prikazujemo najvišje urne koncentracije $SO_2$ v različnih slovenskih krajih v obdobju od 1997–2006.*

*Ugotovimo, ali je kraj meritve odločilen faktor pri koncentraciji $SO_2$! Tveganje je 1-odstotno.*

Preglednica 12.2: Najvišje urne koncentracije $SO_2$ v $\mu g/m^3$

| | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ljubljana-Bežigrad | 1593 | 936 | 786 | 184 | 273 | 157 | 202 | 129 | 94 | 81 |
| Maribor | 211 | 161 | 157 | 117 | 180 | 89 | 70 | 64 | 58 | 60 |
| Celje | 975 | 623 | 228 | 379 | 666 | 224 | 619 | 396 | 157 | 90 |
| Trbovlje | 1806 | 693 | 849 | 634 | 552 | 811 | 758 | 521 | 848 | 379 |
| Hrastnik | 1930 | 978 | 963 | 720 | 731 | 2168 | 507 | 1799 | 549 | 134 |
| Zagorje | 914 | 1092 | 952 | 653 | 1111 | 788 | 693 | 1165 | 954 | 183 |
| Šoštanj | 1536 | 1495 | 2466 | 2855 | 2099 | 2000 | 1392 | 937 | 642 | 1028 |
| Topolšcica | 1050 | 1245 | 1345 | 987 | 835 | 1350 | 812 | 291 | 284 | 288 |
| Veliki Vrh | 1720 | 1530 | 2257 | 1678 | 1569 | 1450 | 1320 | 1329 | 1110 | 771 |
| Zavodnje | 2154 | 2255 | 1963 | 1187 | 954 | 1536 | 947 | 680 | 1106 | 731 |
| Velenje | 672 | 1316 | 709 | 563 | 187 | 725 | 361 | 164 | 210 | 86 |
| Graška Gora | 1579 | 1076 | 1844 | 1505 | 990 | 1024 | 824 | 463 | 497 | 175 |
| Kovk | 3000 | 1916 | 2167 | 1237 | 1451 | 702 | 1806 | 1514 | 1063 | 511 |
| Dobovec | 6072 | 4548 | 3761 | 4073 | 3978 | 4043 | 2910 | 4056 | 1662 | 2290 |
| Ravenska Vas | 2578 | 1846 | 1021 | 1471 | 1397 | 2093 | 1378 | 1779 | 3275 | 590 |

Vir: T. Bolte, s sod., Kakovost zraka v Sloveniji v letu 2006, Agencija RS za okolje, Ljubljana, 2007.

**Rešitev:** Izračunajmo najprej povprečja $\bar{Y}_i$ za posamezne razrede oziroma za posamezne merilne postaje ter skupno povprečje $\bar{Y}$ po enačbah (12.2). Te rezultate podajamo v preglednici 12.3

Preglednica 12.3: Povprečne vrednosti po razredih in skupno povprečje

| $\bar{Y}_1$ | $\bar{Y}_2$ | $\bar{Y}_3$ | $\bar{Y}_4$ | $\bar{Y}_5$ | $\bar{Y}_6$ | $\bar{Y}_7$ | $\bar{Y}_8$ |
|---|---|---|---|---|---|---|---|
| 443.5 | 116.7 | 435.7 | 785.1 | 1047.9 | 850.5 | 1645.0 | 848.7 |
| $\bar{Y}_9$ | $\bar{Y}_{10}$ | $\bar{Y}_{11}$ | $\bar{Y}_{12}$ | $\bar{Y}_{13}$ | $\bar{Y}_{14}$ | $\bar{Y}_{15}$ | $\bar{Y}$ |
| 1473.4 | 1351.3 | 499.3 | 997.7 | 1536.7 | 3739.3 | 1742.8 | 1167.6 |

Nato postavimo ničelno in alternativno domnevo:

$H_0$: Faktor (kraj merilne naprave) ne vpliva na stopnjo onesnaženosti z $SO_2$,
$H_1$: Faktor (kraj merilne naprave) vpliva na stopnjo onesnaženosti z $SO_2$.

## 12.2 Analiza variance za dva faktorja

Pri analizi variance za več faktorjev ugotavljamo ali faktorji ter interakcija med njimi vplivajo na spremenljivko $Y$. Prva neodvisna spremenljivka oziroma faktor $A$ je razporejena v $a$ razredov, druga oziroma faktor $B$ pa v $b$ razredov. Vzemimo, da je število elementov v vsakem razredu enako $n$. To pomeni, da je skupno število elementov v vzorcu enako $abn$. Model tako urejenih podatkov zapišemo s preprosto enačbo

$$
\begin{aligned}
&Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \\
&\qquad i = 1, \ldots, a, \quad j = 1, \ldots, b, \quad k = 1, \ldots, n
\end{aligned}
\tag{12.4}
$$

kjer $Y_{ijk}$ predstavljajo vzorčne vrednosti slučajne spremenljivke $Y$, $\mu$ je pričakovana vrednost, $\alpha_i$ predstavlja vplive posameznih razredov faktorja $A$, $\beta_j$ predstavlja vplive posameznih razredov faktorja $B$, $(\alpha\beta)_{ij}$ predstavlja vpliv interakcije med faktorjema $A$ in $B$, $\varepsilon_{ijk}$ pa predstavlja vzorčna odstopanja od modela. Običajno predpostavimo, da so odstopanja $\varepsilon_{ijk}$ porazdeljena normalno s pričakovano vrednostjo nič in standardno deviacijo $\sigma$.

Podatke pripravimo v preglednici:

| Razredi faktorja $A$ | Razredi faktorja $B$ 1 | 2 | $\cdots$ | $b$ | Povprečja |
|---|---|---|---|---|---|
| 1 | $Y_{111}\dots Y_{11n}$ | $Y_{121}\dots Y_{12n}$ | $\cdots$ | $Y_{1b1}\dots Y_{1bn}$ | $\bar{Y}_{1\cdot}$ |
| 2 | $Y_{211}\dots Y_{21n}$ | $Y_{221}\dots Y_{22n}$ | $\cdots$ | $Y_{2b1}\dots Y_{2bn}$ | $\bar{Y}_{2\cdot}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $a$ | $Y_{a11}\dots Y_{a1n}$ | $Y_{a21}\dots Y_{a2n}$ | $\cdots$ | $Y_{ab1}\dots Y_{abn}$ | $\bar{Y}_{a\cdot}$ |
| Povprečja | $\bar{Y}_{\cdot 1}$ | $\bar{Y}_{\cdot 2}$ | $\cdots$ | $\bar{Y}_{\cdot b}$ | $\bar{Y}$ |

Pri analizi variance za več faktorjev analiziramo vpliv posameznih faktorjev ter vpliv interakcije med posameznimi faktorji. Zato moramo v primeru dveh faktorjev opraviti tri ločena in neodvisna preizkušanja, ki jim ustrezajo tri ničelne in tri alternativne domneve.

$H_0$: $\alpha_i = 0$, za vse $i = 1, \dots, a$ (faktor $A$ ne vpliva),
$H_1$: $\alpha_i \neq 0$, za vsaj en $i = 1, \dots, a$ (faktor $A$ vpliva),

$H_0$: $\beta_j = 0$, za vse $j = 1, \dots, b$ (faktor $B$ ne vpliva),
$H_1$: $\beta_j \neq 0$, za vsaj en $j = 1, \dots, b$ (faktor $B$ vpliva),

$H_0$: $(\alpha\beta_{ij} = 0$, za vse $i = 1, \dots, a$ in $j = 1, \dots, b$ (interakcija med faktorjema $A$ in $B$ ne vpliva),
$H_1$: $(\alpha\beta_{ij} \neq 0$, za vsaj en $i = 1, \dots, a$ in $j = 1, \dots, b$ (interakcija med faktorjema $A$ in $B$ vpliva).

---

Za ta preizkušanje moramo na podoben način kot pri analizi variance za en faktor pripraviti tri statistike. Najprej izračunamo povprečja za posamezne razrede $\bar{Y}_{ij}$ povprečja za razrede faktorja $A$ in $B$ ($\bar{Y}_{i\cdot}$ in $\bar{Y}_{\cdot j}$) in skupno povprečje $\bar{Y}$:

$$\bar{Y}_{ij} = \frac{1}{n}\sum_{k=1}^{n} Y_{ijk} \qquad \bar{Y}_{i\cdot} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}$$

$$\bar{Y}_{\cdot j} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{ijk} \qquad \bar{Y} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}$$

Na osnovi izračunanih povprečij izračunamo še vsote kvadratov:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y})^2,$$

$$SS_A = bn\sum_{i=1}^{a}(\bar{Y}_{i\cdot} - \bar{Y})^2,$$

$$SS_B = an\sum_{j=1}^{b}(\bar{Y}_{\cdot j} - \bar{Y})^2,$$

$$SS_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y})^2,$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{ij})^2,$$

kjer je $SS_T$ celotna vsota kvadratov razlik, $SS_A$ predstavlja vpliv faktorja $A$, $SS_B$ vpliv faktorja $B$, $SS_{AB}$ vpliv interakcije med faktorjema $A$ in $B$, $SS_E$ pa vpliv napake oziroma nepojasnjenih odstopanj. Vsote kvadratov povezuje preprosta zveza

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E.$$

Sedaj lahko tvorimo preglednico analize variance

| Vir odstopanj | Vsota kvadratov | Prostostne stopnje | Povprečni kvadrati | Statistika $F$ |
|---|---|---|---|---|
| Faktor $A$ | $SS_A$ | $n_{ps_A}$ | $MS_A$ | $F_A$ |
| Faktor $B$ | $SS_B$ | $n_{ps_B}$ | $MS_B$ | $F_B$ |
| Interakcija $AB$ | $SS_{AB}$ | $n_{ps_{AB}}$ | $MS_{AB}$ | $F_{AB}$ |
| Napaka | $SS_E$ | $n_{ps_E}$ | $MS_E$ | |
| Skupaj | $SS_T$ | $n_{ps_T}$ | | |

Prostostne stopnje izračunamo po enačbah

$$n_{ps_A} = a - 1, \qquad n_{ps_{AB}} = (a-1)(b-1),$$
$$n_{ps_B} = b - 1, \qquad n_{ps_E} = ab(n-1),$$
$$n_{ps_T} = abn - 1, \qquad n_{ps_T} = n_{ps_A} + n_{ps_B} + n_{ps_{AB}} + n_{ps_E},$$

povprečne kvadrate $MS$ pa po enačbah

$$MS_A = \frac{SS_A}{n_{ps_A}}, \qquad MS_B = \frac{SS_B}{n_{ps_B}},$$
$$MS_{AB} = \frac{SS_{AB}}{n_{ps_{AB}}}, \qquad MS_E = \frac{SS_E}{n_{ps_E}}.$$

Statistike, ki jih izračunamo z enačbami

$$F_A = \frac{MS_A}{MS_E}, \qquad F_B = \frac{MS_B}{MS_E}, \qquad F_{AB} = \frac{MS_{AB}}{MS_E},$$

se porazdeljujejo po porazdelitvi $F$ s prostostnimi stopnjami $\nu_1 = n_{ps_A}$, $n_{ps_B}$ ali $n_{ps_{AB}}$ in $\nu_2 = n_{ps_E}$. Če je statistika $F$ večja od kritične vrednosti $F_{krit} = F_{1-\alpha,\nu_1,\nu_2}$ oziroma presega mejo območja zavrnitve ničelne hipoteze, ničelno hipotezo zavrnemo in trdimo, da je vpliv faktorja $A$, $B$ oziroma interakcije med $A$ in $B$ na spremenljivko $Y$ statistično značilen.

# 13 Simulacije

Simulacije so numerično orodje, s katerim opravimo preizkuse na računalniku. Preizkus vključuje nek matematičen ali logičen model, ki opisuje problem, ki ga rešujemo. S simulacijami inženirji in drugi strokovnjaki rešujejo najrazličnejše probleme: vedenje letala, delovanje telekomunikacijskega sistema, delovanje trga, vojskovanje in drugo.

Če simulacije izvajamo tako, da generiramo slučajne spremenljivke $X$ z določeno porazdelitveno funkcijo $F_X(x)$, takim simulacijam pravimo metoda Monte Carlo. To ime se je prvič pojavilo med drugo svetovno vojno, ko so znanstveniki von Neuman, Ulam, Fermi in drugi v laboratorijih v Los Alamosu tudi s pomočjo simulacij skonstruirali prvo atomsko bombo. Sicer pa je ideja o generiranju slučajnih spremenljivk in njihova uporaba pri reševanju problemov bistveno starejša. Prvi dobro dokumentiran primer je opravil Comte de Buffon leta 1777, ko je v problemu, poznanem kot Buffonova igla, s simulacijami poskušal določiti vrednost števila $\pi$. Pozneje so generiranje slučajnih števil uporabili tudi Lord Kelvin (1901), W.S. Gosset (Student) (1908), Fermi (1930) in drugi (Kalos et al., 1986, Rubinstein, 1981). Z razvojem boljših, hitrejših računalnikov, postaja metoda Monte Carlo uporabna za vse širši krog ljudi.

Generiranje vzorca slučajne spremenljivke $X$ temelji na vzorcu enakomerno porazdeljene slučajne $U$ spremenljivke s porazdelitveno funkcijo $F_U(u) = u$ (za $0 \leq u \leq 1$). Vzorca povsem slučajne spremenljivke ni lahko generirati. Zato z računalniki običajno generiramo vzorce tako imenovanih psevdoslučajnih števil, ki so pravzaprav zaporedja determinističnih števil z zelo veliko periodo ponovitve. Ta zaporedja imajo enake lastnosti kot zaporedja povsem slučajnih števil. Skoraj vsi računalniški programi, namenjeni računanju (na primer prevajalniki: FORTRAN, C, PASCAL, BASIC in drugi programi: MATLAB, MATHEMATICA, EXCEL), vključujejo tudi "generatorje slučajnih števil", s katerimi generiramo zaporedja psevdoslučajnih števil, ki ustrezajo vzorcu slučajne spremenljivke $U$, ki je porazdeljena enakomerno od 0 do 1. Vzorec poljubno porazdeljene slučajne spremenljivke ali vektorja lahko generiramo z različnimi metodami: inverzna metoda, metoda sprejema in zavrnitve, polarna metoda, metoda trakov, mrežna metoda in druge. Najpogosteje uporabljamo inverzno metodo, ki je najprimernejša za generiranje neodvisnih slučajnih spremenljivk.

### Inverzna metoda generiranja vzorca

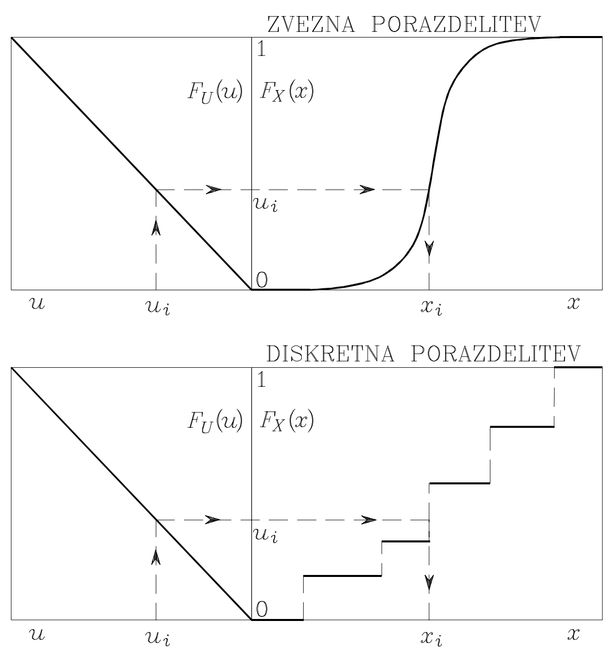
Vzemimo, da imamo enako velika vzorca dveh slučajnih spremenljivk. Prva, $X$, je porazdeljena po porazdelitvi $F_X(x)$, druga, $U$, pa je enakomerno porazdeljena od 0 do 1: $F_U(u) = u$ (za $0 \leq u \leq 1$). Elemente v obeh vzorcih razvrstimo po velikosti in predpostavimo, da je verjetnost, da sta slučajni spremenljivki $X$ in $U$ manjši od $i$-tega elementa ustreznega vzorca enaka, ne glede na porazdelitev. To predpostavko v enačbi zapišemo takole:

$$P[X \leq x_i] = F_X(x_i) = P[U \leq u_i] = F_U(u_i) = u_i$$

Iz zadnje enačbe lahko izpeljemo izraz za določitev elementa vzorca slučajne spremenljivke:

$$F_X(x_i) = u_i \qquad \longrightarrow \qquad x_i = F_X^{-1}(u_i).$$

Če torej poznamo porazdelitveno funkcijo slučajne spremenljivke oziroma njeno inverzno funkcijo in imamo vzorec ($u_i$, $i = 1, \ldots$) enakomerno porazdeljene slučajne spremenljivke $U$, lahko po zadnji enačbi določimo vzorec slučajne spremenljivke $X$. Inverzno metodo nazorno prikažemo tudi grafično na sliki 13.1.




Slika 13.1: Inverzna metoda generiranja vzorca

### Generiranje vzorca normalno porazdeljenih slučajnih spremenljivk

Porazdelitvene funkcije normalne porazdelitve v zaključeni obliki ne poznamo, prav tako ne poznamo njene inverzne funkcije. Zato za generiranje vzorca normalno porazdeljene slučajne spremenljivke ne moremo uporabiti analitične inverzne metode. Če računalniški program omogoča numerični izračun inverzne porazdelitvene funkcije normalne porazdelitve, lahko uporabimo inverzno metodo, četudi funkcije $F_X^{-1}(u_i)$ ne poznamo v eksplicitni obliki. Na primer: v programu EXCEL je funkcija vgrajena kot `norminv`, v programu MATHEMATICA lahko problem rešimo z ukazom `InverseErf`, v programu MATLAB pa lahko uporabimo funkcijo `erfinv`. Če programska oprema ne omogoča uporabe funkcije $F_X^{-1}(u_i)$, moramo vzorec normalno porazdeljene slučajne spremenljivke generirati s kakšno drugo metodo, kot je na primer Box-Müllerjeva polarna metoda. S to metodo vzorec standardno normalno porazdeljene slučajne spremenljivke izračunamo iz dveh vzorcev ($u_{1i}$, $i = 1, \ldots$) in ($u_{2i}$, $i = 1, \ldots$) slučajnih spremenljivk $U_1$ in $U_2$, ki sta neodvisni in porazdeljeni enakomerno od 0 do 1, po naslednji enačbi

$$x_i = \sqrt{-2 \ln u_{1i}} \, \sin(2\, \pi \, u_{2i}).$$

Vzorec $y_i$ poljubne normalno porazdeljene slučajne spremenljivke $Y$ s pričakovano vrednostjo $m_Y$ in standardno deviacijo $\sigma_Y$ iz vzorca $x_i$ določimo s transformacijo

$$y_i = m_Y + x_i \, \sigma_Y .$$

# 14 Dodatek

## 14.1 Varianca povprečja pri končno velikih populacijah

Izpeljati želimo enačbo (8.5) za varianco povprečja vzorca pri končno velikih populacijah in vzorčenjem brez vračanja. Elementi populacije so neodvisne in enako porazdeljene slučajne spremenljivke $X_i$, število vseh teh spremenljivk v populaciji je $N$. Velikost vzorca je enaka $n$. Varianco povprečja izračunamo z enačbo (6.15)

$$\mathrm{var}\left[\bar{X}\right] = \mathrm{E}\left[(\bar{X} - m_X)^2\right] = \frac{1}{n^2}\mathrm{E}\left[\left(\sum_{i=1}^{n}(X_i - m_X)\right)^2\right] = \frac{1}{n^2}\mathrm{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^2\right],$$

kjer smo upoštevali, da je

$$\bar{X} - m_X = \frac{1}{n}\sum_{i=1}^{n} X_i - m_X = \frac{1}{n}\sum_{i=1}^{n}(X_i - m_X)$$

in vpeljali oznako

$$Y_i = X_i - m_X.$$

Vsoto slučajnih spremenljivk $Y_i$ na kvadrat lahko zapišemo v razviti obliki takole:

$$\left(\sum_{i=1}^{n} Y_i\right)^2 = \sum_{i=1}^{n} Y_i^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Y_i Y_j.$$

Ob upoštevanju enačbe (6.28) lahko varianco povprečje sedaj zapišemo z enačbo

$$\mathrm{var}\left[\bar{X}\right] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{E}\left[Y_i^2\right] + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\mathrm{E}\left[Y_i Y_j\right]\right). \tag{14.1}$$

Pričakovana vrednost $\mathrm{E}\left[Y_i^2\right] = \mathrm{E}\left[X_i - m_X\right]$ je zaradi enačbe (6.15) in (8.1) enaka varianci slučajne spremenljivke $X$

$$\mathrm{E}\left[Y_i^2\right] = \sigma_X^2. \tag{14.2}$$

Nekoliko težje je določiti pričakovano vrednost produkta $Y_i\,Y_j$. Izpeljimo najprej pričakovano vrednost produkta dveh diskretnih slučajnih spremenljivk $X$ in $Y$ z zalogama vrednosti $x_i, i = 1, \ldots, n_X$ in $y_j, j = 1, \ldots, n_Y$. Pričakovano vrednost $\mathrm{E}[XY]$ izračunamo po enačbi

$$\mathrm{E}[XY] = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} x_i\,y_j\,p_{XY}(x_i,y_j) = \sum_{i=1}^{n_X} x_i \left(\sum_{j=1}^{n_Y} y_j \frac{p_{XY}(x_i,y_j)}{P_X(x_i)}\right) p_X(x_i) =$$

$$= \sum_{i=1}^{n_X} x_i \left(\sum_{j=1}^{n_Y} y_j\,p_{Y|X}(x_i,y_j)\right) p_X(x_i) = \sum_{i=1}^{n_X} x_i\,\mathrm{E}[Y|X=x_i]\,p_X(x_i) = \mathrm{E}\left[X\,\mathrm{E}[Y|X]\right].$$

pri čemer smo upoštevali definicijo pogojne verjetnostne funkcije (4.20) in definicijo pričakovane vrednosti (6.3). Zadnjo enačbo torej uporabimo pri določitvi pričakovane vrednosti produkta $Y_i\,Y_j$

$$\mathrm{E}[Y_i\,Y_j] = \mathrm{E}\left[Y_i\,\mathrm{E}[Y_j|Y_i]\right]. \tag{14.3}$$

Pogojno pričakovano vrednost $\mathrm{E}[Y_j|Y_i]$ lahko določimo z enačbo, v kateri upoštevamo vse elemente populacije

$$\mathrm{E}[Y_j|Y_i] = \frac{1}{N-1}\sum_{k=1\,(k\neq i)}^{N} Y_k = -\frac{Y_i}{N-1}, \tag{14.4}$$

saj velja, da je vsota vseh $Y_i$ enaka nič

$$\sum_{k=1}^{N} Y_k = \sum_{k=1\,(k\neq i)}^{N} Y_k + Y_i = 0 \qquad \longrightarrow \qquad \sum_{k=1\,(k\neq i)}^{N} Y_k = -Y_i.$$

Sedaj enačbo (14.4) vstavimo v (14.3) in dobimo

$$\mathrm{E}[Y_i\,Y_j] = \mathrm{E}\left[Y_i\left(\frac{Y_i}{N-1}\right)\right] = -\frac{1}{N-1}\mathrm{E}\left[Y_i^2\right] = -\frac{\sigma_X^2}{N-1}. \tag{14.5}$$

Enačbi (14.2) in (14.5) vstavimo v enačbo (14.1) ter zapišemo

$$\mathrm{var}\left[\bar{X}\right] = \frac{1}{n^2}\left(\sum_{i=1}^{n}\sigma_X^2 - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{\sigma_X^2}{N-1}\right).$$

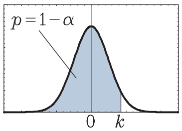
Če upoštevamo, da je v prvi vsoti $n$ elementov, v drugi pa $\binom{n}{2} = \frac{n(n-1)}{2}$ elementov, dobimo končni rezultat

$$\mathrm{var}\left[\bar{X}\right] = \frac{1}{n^2}\left(n\sigma_X^2 - 2\frac{n(n-1)}{2}\frac{\sigma_X^2}{N-1}\right) = \frac{\sigma_X^2}{n}\left(1-\frac{n-1}{N-1}\right) = \frac{\sigma_X^2}{n}\,\frac{N-n}{N-1}.$$

# 15 Preglednice porazdelitev

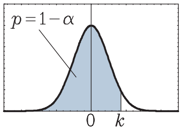
Vse preglednice smo pripravili z računalniškim programom Excel. Namesto teh preglednic lahko uporabite tudi datoteko Statkalk.xls, s katero lahko izračunate vrednosti porazdelitvenih funkcij in njihovih inverznih vrednosti.

## 15.1 Standardizirana normalna porazdelitev



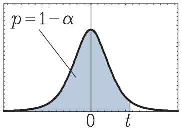

| $u$ ali $k$ | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 |
|---|---|---|---|---|---|
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | 0.51595 |
| 0.1 | 0.53983 | 0.54380 | 0.54776 | 0.55172 | 0.55567 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | 0.59483 |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.62930 | 0.63307 |
| 0.4 | 0.65542 | 0.65910 | 0.66276 | 0.66640 | 0.67003 |
| 0.5 | 0.69146 | 0.69497 | 0.69847 | 0.70194 | 0.70540 |
| 0.6 | 0.72575 | 0.72907 | 0.73237 | 0.73565 | 0.73891 |
| 0.7 | 0.75804 | 0.76115 | 0.76424 | 0.76730 | 0.77035 |
| 0.8 | 0.78814 | 0.79103 | 0.79389 | 0.79673 | 0.79955 |
| 0.9 | 0.81594 | 0.81859 | 0.82121 | 0.82381 | 0.82639 |
| 1.0 | 0.84134 | 0.84375 | 0.84614 | 0.84849 | 0.85083 |
| 1.1 | 0.86433 | 0.86650 | 0.86864 | 0.87076 | 0.87286 |
| 1.2 | 0.88493 | 0.88686 | 0.88877 | 0.89065 | 0.89251 |
| 1.3 | 0.90320 | 0.90490 | 0.90658 | 0.90824 | 0.90988 |
| 1.4 | 0.91924 | 0.92073 | 0.92220 | 0.92364 | 0.92507 |
| 1.5 | 0.93319 | 0.93448 | 0.93574 | 0.93699 | 0.93822 |
| 1.6 | 0.94520 | 0.94630 | 0.94738 | 0.94845 | 0.94950 |
| 1.7 | 0.95543 | 0.95637 | 0.95728 | 0.95818 | 0.95907 |
| 1.8 | 0.96407 | 0.96485 | 0.96562 | 0.96638 | 0.96712 |
| 1.9 | 0.97128 | 0.97193 | 0.97257 | 0.97320 | 0.97381 |
| 2.0 | 0.97725 | 0.97778 | 0.97831 | 0.97882 | 0.97932 |
| 2.1 | 0.98214 | 0.98257 | 0.98300 | 0.98341 | 0.98382 |
| 2.2 | 0.98610 | 0.98645 | 0.98679 | 0.98713 | 0.98745 |
| 2.3 | 0.98928 | 0.98956 | 0.98983 | 0.99010 | 0.99036 |
| 2.4 | 0.99180 | 0.99202 | 0.99224 | 0.99245 | 0.99266 |
| 2.5 | 0.99379 | 0.99396 | 0.99413 | 0.99430 | 0.99446 |
| 2.6 | 0.99534 | 0.99547 | 0.99560 | 0.99573 | 0.99585 |
| 2.7 | 0.99653 | 0.99664 | 0.99674 | 0.99683 | 0.99693 |
| 2.8 | 0.99744 | 0.99752 | 0.99760 | 0.99767 | 0.99774 |
| 2.9 | 0.99813 | 0.99819 | 0.99825 | 0.99831 | 0.99836 |
| 3.0 | 0.99865 | 0.99869 | 0.99874 | 0.99878 | 0.99882 |
| 3.1 | 0.99903 | 0.99906 | 0.99910 | 0.99913 | 0.99916 |
| 3.2 | 0.99931 | 0.99934 | 0.99936 | 0.99938 | 0.99940 |
| 3.3 | 0.99952 | 0.99953 | 0.99955 | 0.99957 | 0.99958 |
| 3.4 | 0.99966 | 0.99968 | 0.99969 | 0.99970 | 0.99971 |
| 3.5 | 0.99977 | 0.99978 | 0.99978 | 0.99979 | 0.99980 |
| 3.6 | 0.99984 | 0.99985 | 0.99985 | 0.99986 | 0.99986 |
| 3.7 | 0.99989 | 0.99990 | 0.99990 | 0.99990 | 0.99991 |

Standardizirana normalna porazdelitev (nadaljevanje)



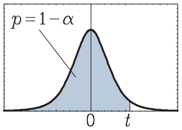

| $u$ ali $k$ | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|
| 0.0 | 0.51994 | 0.52392 | 0.52790 | 0.53188 | 0.53586 |
| 0.1 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57535 |
| 0.2 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.67364 | 0.67724 | 0.68082 | 0.68439 | 0.68793 |
| 0.5 | 0.70884 | 0.71226 | 0.71566 | 0.71904 | 0.72240 |
| 0.6 | 0.74215 | 0.74537 | 0.74857 | 0.75175 | 0.75490 |
| 0.7 | 0.77337 | 0.77637 | 0.77935 | 0.78230 | 0.78524 |
| 0.8 | 0.80234 | 0.80511 | 0.80785 | 0.81057 | 0.81327 |
| 0.9 | 0.82894 | 0.83147 | 0.83398 | 0.83646 | 0.83891 |
| 1.0 | 0.85314 | 0.85543 | 0.85769 | 0.85993 | 0.86214 |
| 1.1 | 0.87493 | 0.87698 | 0.87900 | 0.88100 | 0.88298 |
| 1.2 | 0.89435 | 0.89617 | 0.89796 | 0.89973 | 0.90147 |
| 1.3 | 0.91149 | 0.91308 | 0.91466 | 0.91621 | 0.91774 |
| 1.4 | 0.92647 | 0.92785 | 0.92922 | 0.93056 | 0.93189 |
| 1.5 | 0.93943 | 0.94062 | 0.94179 | 0.94295 | 0.94408 |
| 1.6 | 0.95053 | 0.95154 | 0.95254 | 0.95352 | 0.95449 |
| 1.7 | 0.95994 | 0.96080 | 0.96164 | 0.96246 | 0.96327 |
| 1.8 | 0.96784 | 0.96856 | 0.96926 | 0.96995 | 0.97062 |
| 1.9 | 0.97441 | 0.97500 | 0.97558 | 0.97615 | 0.97670 |
| 2.0 | 0.97982 | 0.98030 | 0.98077 | 0.98124 | 0.98169 |
| 2.1 | 0.98422 | 0.98461 | 0.98500 | 0.98537 | 0.98574 |
| 2.2 | 0.98778 | 0.98809 | 0.98840 | 0.98870 | 0.98899 |
| 2.3 | 0.99061 | 0.99086 | 0.99111 | 0.99134 | 0.99158 |
| 2.4 | 0.99286 | 0.99305 | 0.99324 | 0.99343 | 0.99361 |
| 2.5 | 0.99461 | 0.99477 | 0.99492 | 0.99506 | 0.99520 |
| 2.6 | 0.99598 | 0.99609 | 0.99621 | 0.99632 | 0.99643 |
| 2.7 | 0.99702 | 0.99711 | 0.99720 | 0.99728 | 0.99736 |
| 2.8 | 0.99781 | 0.99788 | 0.99795 | 0.99801 | 0.99807 |
| 2.9 | 0.99841 | 0.99846 | 0.99851 | 0.99856 | 0.99861 |
| 3.0 | 0.99886 | 0.99889 | 0.99893 | 0.99896 | 0.99900 |
| 3.1 | 0.99918 | 0.99921 | 0.99924 | 0.99926 | 0.99929 |
| 3.2 | 0.99942 | 0.99944 | 0.99946 | 0.99948 | 0.99950 |
| 3.3 | 0.99960 | 0.99961 | 0.99962 | 0.99964 | 0.99965 |
| 3.4 | 0.99972 | 0.99973 | 0.99974 | 0.99975 | 0.99976 |
| 3.5 | 0.99981 | 0.99981 | 0.99982 | 0.99983 | 0.99983 |
| 3.6 | 0.99987 | 0.99987 | 0.99988 | 0.99988 | 0.99989 |
| 3.7 | 0.99991 | 0.99992 | 0.99992 | 0.99992 | 0.99992 |

## 15.2 Studentova porazdelitev $t$



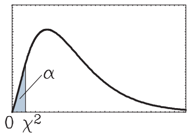

| $\nu$ | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
|---|---|---|---|---|---|---|---|---|
| | $p = 1-\alpha$ | | | | | | | |
| 1 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 | 31.821 | 63.656 |
| 2 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 |
| 11 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 |
| 12 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 |
| 13 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 |
| 14 | 0.692 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 |
| 15 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 |
| 16 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 |
| 17 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 |
| 18 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 |
| 19 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 |
| 20 | 0.687 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 |
| 21 | 0.686 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 |
| 22 | 0.686 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 |
| 23 | 0.685 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 |
| 24 | 0.685 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 |
| 25 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 |
| 26 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 |
| 27 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 |
| 28 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 |
| 29 | 0.683 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 |
| 30 | 0.683 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 |
| 31 | 0.682 | 0.853 | 1.054 | 1.309 | 1.696 | 2.040 | 2.453 | 2.744 |
| 32 | 0.682 | 0.853 | 1.054 | 1.309 | 1.694 | 2.037 | 2.449 | 2.738 |
| 33 | 0.682 | 0.853 | 1.053 | 1.308 | 1.692 | 2.035 | 2.445 | 2.733 |
| 34 | 0.682 | 0.852 | 1.052 | 1.307 | 1.691 | 2.032 | 2.441 | 2.728 |
| 35 | 0.682 | 0.852 | 1.052 | 1.306 | 1.690 | 2.030 | 2.438 | 2.724 |
| 36 | 0.681 | 0.852 | 1.052 | 1.306 | 1.688 | 2.028 | 2.434 | 2.719 |

Studentova porazdelitev $t$ (nadaljevanje)



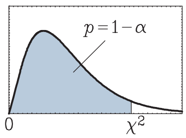

| | $p = 1 - \alpha$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $\nu$ | 0.75 | 0.80 | 0.85 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
| 37 | 0.681 | 0.851 | 1.051 | 1.305 | 1.687 | 2.026 | 2.431 | 2.715 |
| 38 | 0.681 | 0.851 | 1.051 | 1.304 | 1.686 | 2.024 | 2.429 | 2.712 |
| 39 | 0.681 | 0.851 | 1.050 | 1.304 | 1.685 | 2.023 | 2.426 | 2.708 |
| 40 | 0.681 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 |
| 41 | 0.681 | 0.850 | 1.050 | 1.303 | 1.683 | 2.020 | 2.421 | 2.701 |
| 42 | 0.680 | 0.850 | 1.049 | 1.302 | 1.682 | 2.018 | 2.418 | 2.698 |
| 43 | 0.680 | 0.850 | 1.049 | 1.302 | 1.681 | 2.017 | 2.416 | 2.695 |
| 44 | 0.680 | 0.850 | 1.049 | 1.301 | 1.680 | 2.015 | 2.414 | 2.692 |
| 45 | 0.680 | 0.850 | 1.049 | 1.301 | 1.679 | 2.014 | 2.412 | 2.690 |
| 46 | 0.680 | 0.850 | 1.048 | 1.300 | 1.679 | 2.013 | 2.410 | 2.687 |
| 47 | 0.680 | 0.849 | 1.048 | 1.300 | 1.678 | 2.012 | 2.408 | 2.685 |
| 48 | 0.680 | 0.849 | 1.048 | 1.299 | 1.677 | 2.011 | 2.407 | 2.682 |
| 49 | 0.680 | 0.849 | 1.048 | 1.299 | 1.677 | 2.010 | 2.405 | 2.680 |
| 50 | 0.679 | 0.849 | 1.047 | 1.299 | 1.676 | 2.009 | 2.403 | 2.678 |
| 55 | 0.679 | 0.848 | 1.046 | 1.297 | 1.673 | 2.004 | 2.396 | 2.668 |
| 60 | 0.679 | 0.848 | 1.045 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 |
| 65 | 0.678 | 0.847 | 1.045 | 1.295 | 1.669 | 1.997 | 2.385 | 2.654 |
| 70 | 0.678 | 0.847 | 1.044 | 1.294 | 1.667 | 1.994 | 2.381 | 2.648 |
| 75 | 0.678 | 0.846 | 1.044 | 1.293 | 1.665 | 1.992 | 2.377 | 2.643 |
| 80 | 0.678 | 0.846 | 1.043 | 1.292 | 1.664 | 1.990 | 2.374 | 2.639 |
| 85 | 0.677 | 0.846 | 1.043 | 1.292 | 1.663 | 1.988 | 2.371 | 2.635 |
| 90 | 0.677 | 0.846 | 1.042 | 1.291 | 1.662 | 1.987 | 2.368 | 2.632 |
| 95 | 0.677 | 0.845 | 1.042 | 1.291 | 1.661 | 1.985 | 2.366 | 2.629 |
| 100 | 0.677 | 0.845 | 1.042 | 1.290 | 1.660 | 1.984 | 2.364 | 2.626 |
| 110 | 0.677 | 0.845 | 1.041 | 1.289 | 1.659 | 1.982 | 2.361 | 2.621 |
| 120 | 0.677 | 0.845 | 1.041 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 |
| 130 | 0.676 | 0.844 | 1.041 | 1.288 | 1.657 | 1.978 | 2.355 | 2.614 |
| 140 | 0.676 | 0.844 | 1.040 | 1.288 | 1.656 | 1.977 | 2.353 | 2.611 |
| 150 | 0.676 | 0.844 | 1.040 | 1.287 | 1.655 | 1.976 | 2.351 | 2.609 |
| 160 | 0.676 | 0.844 | 1.040 | 1.287 | 1.654 | 1.975 | 2.350 | 2.607 |
| 170 | 0.676 | 0.844 | 1.040 | 1.287 | 1.654 | 1.974 | 2.348 | 2.605 |
| 180 | 0.676 | 0.844 | 1.039 | 1.286 | 1.653 | 1.973 | 2.347 | 2.603 |
| $\infty$ | 0.674 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 |

## 15.3 Porazdelitev $\chi^2$



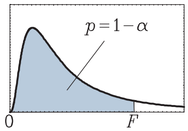

| $\nu$ | $\alpha$ 0.001 | 0.005 | 0.010 | 0.025 | 0.050 | 0.100 | 0.250 |
|---|---|---|---|---|---|---|---|
| 1 | 0.000002 | 0.00004 | 0.0002 | 0.001 | 0.004 | 0.016 | 0.102 |
| 2 | 0.002 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 0.575 |
| 3 | 0.024 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 1.213 |
| 4 | 0.091 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 1.923 |
| 5 | 0.210 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 2.675 |
| 6 | 0.381 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 3.455 |
| 7 | 0.599 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 4.255 |
| 8 | 0.857 | 1.344 | 1.647 | 2.180 | 2.733 | 3.490 | 5.071 |
| 9 | 1.152 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 5.899 |
| 10 | 1.479 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 6.737 |
| 11 | 1.834 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 7.584 |
| 12 | 2.214 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 8.438 |
| 13 | 2.617 | 3.565 | 4.107 | 5.009 | 5.892 | 7.041 | 9.299 |
| 14 | 3.041 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 10.165 |
| 15 | 3.483 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 11.037 |
| 16 | 3.942 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 11.912 |
| 17 | 4.416 | 5.697 | 6.408 | 7.564 | 8.672 | 10.085 | 12.792 |
| 18 | 4.905 | 6.265 | 7.015 | 8.231 | 9.390 | 10.865 | 13.675 |
| 19 | 5.407 | 6.844 | 7.633 | 8.907 | 10.117 | 11.651 | 14.562 |
| 20 | 5.921 | 7.434 | 8.260 | 9.591 | 10.851 | 12.443 | 15.452 |
| 21 | 6.447 | 8.034 | 8.897 | 10.283 | 11.591 | 13.240 | 16.344 |
| 22 | 6.983 | 8.643 | 9.542 | 10.982 | 12.338 | 14.041 | 17.240 |
| 23 | 7.529 | 9.260 | 10.196 | 11.689 | 13.091 | 14.848 | 18.137 |
| 24 | 8.085 | 9.886 | 10.856 | 12.401 | 13.848 | 15.659 | 19.037 |
| 25 | 8.649 | 10.520 | 11.524 | 13.120 | 14.611 | 16.473 | 19.939 |
| 26 | 9.222 | 11.160 | 12.198 | 13.844 | 15.379 | 17.292 | 20.843 |
| 27 | 9.803 | 11.808 | 12.878 | 14.573 | 16.151 | 18.114 | 21.749 |
| 28 | 10.391 | 12.461 | 13.565 | 15.308 | 16.928 | 18.939 | 22.657 |
| 29 | 10.986 | 13.121 | 14.256 | 16.047 | 17.708 | 19.768 | 23.567 |
| 30 | 11.588 | 13.787 | 14.953 | 16.791 | 18.493 | 20.599 | 24.478 |
| 31 | 12.196 | 14.458 | 15.655 | 17.539 | 19.281 | 21.434 | 25.390 |
| 32 | 12.810 | 15.134 | 16.362 | 18.291 | 20.072 | 22.271 | 26.304 |
| 33 | 13.431 | 15.815 | 17.073 | 19.047 | 20.867 | 23.110 | 27.219 |
| 34 | 14.057 | 16.501 | 17.789 | 19.806 | 21.664 | 23.952 | 28.136 |
| 35 | 14.688 | 17.192 | 18.509 | 20.569 | 22.465 | 24.797 | 29.054 |
| 40 | 17.917 | 20.707 | 22.164 | 24.433 | 26.509 | 29.051 | 33.660 |
| 50 | 24.674 | 27.991 | 29.707 | 32.357 | 34.764 | 37.689 | 42.942 |

Porazdelitev $\chi^2$ (nadaljevanje)



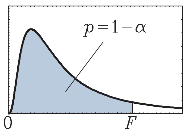

| $\nu$ | $p = 1-\alpha$ 0.500 | 0.750 | 0.900 | 0.950 | 0.975 | 0.990 | 0.995 |
|---|---|---|---|---|---|---|---|
| 1 | 0.455 | 1.323 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 1.386 | 2.773 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 2.366 | 4.108 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 3.357 | 5.385 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 4.351 | 6.626 | 9.236 | 11.070 | 12.832 | 15.086 | 16.750 |
| 6 | 5.348 | 7.841 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 6.346 | 9.037 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 7.344 | 10.219 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 8.343 | 11.389 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 9.342 | 12.549 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 |
| 11 | 10.341 | 13.701 | 17.275 | 19.675 | 21.920 | 24.725 | 26.757 |
| 12 | 11.340 | 14.845 | 18.549 | 21.026 | 23.337 | 26.217 | 28.300 |
| 13 | 12.340 | 15.984 | 19.812 | 22.362 | 24.736 | 27.688 | 29.819 |
| 14 | 13.339 | 17.117 | 21.064 | 23.685 | 26.119 | 29.141 | 31.319 |
| 15 | 14.339 | 18.245 | 22.307 | 24.996 | 27.488 | 30.578 | 32.801 |
| 16 | 15.338 | 19.369 | 23.542 | 26.296 | 28.845 | 32.000 | 34.267 |
| 17 | 16.338 | 20.489 | 24.769 | 27.587 | 30.191 | 33.409 | 35.718 |
| 18 | 17.338 | 21.605 | 25.989 | 28.869 | 31.526 | 34.805 | 37.156 |
| 19 | 18.338 | 22.718 | 27.204 | 30.144 | 32.852 | 36.191 | 38.582 |
| 20 | 19.337 | 23.828 | 28.412 | 31.410 | 34.170 | 37.566 | 39.997 |
| 21 | 20.337 | 24.935 | 29.615 | 32.671 | 35.479 | 38.932 | 41.401 |
| 22 | 21.337 | 26.039 | 30.813 | 33.924 | 36.781 | 40.289 | 42.796 |
| 23 | 22.337 | 27.141 | 32.007 | 35.172 | 38.076 | 41.638 | 44.181 |
| 24 | 23.337 | 28.241 | 33.196 | 36.415 | 39.364 | 42.980 | 45.558 |
| 25 | 24.337 | 29.339 | 34.382 | 37.652 | 40.646 | 44.314 | 46.928 |
| 26 | 25.336 | 30.435 | 35.563 | 38.885 | 41.923 | 45.642 | 48.290 |
| 27 | 26.336 | 31.528 | 36.741 | 40.113 | 43.195 | 46.963 | 49.645 |
| 28 | 27.336 | 32.620 | 37.916 | 41.337 | 44.461 | 48.278 | 50.994 |
| 29 | 28.336 | 33.711 | 39.087 | 42.557 | 45.722 | 49.588 | 52.335 |
| 30 | 29.336 | 34.800 | 40.256 | 43.773 | 46.979 | 50.892 | 53.672 |
| 31 | 30.336 | 35.887 | 41.422 | 44.985 | 48.232 | 52.191 | 55.002 |
| 32 | 31.336 | 36.973 | 42.585 | 46.194 | 49.480 | 53.486 | 56.328 |
| 33 | 32.336 | 38.058 | 43.745 | 47.400 | 50.725 | 54.775 | 57.648 |
| 34 | 33.336 | 39.141 | 44.903 | 48.602 | 51.966 | 56.061 | 58.964 |
| 35 | 34.336 | 40.223 | 46.059 | 49.802 | 53.203 | 57.342 | 60.275 |
| 40 | 39.335 | 45.616 | 51.805 | 55.758 | 59.342 | 63.691 | 66.766 |
| 50 | 49.335 | 56.334 | 63.167 | 67.505 | 71.420 | 76.154 | 79.490 |

## 15.4 Porazdelitev $F$



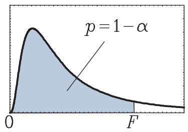

| $p = 1 - \alpha$ | $\nu_2$ | $\nu_1$ = 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 0.900 | 1 | 39.9 | 49.5 | 53.6 | 55.8 | 57.2 | 58.2 |
| 0.950 | 1 | 161.4 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 |
| 0.975 | 1 | 647.8 | 799.5 | 864.2 | 899.6 | 921.8 | 937.1 |
| 0.990 | 1 | 4052.2 | 4999.3 | 5403.5 | 5624.3 | 5764.0 | 5859.0 |
| 0.995 | 1 | 16212.5 | 19997.4 | 21614.1 | 22500.8 | 23055.8 | 23439.5 |
| 0.900 | 2 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 |
| 0.950 | 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 |
| 0.975 | 2 | 38.51 | 39.00 | 39.17 | 39.25 | 39.30 | 39.33 |
| 0.990 | 2 | 98.50 | 99.00 | 99.16 | 99.25 | 99.30 | 99.33 |
| 0.995 | 2 | 198.50 | 199.01 | 199.16 | 199.24 | 199.30 | 199.33 |
| 0.900 | 3 | 5.538 | 5.462 | 5.391 | 5.343 | 5.309 | 5.285 |
| 0.950 | 3 | 10.128 | 9.552 | 9.277 | 9.117 | 9.013 | 8.941 |
| 0.975 | 3 | 17.443 | 16.044 | 15.439 | 15.101 | 14.885 | 14.735 |
| 0.990 | 3 | 34.116 | 30.816 | 29.457 | 28.710 | 28.237 | 27.911 |
| 0.995 | 3 | 55.552 | 49.800 | 47.468 | 46.195 | 45.391 | 44.838 |
| 0.900 | 4 | 4.545 | 4.325 | 4.191 | 4.107 | 4.051 | 4.010 |
| 0.950 | 4 | 7.709 | 6.944 | 6.591 | 6.388 | 6.256 | 6.163 |
| 0.975 | 4 | 12.218 | 10.649 | 9.979 | 9.604 | 9.364 | 9.197 |
| 0.990 | 4 | 21.198 | 18.000 | 16.694 | 15.977 | 15.522 | 15.207 |
| 0.995 | 4 | 31.332 | 26.284 | 24.260 | 23.154 | 22.456 | 21.975 |
| 0.900 | 5 | 4.060 | 3.780 | 3.619 | 3.520 | 3.453 | 3.405 |
| 0.950 | 5 | 6.608 | 5.786 | 5.409 | 5.192 | 5.050 | 4.950 |
| 0.975 | 5 | 10.007 | 8.434 | 7.764 | 7.388 | 7.146 | 6.978 |
| 0.990 | 5 | 16.258 | 13.274 | 12.060 | 11.392 | 10.967 | 10.672 |
| 0.995 | 5 | 22.785 | 18.314 | 16.530 | 15.556 | 14.939 | 14.513 |
| 0.900 | 6 | 3.776 | 3.463 | 3.289 | 3.181 | 3.108 | 3.055 |
| 0.950 | 6 | 5.987 | 5.143 | 4.757 | 4.534 | 4.387 | 4.284 |
| 0.975 | 6 | 8.813 | 7.260 | 6.599 | 6.227 | 5.988 | 5.820 |
| 0.990 | 6 | 13.745 | 10.925 | 9.780 | 9.148 | 8.746 | 8.466 |
| 0.995 | 6 | 18.635 | 14.544 | 12.917 | 12.028 | 11.464 | 11.073 |

Porazdelitev $F$ (nadaljevanje)



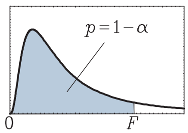

| $p = 1 - \alpha$ | $\nu_2$ | $\nu_1$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 0.900 | 7 | 3.589 | 3.257 | 3.074 | 2.961 | 2.883 | 2.827 |
| 0.950 | 7 | 5.591 | 4.737 | 4.347 | 4.120 | 3.972 | 3.866 |
| 0.975 | 7 | 8.073 | 6.542 | 5.890 | 5.523 | 5.285 | 5.119 |
| 0.990 | 7 | 12.246 | 9.547 | 8.451 | 7.847 | 7.460 | 7.191 |
| 0.995 | 7 | 16.235 | 12.404 | 10.883 | 10.050 | 9.522 | 9.155 |
| 0.900 | 8 | 3.458 | 3.113 | 2.924 | 2.806 | 2.726 | 2.668 |
| 0.950 | 8 | 5.318 | 4.459 | 4.066 | 3.838 | 3.688 | 3.581 |
| 0.975 | 8 | 7.571 | 6.059 | 5.416 | 5.053 | 4.817 | 4.652 |
| 0.990 | 8 | 11.259 | 8.649 | 7.591 | 7.006 | 6.632 | 6.371 |
| 0.995 | 8 | 14.688 | 11.043 | 9.597 | 8.805 | 8.302 | 7.952 |
| 0.900 | 9 | 3.360 | 3.006 | 2.813 | 2.693 | 2.611 | 2.551 |
| 0.950 | 9 | 5.117 | 4.256 | 3.863 | 3.633 | 3.482 | 3.374 |
| 0.975 | 9 | 7.209 | 5.715 | 5.078 | 4.718 | 4.484 | 4.320 |
| 0.990 | 9 | 10.562 | 8.022 | 6.992 | 6.422 | 6.057 | 5.802 |
| 0.995 | 9 | 13.614 | 10.107 | 8.717 | 7.956 | 7.471 | 7.134 |
| 0.900 | 10 | 3.285 | 2.924 | 2.728 | 2.605 | 2.522 | 2.461 |
| 0.950 | 10 | 4.965 | 4.103 | 3.708 | 3.478 | 3.326 | 3.217 |
| 0.975 | 10 | 6.937 | 5.456 | 4.826 | 4.468 | 4.236 | 4.072 |
| 0.990 | 10 | 10.044 | 7.559 | 6.552 | 5.994 | 5.636 | 5.386 |
| 0.995 | 10 | 12.827 | 9.427 | 8.081 | 7.343 | 6.872 | 6.545 |
| 0.900 | 11 | 3.225 | 2.860 | 2.660 | 2.536 | 2.451 | 2.389 |
| 0.950 | 11 | 4.844 | 3.982 | 3.587 | 3.357 | 3.204 | 3.095 |
| 0.975 | 11 | 6.724 | 5.256 | 4.630 | 4.275 | 4.044 | 3.881 |
| 0.990 | 11 | 9.646 | 7.206 | 6.217 | 5.668 | 5.316 | 5.069 |
| 0.995 | 11 | 12.226 | 8.912 | 7.600 | 6.881 | 6.422 | 6.102 |
| 0.900 | 15 | 3.073 | 2.695 | 2.490 | 2.361 | 2.273 | 2.208 |
| 0.950 | 15 | 4.543 | 3.682 | 3.287 | 3.056 | 2.901 | 2.790 |
| 0.975 | 15 | 6.200 | 4.765 | 4.153 | 3.804 | 3.576 | 3.415 |
| 0.990 | 15 | 8.683 | 6.359 | 5.417 | 4.893 | 4.556 | 4.318 |
| 0.995 | 15 | 10.798 | 7.701 | 6.476 | 5.803 | 5.372 | 5.071 |

Porazdelitev $F$ (nadaljevanje)



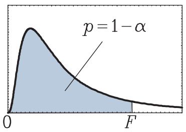

| $p = 1 - \alpha$ | $\nu_2$ | $\nu_1$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| 0.900 | 20 | 2.975 | 2.589 | 2.380 | 2.249 | 2.158 | 2.091 |
| 0.950 | 20 | 4.351 | 3.493 | 3.098 | 2.866 | 2.711 | 2.599 |
| 0.975 | 20 | 5.871 | 4.461 | 3.859 | 3.515 | 3.289 | 3.128 |
| 0.990 | 20 | 8.096 | 5.849 | 4.938 | 4.431 | 4.103 | 3.871 |
| 0.995 | 20 | 9.944 | 6.987 | 5.818 | 5.174 | 4.762 | 4.472 |
| 0.900 | 25 | 2.918 | 2.528 | 2.317 | 2.184 | 2.092 | 2.024 |
| 0.950 | 25 | 4.242 | 3.385 | 2.991 | 2.759 | 2.603 | 2.490 |
| 0.975 | 25 | 5.686 | 4.291 | 3.694 | 3.353 | 3.129 | 2.969 |
| 0.990 | 25 | 7.770 | 5.568 | 4.675 | 4.177 | 3.855 | 3.627 |
| 0.995 | 25 | 9.475 | 6.598 | 5.462 | 4.835 | 4.433 | 4.150 |
| 0.900 | 30 | 2.881 | 2.489 | 2.276 | 2.142 | 2.049 | 1.980 |
| 0.950 | 30 | 4.171 | 3.316 | 2.922 | 2.690 | 2.534 | 2.421 |
| 0.975 | 30 | 5.568 | 4.182 | 3.589 | 3.250 | 3.026 | 2.867 |
| 0.990 | 30 | 7.562 | 5.390 | 4.510 | 4.018 | 3.699 | 3.473 |
| 0.995 | 30 | 9.180 | 6.355 | 5.239 | 4.623 | 4.228 | 3.949 |
| 0.900 | 50 | 2.809 | 2.412 | 2.197 | 2.061 | 1.966 | 1.895 |
| 0.950 | 50 | 4.034 | 3.183 | 2.790 | 2.557 | 2.400 | 2.286 |
| 0.975 | 50 | 5.340 | 3.975 | 3.390 | 3.054 | 2.833 | 2.674 |
| 0.990 | 50 | 7.171 | 5.057 | 4.199 | 3.720 | 3.408 | 3.186 |
| 0.995 | 50 | 8.626 | 5.902 | 4.826 | 4.232 | 3.849 | 3.579 |
| 0.900 | 100 | 2.756 | 2.356 | 2.139 | 2.002 | 1.906 | 1.834 |
| 0.950 | 100 | 3.936 | 3.087 | 2.696 | 2.463 | 2.305 | 2.191 |
| 0.975 | 100 | 5.179 | 3.828 | 3.250 | 2.917 | 2.696 | 2.537 |
| 0.990 | 100 | 6.895 | 4.824 | 3.984 | 3.513 | 3.206 | 2.988 |
| 0.995 | 100 | 8.241 | 5.589 | 4.542 | 3.963 | 3.589 | 3.325 |
| 0.900 | $\infty$ | 2.706 | 2.303 | 2.084 | 1.945 | 1.847 | 1.774 |
| 0.950 | $\infty$ | 3.841 | 2.996 | 2.605 | 2.372 | 2.214 | 2.099 |
| 0.975 | $\infty$ | 5.024 | 3.689 | 3.116 | 2.786 | 2.566 | 2.408 |
| 0.990 | $\infty$ | 6.635 | 4.605 | 3.782 | 3.319 | 3.017 | 2.802 |
| 0.995 | $\infty$ | 7.879 | 5.298 | 4.279 | 3.715 | 3.350 | 3.091 |

Porazdelitev $F$ (nadaljevanje)



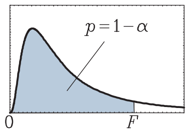

| $p=1-\alpha$ | $\nu_2$ | $\nu_1$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 7 | 8 | 9 | 10 | 11 | 15 |
| 0.900 | 1 | 58.9 | 59.4 | 59.9 | 60.2 | 60.5 | 61.2 |
| 0.950 | 1 | 236.8 | 238.9 | 240.5 | 241.9 | 243.0 | 245.9 |
| 0.975 | 1 | 948.2 | 956.6 | 963.3 | 968.6 | 973.0 | 984.9 |
| 0.990 | 1 | 5928.3 | 5981.0 | 6022.4 | 6055.9 | 6083.4 | 6157.0 |
| 0.995 | 1 | 23715.2 | 23923.8 | 24091.5 | 24221.8 | 24333.6 | 24631.6 |
| 0.900 | 2 | 9.35 | 9.37 | 9.38 | 9.39 | 9.40 | 9.42 |
| 0.950 | 2 | 19.35 | 19.37 | 19.38 | 19.40 | 19.40 | 19.43 |
| 0.975 | 2 | 39.36 | 39.37 | 39.39 | 39.40 | 39.41 | 39.43 |
| 0.990 | 2 | 99.36 | 99.38 | 99.39 | 99.40 | 99.41 | 99.43 |
| 0.995 | 2 | 199.36 | 199.38 | 199.39 | 199.39 | 199.42 | 199.43 |
| 0.900 | 3 | 5.266 | 5.252 | 5.240 | 5.230 | 5.222 | 5.200 |
| 0.950 | 3 | 8.887 | 8.845 | 8.812 | 8.785 | 8.763 | 8.703 |
| 0.975 | 3 | 14.624 | 14.540 | 14.473 | 14.419 | 14.374 | 14.253 |
| 0.990 | 3 | 27.671 | 27.489 | 27.345 | 27.228 | 27.132 | 26.872 |
| 0.995 | 3 | 44.434 | 44.125 | 43.881 | 43.685 | 43.525 | 43.085 |
| 0.900 | 4 | 3.979 | 3.955 | 3.936 | 3.920 | 3.907 | 3.870 |
| 0.950 | 4 | 6.094 | 6.041 | 5.999 | 5.964 | 5.936 | 5.858 |
| 0.975 | 4 | 9.074 | 8.980 | 8.905 | 8.844 | 8.794 | 8.657 |
| 0.990 | 4 | 14.976 | 14.799 | 14.659 | 14.546 | 14.452 | 14.198 |
| 0.995 | 4 | 21.622 | 21.352 | 21.138 | 20.967 | 20.824 | 20.438 |
| 0.900 | 5 | 3.368 | 3.339 | 3.316 | 3.297 | 3.282 | 3.238 |
| 0.950 | 5 | 4.876 | 4.818 | 4.772 | 4.735 | 4.704 | 4.619 |
| 0.975 | 5 | 6.853 | 6.757 | 6.681 | 6.619 | 6.568 | 6.428 |
| 0.990 | 5 | 10.456 | 10.289 | 10.158 | 10.051 | 9.963 | 9.722 |
| 0.995 | 5 | 14.200 | 13.961 | 13.772 | 13.618 | 13.491 | 13.146 |
| 0.900 | 6 | 3.014 | 2.983 | 2.958 | 2.937 | 2.920 | 2.871 |
| 0.950 | 6 | 4.207 | 4.147 | 4.099 | 4.060 | 4.027 | 3.938 |
| 0.975 | 6 | 5.695 | 5.600 | 5.523 | 5.461 | 5.410 | 5.269 |
| 0.990 | 6 | 8.260 | 8.102 | 7.976 | 7.874 | 7.790 | 7.559 |
| 0.995 | 6 | 10.786 | 10.566 | 10.391 | 10.250 | 10.133 | 9.814 |

Porazdelitev $F$ (nadaljevanje)



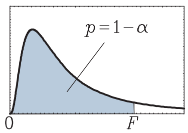

| $p = 1 - \alpha$ | $\nu_2$ | $\nu_1$ = 7 | 8 | 9 | 10 | 11 | 15 |
|---|---|---|---|---|---|---|---|
| 0.900 | 7 | 2.785 | 2.752 | 2.725 | 2.703 | 2.684 | 2.632 |
| 0.950 | 7 | 3.787 | 3.726 | 3.677 | 3.637 | 3.603 | 3.511 |
| 0.975 | 7 | 4.995 | 4.899 | 4.823 | 4.761 | 4.709 | 4.568 |
| 0.990 | 7 | 6.993 | 6.840 | 6.719 | 6.620 | 6.538 | 6.314 |
| 0.995 | 7 | 8.885 | 8.678 | 8.514 | 8.380 | 8.270 | 7.968 |
| 0.900 | 8 | 2.624 | 2.589 | 2.561 | 2.538 | 2.519 | 2.464 |
| 0.950 | 8 | 3.500 | 3.438 | 3.388 | 3.347 | 3.313 | 3.218 |
| 0.975 | 8 | 4.529 | 4.433 | 4.357 | 4.295 | 4.243 | 4.101 |
| 0.990 | 8 | 6.178 | 6.029 | 5.911 | 5.814 | 5.734 | 5.515 |
| 0.995 | 8 | 7.694 | 7.496 | 7.339 | 7.211 | 7.105 | 6.814 |
| 0.900 | 9 | 2.505 | 2.469 | 2.440 | 2.416 | 2.396 | 2.340 |
| 0.950 | 9 | 3.293 | 3.230 | 3.179 | 3.137 | 3.102 | 3.006 |
| 0.975 | 9 | 4.197 | 4.102 | 4.026 | 3.964 | 3.912 | 3.769 |
| 0.990 | 9 | 5.613 | 5.467 | 5.351 | 5.257 | 5.178 | 4.962 |
| 0.995 | 9 | 6.885 | 6.693 | 6.541 | 6.417 | 6.314 | 6.032 |
| 0.900 | 10 | 2.414 | 2.377 | 2.347 | 2.323 | 2.302 | 2.244 |
| 0.950 | 10 | 3.135 | 3.072 | 3.020 | 2.978 | 2.943 | 2.845 |
| 0.975 | 10 | 3.950 | 3.855 | 3.779 | 3.717 | 3.665 | 3.522 |
| 0.990 | 10 | 5.200 | 5.057 | 4.942 | 4.849 | 4.772 | 4.558 |
| 0.995 | 10 | 6.303 | 6.116 | 5.968 | 5.847 | 5.746 | 5.471 |
| 0.900 | 11 | 2.342 | 2.304 | 2.274 | 2.248 | 2.227 | 2.167 |
| 0.950 | 11 | 3.012 | 2.948 | 2.896 | 2.854 | 2.818 | 2.719 |
| 0.975 | 11 | 3.759 | 3.664 | 3.588 | 3.526 | 3.474 | 3.330 |
| 0.990 | 11 | 4.886 | 4.744 | 4.632 | 4.539 | 4.462 | 4.251 |
| 0.995 | 11 | 5.865 | 5.682 | 5.537 | 5.418 | 5.320 | 5.049 |
| 0.900 | 15 | 2.158 | 2.119 | 2.086 | 2.059 | 2.037 | 1.972 |
| 0.950 | 15 | 2.707 | 2.641 | 2.588 | 2.544 | 2.507 | 2.403 |
| 0.975 | 15 | 3.293 | 3.199 | 3.123 | 3.060 | 3.008 | 2.862 |
| 0.990 | 15 | 4.142 | 4.004 | 3.895 | 3.805 | 3.730 | 3.522 |
| 0.995 | 15 | 4.847 | 4.674 | 4.536 | 4.424 | 4.329 | 4.070 |

Porazdelitev $F$ (nadaljevanje)



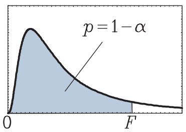

| $p = 1 - \alpha$ | $\nu_2$ | $\nu_1$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 7 | 8 | 9 | 10 | 11 | 15 |
| 0.900 | 20 | 2.040 | 1.999 | 1.965 | 1.937 | 1.913 | 1.845 |
| 0.950 | 20 | 2.514 | 2.447 | 2.393 | 2.348 | 2.310 | 2.203 |
| 0.975 | 20 | 3.007 | 2.913 | 2.837 | 2.774 | 2.721 | 2.573 |
| 0.990 | 20 | 3.699 | 3.564 | 3.457 | 3.368 | 3.294 | 3.088 |
| 0.995 | 20 | 4.257 | 4.090 | 3.956 | 3.847 | 3.756 | 3.502 |
| 0.900 | 25 | 1.971 | 1.929 | 1.895 | 1.866 | 1.841 | 1.771 |
| 0.950 | 25 | 2.405 | 2.337 | 2.282 | 2.236 | 2.198 | 2.089 |
| 0.975 | 25 | 2.848 | 2.753 | 2.677 | 2.613 | 2.560 | 2.411 |
| 0.990 | 25 | 3.457 | 3.324 | 3.217 | 3.129 | 3.056 | 2.850 |
| 0.995 | 25 | 3.939 | 3.776 | 3.645 | 3.537 | 3.447 | 3.196 |
| 0.900 | 30 | 1.927 | 1.884 | 1.849 | 1.819 | 1.794 | 1.722 |
| 0.950 | 30 | 2.334 | 2.266 | 2.211 | 2.165 | 2.126 | 2.015 |
| 0.975 | 30 | 2.746 | 2.651 | 2.575 | 2.511 | 2.458 | 2.307 |
| 0.990 | 30 | 3.305 | 3.173 | 3.067 | 2.979 | 2.906 | 2.700 |
| 0.995 | 30 | 3.742 | 3.580 | 3.451 | 3.344 | 3.255 | 3.006 |
| 0.900 | 50 | 1.840 | 1.796 | 1.760 | 1.729 | 1.703 | 1.627 |
| 0.950 | 50 | 2.199 | 2.130 | 2.073 | 2.026 | 1.986 | 1.871 |
| 0.975 | 50 | 2.553 | 2.458 | 2.381 | 2.317 | 2.263 | 2.109 |
| 0.990 | 50 | 3.020 | 2.890 | 2.785 | 2.698 | 2.625 | 2.419 |
| 0.995 | 50 | 3.376 | 3.219 | 3.092 | 2.988 | 2.900 | 2.653 |
| 0.900 | 100 | 1.778 | 1.732 | 1.695 | 1.663 | 1.636 | 1.557 |
| 0.950 | 100 | 2.103 | 2.032 | 1.975 | 1.927 | 1.886 | 1.768 |
| 0.975 | 100 | 2.417 | 2.321 | 2.244 | 2.179 | 2.124 | 1.968 |
| 0.990 | 100 | 2.823 | 2.694 | 2.590 | 2.503 | 2.430 | 2.223 |
| 0.995 | 100 | 3.127 | 2.972 | 2.847 | 2.744 | 2.657 | 2.411 |
| 0.900 | $\infty$ | 1.717 | 1.670 | 1.632 | 1.599 | 1.570 | 1.487 |
| 0.950 | $\infty$ | 2.010 | 1.938 | 1.880 | 1.831 | 1.789 | 1.666 |
| 0.975 | $\infty$ | 2.288 | 2.192 | 2.114 | 2.048 | 1.993 | 1.833 |
| 0.990 | $\infty$ | 2.639 | 2.511 | 2.407 | 2.321 | 2.248 | 2.039 |
| 0.995 | $\infty$ | 2.897 | 2.744 | 2.621 | 2.519 | 2.432 | 2.187 |

Porazdelitev $F$ (nadaljevanje)



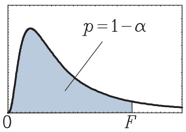

| $p=1-\alpha$ | $\nu_2$ | $\nu_1$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | 20 | 25 | 30 | 50 | 100 | $\infty$ |
| 0.900 | 1 | 61.7 | 62.1 | 62.3 | 62.7 | 63.0 | 63.3 |
| 0.950 | 1 | 248.0 | 249.3 | 250.1 | 251.8 | 253.0 | 254.3 |
| 0.975 | 1 | 993.1 | 998.1 | 1001.4 | 1008.1 | 1013.2 | 1018.3 |
| 0.990 | 1 | 6208.7 | 6239.9 | 6260.4 | 6302.3 | 6333.9 | 6365.6 |
| 0.995 | 1 | 24836.5 | 24959.4 | 25041.4 | 25212.8 | 25339.4 | 25466.1 |
| 0.900 | 2 | 9.44 | 9.45 | 9.46 | 9.47 | 9.48 | 9.49 |
| 0.950 | 2 | 19.45 | 19.46 | 19.46 | 19.48 | 19.49 | 19.50 |
| 0.975 | 2 | 39.45 | 39.46 | 39.46 | 39.48 | 39.49 | 39.50 |
| 0.990 | 2 | 99.45 | 99.46 | 99.47 | 99.48 | 99.49 | 99.50 |
| 0.995 | 2 | 199.45 | 199.45 | 199.48 | 199.48 | 199.48 | 199.51 |
| 0.900 | 3 | 5.184 | 5.175 | 5.168 | 5.155 | 5.144 | 5.134 |
| 0.950 | 3 | 8.660 | 8.634 | 8.617 | 8.581 | 8.554 | 8.526 |
| 0.975 | 3 | 14.167 | 14.115 | 14.081 | 14.010 | 13.956 | 13.902 |
| 0.990 | 3 | 26.690 | 26.579 | 26.504 | 26.354 | 26.241 | 26.125 |
| 0.995 | 3 | 42.779 | 42.590 | 42.466 | 42.211 | 42.022 | 41.829 |
| 0.900 | 4 | 3.844 | 3.828 | 3.817 | 3.795 | 3.778 | 3.761 |
| 0.950 | 4 | 5.803 | 5.769 | 5.746 | 5.699 | 5.664 | 5.628 |
| 0.975 | 4 | 8.560 | 8.501 | 8.461 | 8.381 | 8.319 | 8.257 |
| 0.990 | 4 | 14.019 | 13.911 | 13.838 | 13.690 | 13.577 | 13.463 |
| 0.995 | 4 | 20.167 | 20.003 | 19.892 | 19.667 | 19.497 | 19.325 |
| 0.900 | 5 | 3.207 | 3.187 | 3.174 | 3.147 | 3.126 | 3.105 |
| 0.950 | 5 | 4.558 | 4.521 | 4.496 | 4.444 | 4.405 | 4.365 |
| 0.975 | 5 | 6.329 | 6.268 | 6.227 | 6.144 | 6.080 | 6.015 |
| 0.990 | 5 | 9.553 | 9.449 | 9.379 | 9.238 | 9.130 | 9.020 |
| 0.995 | 5 | 12.903 | 12.756 | 12.656 | 12.454 | 12.300 | 12.144 |
| 0.900 | 6 | 2.836 | 2.815 | 2.800 | 2.770 | 2.746 | 2.722 |
| 0.950 | 6 | 3.874 | 3.835 | 3.808 | 3.754 | 3.712 | 3.669 |
| 0.975 | 6 | 5.168 | 5.107 | 5.065 | 4.980 | 4.915 | 4.849 |
| 0.990 | 6 | 7.396 | 7.296 | 7.229 | 7.091 | 6.987 | 6.880 |
| 0.995 | 6 | 9.589 | 9.451 | 9.358 | 9.170 | 9.026 | 8.879 |

Porazdelitev $F$ (nadaljevanje)




| | | $\nu_1$ | | | | | |
|---|---|---|---|---|---|---|---|
| $p = 1-\alpha$ | $\nu_2$ | 20 | 25 | 30 | 50 | 100 | $\infty$ |
| 0.900 | 7 | 2.595 | 2.571 | 2.555 | 2.523 | 2.497 | 2.471 |
| 0.950 | 7 | 3.445 | 3.404 | 3.376 | 3.319 | 3.275 | 3.230 |
| 0.975 | 7 | 4.467 | 4.405 | 4.362 | 4.276 | 4.210 | 4.142 |
| 0.990 | 7 | 6.155 | 6.058 | 5.992 | 5.858 | 5.755 | 5.650 |
| 0.995 | 7 | 7.754 | 7.623 | 7.534 | 7.354 | 7.217 | 7.076 |
| 0.900 | 8 | 2.425 | 2.400 | 2.383 | 2.348 | 2.321 | 2.293 |
| 0.950 | 8 | 3.150 | 3.108 | 3.079 | 3.020 | 2.975 | 2.928 |
| 0.975 | 8 | 3.999 | 3.937 | 3.894 | 3.807 | 3.739 | 3.670 |
| 0.990 | 8 | 5.359 | 5.263 | 5.198 | 5.065 | 4.963 | 4.859 |
| 0.995 | 8 | 6.608 | 6.482 | 6.396 | 6.222 | 6.087 | 5.951 |
| 0.900 | 9 | 2.298 | 2.272 | 2.255 | 2.218 | 2.189 | 2.159 |
| 0.950 | 9 | 2.936 | 2.893 | 2.864 | 2.803 | 2.756 | 2.707 |
| 0.975 | 9 | 3.667 | 3.604 | 3.560 | 3.472 | 3.403 | 3.333 |
| 0.990 | 9 | 4.808 | 4.713 | 4.649 | 4.517 | 4.415 | 4.311 |
| 0.995 | 9 | 5.832 | 5.708 | 5.625 | 5.454 | 5.322 | 5.188 |
| 0.900 | 10 | 2.201 | 2.174 | 2.155 | 2.117 | 2.087 | 2.055 |
| 0.950 | 10 | 2.774 | 2.730 | 2.700 | 2.637 | 2.588 | 2.538 |
| 0.975 | 10 | 3.419 | 3.355 | 3.311 | 3.221 | 3.152 | 3.080 |
| 0.990 | 10 | 4.405 | 4.311 | 4.247 | 4.115 | 4.014 | 3.909 |
| 0.995 | 10 | 5.274 | 5.153 | 5.071 | 4.902 | 4.772 | 4.639 |
| 0.900 | 11 | 2.123 | 2.095 | 2.076 | 2.036 | 2.005 | 1.972 |
| 0.950 | 11 | 2.646 | 2.601 | 2.570 | 2.507 | 2.457 | 2.404 |
| 0.975 | 11 | 3.226 | 3.162 | 3.118 | 3.027 | 2.956 | 2.883 |
| 0.990 | 11 | 4.099 | 4.005 | 3.941 | 3.810 | 3.708 | 3.602 |
| 0.995 | 11 | 4.855 | 4.736 | 4.654 | 4.488 | 4.359 | 4.226 |
| 0.900 | 15 | 1.924 | 1.894 | 1.873 | 1.828 | 1.793 | 1.755 |
| 0.950 | 15 | 2.328 | 2.280 | 2.247 | 2.178 | 2.123 | 2.066 |
| 0.975 | 15 | 2.756 | 2.689 | 2.644 | 2.549 | 2.474 | 2.395 |
| 0.990 | 15 | 3.372 | 3.278 | 3.214 | 3.081 | 2.977 | 2.868 |
| 0.995 | 15 | 3.883 | 3.766 | 3.687 | 3.523 | 3.394 | 3.260 |

Porazdelitev $F$ (nadaljevanje)




| $p=1-\alpha$ | $\nu_2$ | $\nu_1$ = 20 | 25 | 30 | 50 | 100 | $\infty$ |
|---|---|---|---|---|---|---|---|
| 0.900 | 20 | 1.794 | 1.761 | 1.738 | 1.690 | 1.650 | 1.607 |
| 0.950 | 20 | 2.124 | 2.074 | 2.039 | 1.966 | 1.907 | 1.843 |
| 0.975 | 20 | 2.464 | 2.396 | 2.349 | 2.249 | 2.170 | 2.085 |
| 0.990 | 20 | 2.938 | 2.843 | 2.778 | 2.643 | 2.535 | 2.421 |
| 0.995 | 20 | 3.318 | 3.203 | 3.123 | 2.959 | 2.828 | 2.690 |
| 0.900 | 25 | 1.718 | 1.683 | 1.659 | 1.607 | 1.565 | 1.518 |
| 0.950 | 25 | 2.007 | 1.955 | 1.919 | 1.842 | 1.779 | 1.711 |
| 0.975 | 25 | 2.300 | 2.230 | 2.182 | 2.079 | 1.996 | 1.906 |
| 0.990 | 25 | 2.699 | 2.604 | 2.538 | 2.400 | 2.289 | 2.169 |
| 0.995 | 25 | 3.013 | 2.898 | 2.819 | 2.652 | 2.519 | 2.377 |
| 0.900 | 30 | 1.667 | 1.632 | 1.606 | 1.552 | 1.507 | 1.456 |
| 0.950 | 30 | 1.932 | 1.878 | 1.841 | 1.761 | 1.695 | 1.622 |
| 0.975 | 30 | 2.195 | 2.124 | 2.074 | 1.968 | 1.882 | 1.787 |
| 0.990 | 30 | 2.549 | 2.453 | 2.386 | 2.245 | 2.131 | 2.006 |
| 0.995 | 30 | 2.823 | 2.708 | 2.628 | 2.459 | 2.323 | 2.176 |
| 0.900 | 50 | 1.568 | 1.529 | 1.502 | 1.441 | 1.388 | 1.327 |
| 0.950 | 50 | 1.784 | 1.727 | 1.687 | 1.599 | 1.525 | 1.438 |
| 0.975 | 50 | 1.993 | 1.919 | 1.866 | 1.752 | 1.656 | 1.545 |
| 0.990 | 50 | 2.265 | 2.167 | 2.098 | 1.949 | 1.825 | 1.683 |
| 0.995 | 50 | 2.470 | 2.353 | 2.272 | 2.097 | 1.951 | 1.786 |
| 0.900 | 100 | 1.494 | 1.453 | 1.423 | 1.355 | 1.293 | 1.214 |
| 0.950 | 100 | 1.676 | 1.616 | 1.573 | 1.477 | 1.392 | 1.283 |
| 0.975 | 100 | 1.849 | 1.770 | 1.715 | 1.592 | 1.483 | 1.347 |
| 0.990 | 100 | 2.067 | 1.965 | 1.893 | 1.735 | 1.598 | 1.427 |
| 0.995 | 100 | 2.227 | 2.108 | 2.024 | 1.840 | 1.681 | 1.485 |
| 0.900 | $\infty$ | 1.421 | 1.375 | 1.342 | 1.263 | 1.185 | 1.000 |
| 0.950 | $\infty$ | 1.571 | 1.506 | 1.459 | 1.350 | 1.243 | 1.000 |
| 0.975 | $\infty$ | 1.708 | 1.626 | 1.566 | 1.428 | 1.296 | 1.000 |
| 0.990 | $\infty$ | 1.878 | 1.773 | 1.696 | 1.523 | 1.358 | 1.000 |
| 0.995 | $\infty$ | 2.000 | 1.877 | 1.789 | 1.590 | 1.402 | 1.000 |

## 15.5 Preverjanje Kolmogorov-Smirnov

| $n$ | $D_{n,\alpha=0.10}$ | $D_{n,\alpha=0.05}$ | $D_{n,\alpha=0.02}$ | $D_{n,\alpha=0.01}$ |
|---|---|---|---|---|
| 5 | 0.5088 | 0.5646 | 0.6309 | 0.6767 |
| 6 | 0.4681 | 0.5195 | 0.5804 | 0.6226 |
| 7 | 0.4359 | 0.4838 | 0.5405 | 0.5798 |
| 8 | 0.4097 | 0.4546 | 0.5080 | 0.5448 |
| 9 | 0.3877 | 0.4302 | 0.4807 | 0.5156 |
| 10 | 0.3690 | 0.4094 | 0.4575 | 0.4907 |
| 11 | 0.3527 | 0.3914 | 0.4373 | 0.4691 |
| 12 | 0.3385 | 0.3756 | 0.4197 | 0.4501 |
| 13 | 0.3258 | 0.3616 | 0.4040 | 0.4333 |
| 14 | 0.3145 | 0.3490 | 0.3900 | 0.4183 |
| 15 | 0.3043 | 0.3377 | 0.3773 | 0.4047 |
| 16 | 0.2951 | 0.3274 | 0.3659 | 0.3924 |
| 17 | 0.2866 | 0.3181 | 0.3554 | 0.3812 |
| 18 | 0.2789 | 0.3095 | 0.3458 | 0.3709 |
| 19 | 0.2717 | 0.3015 | 0.3369 | 0.3614 |
| 20 | 0.2651 | 0.2942 | 0.3287 | 0.3525 |
| 21 | 0.2589 | 0.2873 | 0.3210 | 0.3444 |
| 22 | 0.2532 | 0.2810 | 0.3139 | 0.3367 |
| 23 | 0.2478 | 0.2750 | 0.3072 | 0.3296 |
| 24 | 0.2428 | 0.2694 | 0.3010 | 0.3228 |
| 25 | 0.2380 | 0.2641 | 0.2951 | 0.3165 |
| 30 | 0.2179 | 0.2418 | 0.2701 | 0.2898 |
| 35 | 0.2021 | 0.2243 | 0.2506 | 0.2688 |
| 40 | 0.1894 | 0.2102 | 0.2348 | 0.2519 |
| 50 | 0.1698 | 0.1885 | 0.2106 | 0.2259 |
| 60 | 0.1553 | 0.1723 | 0.1926 | 0.2065 |
| 70 | 0.1440 | 0.1598 | 0.1785 | 0.1915 |
| 80 | 0.1348 | 0.1496 | 0.1672 | 0.1793 |
| 90 | 0.1272 | 0.1412 | 0.1578 | 0.1692 |
| 100 | 0.1208 | 0.1341 | 0.1498 | 0.1607 |
| 120 | 0.1104 | 0.1225 | 0.1369 | 0.1468 |
| 150 | 0.0989 | 0.1097 | 0.1226 | 0.1315 |
| 200 | 0.0858 | 0.0952 | 0.1063 | 0.1141 |
| 250 | 0.0768 | 0.0852 | 0.0952 | 0.1021 |
| 300 | 0.0701 | 0.0778 | 0.0870 | 0.0933 |
| 350 | 0.0650 | 0.0721 | 0.0806 | 0.0864 |
| 400 | 0.0608 | 0.0675 | 0.0754 | 0.0809 |
| 500 | 0.0544 | 0.0604 | 0.0675 | 0.0724 |

# Literatura

Literature, ki obravnavajo verjetnostni račun, statistiko in sorodne vede, je mnogo, napisana je v različnih jezikih, tudi slovenščini. Tukaj prikazani spisek, ki vsebuje le slovenske in angleške vire, ni popoln, čeprav obsega mnogo več literature, kot bi je potreboval tudi najbolj znanja željni študent. Poleg običajnih podatkov o knjigi so tukaj podane še oznake knjižnic, v katerih lahko knjigo dobimo (CTK - Centralna tehniška knjižnica, FGG - knjižnica Fakultete za gradbeništvo in geodezijo, PF - knjižnica Pedagoške fakultete, IG - knjižnica Inštituta za geologijo, KI - knjižnica Kemijskega inštituta).


[1] Barnet, V.; Turkman, K. F., Statistics for the Environment (Water Related Issues), John Wiley&Sons, 1994. (IG)

[2] Benjamin, J. R.; Cornell, C. A., Probability, Statistics, and Decision for Civil Engineers, McGraw-Hill, 1970. (FGG)

[3] Blejec, M.; Statistične metode za ekonomiste, Univerza v Ljubljani, Ljubljana, 1976.

[4] Bury, K., Statistical Distributions in Engineering, Cambridge University Press, 1999. (CTK)

[5] Cryer, A.; Davies, M.; Francis, B.; Goodall, G., Statistics 5&6, Hodder&Staughton, 1998. (FGG)

[6] Davies, M.; Francis, B., Statistics 4, Hodder&Staughton, 1997. (FGG)

[7] Drnovšek, R.; Košir, T.; Kramar, E.; Lešnjak, G., Zbirka rešenih nalog iz verjetnostnega računa, DMFA Slovenije, 1998. (CTK)

[8] Everitt, B. S., The Cambridge dictionary of Statistics, Cambridge University Press, 1998. (CTK)

[9] Gumbel, E. J., Statistics of Extrems, Columbia University Press, 1958.

[10] Hayes, W. L., Statistics, Harcourt Brace College Publishers, 1994. (CTK)

[11] Hines, W. W.; Montgomery, D.C., Probability and Statistics in Engineering and Management Science, John Wiley&Sons, 1990. (CTK)

[12] Huff, D., How to lie with statistics, W.W. Norton&Company, 1993.

[13] Jamnik, R., Uvod v matematično statistiko, DMFA Slovenije, 1976. (FGG)

[14] Jamnik, R., Verjetnostni račun, Mladinska knjiga, 1971. (FGG)

[15] Jamnik, R., Verjetnostni račun in statistika, DMFA Slovenije, 1986. (FGG)

[16] Kinney, J. J., Probability, An Introduction with Statistical Applications, John Wiley&Sons, 1997. (CTK)

[17] Košmelj, K., Opisna statistika na zgledih, Didakta, 1995. (CTK)

[18] Košmelj, B., Arh, F., Doberšek Urbanc, A., Ferligoj, A., Omladič M., Statistični terminološki slovar, Statistično društvo Slovenije, Statistični urad Republike Slovenije, 2001. (CTK)

[19] Kottegoda, N. T.; Rosso, R., Statistics, Probability and Reliability for Civil and Environmental Engineering, McGraw-Hill, 1997. (IG)

[20] McCuen, R. H.; Snyder, W. M., Hydrologic Modeling: Statistical Methods and Applications, Prentice-Hall, 1986.

[21] Metcalfe, A. V., Statistics in Civil Engineering, Arnold, 1997. (CTK)

[22] Montgomery, D. C.; Runger, G. C., Applied Statistics and Probability for Engineers, John Wiley&Sons, 1994. (CTK)

[23] Moore, D. S., Statistics, Concepts and Contraversies, W.H. Freeman and Comp., 1996. (CTK)

[24] Mosteller, F., Fifty challanging problems in probability, Dover Publications, 1965.

[25] Phillips, J. L., How to Think about Statistics, W.H. Freeman and Comp., 1996. (CTK)

[26] Rao, P. S. R. S., Variance Components Estimation, Chapman&Hall, 1997. (CTK)

[27] Reiss, R. D.; Thomas, M., Statistical Analysis od Extreme Values, Birkhaeuser Verlag, 1997. (PF)

[28] Ross, S. M., Introduction to Probability Models, Academic Press, 1997. (CTK)

[29] Rozanov, Y. A., Probability Theory, A Concise Course, Dover Publications, 1969.

[30] Sadovskii, L. E.; Sadovskii, A. L. Mathematics and Sport, American Mathematical Society, 1991.

[31] Sanders, D. H., Statistics, A Fresh Approach, McGraw-Hill, 1990. (KI)

[32] Spiegel, M. R., Theory and Problems of Probability and Statistics, McGraw-Hill, 1997. (CTK)

[33] Spiegel, M. R., Theory and Problems of Statistics, McGraw-Hill, 1992. (BTF, CTK)

[34] Vadnal, A., Elementarni uvod v verjetnostni račun, DZS, 1979. (CTK)

[35] Vining, G. G., Statistical Methods for Engineers, Duxbury Press, 1998. (CTK)

## Nekaj zanimivih internetnih naslovov

[1] Cedilnik, A., Valantič,T., Statistični urad republike Slovenije. Internet: http://www.sigov.si/zrs/slo/

[2] Grinstead, C. M.; Snell, J. L., Introduction to Probability, 1998. Internet: http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

[3] Hopkins, W. G., A New View of Statistics, 2002. Internet: http://www.sportsci.org/resource/stats/index.html

[4] NIST/SEMATECH e-Handbook of Statistical Methods, Engineering statistics Handbook, 2002. Internet: http://www.itl.nist.gov/ div898/handbook/index.htm

[5] Piele D., Introduction to probability, Mathematica notebooks. Internet: http://www.uwp.edu/academic/mathematics/probability/ index.htm

[6] Pollett, P., Bob Dobrow, B., The probability web, 1995-2002. Internet: http://www.mathcs.carleton.edu/probweb/probweb.html

[7] Siegrist, K., Virtual Laboratories in Probability and Statistics, 1997-2001. Internet: http://www.math.uah.edu/stat/index.html

[8] StatSoft, Inc. Electronic Statistics Textbook. Tulsa, OK: StatSoft, 2002. Internet: http://www.statsoft.com/textbook/stathome.html

[9] Wolfram Research, Statistics with Mathematica. Internet: http:// www.wri.com/solutions/statistics/functions/functions_testing.html

# Stvarno kazalo