VAJE IZ VERJETNOSTI IN STATISTIKE

Martin Raič

1. Osnove kombinatorike

Pravilo vsote, pravilo produkta. Variacije, kombinacije in permutacije.

1. Na koliko načinov lahko opremimo dnevno sobo, če imamo na voljo 4 vrste parketa, 3 vrste nelesnih talnih oblog in 5 vrst pohištva?

2. Koliko je:

   a) vseh trimestnih števil?
   b) vseh sodih trimestnih števil?
   c) vseh trimestnih števil s sodo prvo števko?
   d) vseh trimestnih števil s samimi enakimi števkami?
   e) vseh trimestnih števil s samimi različnimi števkami?
   f) vseh trimestnih števil, ki so palindromi?

3. Na koliko načinov lahko iz škatle s petimi različnimi kroglicami vzamemo:

   a) eno kroglico
   b) dve kroglici
   c) tri kroglice

   Pri tem ločite primer, ko kroglice vračamo, in primer, ko jih ne vračamo. Poleg tega ločite primer, ko je vrstni red jemanja pomemben, in primer, ko ni pomemben.

   Pri različici, ko kroglic ne vračamo in vrstni red jemanja ni pomemben, primerjajte rezultata iz točk b) in c).

4. Na koliko načinov lahko na ravno polico razporedimo 3 begonije in 4 fuksije? Pri tem ločite primer, ko razločujemo vse cvetlice, in primer, ko cvetlic iste vrste med seboj ne razločujemo.

Splošneje: iz škatle z n različnimi kroglicami lahko izvlečemo k kroglic na naslednje število načinov:

	vrstni red vlečenja	
	pomemben	ni pomemben
vračamo	${}^{(\mathrm{p})}V_n^k = n^k$	${}^{(\mathrm{p})}C_n^k = \binom{n+k-1}{k}$
ne vračamo	$V_n^k = n(n-1)\ldots(n-k+1)$	$C_n^k = \binom{n}{k} = \dfrac{V_n^k}{k!}$

Velja še $V_n^k = \dfrac{n!}{(n-k)!}$ *in* $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!} = \binom{n}{n-k}$.

5. V posodi je 6 rdečih in 4 modre kroglice, vse kroglice so različne. Na koliko načinov lahko iz posode brez vračanja vzamemo (vrstni red ni pomemben):

   a) 4 rdeče in 2 modri kroglici?

   b) 4 kroglice, a od tega vsaj eno rdečo in vsaj eno modro?

6. Na koliko načinov lahko v ravno vrsto položimo tri brezove, dve leskovi in štiri vrbove šibe, če:

   a) vse šibe razločujemo in ni omejitev?

   b) vse šibe razločujemo ter morajo priti najprej brezove, nato leskove in nazadnje vrbove?

   c) vse šibe razločujemo in morajo biti šibe posamezne vrste skupaj?

   d) šib iste vrste med seboj ne razločujemo in ni omejitev?

7. Na koliko načinov lahko razvrstimo šest otrok (ki jih razločujemo) na vrtiljak s šestimi sedeži (ki jih ločimo le glede na njihovo medsebojno lego)? Kaj pa na vrtiljak z desetimi sedeži? Na vsak sedež gre največ en otrok.

2. Elementarna verjetnost

Diskretni pogled na verjetnost v povezavi s kombinatoriko, geometrijski pogled na verjetnost. Računanje z dogodki.

Če so vsi izidi enako verjetni, velja:

$$P = \frac{\textit{število ugodnih izidov}}{\textit{število vseh izidov}}$$

1. Vržemo dve standardni kocki, vsi izidi (kjer kocki ločimo) so enako verjetni. Kolikšna je verjetnost, da je skupno število pik enako 8?
2. Zakonca načrtujeta štiri otroke. Kaj je verjetneje: da bosta oba spola enako zastopana ali da bodo trije enega, eden pa nasprotnega spola? Vse razporeditve spolov so enako verjetne.
3. Vržemo pet kock, vse kombinacije pik (kjer kocke ločimo) so enako verjetne. Kolikšna je verjetnost, da je na vsaj eni kocki padla šestica?
4. V posodi je 5 belih, 4 črne in 3 rdeče kroglice. Iz posode potegnemo tri kroglice. Kolikšna je verjetnost, da bo med njimi po ena kroglica vsake barve, če:
   a) kroglice vračamo?
   b) kroglic ne vračamo?

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

5. Iz posode, v kateri so 3 rdeče, 2 zeleni in 5 belih kroglic, na slepo in brez vračanja izvlečemo dve kroglici. Kolikšna je verjetnost, da je prva rdeča ali pa druga zelena?
6. Kolikšna je verjetnost, da v skupini n ljudi obstajata dva, ki imata rojstni dan na isti dan? Prestopna leta zanemarite.
7. Dan je dobro premešan kup 16 kart, med katerimi so štirje piki. Kolikšna je verjetnost, da sta med prvimi osmimi kartami natanko dva pika?
8. Med 100 izdelki v seriji je 10 okvarjenih. Iz serije na slepo izberemo 10 izdelkov. Če je med njimi več kot en okvarjen, serijo zavrnemo. Kolikšna je verjetnost, da se bo to zgodilo?
9. V posodi je 8 belih, 4 črne in 2 rdeči kroglici. Iz posode brez vračanja potegnemo sedem kroglic. Kolikšna je verjetnost, da bo razmerje barv enako kot v posodi?

Računanje z dogodki:

$$
\begin{array}{rlrl}
A\cup B &= B\cup A & A\cap B &= B\cap A\\
A\cup(B\cup C) &= (A\cup B)\cup C & A\cap(B\cap C) &= (A\cap B)\cap C\\
A\cap(B\cup C) &= (A\cap B)\cup(A\cap C) & A\cup(B\cap C) &= (A\cup B)\cap(A\cup C)\\
A\cup G &= G & A\cap G &= A\\
A\cup N &= A & A\cap N &= N\\
\overline{A\cup B} &= \bar A\cap\bar B & \overline{A\cap B} &= \bar A\cup\bar B\\
A\cup\bar A &= G & A\cap\bar A &= N\\
\overline{\overline{A}} &= A & &
\end{array}
$$

10. Poenostavite naslednji izraz z dogodki:

$$(B\cup C)\cap(B\cup\overline{C})\cap(\overline{B}\cup C)$$

11. Dani so dogodki A, B in C. Matematično zapišite:

    a) dogodek, da se ne zgodi niti A niti B niti C;

    b) dogodek, da se zgodi natanko eden od teh treh dogodkov;

    c) dogodek, da se zgodita vsaj dva od teh treh dogodkov.

    Izračunajte še verjetnosti zgornjih dogodkov, če veste, da je $P(A) = 0\dot{}3$, $P(B) = 0\dot{}45$, $P(C) = 0\dot{}6$, $P(A\cap B) = 0\dot{}2$, $P(A\cap C) = 0\dot{}2$, $P(B\cap C) = 0\dot{}3$ in $P(A\cap B\cap C) = 0\dot{}1$.

    Opomba. Dogodki pod a), b) in c) tvorijo popoln sistem dogodkov.

Načelo vključitev in izključitev:

$$
\begin{aligned}
P(A_1\cup\cdots\cup A_n) = {} & P(A_1)+\cdots+P(A_n) - \\
& - P(A_1\cap A_2) - P(A_1\cap A_3) - \cdots - P(A_{n-1}\cap A_n) + \\
& \cdots \\
& + (-1)^{n+1}\,P(A_1\cap\cdots\cap A_n)
\end{aligned}
$$

12. Mama napiše pet različnih pisem in pripravi pet kuvert za ta pisma s samimi različnimi naslovi. Mimo pride navihani Petrček in povsem slučajno vtakne pisma v kuverte, v vsako kuverto po eno pismo. Kolikšna je verjetnost, da je vsaj eno pismo v pravi kuverti?

Geometrijska verjetnost:

$$P = \frac{\textit{mera ugodnih izidov}}{\textit{mera vseh izidov}}$$

Pri tem je mera lahko dolžina, ploščina itd.

13. Do šole je štiri minute hoda, vmes pa je semafor, na katerem dve minuti gori zelena, dve minuti pa rdeča luč. Od doma se odpravim pet minut pred začetkom pouka. Kolikšna je verjetnost, da pridem še pravočasno, če se držim predpisov? Kaj pa, če sta na poti dva semaforja?

14. Avtobus se ustavi na postaji med 6:55 in 7:05, in sicer z enakomerno porazdelitvijo. Sam pa sem nagnjen k zamujanju in pridem na postajo med 7:00 in 7:07, tudi z enakomerno porazdelitvijo in neodvisno od avtobusa.

    a) Kolikšna je verjetnost, da ujamem ta avtobus?

    b) Avtobus vozi do postaje, na kateri izstopim, še 55 minut, sam pa potem potrebujem še tri minute, da pridem do predavalnice. Predavanje se začne ob 8:00. Kolikšna je verjetnost, da pridem še pravočasno na predavanje? Če tega avtobusa ne ujamem, seveda zamudim.

15. *Buffonova igla.* Na list papirja z ravnimi vzporednimi črtami, razmaknjenimi za a, vržemo iglo dolžine b. Kolikšna je verjetnost, da igla seka katero od črt?

3. Pogojna verjetnost

Računanje pogojne verjetnosti po definiciji. Izrek o popolni verjetnosti, Bayesova formula. Neodvisnost. Zapletenejši primeri pogojne verjetnosti.

Definicija pogojne verjetnosti:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

(slika). Če je dogodek B sestavljen iz samih enako verjetnih izidov, pa je tudi:

$$P(A \mid B) = \frac{|A \cap B|}{|B|} .$$

1. Vržemo standardno kocko. Naj bo A dogodek, da padejo vsaj štiri pike, B dogodek, da pade šest pik, L pa dogodek, da pade liho mnogo pik. Izračunajte $P(A \mid L)$ in $P(B \mid L)$. Kaj pa, če kocka ni poštena, tako da ena pika pade z verjetnostjo $0{\cdot}3$, izidi z dvema, tremi, štirimi in petimi pikami imajo verjetnost $0{\cdot}15$, šest pik pa pade z verjetnostjo $0{\cdot}1$?

2. Iz dobro premešanega kupa 16 kart, med katerimi so štirje piki, izvlečemo štiri karte. Kolikšna je pogojna verjetnost, da je prva med njimi pik, če vemo, da sta med njimi natanko dva pika?

3. *Bertrandov paradoks*. Dane so tri škatle. V eni sta dva zlata kovanca, v drugi en zlat in en srebrn, v tretji pa dva srebrna. Dovoljeno nam je, da na slepo izberemo en kovanec (t. j. vseh šest z enako verjetnostjo). Če uganemo, kakšen je drugi kovanec v škatli, ki smo jo izbrali, dobimo kovanec. Kolikšna je verjetnost, da dobimo kovanec?

   Razmislek: Recimo, da je kovanec zlat. Potem vemo, da je prišel ali iz škatle z dvema zlatima kovancema ali pa iz škatle z enim zlatim in enim srebrnim kovancem. Ker sta obe škatli enako verjetni, je verjetnost, da bomo uganili, enaka $1/2$, ne glede na to, kaj rečemo.

   Je s tem razmislekom kaj narobe?

4. *Monty–Hallov paradoks*. Dana so tri vrata, za enimi je skrit porsche, za dvema pa koza. Najprej izberemo ena vrata, ne da bi jih odprli, nakar vodja igre odpre ena izmed vrat, za katerima je koza in ki jih nismo izbrali. Nato nam ponovno ponudi, da izberemo ena izmed še zaprtih vrat. Tisto, kar se skriva za njimi, dobimo. Kaj je bolje narediti? Privzamemo, da so vse možnosti za vrata, za katerimi stoji porsche, enako verjetne. Prav tako privzamemo, da vodja igre v primeru, ko ima možnost izbire, izbere povsem naključno.

5. Janez in Peter igrata namizni tenis. V vsaki rundi nekdo zmaga in oba sta enakovredna (ne glede na zgodovino), runde so med seboj neodvisne, igrata pa, dokler eden od njiju ne dobi šest rund. Trenutni izid je 4:2 za Janeza. Kolikšna je verjetnost, da bo Janez na koncu tudi zmagal?

6. Janez in Peter spet igrata namizni tenis. Spet v vsaki rundi nekdo zmaga in runde so prav tako neodvisne, a tokrat Janez dobi posamezno rundo z verjetnostjo $1/3$, igrata pa na dve točki razlike. Kolikšna je zdaj verjetnost, da bo Janez na koncu zmagal? Le-to zdaj računamo od začetka, t. j. izida 0:0.

   Namig: Rekurzivna formula

Izrek o popolni verjetnosti. *Če $H_1, H_2, \ldots H_n$ tvorijo popoln sistem dogodkov (t. j. vedno se zgodi natanko eden izmed njih), velja:*

$$P(A) = P(H_1)\ P(A \mid H_1) + P(H_2)\ P(A \mid H_2) + \cdots + P(H_n)\ P(A \mid H_n)$$

7. V prvi posodi je 5 belih, 3 rdeče in 2 črni kroglici, v drugi posodi pa so tri bele in tri črne kroglice. Iz prve posode v drugo na slepo premestimo eno kroglico, nato pa iz druge na slepo in brez vračanja potegnemo dve kroglici. Kolikšna je verjetnost, da je med njima ena bela in ena črna?

Bayesova formula. *Če $H_1, H_2, \ldots H_n$ tvorijo popoln sistem dogodkov, velja:*

$$P(H_i \mid A) = \frac{P(H_i)\ P(A \mid H_i)}{P(H_1)\ P(A \mid H_1) + \cdots + P(H_n)\ P(A \mid H_n)}$$

8. Žena pošlja moža na trg po solato, ki jo prodajata dve branjevki, Francka in Micka. Verjetnost, da mož kupi solato pri Francki, je 40%, verjetnost, da kupi pri Micki, pa 60%. Francka ima 10%, Micka pa 20% nagnite solate. Mož prinese domov nagnito glavo solate in žena ga nahruli: "Drugič raje glej solato, ne pa Micke!" Kolikšna je pogojna verjetnost, da je mož res kupil solato pri Micki? Privzamemo, da branjevki solato izbirata na slepo.

9. Matičnemu podjetju dobavljajo trije kooperanti: kooperant Alfa Deli dobavlja 20%, kooperant Bobo Deli 50%, kooperant Centro Deli pa 30% vseh delov. Kooperant Alfa Deli ima 5%, Bobo Deli 1%, Centro Deli pa 2% okvarjenih delov. Kontrolor v matičnem podjetju testira na slepo izbran del in izkaže se, da je okvarjen, zato zavzdihne: "Oh, že spet ti Alfa Deli!" Kolikšna je verjetnost, da je bil del dobavil kooperant Alfa Deli?

10. V prvi posodi je 6 belih in 4 rdeče kroglice, v drugi pa ena bela in ena rdeča. Najprej na slepo premestimo 3 kroglice iz prve posode v drugo, nato pa iz druge posode potegnemo dve kroglici (brez vračanja). Obe sta rdeči. Kolikšna je pogojna verjetnost, da so bile vse tri premeščene kroglice rdeče?

Dogodka A in B sta neodvisna, če velja:

$$P(A \cap B) = P(A)\ P(B)$$

(t. j. $P(A \mid B) = P(A)$). Splošneje, dogodki $A_1, A_2, A_3 \ldots$ so neodvisni, če za poljubne različne indekse $i_1, i_2, \ldots i_k$ velja:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})\ P(A_{i_2}) \ldots P(A_{i_k})\,.$$

11. Na kupu so štiri karte: pikov kralj, pikova dama, srčev kralj in srčeva dama. Na slepo izvlečemo eno izmed kart. Definirajmo naslednje dogodke:

$$\begin{aligned} A &:= [\text{izvlekli smo pika}] \\ B &:= [\text{izvlekli smo damo}] \\ C &:= [\text{izvlekli smo srčevega kralja ali pikovo damo}] \end{aligned}$$

    Sta dogodka A in B neodvisna? Kaj pa A in C? Kaj pa B in C? Kako pa je z dogodki A, B in C, so neodvisni?

12. Danih je osem kart: as, kralj, dama, fant, 10, 9, 8 in 7. Slučajno potegnemo eno karto. Definirajmo naslednje dogodke:

$$\begin{aligned} A &:= [\text{karta je as, kralj, dama ali fant}] \\ B &:= [\text{karta je as, 9, 8 ali 7}] \\ C &:= [\text{karta je as, kralj, dama ali 10}] \end{aligned}$$

    So dogodki A, B in C neodvisni?

13. Vržemo tri kovance. Meti so med seboj neodvisni, verjetnosti, da pade grb, pa niso nujno enake. Naj bo A dogodek, da se na prvem kovancu pojavi grb, B pa dogodek, da se grb pojavi na natanko dveh kovancih.
    a) Recimo, da so vsi trije kovanci pošteni, se pravi, da je verjetnost za grb pri vseh kovancih enaka $1/2$. Sta dogodka A in B neodvisna?
    b) Recimo, da je prvi kovanec pošten, druga dva pa ne: na vsakem od njiju se grb pojavi z verjetnostjo p. Pri katerih p sta A in B neodvisna?

14. *Simpsonov paradoks.* Dve zdravili so preizkušali na ženskah in moških. Rezultati so naslednji:

zdravljenje	ženske		moški	
	Prvo zdravilo	Drugo zdravilo	Prvo zdravilo	Drugo zdravilo
uspelo	200	10	190	1000
ni uspelo	1800	190	10	1000

Katero zdravilo je bilo uspešnejše:

- pri ženskah?
- pri moških?
- pri obojih skupaj?

Komentirajte!

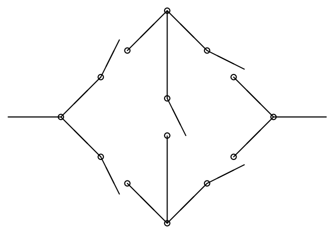
Neodvisnost izpeljanih dogodkov. *Naj bo $\mathcal{F}$ družina dogodkov. S $\sigma(\mathcal{F})$ označimo najmanjšo σ-algebro, ki vsebuje $\mathcal{F}$, t. j. družino vseh dogodkov, ki jih dobimo iz dogodkov iz $\mathcal{F}$ s števnimi unijami in komplementi.*

Naj bodo:

$$\begin{array}{cccc} A_{11}, & A_{12}, & A_{13}, & \ldots \\ A_{21}, & A_{22}, & A_{23}, & \ldots \\ \vdots & \vdots & \vdots & \\ A_{n1}, & A_{n2}, & A_{n3}, & \ldots \end{array}$$

neodvisni dogodki. Tedaj so tudi poljubni dogodki $B_1 \in \sigma(A_{11}, A_{12}, \ldots)$, $B_2 \in \sigma(A_{21}, A_{22}, \ldots)$, $\ldots B_n \in \sigma(A_{n1}, A_{n2}, \ldots)$ neodvisni.

15. V vezju, ki ga prikazuje spodnja skica, vsako stikalo prepušča električni tok z verjetnostjo 1/3, posamezna stikala pa so med seboj neodvisna. Kolikšna je verjetnost, da vezje prepušča tok?



16. Janez, Francelj in Tone gredo streljat zajce. Janez zadene z verjetnostjo 0˙1, Francelj z verjetnostjo 0˙2, Tone pa z verjetnostjo 0˙3, neodvisno drug od drugega.

    a) Vsi pomerijo, ustrelijo in zajec je zadet. Kolikšna je pogojna verjetnost, da ga je Janez zadel?

    b) Ko pridejo do zajca, se izkaže, da ga je zadel natanko eden. Kolikšna je pogojna verjetnost, da je bil to Janez?

17. Andraž, Bojan, Cilka in Darja streljajo v tarčo. Andraž in Bojan streljata z modrimi, Cilka in Darja pa z rdečimi puščicami. Andraž zadene z verjetnostjo 0˙6, Bojan z verjetnostjo 0˙7, Cilka z verjetnostjo 0˙5, Darja pa z verjetnostjo 0˙9. Vsi hkrati pomerijo in ustrelijo, neodvisno drug od drugega. V tarči se znajdeta ena modra in ena rdeča puščica. Kolikšna je pogojna verjetnost, da sta to Bojanova in Darjina?

18. Študent se od 50 izpitnih vprašanj nauči le 30. Za vsako vprašanje, ki se ga nauči, je potem še 30% verjetnosti, da pozabi odgovor, za vsako vprašanje, ki se ga ne nauči, pa je še 10% verjetnosti, da odgovor ugane. Privzamemo, da so dogodki, da študent posamezno vprašanje pozabi oz. ugane odgovor nanj, neodvisni. Na izpitu dobi tri na slepo izbrana vprašanja in izpit naredi, če pravilno odgovori na vsaj dve vprašanji. Kolikšna je verjetnost, da bo naredil izpit?

19. Miha se odpravi na obisk k vinogradnikom Janezu, Lojzu in Štefanu. Vsak mu ponudi kozarec vina. Janez mu ponudi šmarnico z verjetnostjo 60%, Lojz z verjetnostjo 40%, Štefan pa z verjetnostjo 10%. Verjetnost, da Miho boli glava, ne da bi pil šmarnico, je 10%, verjetnost, da ga boli po kozarcu šmarnice (ne glede na to, čigave), 40%, po dveh kozarcih (ne glede na to, čigave šmarnice) 70% in po treh kozarcih 100%. Naslednji dan Miho boli glava in prijatelj Tone mu pravi: "Janez in Lojz sta ti gotovo dala šmarnico!" Kolikšna je pogojna verjetnost, da ima prav? Privzamemo, da vinogradniki izberejo vrsto vina neodvisno drug od drugega.

4. Slučajne spremenljivke

Pojem porazdelitve, porazdelitvena funkcija, porazdelitvena shema diskretno porazdeljene slučajne spremenljivke, porazdelitvena gostota zvezno porazdeljene slučajne spremenljivke. Ugotavljanje in prepoznavanje porazdelitev. Približni obrazci za binomsko porazdelitev. Transformacije (funkcije) slučajnih spremenljivk. Slučajni vektorji. Robne porazdelitve. Neodvisnost slučajnih spremenljivk. Transformacije slučajnih vektorjev. Pogojne porazdelitve.

Porazdelitev diskretne slučajne spremenljivke (t. j. take, ki svoje vrednosti zavzema le na končni ali števno neskončni množici) lahko opišemo s porazdelitveno shemo:

$$X \sim \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{pmatrix}$$

ki pomeni $P(X = a_1) = p_1$, $P(X = a_2) = p_2$ itd.
Porazdelitev vsake realne slučajne spremenljivke lahko opišemo s porazdelitveno funkcijo:

$$F_X(x) = P(X < x)$$

1. Vržemo standardno kocko in število pik, ki padejo, označimo z X. Zapišite in poimenujte porazdelitev te slučajne spremenljivke ter narišite graf njene porazdelitvene funkcije.

2. Neodvisno vržemo dva poštena kovanca in standardno kocko. Za vsako piko dobimo en evro, za vsako cifro, ki pade, pa dva evra. Slučajna spremenljivka S naj predstavlja skupni znesek v evrih, ki ga dobimo. Zapišite njeno porazdelitev.

Porazdelitev zvezne slučajne spremenljivke lahko opišemo s poradelitveno gostoto p_X, za katero velja:

$$P(a < X < b) = P(a \leq X \leq b) = \int_a^b p_X(x)\,\mathrm{d}x$$

Tako velja:

$$F_X(x) = \int_{-\infty}^{x} p_X(t)\,\mathrm{d}t$$

in za skoraj vsak x velja $p_X(x) = F'_X(x)$.
Kakor pri diskretnih slučajnih spremenljivkah velja:

$$p_1 + p_2 + \cdots = 1$$

tudi pri zveznih velja:

$$\int_{-\infty}^{\infty} p_X(x)\,\mathrm{d}x = 1$$

3. Recimo, da hitro zavrtimo povsem simetrično kolo z označeno točko na obodu. Kolo se zelo počasi ustavlja. Slučajna spremenljivka X naj predstavlja kot, pri katerem se je ustavila označena točka (kot merimo od 0 do 2π). Zapišite porazdelitveno funkcijo te slučajne spremenljivke, poimenujte porazdelitev in zapišite še njeno gostoto.

4. Rok in Simona se zmenita na določenem mestu točno ob 18:00, prideta pa enkrat med 18:00 in 18:10, in sicer z enakomerno porazdelitvijo in neodvisno drug od drugega. Ljubosumni Maks vse od 18:00 opreza za vogalom in čaka, dokler ne prideta obadva. Slučajna spremenljivka M naj predstavlja, koliko časa (v minutah) je čakal Maks. Zapišite porazdelitveno funkcijo in gostoto te slučajne spremenljivke.

5. Naj bo X število šestic, ki padejo v desetih neodvisnih metih standardne kocke. Zapišite in poimenujte porazdelitev te slučajne spremenljivke.

6. Šestkrat vržemo nepošten kovanec, pri katerem grb pade z verjetnostjo $1/3$. Meti so med seboj neodvisni. Kolikšna je verjetnost, da padeta več kot dva grba?

Naj bo $X \sim \mathrm{B}(n,p)$ in $n \to \infty$ ter še $k \in \mathbb{N}_0$. Če je $p \ll 1/\sqrt{n}$ in $k \ll \sqrt{n}$, velja **Poissonov obrazec:**

$$P(X = k) \sim \frac{(np)^k e^{-np}}{k!} \,.$$

Če pa je $p, 1-p \gg 1/n$ (ali, ekvivalentno, $\sigma := \sqrt{np(1-p)} \to \infty$) in še $|k - np| \ll \sigma^{4/3}$, velja **Laplaceova lokalna formula:**

$$P(X = k) \sim \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(k-np)^2/(2\sigma^2)} \,.$$

Meja med smotrnostjo uporabe Poissonovega obrazca in Laplaceove lokalne formule je približno pri $p = 0\text{'}6/\sqrt[3]{n}$.

7. Naj bo X spet število šestic, ki padejo v desetih neodvisnih metih standardne kocke. Preverite, kako natančna sta Poissonov obrazec in Laplaceova lokalna formula pri izračunu $P(X = 1)$.

8. Naj bo $X \sim \mathrm{B}(50, 0\text{'}4)$. Izračunajte $P(X = 20)$. Točen rezultat primerjajte z rezultatoma, dobljenima po Poissonovem obrazcu in po Laplaceovi lokalni formuli.

9. Naj bo $X \sim \mathrm{B}(1000, 0\text{'}99)$. Izračunajte $P(X = 990)$.

Če je $X \sim \mathrm{B}(n,p)$, $a \le b$, $\sigma := \sqrt{np(1-p)} \to \infty$ in je $|a - np| \ll \sigma^{4/3}$ **ali** *$|b - np| \ll \sigma^{4/3}$, velja* **Laplaceova integralska formula**, *in sicer za $a, b \in \mathbb{Z}$ v obliki:*

$$P(a \le X \le b) \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sigma}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sigma}\right) ,$$

za $a, b \in \mathbb{Z} + \frac{1}{2}$ ali $b - a \gg 1$ pa v obliki:

$$P(a < X < b) \sim P(a \le X \le b) \sim \Phi\left(\frac{b - np}{\sigma}\right) - \Phi\left(\frac{a - np}{\sigma}\right) .$$

Funkcija Φ je **Gaussov verjetnostni integral:**

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,\mathrm{d}t$$

in je liha, velja pa še $\lim_{x\to\infty} \Phi(x) = 1/2$ (graf!).
V literaturi so definicije funkcije Φ različne, zato je treba paziti!

10. V tovarni vsak dan proizvedejo 1600 izdelkov. Za vsakega je verjetnost, da bo okvarjen, enaka 10%.

a) Kolikšna je verjetnost, da bo okvarjenih natanko 160 izdelkov? Kolikšna pa je verjetnost, da bo okvarjenih natanko 175 izdelkov?

b) Kolikšna je verjetnost, da bo okvarjenih več kot 175 izdelkov? Kaj pa, da bo okvarjenih manj kot 150 izdelkov?

c) Okvarjene izdelke spravijo v skladišče, kjer jih popravijo in ki se dnevno prazni. Najmanj kako veliko mora biti skladišče, če naj bo verjetnost, da bo premajhno, največ $0\dot{}05$?

11. Zavarovalnica je proti nezgodi zavarovala 1000 oseb. Vsako od njih doleti nezgoda z verjetnostjo $0\dot{}0015$ in osebe so med seboj neodvisne. Kolikšna je verjetnost, da se noben zavarovanec ne ponesreči? Kolikšna pa je verjetnost, da se ponesrečita več kot dva? Točen rezultat primerjajte z rezultati, dobljenimi po Poissonovem obrazcu, Laplaceovi lokalni in Laplaceovi integralski formuli.

12. Verjetnost, da je izdelek prvovrsten, je 60%. Najmanj koliko izdelkov moramo naročiti, če naj bo med naročenimi izdelki z verjetnostjo najmanj $0\dot{}99$ vsaj 59% izdelkov prvovrstnih? Seveda privzamemo, da so posamezni izdelki med seboj neodvisni.

13. Verjetnost, da je izdelek prvovrsten, je 10%. Najmanj koliko izdelkov moramo naročiti, če naj bo med naročenimi izdelki z verjetnostjo najmanj $0\dot{}95$ vsaj 100 izdelkov prvovrstnih?

14. Naj bo X število metov standardne kocke, ki jo mečemo, dokler ne pade šestica. Meti so med seboj neodvisni. Zapišite in poimenujte porazdelitev slučajne spremenljivke X.

15. Naj bo X število metov standardne kocke, ki jo mečemo, dokler šestica ne pade desetkrat. Meti so med seboj neodvisni. Zapišite in poimenujte porazdelitev slučajne spremenljivke X.

16. Naj bo X število metov poštenega kovanca, ki ga mečemo, dokler ne pade cifra, takoj za njo pa še grb. Meti so med seboj neodvisni. Zapišite in poimenujte porazdelitev slučajne spremenljivke X.

17. Naj bo X število metov poštenega kovanca, ki ga mečemo, dokler ne pade tako cifra kot tudi grb. Meti so med seboj neodvisni. Zapišite porazdelitev slučajne spremenljivke X. Je le-ta kaj povezana s kako znano porazdelitvijo?

18. Med 16 kartami so štirje piki. Na slepo in brez vračanja izvlečemo sedem kart. Naj bo X število pikov med njimi. Zapišite in poimenujte porazdelitev slučajne spremenljivke X.

19. Danih je 12 praznih škatel. Mimo pride Janezek, slučajno izbere tri škatle in v vsako vrže po eno kroglico. Mimo pride še Marička, slučajno in neodvisno od Janezka izbere štiri škatle in prav tako v vsako vrže po eno kroglico. Naj bo X število škatel, v katerih sta dve kroglici. Zapišite in poimenujte porazdelitev te slučajne spremenljivke.

20. Slučajna spremenljivka X ima diskretno porazdelitev, podano po predpisu:

$$P(X = k) = ck \quad k = 1, 2, \ldots 10$$

(in $P(X = k) = 0$ za $k \notin \{1, 2, \ldots 10\}$). Izračunajte konstanto c in še $P(X > 3)$.

21. Naj bo $\lambda > 0$. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

$$p_X(x) = \begin{cases} ce^{-\lambda x} & ; \ x > 0 \\ 0 & ; \ \text{sicer} \end{cases}$$

Izračunajte konstanto c, poimenujte porazdelitev in določite porazdelitveno funkcijo $F_X(x)$. Izračunajte še $P(1 < X < 2)$.

22. Slučajna spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0\text{'}1 & 0\text{'}3 & 0\text{'}2 & 0\text{'}4 \end{pmatrix}$$

Zapišite porazdelitev slučajne spremenljivke $Y = X^2$.

23. Slučajna spremenljivka X je porazdeljena geometrijsko $\mathrm{G}(p)$. Zapišite porazdelitev slučajne spremenljivke $Y := \sin(\pi X/2)$.

24. Naj bo $b \geq 2$ naravno število in naj bo slučajna spremenljivka U porazdeljena enakomerno $\mathrm{E}(0, 1)$. Določite porazdelitev slučajne spremenljivke $D := \lfloor b^U \rfloor$.

25. Na razpolago imamo generator slučajnih števil, ki generira enakomerno porazdelitev $\mathrm{E}(0, 1)$. Kako bi generirali porazdelitev slučajne spremenljivke X iz 22. naloge? Kaj pa eksponentno porazdelitev $\mathrm{Exp}(\lambda)$?

Naj bo $\mu \in \mathbb{R}$ in $\sigma > 0$. Normalna (Gaussova) porazdelitev $\mathrm{N}(\mu, \sigma)$ je porazdelitev z gostoto:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Normalna porazdelitev $\mathrm{N}(\mu, 0)$ je porazdelitev, ki je skoncentrirana v μ ($X \sim \mathrm{N}(\mu, 0)$ pomeni $P(X = \mu) = 1$). Standardizirana normalna porazdelitev $\mathrm{N}(0, 1)$ ima potemtakem gostoto:

$$p(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

in porazdelitveno funkcijo $F(x) = \frac{1}{2} + \Phi(x)$.

26. Slučajna spremenljivka Z je porazdeljena standardizirano normalno. Izračunajte $P(Z < 1\text{'}5)$.

Naj bo X zvezno porazdeljena slučajna spremenljivka z zalogo vrednosti na (a, b) in gostoto p_X. Nadalje naj bo $f\colon (a,b) \to (c,d)$ bijektivna zvezno odvedljiva preslikava, katere odvod ni nikjer enak 0. Tedaj ima slučajna spremenljivka $Y := f(X)$ gostoto:

$$p_Y(y) = p_X\big(f^{-1}(y)\big)\big|(f^{-1})'(y)\big|\,, \quad c < y < d\,.$$

Splošneje, če je f dovolj lepa funkcija in

$$P\big(f \text{ v } X \textit{ ni odvedljiva ali } f'(X) = 0\big) = 0\,,$$

velja:

$$p_Y(y) = \sum_{x; f(x)=y} \frac{p_X(x)}{\big|f'(x)\big|}\,.$$

27. Slučajna spremenljivka X je porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Kako je porazdeljena slučajna spremenljivka $Y := aX + b$?

Če je $X \sim \mathrm{N}(\mu, \sigma)$, velja:

$$P(a < X < b) = P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Laplaceova integralska formula tako ne pomeni nič drugega kot to, da za velike n in p, ki ni preblizu 0 ali 1, velja:

$$\mathrm{B}(n, p) \approx \mathrm{N}\big(np, \sqrt{npq}\big)$$

kjer je $q = 1 - p$.

28. Slučajna spremenljivka X je porazdeljena normalno $\mathrm{N}(9, 5)$. Izračunajte $P(X < 0)$.

29. Slučajna spremenljivka Z je porazdeljena standardizirano normalno. Poiščite porazdelitvi slučajnih spremenljivk $X = e^Z$ in $Y = Z^2$. Poimenujte porazdelitev slučajne spremenljivke Y.

Diskretni slučajni vektorji:

– *Porazdelitev slučajnega vektorja* (X, Y) *podamo s* $P(X = x, Y = y)$ *(navzkrižna porazdelitev slučajnih spremenljivk* X *in* Y*).*

– *Porazdelitve komponent imenujemo robne porazdelitve:*

$$P(X = x) = \sum_y P(X = x, Y = y)$$
$$P(Y = y) = \sum_x P(X = x, Y = y)\,.$$

– X *in* Y *sta neodvisni, brž ko za poljubna* x *in* y *velja* $P(X = x, Y = y) = P(X = x)\,P(Y = y)$*.*

30. Iz posode, v kateri so 3 rdeče, 2 modri in 5 belih kroglic, na slepo in brez vračanja izvlečemo tri kroglice. Naj bo R število rdečih, M pa število modrih med njimi. Zapišite porazdelitev slučajnega vektorja (R, M) ter določite in poimenujte še robni porazdelitvi. Sta slučajni spremenljivki R in M neodvisni? Zapišite in poimenujte še porazdelitev slučajne spremenljivke $R + M$.

31. Slučajni vektor (X, Y) je porazdeljen po naslednji shemi:

	$Y = 0$	$Y = 1$	$Y = 2$
$X = -1$	0'05	0'1	
$X = 0$	0'1		
$X = 1$	0'05		

Dopolnite tabelo tako, da bosta slučajni spremenljivki X in Y neodvisni. Poiščite še robni porazdelitvi in porazdelitev razlike $Y - X$.

32. Dane so neodvisne slučajne spremenljivke $X_1, X_2, X_3 \sim \mathrm{Be}(0'3)$. Določite porazdelitev njihove vsote $S := X_1 + X_2 + X_3$.

    Splošneje, naj bodo $X_1, \ldots X_n \sim \mathrm{Be}(p)$ neodvisne slučajne spremenljivke. Kako je porazdeljena vsota $S := X_1 + \cdots + X_n$?

Posledica. Če sta $S_1 \sim \mathrm{B}(n_1, p)$ in $S_2 \sim \mathrm{B}(n_2, p)$ neodvisni slučajni spremenljivki, je $S := S_1 + S_2 \sim \mathrm{B}(n_1 + n_2, p)$.

33. Naj bosta $S_1 \sim \mathrm{P}(\lambda_1)$ in $S_2 \sim \mathrm{P}(\lambda_2)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $S := S_1 + S_2$?

34. Naj bodo $X_1, \ldots X_n \sim \mathrm{G}(p)$ neodvisne slučajne spremenljivke. Kako je porazdeljena vsota $S := X_1 + \cdots + X_n$?

Posledica. Če sta $S_1 \sim \mathrm{P}(n_1, p)$ in $S_2 \sim \mathrm{P}(n_2, p)$ neodvisni slučajni spremenljivki, je $S := S_1 + S_2 \sim \mathrm{P}(n_1 + n_2, p)$.

35. Ministrstvo dobi vsak dan na mizo slučajno število prošenj, ki je porazdeljeno po Poissonovem zakonu $\mathrm{P}(\lambda)$. Dnevi so med seboj neodvisni. Ministrstvo vsak dan razreši eno prošnjo, če jo seveda ima, morebitne preostale pa pusti za kasneje. Kolikšna je verjetnost, da ministrstvu po dveh dneh dela preostaneta vsaj še dve nerazrešeni prošnji?

36. V avtobusu, ki ima 32 sedežev, je 30 potnikov. Vsak potnik bo z verjetnostjo $1/30$ na naslednji postaji izstopil, neodvisno od drugih potnikov. Število potnikov, ki bodo na naslednji postaji vstopili, pa je porazdeljeno po Poissonu $\mathrm{P}(1)$ in seveda neodvisno od izstopnih namer trenutnih potnikov. Približno izračunajte verjetnost, da ne bodo mogli vsi potniki sedeti, ko bo avtobus speljal s postaje.

    Namig: Dogodek, da izstopita več kot dva potnika in vseeno ne bo dovolj sedežev za vse, je tako malo verjeten, da ga lahko zanemarimo.

Porazdelitev zveznega dvorazsežnega slučajnega vektorja (X,Y) lahko opišemo z dvorazsežno (navzkrižno) gostoto $p_{X,Y}$, za katero velja:

$$P\big((X,Y) \in A\big) = \iint_A p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y\,.$$

Seveda velja:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1\,.$$

Komponente zveznih slučajnih vektorjev so tudi zvezno porazdeljene. Robni gostoti slučajnega vektorja (X,Y):

$$p_X(x) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}y\,, \qquad p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x,y)\,\mathrm{d}x$$

Zvezno porazdeljeni slučajni spremenljivki sta neodvisni natanko tedaj, ko je tudi slučajni vektor (X,Y) porazdeljen zvezno z gostoto:

$$p_{X,Y}(x,y) = p_X(x)\,p_Y(y)\,.$$

37. Slučajni vektor (X,Y) je porazdeljen zvezno z naslednjo gostoto:

$$p_{X,Y}(x,y) = \begin{cases} c\,e^{-x} & ;\ x > y > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

    Poiščite konstanto c ter robni gostoti p_X in p_Y. Sta slučajni spremenljivki X in Y neodvisni? Izračunajte še $P(2 < Y < 3)$, $P(Y < X)$ in $P(2Y < X)$.

Če velja $Y = g(X, Z)$ ter je $g(x, z)$ dovolj lepa funkcija, ki je za vsak x monotona v z, velja:

$$p_Z(z) = \int_{-\infty}^{\infty} p_{X,Y}\big(x, g(x,z)\big) \left|\frac{\partial g}{\partial z}(x,z)\right| \mathrm{d}x\,.$$

Še splošneje, če je $Z = f(X, Y)$, f dovolj lepa funkcija in

$$P\left(f \text{ v } (X,Y) \text{ ni odvedljiva ali } \frac{\partial f}{\partial y}(X,Y) = 0\right) = 0\,,$$

velja:

$$p_Z(z) = \int_{-\infty}^{\infty} \sum_{y; f(x,y)=z} \frac{p_{X,Y}(x,y)}{\left|\frac{\partial f}{\partial y}(x,y)\right|} \mathrm{d}x\,.$$

38. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$p_{X,Y}(x,y) = \begin{cases} e^{-x} & ;\ x > y > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

Zapišite porazdelitev slučajne spremenljivke $Z := X + Y$.

39. Slučajni vektor (X, Y) je porazdeljen zvezno z gostoto:

$$p_{X,Y}(x,y) = \begin{cases} \dfrac{c}{(1+x+y)^3} & ;\ x, y > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

Zapišite porazdelitev slučajne spremenljivke $Z := Y/X$ in izračunajte konstanto c. Izračunajte še $P(X < 2Y)$.

40. Slučajni spremenljivki X in Y sta neodvisni in porazdeljeni enakomerno na intervalu $[0, 1]$. Zapišite porazdelitveno funkcijo slučajne spremenljivke $Z := XY$.

41. Slučajni vektor (X, Y) ima naslednjo porazdelitveno gostoto:

$$p_{X,Y}(x,y) = \frac{Cx^2y^2}{(1+x^2+y^2)(x^2+y^2)^{5/2}}$$

Izračunajte porazdelitveno gostoto slučajne spremenljivke $Z := X^2+Y^2$ (s konstanto vred).

42. Naj bosta $X_1 \sim \mathrm{N}(\mu_1, \sigma_1)$ in $X_2 \sim \mathrm{N}(\mu_2, \sigma_2)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $X := X_1 + X_2$?

43. Avtobus pride na postajo ob času, ki je porazdeljen normalno $\mathrm{N}(7{:}00, 3\,\mathrm{min})$, sam pa sem nagnjen k zamujanju in pridem na postajo ob času, ki je porazdeljen normalno $\mathrm{N}(7{:}01, 4\,\mathrm{min})$. Kolikšna je verjetnost, da še ujamem avtobus?

44. Določeni pojavi, ki se pojavljajo v času (npr. telefonski klici, nesreče, radioaktivni razpadi), tvorijo *Poissonov proces* z intenziteto λ, če je število pojavov, ki se zgodijo v katerem koli časovnem intervalu dolžine t, porazdeljeno po Poissonu $\mathrm{P}(\lambda t)$, in če za poljubna disjunktna časovna intervala velja, da je število pojavov, ki se zgodijo v prvem, neodvisno od števila pojavov, ki se zgodijo v drugem časovnem intervalu. Naj bo T_n slučajna spremenljivka, ki pove čas, ki mine od izbranega trenutka do n-tega pojava, ki se zgodi po izbranem trenutku. Kako je porazdeljena ta slučajna spremenljivka? Kako za $n = 1$ še imenujemo to porazdelitev?

45. Naj bosta $S_1 \sim \mathrm{Gama}(a_1, \lambda)$ in $S_2 \sim \mathrm{Gama}(a_2, \lambda)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $S := S_1 + S_2$?

Posledica. Če so $X_1, \dots X_n \sim \mathrm{Exp}(\lambda)$ neodvisne slučajne spremenljivke, je $S := X_1 + \cdots + X_n \sim \mathrm{Gama}(n, \lambda)$.

Posledica. Če so $Z_1, \dots Z_n$ neodvisne slučajne spremenljivke, porazdeljene standardizirano normalno, je $Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n)$.

46. Slučajni spremenljivki X in Y sta neodvisni, pri čemer je X porazdeljena normalno $\mathrm{N}(0, \sigma)$, Y pa ima porazdelitev $\mathrm{Gama}(\lambda, a)$. Določite porazdelitev slučajne spremenljivke $T = X/\sqrt{Y}$.

47. Slučajni spremenljivki X_1 in X_2 sta neodvisni s porazdelitvijo gama: $X_1 \sim \mathrm{Gama}(\lambda_1, a_1)$ in $X_2 \sim \mathrm{Gama}(\lambda_2, a_2)$. Določite porazdelitev slučajne spremenljivke $Y = X_1/X_2$.

48. Porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y=0$	$Y=1$	$Y=2$
$X=-1$	$\frac{1}{12}$	0	$\frac{1}{6}$
$X=0$	$\frac{1}{4}$	$\frac{1}{12}$	0
$X=1$	0	$\frac{1}{4}$	$\frac{1}{6}$

Določite pogojno porazdelitev slučajne spremenljivke X glede na $Y = 0$ in Y glede na $X = -1$.

49. Dani sta neodvisni slučajni spremenljivki $X \sim \mathrm{P}(\lambda)$ in $Y \sim \mathrm{P}(\mu)$. Določite pogojno porazdelitev slučajne spremenljivke X glede na $Z := X + Y$.

Zvezne pogojne porazdelitve:

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

50. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$p_{X,Y}(x, y) = \begin{cases} e^{-x} & ; x > y > 0 \\ 0 & ; \text{sicer} \end{cases}$$

Določite pogojno porazdelitev slučajne spremenljivke X glede na Y in Y glede na X.

5. Številske karakteristike

Matematično upanje, disperzija, pogojno matematično upanje. Neposreden izračun matematičnega upanja (metoda indikatorjev). Kovarianca, kovariančna matrika, Pearsonov korelacijski koeficient. Višji momenti, asimetrija, sploščenost. Vrstilne karakteristike.

Matematično upanje:

$$E(X) = \sum_x x \; P(X = x) \qquad E(X) = \int_{-\infty}^{\infty} x \, p_X(x) \, \mathrm{d}x$$

$$E\big(f(X)\big) = \sum_x f(x) \; P(X = x) \qquad E\big(f(X)\big) = \int_{-\infty}^{\infty} f(x) \, p_X(x) \, \mathrm{d}x$$

Disperzija (varianca):

$$D(X) = E\big((X - E(X))^2\big) = E(X^2) - \big(E(X)\big)^2$$

U-metoda:

$$\begin{aligned} E(X) &= u + E(X - u) \\ D(X) &= D(X - u) = E\big((X - u)^2\big) - \big(E(X - u)\big)^2 \end{aligned}$$

Standardni odklon: $\sigma(X) = \sqrt{D(X)}$.

1. Slučajna spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 2 & 3 & 4 \\ 0\dot{}1 & 0\dot{}3 & 0\dot{}1 & 0\dot{}1 & 0\dot{}4 \end{pmatrix}$$

   Izračunajte $E(X)$, $D(X)$ in $\sigma(X)$.

2. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$, kjer je $\lambda > 0$. Izračunajte $E(X)$, $D(X)$ in $E(e^{-X})$.

$$\begin{aligned} E(aX + b) &= a \, E(X) + b \\ E(X + Y) &= E(X) + E(Y) \end{aligned}$$

3. Slučajna spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$p_X(x) = \frac{2}{\pi(1 + x^2)^2}$$

   Izračunajte $E(1 + X^2)$ in $D(X)$.

$$E\big(f(X,Y)\big) = \sum_x \sum_y f(x,y)\ P(X=x, Y=y)$$
$$E\big(f(X,Y)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, p_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

4. Diskretna porazdelitev slučajnega vektorja (X,Y) je podana z naslednjo tabelo:

	$Y=0$	$Y=1$	$Y=3$
$X=0$	0˙05	0˙1	0˙2
$X=1$	0˙15	0	0˙15
$X=4$	0˙2	0˙1	0˙05

Izračunajte $E(XY^2)$.

5. Slučajni vektor (X,Y) je porazdeljen zvezno z naslednjo gostoto:

$$p_{X,Y}(x,y) = \begin{cases} e^{-x} & ;\ x > y > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

Izračunajte $E\big(e^{Y-X}\big)$.

6. Slučajni vektor (X,Y) je porazdeljen zvezno z naslednjo gostoto:

$$p_{X,Y}(x,y) = \frac{c}{(1+x^2+y^2)^3}$$

Izračunajte konstanto c ter še $E(X)$, $E(Y)$, $E(X^2)$, $E(Y^2)$ in $E(X^2+Y^2)$.

7. Slučajna spremenljivka Z naj bo porazdeljena standardizirano normalno. Za vsak $n \in \mathbb{N}$ izračunajte $E(Z^n)$. *Namig: indukcija.*

Posledica. Če je $Z \sim \mathrm{N}(0,1)$, je $D(Z) = 1$.

$$\boxed{D(aX+b) = a^2\, D(X)}$$

Posledica. Če je $X \sim \mathrm{N}(\mu,\sigma)$, je $E(X) = \mu$ in $D(X) = \sigma^2$.

Če sta X in Y neodvisni, velja:

$$E(XY) = E(X)\,E(Y)$$
$$D(X+Y) = D(X) + D(Y)$$

Pozor! *Zakaj ne velja:*

$$D(X+X) = D(X) + D(X) = 2\,D(X)\,?$$

8. *Matematično upanje in disperzija binomske porazdelitve.*

9. *Matematično upanje in disperzija porazdelitve gama.*

10. Dani sta neodvisni slučajni spremenljivki $X \sim \chi^2(3)$ in $Y \sim \chi^2(5)$. Izračunajte $D(3X - Y)$ in $E\big((X + 2Y)^2\big)$.

11. Za okroglo mizo sedi 8 gostov, ki igrajo karte. Vsak ima v rokah eno karto, razdeljeno z dobro premešanega kupa standardnih 52 kart. Vsak lahko vidi svojo karto ter karti svojih sosedov na levi in na desni. Posamezen igralec stavi svojo ženo, če ima asa in hkrati nobeden od njegovih sosedov nima asa. Označimo z S število igralcev, ki stavijo svojo ženo. Izračunajte $E(S)$ in $D(S)$.

12. *Matematično upanje in disperzija geometrijske in Pascalove porazdelitve.*

13. *Izračun matematičnega upanja hipergeometrijske porazdelitve z indikatorji.*

Pogojno matematično upanje:

$$
\begin{aligned}
E\big[f(X,Y) \mid Y = y\big] &= \sum_x f(x,y)\, P(X = x \mid Y = y) = \\
&= \frac{1}{P(Y = y)} \sum_x f(x,y)\, P(X = x, Y = y) \\
E\big[f(X,Y) \mid Y = y\big] &= \int_{-\infty}^{\infty} f(x,y)\, p_{X|Y}(x \mid y)\, \mathrm{d}x = \\
&= \frac{1}{p_Y(y)} \int_{-\infty}^{\infty} f(x,y)\, p_{X,Y}(x,y)\, \mathrm{d}x
\end{aligned}
$$

14. Diskretna porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y = 0$	$Y = 1$	$Y = 3$
$X = 0$	0'05	0'1	0'2
$X = 1$	0'15	0	0'15
$X = 4$	0'2	0'1	0'05

Izračunajte $E(Y \mid X = 4)$, $E(Y^2 \mid X = 4)$ in $E(XY^2 \mid X = 4)$.

15. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$
p_{X,Y}(x,y) = \begin{cases} 3(x^2y + xy^2) & ;\ x, y \in [0,1] \\ 0 & ;\ \text{sicer} \end{cases}
$$

Izračunajte $E(Y \mid X = x)$ in $E(XY \mid X = x)$.

Kovarianca:

$$
\begin{aligned}
K(X,Y) &:= E\big((X - E(X))\,(Y - E(Y))\big) = \\
&= E(XY) - E(X)\,E(Y)
\end{aligned}
$$

Velja $K(X,X) = D(X)$ *in* $K(X,Y) = K(Y,X)$. *Če sta* a *in* b *konstanti, velja* $K(X+a, Y+b) = K(X,Y)$ *in* $K(aX+bY, Z) = a\,K(X,Z) + b\,K(Y,Z)$.
Če je $K(X,Y) = 0$, *pravimo, da sta slučajni spremenljivki* X *in* Y *nekorelirani.*
X, Y *neodvisni* $\Longrightarrow E(XY) = E(X)\,E(Y) \Longleftrightarrow X, Y$ *neodvisni*
Nekoreliranost še ne pomeni neodvisnosti.

Korelacijski koeficient:

$$
r(X,Y) = \frac{K(X,Y)}{\sigma(X)\,\sigma(Y)}
$$

Velja $-1 \le r(X,Y) \le 1$. *Če so* a, b, c *in* d *konstante ter* $a, c > 0$, *velja* $r(aX+b, cY+d) = r(X,Y)$.

16. Diskretna porazdelitev slučajnega vektorja (X,Y) je podana z naslednjo tabelo:

	$Y=0$	$Y=1$	$Y=3$
$X=0$	0'05	0'1	0'2
$X=1$	0'15	0	0'15
$X=4$	0'2	0'1	0'05

Izračunajte $K(X,Y)$ in $r(X,Y)$.

17. Diskretna porazdelitev slučajnega vektorja (X,Y) je podana z naslednjo tabelo:

	$Y=-1$	$Y=0$	$Y=1$
$X=-1$	0'1	0	0'1
$X=0$	0	0'6	0
$X=1$	0'1	0	0'1

Dokažite, da sta slučajni spremenljivki X in Y nekorelirani. Sta tudi neodvisni?

Slučajni spremenljivki X *in* Y *sta zagotovo neodvisni v naslednjih treh primerih:*

- *če sta nekorelirani in dihotomni, t. j. posamezna slučajna spremenljivka lahko zavzame kvečjemu dve vrednosti;*
- *če sta nekorelirani in je njuna navzkrižna porazdelitev dvorazsežna normalna;*
- *če za poljubni omejeni merljivi funkciji* f *in* g *velja, da sta slučajni spremenljivki* $f(X)$ *in* $g(Y)$ *nekorelirani.*

18. Slučajni vektor (X, Y) je porazdeljen zvezno z naslednjo gostoto:

$$p_{X,Y}(x, y) = \begin{cases} 3(x^2y + xy^2) & ; x, y \in [0, 1] \\ 0 & ; \text{sicer} \end{cases}$$

Izračunajte $K(X, Y)$ in $r(X, Y)$.

Za slučajni vektor $\mathbf{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ *definiramo* **kovariančno matriko**:

$$K(\mathbf{X}) := \begin{bmatrix} K(X_1, X_1) & \cdots & K(X_1, X_n) \\ \vdots & & \vdots \\ K(X_n, X_1) & \cdots & K(X_n, X_n) \end{bmatrix} = E(\mathbf{X}\mathbf{X}^T) - E(\mathbf{X})\,E(\mathbf{X})^T$$

Za poljubno deterministično matriko A velja $K(A\mathbf{X}) = A\,K(\mathbf{X})\,A^T$.

19. Slučajni vektor (X, Y, Z) ima naslednjo kovariančno matriko:

$$K = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

Določite parameter a tako, da bosta slučajni spremenljivki $U := aX + 2Y + Z$ in $V := X - Y + aZ$ nekorelirani.

Asimetrija:

$$A(X) = \frac{E\big((X - E(X))^3\big)}{\big(D(X)\big)^{3/2}}$$

Če sta $a > 0$ in b konstanti, je $A(aX + b) = A(X)$.

Sploščenost (kurtozis):

$$K(X) = \frac{E\big((X - E(X))^4\big)}{\big(D(X)\big)^2} - 3$$

Če sta $a \neq 0$ in b konstanti, je $K(aX + b) = K(X)$.

Normalno porazdeljena slučajna spremenljivka ima sploščenost nič. Skica porazdelitev z različno predznačenimi asimetrijami in sploščenostmi.

20. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Izračunajte njeno asimetrijo $A(X)$ in sploščenost $K(X)$.

21. Slučajna spremenljivka X je porazdeljena po Poissonu $\mathrm{P}(\lambda)$. Izračunajte njeno asimetrijo $A(X)$ in sploščenost $K(X)$.

Število q_α je **kvantil** *slučajne spremenljivke X za verjetnost α, če velja:*

$$P(X < q_\alpha) \le \alpha, \ \ P(X \le q_\alpha) \ge \alpha\,.$$

Če je X zvezno porazdeljena in ima v okolici točke q_α strogo pozitivno gostoto, pa velja:

$$F_X(q_\alpha) = \alpha\,.$$

Mediano $m = q_{1/2}$ lahko razumemo kot nadomestek za matematično upanje, semiinterkvartilni razmik:

$$s = \frac{q_{3/4} - q_{1/4}}{2}$$

pa za varianco. Pozor: porazdelitvi:

$$\begin{pmatrix} -1 & 0 & 1 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \quad \textit{in} \quad \begin{pmatrix} -1 & 0 & 2 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

imata enaki mediani, ne pa tudi matematičnih upanj.

22. Slučajna spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$p_X(x) = \begin{cases} \dfrac{1}{(1+x)^2} & ; x \ge 0 \\ 0 & ; \text{sicer} \end{cases}$$

Pokažite, da $E(X)$ ne obstaja, ter izračunajte njeno mediano in semiinterkvartilni razmik.

6. Rodovne in karakteristične funkcije

Rodovne funkcije: osnove, konvolucijski izrek, procesi razvejanja. Karakteristične funkcije: osnove, povezava z rodovnimi funkcijami, konvolucijski izrek, inverzna formula.

Rodovna funkcija:

$$G_X(s) = E\big(s^X\big)$$

Če je $X \geq 0$, je $G_X(s)$ definirana za $0 \leq s \leq 1$ ali še splošneje za vse kompleksne s z $|s| \leq 1$, ki ne pripadajo $[-1, 0)$.
Če je X celoštevilska, je $G_X(s)$ definirana za vse kompleksne s z $|s| = 1$.
Če ima X vrednosti v $\mathbb{N}_0$, je $G_X(s)$ definirana za vse kompleksne s z $|s| \leq 1$.
Če je $X \sim \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ p_1 & p_2 & p_3 & \cdots \end{pmatrix}$, je $G_X(s) = p_1 s^{a_1} + p_2 s^{a_2} + p_3 s^{a_3} + \cdots$.
Če ima X vrednosti v $\mathbb{N}_0$, je $P(X = k) = \dfrac{G_X^{(k)}(0)}{k!}$.
Velja še $E(X) = G'_X(1)$. Splošneje,

$$E\big[X(X-1)\ldots(X-k+1)\big] = G^{(k)}(1)\,.$$

1. Izračunajte rodovno funkcijo slučajne spremenljivke X, porazdeljene po Poissonu $\mathrm{P}(\lambda)$. Izračunajte še $E(X)$ in $D(X)$.

Če sta X in Y neodvisni, je $G_{X+Y} = G_X\, G_Y$.

2. Naj bo $X \sim \mathrm{B}(n, p)$. S pomočjo rodovne funkcije izračunajte $E\left(\dfrac{1}{1+X}\right)$.

3. Janez ima dva otroka, ki še nimata svojih otrok. Za vsakega od njiju je porazdelitev števila otrok, ki jih bo imel, enaka:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Privzamemo še, da sta števili otrok Janezovih otrok neodvisni. Določite porazdelitev števila njegovih vnukov.

4. Nika še nima otrok. Porazdelitev števila njenih bodočih otrok je enaka:

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix},$$

porazdelitev števila otrok vsakega eventuelnega Nikinega otroka pa je enaka:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Privzamemo še, da so števila otrok Nikinih otrok neodvisna. Določite matematično upanje števila Nikinih vnukov in še verjetnost, da Nika ostane brez vnukov.

Če so $X_1, X_2, \ldots$ neodvisne in enako porazdeljene slučajne spremenjivke z rodovno funkcijo G_2 in je N slučajna spremenljivka z vrednostmi v $\mathbb{N}_0$, neodvisna od slučajnih spremenljivk X_i in z rodovno funkcijo G_1, ima vsota $X_1 + X_2 + \ldots X_N$ rodovno funkcijo $G(s) = G_1\big(G_2(s)\big)$.

5. Maks prav tako še nima otrok. Porazdelitev število njegovih otrok in števila otrok posameznega njegovega potomca je enaka:

$$\left(\begin{array}{ccc} 0 & 1 & 2 \\ 1/6 & 1/3 & 1/2 \end{array}\right).$$

Privzamemo še, da so števila otrok vsakega posameznika neodvisna. Določite verjetnost, da bo Maksovo potomstvo nekoč izumrlo.

Pri procesu razvejanja, pri katerem so števila neposrednih potomcev vsakega posameznika neodvisna in enako porazdeljena z rodovno funkcijo G, je verjetnost, da proces izumre, enaka $\min\{s \in [0,1] \; ; G(s) = s\}$.
Če je $E(X) < 1$, proces izumre z verjetnostjo ena.
Če je $X \geq 1$, proces izumre z verjetnostjo nič.

Karakteristična funkcija:

$$\phi_X(t) = E\big(e^{itX}\big)$$

Definirana je za vsak $t \in \mathbb{R}$.
Če je X celoštevilska, je tudi:

$$\phi_X(t) = G_X(e^{it}).$$

6. Izračunajte karakteristično funkcijo slučajne spremenljivke, ki je porazdeljena geometrijsko $\mathrm{G}(p)$.

7. Izračunajte karakteristično funkcijo slučajne spremenljivke, ki je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$.

8. Izračunajte karakteristično funkcijo slučajne spremenljivke, ki je porazdeljena zvezno z gostoto $p_X(x) = \frac{1}{2}e^{-|x|}$.

Če sta X in Y neodvisni, je $\phi_{X+Y} = \phi_X \phi_Y$.

9. Karakteristična funkcija binomske porazdelitve.

10. Karakteristična funkcija Pascalove porazdelitve.

11. Karakteristična funkcija porazdelitve gama.

Inverzna formula. *Če je slučajna spremenljivka X porazdeljena zvezno z gostoto p, ki je v dani točki x odvedljiva, velja formula:*

$$p(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(t)e^{-itx}\,\mathrm{d}t$$

kjer je ϕ karakteristična funkcija slučajne spremenljivke X (pod pogojem, da zgornji integral obstaja).

12. Slučajna spremenljivka X ima naslednjo karakteristično funkcijo:

$$\phi_X(t) = \begin{cases} 1-|x| & ; |x| \le 1 \\ 0 & ; \text{sicer} \end{cases}$$

Rekonstruirajte njeno porazdelitev.

13. Rekonstruirajte porazdelitev slučajne spremenljivke s karakteristično funkcijo $\phi_X(t) = e^{-|t|}$.

14. Rekonstruirajte porazdelitev slučajne spremenljivke s karakteristično funkcijo:

$$\phi_X(t) = \frac{\cos t + 2\cos^2 t + 2i\sin t\cos t}{3}$$

15. Neodvisne slučajne spremenljivke $X_1, \ldots X_n$ imajo *Cauchyjevo porazdelitev*, t. j. zvezno porazdelitev z gostoto:

$$p(x) = \frac{1}{\pi(1+x^2)}$$

Določite porazdelitev slučajne spremenljivke:

$$\bar{X} := \frac{X_1 + \cdots + X_n}{n}$$

7. Limitni izreki

Šibki in krepki zakon velikih števil. Centralni limitni izrek.

Zaporedje slučajnih spremenljivk $X_1, X_2, \ldots$ konvergira proti slučajni spremenljivki X:

- **skoraj gotovo** *($X_n \xrightarrow[n\to\infty]{\text{s.g.}} X$), če velja $P(\lim_{n\to\infty} X_n = X) = 1$;*
- **v verjetnosti** *($X_n \xrightarrow[n\to\infty]{P} X$), če za vsak $\varepsilon > 0$ velja* $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$.

Iz skoraj gotove konvergence sledi konvergenca v verjetnosti.

Krepki zakon velikih števil Kolmogorova. *Če so slučajne spremenljivke $X_1, X_2, \ldots$ neodvisne in enako porazdeljene ter če obstaja $\mu = E(X_n)$ (t. j. $E\big(|X_n|\big) < \infty$), velja:*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{s.g.}} \mu$$

1. *Standardni slučajni sprehod* je zaporedje slučajnih spremenljivk $S_n = X_1 + \cdots + X_n$, kjer so $X_1, X_2, \ldots$ neodvisne in:

$$X_i \sim \begin{pmatrix} -1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$$

   Ali zaporedje S_n/n konvergira proti nič v verjetnosti? Kaj pa skoraj gotovo?

2. Naj bo spet S_n standardni slučajni sprehod. Pokažite, da se le-ta skoraj gotovo vrne v izhodišče, t. j. da z verjetnostjo 1 obstaja tak $n \in \mathbb{N}$, da je $S_n = 0$.

3. Naj bo S_n *nesimetrični* slučajni sprehod, torej $S_n = X_1 + \cdots + X_n$, kjer so $X_1, X_2, \ldots$ neodvisne in:

$$X_i \sim \begin{pmatrix} -1 & 1 \\ 1-p & p \end{pmatrix}.$$

   Za vsak $k \in \mathbb{Z}$ izračunajte verjetnost, da sprehod (še) kdaj obišče stanje k.

4. Naj bo spet S_n standardni slučajni sprehod. Definirajmo slučajne spremenljivke T_n po predpisu:

$$T_n = \begin{cases} \dfrac{S_n}{n} & ; S_n \neq 0 \\ 2 & ; S_n = 0 \end{cases}$$

   Pokažite, da zaporedje T_n konvergira proti nič v verjetnosti, vendar pa je skoraj gotovo divergentno.

Centralni limitni izrek. *Če je S vsota veliko neodvisnih slučajnih spremenljivk z dovolj lepimi porazdelitvami, od katerih nobena posebej ne izstopa, je približno $S \sim \mathrm{N}(\mu, \sigma)$, kjer je $\mu = E(S)$ in $\sigma^2 = D(S)$, torej je:*

$$P(a < X < b),\ P(a \le X \le b) \approx \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Če je $S = X_1 + X_2 + \cdots + X_n$ in so $X_1, \ldots X_n$ neodvisne, je seveda:

$$\begin{aligned} \mu &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ \sigma^2 &= D(X_1) + D(X_2) + \cdots + D(X_n)\,. \end{aligned}$$

5. Slučajne spremenljivke $X_1, \ldots X_{100}$ so neodvisne in porazdeljene diskretno po naslednji shemi:
$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0\text{'}5 & 0\text{'}1 & 0\text{'}3 & 0\text{'}1 \end{pmatrix}$$
Ocenite $P(170 < X_1 + \cdots + X_{100} < 210)$ in utemeljite odgovor.

6. Slučajna spremenljivka X ima porazdelitev hi kvadrat s 100 prostostnimi stopnjami. Približno izračunajte $P(X > 110)$ in $P(90 < X < 110)$. Odgovor utemeljite!

7. Potapljač se potaplja, dokler ne nabere 80 biserov. Pri vsakem potopu nabere največ en biser, pa še tega le z verjetnostjo 20%. Potopi so med seboj neodvisni.
   a) Kolikšno je pričakovano število potopov (matematično upanje)?
   b) Ocenite verjetnost, da se bo moral potopiti več kot 450-krat. Odgovor utemeljite (možni sta dve uporabi centralnega limitnega izreka)!

8. Naj bodo $X_1, X_2, X_3, \ldots$ neodvisne slučajne spremenljivke, porazdeljene diskretno po predpisih:
$$X_n \sim \begin{pmatrix} -n & 0 & n \\ \frac{1}{2n^2} & 1 - \frac{1}{n^2} & \frac{1}{2n^2} \end{pmatrix}.$$
Pokažite, da za njihove delne vsote centralni limitni izrek ne velja, čeprav imajo slučajne spremenljivke X_n enaka matematična upanja in disperzije.

8. Zadostne in postranske statistike

Zadostnost statistik. Fisher–Neymanov faktorizacijski izrek. Eksponentne družine porazdelitev in minimalne zadostne statistike. Postranske statistike, uporaba Basujevega izreka.

> *Statistika T, ki temelji na opažanju X, t. j. $T = t(X)$, je zadostna za model, v katerem je verjetnost odvisna od parametra θ, če je pogojna porazdelitev opažanja X glede na T neodvisna od θ (pri klasični statistiki je to neodvisnost v funkcijskem smislu).*

1. Dokažite, da je pri Bernoullijevem zaporedju poskusov, od katerih vsak uspe z verjetnostjo θ, število uspešnih poskusov zadostna statistika za θ.

2. Denimo, da zaporedje poskusov tvori *markovsko verigo* z začetno verjetnostjo uspeha θ in prehodno matriko:

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\theta}{2} & \frac{\theta}{2} \\ \frac{1-\theta}{2} & \frac{1+\theta}{2} \end{bmatrix} .$$

   To pomeni, če $X_k = 1$ pomeni, da k-ti poskus uspe, $X_k = 0$ pa, da ne uspe, velja:

$$\begin{aligned} &P_\theta(X_1 = 1) = \theta \,, \\ &P_\theta(X_{k+1} = x_{k+1} \mid X_1 = x_1, \ldots X_k = x_k) = p_{x_k x_{k+1}} \,; \quad k = 1, 2, \ldots n-1 \,. \end{aligned}$$

   a) Dokažite, da vsak poskus uspe z verjetnostjo θ.
   b) Dokažite, da število uspelih poskusov ni zadostna statistika za θ.
   c) Poiščite kakšno zadostno statistiko, katere razsežnost ni odvisna od velikosti vzorca.

3. Dokažite *Fisher–Neymanov faktorizacijski izrek* za diskretne porazdelitve: če opažanje X zavzame vrednosti v števni množici, je statistika $T = t(X)$ zadostna natanko tedaj, ko obstajata taki funkciji f in g, da je:

$$P_\theta(X = x) = f(x)\, g\big(t(x), \theta\big) \tag{$*$}$$

   za vse x in Θ.

Fisher–Neymanov faktorizacijski izrek v splošnem. Če je $p_X(x, \theta)$ verjetnostna funkcija ali pa gostota opažanja X, je statistika $T = t(X)$ zadostna za parameter θ natanko tedaj, ko obstajata taki funkciji f in g, da je:

$$p_X(x, \theta) = f(x)\, g\big(t(x), \theta\big)$$

za vse x in Θ.

V vseh nadaljnjih nalogah tega razdelka je na populaciji definirana statistična spremenljivka X s porazdelitvijo iz dane družine, na voljo pa imamo vzorec, na katerem ima dana spremenljivka vrednosti $X_1, X_2, \ldots X_n$, pri čemer so vrednosti neodvisne in porazdeljene enako kot X.

4. Poiščite kakšno enorazsežno zadostno statistiko za Poissonovo porazdelitev $\mathrm{P}(\lambda)$.

Eksponentna družina porazdelitev je tista, pri kateri se da verjetnostna funkcija ali gostota zapisati v obliki:

$$p_X(x, \theta) = f(x)\, g(\theta)\, e^{f_1(x)\, g_1(\theta) + f_2(x)\, g_2(\theta) + \cdots + f_m(x)\, g_m(\theta)}\,.$$

Pri zgoraj opisani družini je $\big(f_1(X), \ldots f_m(X)\big)$ zadostna statistika za to družino. Če so funkcije $1, g_1, \ldots g_m$ linearno neodvisne, je ta statistika minimalna, t. j. da se jo izraziti z vsako zadostno statistiko.

5. Pokažite, da je vzorčna vsota minimalna zadostna statistika za parameter pri Poissonovi porazdelitvi.

6. Poiščite minimalno zadostno statistiko za:

   a) normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, kjer je μ neznan, σ pa znan;
   b) normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, kjer je σ neznan, μ pa znan;
   c) normalno porazdelitev $\mathrm{N}(\mu, \sigma)$, kjer sta oba parametra neznana;
   d) normalno porazdelitev $\mathrm{N}\big(a, \sqrt{a}\big)$;
   e) normalno porazdelitev $\mathrm{N}(a, a)$.

7. Poiščite minimalno zadostno statistiko za porazdelitev gama, kjer sta oba parametra neznana.

Statistika U je **postranska** *(angl. ancillary), če njena porazdelitev ni odvisna od parametra.*
Postranske statistike so med drugim pomembne pri testiranju statističnega modela, v katerem so postranske, proti širšemu modelu (več o tem v 11. razdelku).

8. Denimo, da statistični model predvideva normalno porazdelitev $\mathrm{N}(1,1)$ ali pa $\mathrm{N}(-1,1)$. Poiščite kakšno netrivialno postransko statistiko.

9. Poiščite postransko statistiko za normalno porazdelitev $\mathrm{N}(\mu,\sigma)$, katere zaloga vrednosti ima čimvečjo dimenzijo:

   a) če je μ neznan, σ pa znan;
   b) če je σ znan, μ pa neznan;
   c) če sta μ in σ oba neznana.

Parametrični prostor *eksponentne družine z verjetnostno funkcijo ali gostoto:*
$$p_X(x,\theta) = f(x)\,g(\theta)\,e^{f_1(x)\,g_1(\theta)+f_2(x)\,g_2(\theta)+\cdots+f_m(x)\,g_m(\theta)}\,,$$
kjer parameter θ preteče množico Θ, je množica $\Gamma := \big\{(g_1(\theta), g_2(\theta), \ldots g_m(\theta))\ ;\ \theta\in\Theta\big\}$.
Brž ko parametrični prostor vsebuje kakšno odprto množico v $\mathbb{R}^m$, je naravna zadostna statistika $\big(f_1(X), \ldots f_m(X)\big)$ neodvisna od katere koli postranske statistike pri vseh $\theta\in\Theta$.

10. Iz postranskih statistik iz 9. naloge potegnite zaključke o neodvisnosti pomembnih statistik pri Gaussovih modelih.

11. Za testiranje hipoteze, da je slučajna spremenljivka porazdeljena normalno $\mathrm{N}(\mu,\sigma)$, kjer je σ dan, μ pa neznan, lahko uporabimo postransko statistiko $\sum_{i=1}^n (X_i-\bar{X})^2$. Določite njeno porazdelitev.

12. Za testiranje hipoteze, da je slučajna spremenljivka porazdeljena normalno $\mathrm{N}(\mu,\sigma)$, kjer je μ dan, σ pa neznan, lahko uporabimo postransko statistiko:
$$U := \frac{\bar{X}-\mu}{\sqrt{\sum_{i=1}^n (X_i-\mu)^2}}\,.$$
Vendar pa navadno uporabimo *Studentovo statistiko*:
$$T := \frac{\bar{X}-\mu}{\sqrt{\sum_{i=1}^n (X_i-\bar{X})^2}}\,\sqrt{n(n-1)}\,.$$
Dokažite, da sta statistiki T in U v deterministični bijektivni korespondenci in zato ekvivalentni (če smo natančni, se moramo v resnici omejiti na dogodek z verjetnostjo ena). Določite še porazdelitev statistike T.

9. Točkasto ocenjevanje parametrov

Doslednost in nepristranskost. Srednja kvadratična napaka. Pridobivanje cenilk: metoda momentov in metoda največjega verjetja.

Cenilka $\hat{\zeta}$ parametra ζ je nepristranska, če je $E(\hat{\zeta}) = \zeta$. Pristranskost cenilke $\hat{\zeta}$ definiramo kot:

$$B(\hat{\zeta}) := E(\hat{\zeta}) - \zeta$$

Cenilka je dosledna, če za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} P(|\hat{\zeta} - \zeta| < \varepsilon) = 1$$

Zadosten pogoj za doslednost je, da velja $\lim_{n\to\infty} q(\hat{\zeta}) = 0$, kjer je $q(\hat{\zeta})$ srednja kvadratična napaka, definirana po predpisu:

$$q(\hat{\zeta}) := E\big[(\hat{\zeta} - \zeta)^2\big] = D(\hat{\zeta}) + \big(B(\hat{\zeta})\big)^2$$

Manjša kot je srednja kvadratična napaka, učinkovitejša je cenilka.

1. Iz populacije velikosti N vzamemo vzorec velikosti n. Na populaciji je definirana statistična spremenljivka X, njene vrednosti na populaciji označimo z $x_1, x_2, \ldots x_N$, na vzorcu pa z $X_1, X_2, \ldots X_n$. Tedaj lahko definiramo *populacijsko povprečje*:

   $$\mu := E(X) = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

   in *vzorčno povprečje*:

   $$\bar{X} := \frac{X_1 + X_2 + \cdots + X_n}{n}\,.$$

   Dokažite, da je $\bar{X}$ nepristranska cenilka za μ, in izračunajte srednjo kvadratično napako. Je to dosledna cenilka? Ločite dve možnosti, in sicer, da gre za enostavni slučajni vzorec s ponavljanjem in brez ponavljanja.

2. Pri *stratificiranem vzorčenju* populacijo razdelimo na več podpopulacij – *stratumov* – in iz vsakega vzamemo vzorec predpisane velikosti. Recimo, da je populacija velika in da so razmerja med velikosti stratumov in celotne populacije enaka $p_1, p_2, \ldots p_r$.

   Na populaciji imamo spet definirano populacijsko spremenljivko X in želimo oceniti njeno populacijsko povprečje μ. Seveda velja:

   $$\mu = p_1\mu_1 + p_2\mu_2 + \cdots + p_r\mu_r\,,$$

   kjer so $\mu_1, \mu_2, \ldots \mu_r$ povprečja stratumov. Če iz vsakega stratuma i vzamemo enostavni slučajni vzorec velikosti n_i in z $\bar{X}_1, \bar{X}_2, \ldots \bar{X}_r$ označimo vzorčna povprečja na stratumih, je torej:

   $$\bar{X} := p_1\bar{X}_1 + p_2\bar{X}_2 + \cdots + p_r\bar{X}_r$$

   nepristranska cenilka za μ.

a) Izračunajte njeno disperzijo.

b) Dokažite, da je le-ta pri *proporcionalnem stratificiranem vzorčenju*, t. j. $n_i = np_i$, kjer je n velikost celotnega vzorca, manjša ali enaka disperziji cenilke, ki bi jo dobili iz enostavnega slučajnega vzorca brez stratifikacije. *Namig*: pomagajte si z *disperzijo znotraj skupin*:

$$\sigma_{\mathrm{W}}^2 := p_1\sigma_1^2 + p_2\sigma_2^2 + \ldots p_r\sigma_r^2\,,$$

kjer je σ_i^2 disperzija znotraj i-tega stratuma, in *disperzijo med skupinami*:

$$\sigma_{\mathrm{B}}^2 := p_1(\mu_1 - \mu)^2 + p_2(\mu_2 - \mu)^2 + \cdots + p_r(\mu_r - \mu)^2\,.$$

c) Recimo, da vzamemo dovolj velik vzorec, in da poznamo variance znotraj stratumov. Približno kako velike vzorce moramo vzeti iz posameznih stratumov pri določeni velikosti celotnega vzorca, da bo disperzija cenilke $\bar{X}$ najmanjša?

V vseh nadaljnjih nalogah tega razdelka je na populaciji definirana statistična spremenljivka X s porazdelitvijo iz dane družine, na voljo pa imamo vzorec, na katerem ima dana spremenljivka vrednosti $X_1, X_2, \ldots X_n$, pri čemer so vrednosti neodvisne in porazdeljene enako kot X.

3. Je vzorčno povprečje dosledna cenilka parametra a pri zvezni porazdelitvi z gostoto:

$$p(x) = \frac{1}{\pi\big(1 + (x-a)^2\big)}\,?$$

Metoda momentov temelji na tem, da za cenilke populacijskih momentov $z_k = E(X^k)$ vzamemo:

$$\hat{z}_k = \frac{X_1^k + X_2^k + \cdots + X_n^k}{n}$$

Tako neznane parametre $\zeta_1, \ldots \zeta_r$ najprej izrazimo z $z_1, \ldots z_r$; na isti način se potem cenilke $\hat{\zeta}_1, \ldots \hat{\zeta}_r$ izražajo z $\hat{z}_1, \ldots \hat{z}_r$. Cenilke $\hat{z}_k$ so nepristranske in dosledne, zato so tudi cenilke $\hat{\zeta}_k$, dobljene po metodi momentov, vedno dosledne. Niso pa nujno nepristranske.

4. Statistična spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(1/\alpha)$, kjer je $\alpha > 0$ neznan parameter. Po metodi momentov poiščite cenilko za α. Je cenilka nepristranska? Je dosledna?

5. Statistična spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 1 & 2 \\ 1-2a-b & a & b & a \end{pmatrix}$$

Po metodi momentov poiščite cenilki za parametra a in b. Ocenite ju iz naslednjega vzorca:

$$-1, 0, 1, 1, 0, 0, -1, 2, 1, 2$$

6. Statistična spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$p_X(x) = \frac{1}{2\alpha} \exp\left(-\frac{|x|}{\alpha}\right)$$

Po metodi momentov poiščite cenilko za neznani parameter α.

7. Statistična spremenljivka X naj ima končno disperzijo σ^2.

   a) Po metodi momentov poiščite cenilko za σ^2 in pokažite, da je le-ta vedno pristranska.
   b) Označimo cenilko iz prejšnje točke s $\hat{\sigma}$. Pokažite, da obstaka tak faktor k', odvisen le od velikosti vzorca, da je $S^2 := k'\hat{\sigma}^2$ nepristranska cenilka za σ^2.
   c) Privzemimo, da ima X končen četrti moment. Izračunajte srednji kvadratični napaki cenilk $\hat{\sigma}^2$ in S^2.
   d) Naj bo X porazdeljena normalno. Pokažite, da obstaja tak faktor k^*, odvisen le od velikosti vzorca, da je $k^*\hat{\sigma}^2$ najučinkovitejša izmed cenilk oblike $k\hat{\sigma}^2$, $k > 0$. Izračunajte še $q(k^*\hat{\sigma}^2)$.

8. Statistična spremenljivka X je porazdeljena po Poissonu $\mathrm{P}(\lambda)$. Dokažite, da sta $\bar{X}$ in:

$$S^2 := \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

obe nepristranski cenilki za λ. Katera ima manjšo disperzijo?

Naj bo $p(x, \theta)$ verjetnostna funkcija ali gostota (verjetje) neke porazdelitve z neznanim parametrom θ. Metoda največjega verjetja (maksimalne zanesljivosti) išče cenilko $\hat{\theta}$ za θ tam, kjer je vrednost $p(X, \hat{\theta})$ maksimalna. Če imamo dan enostavni slučajni vzorec, se verjetje izraža na naslednji način:

$$L = p(x_1, \theta)\, p(x_2, \theta) \ldots p(x_n, \theta)\,.$$

Za iskanje maksimuma navadno rešujemo enačbo:

$$\frac{\partial \ln L}{\partial \theta} = 0\,.$$

9. Statistična spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(1/\alpha)$, kjer je $\alpha > 0$ neznan parameter. Po metodi največjega verjetja poiščite cenilko za α. Dobimo isto cenilko kot po metodi momentov?

10. Statistična spremenljivka X je porazdeljena diskretno po naslednji shemi:

$$\begin{pmatrix} -1 & 0 & 1 & 2 \\ 1-2a-b & a & b & a \end{pmatrix}$$

Po metodi največjega verjetja iz vzorca:

$$-1, 0, 1, 1, 0, 0, -1, 2, 1, 2$$

ocenite parametra a in b. Dobimo isto oceno kot po metodi momentov?

11. Statistična spremenljivka X je porazdeljena enakomerno $\mathrm{E}(0, a)$.

    a) Naj bo A cenilka za a, dobljena po metodi momentov, M pa cenilka, dobljena po metodi največjega verjetja. Katera od cenilk je nepristranska? Katera je učinkovitejša?

    b) Dokažite, da je možno določiti tak k', odvisen le od velikosti vzorca, da je cenilka $M' = k'M$ nepristranska.

    c) Dokažite, da je možno določiti tak k^*, odvisen le od velikosti vzorca, da je cenilka $M^* = k^*M$ najučinkovitejša izmed cenilk oblike kM, kjer je $k > 0$, in sicer ne glede na vrednost parametra a. Primerjajte srednjo kvadratično napako cenilk M, M' in M^*.

10. Intervali zaupanja

Bernoullijevo zaporedje poskusov. Normalna porazdelitev z znanim σ. Normalna porazdelitev z obema neznanima parametroma, iščemo μ ali σ.

Interval zaupanja $[\zeta_{\min}, \zeta_{\max}]$ za ζ pri stopnji zaupanja β je določen z neenačbo:

$$P\big(\zeta_{\min} \le \zeta \le \zeta_{\max}\big) \ge \beta$$

ki mora veljati za vse verjetnostne mere P iz našega statističnega modela, $\zeta_{\min}$ in $\zeta_{\max}$ pa morata biti opazljivi. Če se da, interval izberemo tako, da je ζ natančna spodnja meja. Tipično vzamemo $\beta = 0\dot{}90$, $0\dot{}95$ ali $0\dot{}99$.

Interval zaupanja za delež. *Naj bo θ delež enot v populaciji, ki imajo določeno lastnost. Vzamemo reprezentativen vzorec $n > 30$ neodvisnih enot in k jih ima dano lastnost. Izračunamo:*

$$\hat{\theta} = \frac{k}{n}$$

kar je najučinkovitejša cenilka za θ. Nato poiščemo c, za katerega velja:

$$\Phi(c) = \frac{\beta}{2} \quad \textit{oz.} \quad c = z_{(1+\beta)/2}$$

Tu je z_p kvantil standardne normalne porazdelitve z indeksom p. Velja $z_p = t_p$ pri $df = \infty$. Nadalje izračunamo:

$$\Delta = c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Interval zaupanja je potem $\hat{\theta} - \Delta \le \theta \le \hat{\theta} + \Delta$.

1. Izmed 100 poskusov jih je uspelo 20. Določite 95% interval zaupanja za verjetnost, da poskus uspe.

2. Pri 10000 metih kovanca je padlo 5048 grbov. Določite 90% interval zaupanja za verjetnost, da pade grb.

V vseh nadaljnjih nalogah tega razdelka je na populaciji definirana statistična spremenljivka X s porazdelitvijo iz dane družine, na voljo pa imamo vzorec, na katerem ima ustrezna spremenljivka vrednosti $X_1, X_2, \ldots X_n$, pri čemer so vrednosti neodvisne in porazdeljene enako kot X.

Ocenjujemo parameter μ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, *kjer je σ znan. Izračunamo:*

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c &= z_{(1+\beta)/2} \\
\Delta &= \frac{c\sigma}{\sqrt{n}}
\end{aligned}$$

Interval zaupanja: $\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta$.

3. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, 5)$. Vrednosti na vzorcu so:

   101, 91, 93, 103, 91, 101, 103, 95, 95

   Določite 95% interval zaupanja za μ.

Ocenjujemo parameter μ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, *kjer σ ni znan. Izračunamo:*

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c &= t_{(1+\beta)/2} \ \textit{pri}\ df = n - 1 \\
S &= \sqrt{\frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}} \\
\Delta &= \frac{cS}{\sqrt{n}}
\end{aligned}$$

Interval zaupanja: $\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta$.

4. Isto kot prejšnja naloga, le da σ zdaj ni znan.

Ocenjujemo parameter σ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, *kjer μ ni znan. Izračunamo:*

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c_1 &= \chi^2_{(1-\beta)/2} \ \textit{pri}\ df = n - 1 \\
c_2 &= \chi^2_{(1+\beta)/2} \ \textit{pri}\ df = n - 1 \\
S^2 &= \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}
\end{aligned}$$

Interval zaupanja:

$$S\sqrt{\frac{n-1}{c_2}} \leq \sigma \leq S\sqrt{\frac{n-1}{c_1}}$$

oziroma:

$$\sqrt{\frac{1}{c_2}\sum_{i=1}^{n}(X_i - \bar{X})^2} \leq \sigma \leq \sqrt{\frac{1}{c_1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

5. Statistična spremenljivka je porazdeljena normalno $N(\mu, \sigma)$. Vrednosti na vzorcu so:

$$124, 129, 126, 122, 124$$

Ocenite σ po občutku. Ali pride v 90% interval zaupanja?

6. Telesna teža v skupini 75 učencev ima naslednjo frekvenčno porazdelitev:

teža [kg]	39	40	41	42	43	44	45	46	47	48	49	50
št. učencev	1	3	5	8	8	7	9	8	6	6	4	3

teža [kg]	51	52	53	54	59
št. učencev	2	2	1	1	1

Privzemimo, da ta skupina predstavlja enostavni slučajni vzorec iz populacije, kjer je telesna teža porazdeljena normalno $N(\mu, \sigma)$. Poiščite 95% interval zaupanja za σ.

11. Testi značilnosti

Z-test. T-test sredine in razlike sredin. Testiranje disperzije. Pearsonov test hi kvadrat. Kontingenčni test. Test z znaki. Wilcoxon–Mann–Whitneyev test.

Želeli bi testirati, ali so podatki v vzorcu v skladu z ničelno hipotezo H_0 o porazdelitvi populacije ali pa so morda bolj v skladu z alternativno hipotezo H_1. Pri testih značilnosti bodisi zavrnemo ničelno hipotezo bodisi pravimo, da odstopanja niso statistično dovolj značilna, da bi jo zavrnili.
Postopku, po katerem se odločimo, ali bomo ničelno hipotezo zavrnili ali ne, pravimo test. Test ima stopnjo značilnosti α, če za vsako verjetnostno mero P, za katero velja H_0, velja:

$$P(H_0 \text{ zavrnemo}) \leq \alpha\,.$$

Če se da, test načtrujemo tako, da je α natančna zgornja meja.
Ena možnost je, da vzorce razvrstimo glede na to, kateri bolj "ustreza" ničelni hipotezi. Hipotezo zavrnemo, če je pri veljavnosti ničelne hipoteze verjetnost, da bo vzorec toliko ali še manj "ustrezen", kot je dobljeni vzorec, vselej največ α. Slednji verjetnosti pravimo tudi p-vrednost.
Če ničelno hipotezo zavrnemo pri $\alpha = 0{\cdot}05$ oz. če je p-vrednost manjša ali enaka $0{\cdot}05$, pravimo, da so odstopanja statistično značilna. Če se to zgodi pri pragu $0{\cdot}01$, pa pravimo, da so statistično zelo značilna.
"Ustreznost" določimo glede na ničelno in alternativno hipotezo. "Ustreznejši" vzorci so bolj verjetni pri ničelni, manj "ustrezni" pa pri alternativni hipotezi.

1. Loterija za neko vrsto srečke trdi, da jih je več kot pol dobitnih. Kupili smo osem srečk in le dve sta bili dobitni. Lahko začnemo sumiti, da loterija laže?

2. Pri 10000 metih kovanca je padlo 4875 grbov. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je verjetnost, da pade grb, enaka 1/2, proti alternativni hipotezi, da je različna od 1/2. Kaj pa, če bi vzeli $\alpha = 0{\cdot}01$?

3. Prireditelj neke igre na srečo navaja, da je verjetnost, da bo v posamezni igri izžreban glavni dobitek, enaka 1/50. Recimo, da po 100 igrah glavni dobitek še vedno ni bil izžreban. Ali pri stopnji značilnosti $\alpha = 0{\cdot}1$ lahko začnemo sumiti prireditelja, da goljufa?

Veliko testov poteka tako, da izračunamo testno statistiko T in H_0 zavrnemo, če T pade v kritično območje K_α. Tipična ničelna hipoteza je $\zeta = \zeta_0$, kjer je ζ kak parameter porazdelitve, tipične alternativne hipoteze pa so $\zeta \neq \zeta_0$ (dvostranski test) ali $\zeta < \zeta_0$ oz. $\zeta > \zeta_0$ (enostranski test).

Testiranje deleža. *Naj bo θ delež enot v populaciji, ki imajo določeno lastnost. Testiramo ničelno hipotezo H_0, ki pravi, da je $\theta = \theta_0$. Vzamemo reprezentativen vzorec $n > 30$ neodvisnih enot in k jih ima dano lastnost. Ker pri veljavnosti H_0 približno velja:*

$$Z := \sqrt{\frac{n}{\theta_0(1-\theta_0)}}\left(\frac{k}{n} - \theta_0\right) \sim \mathrm{N}(0,1)$$

glede na H_1 postavimo:

$$\begin{array}{ll} H_1\colon\ \theta \neq \theta_0 : & K_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty) \\ H_1\colon\ \theta < \theta_0 : & K_\alpha = (-\infty, -z_{1-\alpha}) \\ H_1\colon\ \theta > \theta_0 : & K_\alpha = (z_{1-\alpha}, \infty) \end{array}$$

4. Tovarna jamči, da je delež izdelkov z napako enak 20%. V vzorcu 100 izdelkov pa jih je 23 z napako. Pri stopnji značilnosti $\alpha = 0\text{˙}05$ testirajte hipotezo, da je verjetnost, da ima izdelek napako, enaka $0\text{˙}2$, proti alternativni hipotezi, da je večja od $0\text{˙}2$. Kaj pa, če bi imelo napako 230 izdelkov izmed 300?

V vseh nadaljnjih nalogah tega razdelka imamo za vsako statistično spremenljivko X, definirano na populaciji, na voljo vzorec, na katerem ima ustrezna spremenljivka vrednosti $X_1, X_2, \ldots X_n$, pri čemer so vrednosti neodvisne in porazdeljene enako kot X.

Testiranje parametra μ pri normalni porazdelitvi z znanim σ. *Testiramo ničelno hipotezo $H_0\colon\ \mu = \mu_0$. Ker pri H_0 velja:*

$$Z := \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)$$

glede na H_1 spet postavimo:

$$\begin{array}{ll} H_1\colon\ \mu \neq \mu_0 : & K_\alpha = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty) \\ H_1\colon\ \mu < \mu_0 : & K_\alpha = (-\infty, -z_{1-\alpha}) \\ H_1\colon\ \mu > \mu_0 : & K_\alpha = (z_{1-\alpha}, \infty) \end{array}$$

5. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, 5)$, dajo naslednje vrednosti:

   101, 91, 93, 103, 91, 101, 103, 95, 95

   Pri stopnji značilnosti $\alpha = 0\text{˙}05$ testirajte ničelno hipotezo, da je $\mu = 100$, proti alternativni hipotezi, da je $\mu \neq 100$. Kaj pa, če bi za alternativno hipotezo vzeli $\mu < 100$ ali $\mu > 100$?

Testiranje parametra μ pri normalni porazdelitvi z neznanim σ. *Testiramo ničelno hipotezo $H_0\colon\ \mu = \mu_0$. Ker pri H_0 velja:*

$$T := \frac{\bar{X} - \mu_0}{S}\sqrt{n} \sim \mathrm{S}(n-1)$$

kjer je S definiran tako kot pri konstrukciji intervala zaupanja za ta primer, glede na H_1 tokrat postavimo:

$$\begin{array}{ll} H_1\colon\ \mu \neq \mu_0: & K_\alpha = (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, \infty) \\ H_1\colon\ \mu < \mu_0: & K_\alpha = (-\infty, -t_{1-\alpha}) \\ H_1\colon\ \mu > \mu_0: & K_\alpha = (t_{1-\alpha}, \infty) \end{array}$$

in sicer pri $df = n - 1$.

6. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

   109, 107, 95, 121, 107, 97, 111, 121, 95

   Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte ničelno hipotezo, da je $\mu = 100$, proti alternativni hipotezi, da je $\mu \neq 100$. Kaj pa, če bi vedeli, da je $\sigma = 10$?

7. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

   46, 51, 48, 46, 52, 47, 51, 44, 47, 48

   Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte ničelno hipotezo, da je $\mu = 50$, proti alternativni hipotezi, da je $\mu \neq 50$. Kaj pa, če bi za alternativno hipotezo vzeli, da je $\mu < 50$?

Testiranje enakosti sredin dveh količin na isti populaciji. *Če sta na isti populaciji definirani statistični spremenljivki X in Y, za kateri velja $(X, Y) \sim \mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, lahko ničelno hipotezo $H_0\colon\ \mu_1 = \mu_2$ testiramo tako, da ustrezno testiramo razliko $X - Y$. Seveda lahko to storimo, brž ko je razlika $X - Y$ porazdeljena normalno.*

8. Na desetih osebah so preizkušali učinek neke diete proti debelosti. Osebe so stehtali pred začetkom in po koncu diete. Podatki so naslednji:

Pred dieto	125	131	126	117	114
Po dieti	121	118	119	121	113

Pred dieto	134	123	135	100	117
Po dieti	118	111	130	97	118

   Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte ničelno hipotezo, da dieta nima učinka, proti alternativni hipotezi, da ima shujševalni učinek. Privzeti smete, da je vektor telesne teže pred in po dieti porazdeljen dvorazsežno normalno.

Testiranje enakosti sredin na dveh populacijah. *Na prvi populaciji je dana statistična spremenljivka* $X \sim \mathrm{N}(\mu_1, \sigma)$, *na drugi pa* $Y \sim \mathrm{N}(\mu_2, \sigma)$. *Iz vsake populacije vzamemo ustrezen vzorec, pri čemer sta vzorca tudi med seboj neodvisna. Vrednosti spremenljivke* X *na prvem vzorcu pridejo* $X_1, \ldots X_m$, *vrednosti spremenljivke* Y *na drugem vzorcu pa* $Y_1, \ldots Y_n$. *Testiramo hipotezo* $H_0\colon\ \mu_1 = \mu_2$. *Ker pri* H_0 *velja:*

$$T := \frac{\bar{X} - \bar{Y}}{S} \sqrt{\frac{mn}{m+n}} \sim \mathrm{S}(m+n-2)$$

kjer je:

$$S = \sqrt{\frac{(X_1 - \bar{X})^2 + \cdots + (X_m - \bar{X})^2 + (Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2}{m+n-2}}$$

glede na H_1 *tokrat postavimo:*

$$\begin{array}{ll} H_1\colon\ \mu_1 \neq \mu_2 : & K_\alpha = (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, \infty) \\ H_1\colon\ \mu_1 < \mu_2 : & K_\alpha = (-\infty, -t_{1-\alpha}) \\ H_1\colon\ \mu_1 > \mu_2 : & K_\alpha = (t_{1-\alpha}, \infty) \end{array}$$

pri $df = m+n-2$.

9. Vzorčne vrednosti statistične spremenljivke $X \sim \mathrm{N}(\mu_1, \sigma)$ iz prve populacije pridejo:

   $$25,\ 16,\ 23,\ 17,\ 22,\ 18,\ 18,\ 21,\ 20\,,$$

   vzorčne vrednosti statistične spremenljivke $Y \sim \mathrm{N}(\mu_2, \sigma)$ iz druge populacije pa pridejo:

   $$19,\ 21,\ 23,\ 21,\ 25,\ 21,\ 24$$

   Pri stopnji značilnosti $\alpha = 0{\dot{}}05$ testirajte hipotezo, da je $\mu_1 = \mu_2$, proti alternativni hipotezi, da je $\mu_1 \neq \mu_2$.

10. Vzorčne vrednosti statistične spremenljivke $X \sim \mathrm{N}(\mu_1, \sigma)$ iz prve populacije pridejo:

    $$102,\ 96,\ 103,\ 98,\ 105,\ 97,\ 103,\ 98,\ 100,\ 98,\ 99,\ 101$$

    vzorčne vrednosti statistične spremenljivke $Y \sim \mathrm{N}(\mu_2, \sigma)$ iz druge populacije pa pridejo:

    $$95,\ 97,\ 95,\ 99,\ 95,\ 95$$

    Pri stopnji značilnosti $\alpha = 0{\dot{}}01$ testirajte hipotezo, da je $\mu_1 = \mu_2$, proti alternativni hipotezi, da je $\mu_1 > \mu_2$.

Testiranje enakosti sredin več populacij z analizo variance (ANOVA). *Danih je k populacij, na vsaki je definirana statistična spremenljivka, naj bodo to $X_1 \sim \mathrm{N}(\mu_1, \sigma)$, $\ldots X_k \sim \mathrm{N}(\mu_k, \sigma)$. Iz vsake populacije vzamemo vzorec, pri čemer so vse enote vzorcev med seboj neodvisne. Vrednosti na vzorcu iz i-te populacije označimo z $X_{i1}, \ldots X_{in_i}$. Testiramo hipotezo $H_0\colon\ \mu_1 = \mu_2 = \ldots = \mu_k$, alternativna hipoteza H_1 pa je nasprotje H_0. Izračunajmo:*

$$\bar{X}_i := \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}\,, \quad n := \sum_{i=1}^{k} n_i\,, \quad \bar{X} := \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} = \sum_{i=1}^{k}\frac{n_i}{n}\,\bar{X}_i\,,$$

$$S_{\mathrm{B}}^2 := \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2\,, \quad S_{\mathrm{W}}^2 := \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2\,.$$

Če velja H_0, sta S_{B}^2 in S_{W}^2 neodvisni ter $S_{\mathrm{B}}^2/\sigma^2 \sim \chi^2(k-1)$ in $S_{\mathrm{W}}^2/\sigma^2 \sim \chi^2(n-k)$, zato je:

$$F := \frac{S_{\mathrm{B}}^2/(k-1)}{S_{\mathrm{W}}^2/(n-k)} \sim \mathrm{F}(k-1, n-k)$$

kjer je $\mathrm{F}(k-1, n-k)$ Snedecorjeva porazdelitev. Zato postavimo $K_\alpha := (F_{1-\alpha}, \infty)$ pri $df_1 = k-1$ in $df_2 = n-k$.

11. Pacientom, ki so jim dajali določena zdravila, so merili neki parameter. Meritve so dale naslednje vrednosti:

Aspirin:	3, 5, 3, 5
Tilenol:	2, 2, 4, 4
Placebo:	2, 1, 2

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je vrednost parametra neodvisna od tega, ali pacient jemlje katero izmed obeh zdravil ali pa sploh nobenega.

Testiranje parametra σ pri normalni porazdelitvi z neznanim μ. *Testiramo ničelno hipotezo $H_0\colon\ \sigma = \sigma_0$. Ker pri H_0 velja:*

$$\chi^2 := (n-1)\,\frac{S^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$$

glede na H_1 postavimo:

$$\begin{aligned} &H_1\colon\ \sigma \neq \sigma_0: && K_\alpha = [0, \chi^2_{\alpha/2}) \cup (\chi^2_{1-\alpha/2}, \infty)\\ &H_1\colon\ \sigma < \sigma_0: && K_\alpha = [0, \chi^2_{\alpha})\\ &H_1\colon\ \sigma > \sigma_0: && K_\alpha = (\chi^2_{1-\alpha}, \infty) \end{aligned}$$

pri $df = n-1$.

12. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$99,\ 90,\ 108,\ 111,\ 97,\ 93,\ 90,\ 106,\ 104,\ 102$$

Pri $\alpha = 0{\text{'}}05$ testirajte:

a) ničelno hipotezo, da je $\sigma = 5$, proti alternativni hipotezi, da je $\sigma \neq 5$;

b) ničelno hipotezo, da je $\sigma = 10$, proti alternativni hipotezi, da je $\sigma < 10$.

Testiranje čisto določene porazdelitve. *Testiramo, ali je porazdelitev dane statistične spremenljivke enaka:*

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & \cdots & p_r \end{pmatrix}$$

Pri tem so lahko $a_1, \ldots a_r$ dejanske vrednosti spremenljivke ali pa le razredi, v katere pade. Vzorec velikosti n ima frekvenčno porazdelitev:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ N_1 & N_2 & \cdots & N_r \end{pmatrix}$$

Izračunamo teoretične frekvence $\tilde{N}_k := np_k$, za katere mora veljati $\tilde{N}_k \geq 5$ (sicer razrede združimo). Ker tedaj pri veljavnosti ničelne hipoteze približno velja:

$$\chi^2 := \sum_{k=1}^{r} \frac{(N_k - \tilde{N}_k)^2}{\tilde{N}_k} \sim \chi^2(r-1)$$

postavimo:

$$K_\alpha = (\chi^2_{1-\alpha}, \infty)$$

pri $df = r - 1$.

13. V vzorcu je 20 osebkov tipa RR, 45 tipa Rr in 35 tipa rr. Pri stopnji značilnosti $\alpha = 0{\text{'}}05$ testirajte hipotezo, da je v populaciji 25% osebkov tipa RR, 50% tipa Rr in 25% tipa rr. Kaj pa, če bi bilo v vzorcu 200 osebkov tipa RR, 450 tipa Rr in 350 tipa rr?

14. Pri kvizu Lepo je biti milijonar od 22. novembra do 28. decembra 2003 je bil 21-krat pravilen odgovor A, 42-krat B, 77-krat C in 116-krat D. Pri stopnji značilnosti $\alpha = 0{\text{'}}01$ testirajte ničelno hipotezo, da so odgovori enakomerno porazdeljeni, proti alternativni hipotezi, da niso.

    Če izvzamemo prvih pet vprašanj, je bil A pravilen 21-krat, B 37-krat, C 53-krat in D 25-krat. Naredite isti test.

15. Podatki imajo naslednjo frekvenčno porazdelitev:

pod 10	10–20	20–30	30–40	nad 40
5	20	35	25	15

Pri stopnji značilnosti $\alpha = 0\text{˙}05$ testirajte hipotezo, da podatki izhajajo iz populacije, porazdeljene normalno $\mathrm{N}(25, 10)$.

Kontingenčni test neodvisnosti. *Testiramo, ali sta spremenljivki X in Y, definirani na isti populaciji, neodvisni, pri čemer v vzorcu kombinacija $X = a_i, Y = b_j$ nastopa N_{ij}-krat ($i = 1, \ldots r$, $j = 1, \ldots s$). Ker pri neodvisnosti približno velja:*

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(N_{ij} - \tilde{N}_{ij})^2}{\tilde{N}_{ij}} \sim \chi^2\big((r-1)(s-1)\big)$$

kjer je:

$$N_{i.} = \sum_{j=1}^{s} N_{ij}\ , \quad N_{.j} = \sum_{i=1}^{r} N_{ij}\ , \quad n = \sum_{i=1}^{n}\sum_{j=1}^{n} N_{ij}\ , \quad \tilde{N}_{ij} = \frac{N_{i.}N_{.j}}{n}$$

postavimo:

$$K_\alpha = (\chi^2_{1-\alpha}, \infty)$$

pri $df = (r-1)(s-1)$.

16. Na vzorcu 100 oseb dobimo naslednjo navzkrižno frekvenčno porazdelitev barve oči in las:

lasje \ oči	modre	zelene	rjave
blond	15	8	2
rdeči	5	2	3
rjavi	3	30	7
črni	2	10	13

Pri stopnji značilnosti $\alpha = 0\text{˙}01$ testirajte hipotezo, da sta barva las in barva oči neodvisni.

17. Na neki internetni strani je 1990 ljudi glasovalo, kateri film bo najverjetneje dobil oskarja. Anketirance so razdelili v tri razrede glede na starost. Rezultati ankete so naslednji:

	do 20	od 20 do 40	nad 40
Gospodar prstanov	350	250	180
Skrivnostna reka	80	100	100
Seabiscuit	70	90	130
Zgubljeno s prevodom	50	80	110
Gospodar in bojevnik	200	150	50

S kontingenčnim testom preizkusite domnevo, da je mnenje o oskarjih neodvisno od starosti. Stopnja značilnosti naj bo $0\text{˙}05$.

Test z znaki. *Testiramo, ali sta urejenostni statistični spremenljivki X in Y na isti populaciji enako porazdeljeni. Vzamemo vzorec velikosti n in z S^+ označimo število enot, pri katerih je $X > Y$, z S^- pa število enot, pri katerih je $X < Y$. Če sta X in Y porazdeljeni zvezno in enako ter vzorec dovolj velik, približno velja:*

$$Z := \frac{S^+ - S^-}{\sqrt{n}} \sim \mathrm{N}(0,1)$$

in tako za kritično območje vzamemo:

$$K_\alpha := (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, \infty)$$

pri $df = \infty$.

18. Meritve količin X in Y so zapisane v naslednji tabeli:

X	66	37	63	70	76	62	97	51	14	83	65	95	68	87	94
Y	24	51	21	51	6	77	5	48	100	31	23	61	61	74	76

X	62	65	62	65	83	19	61	45	70	24	52	81	80	18	81
Y	97	22	16	20	18	54	54	86	22	78	15	27	35	39	79

X	5	29	13	91	47	6	53	43	71	72	49	98	80	91	14
Y	15	24	37	58	23	43	58	46	59	49	1	18	23	32	55

X	26	85	70	75	35
Y	36	28	99	68	16

Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte hipotezo, da sta X in Y enako porazdeljeni.

Wilcoxon–Mann–Whitneyev test. *Testiramo, ali sta urejenostni statistični spremenljivki X in Y na različnih populacijah enako porazdeljeni. Iz njiju vzamemo neodvisna vzorca, iz prve velikosti m, iz druge pa velikosti n. Vzorca združimo in elemente uredimo po velikosti glede na vrednosti spremenljivk X in Y, recimo od namanjše do največje. Naj bodo $R_1, \ldots R_m$ mesta (rangi) elementov prvega vzorca. Če sta X in Y res enako porazdeljeni in vzorca dovolj velika, približno velja:*

$$Z := \sqrt{\frac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^{m} R_i - m(m+n+1)\right) \sim \mathrm{N}(0,1)$$

Navadno alternativna hipoteza trdi, da sta X in Y porazdeljeni različno. V tem primeru postavimo:

$$K_\alpha := (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

Alternativna hipoteza pa lahko trdi tudi, da je X stohastično strogo večja od Y, t. j. $P(X < w) \le P(Y < w)$ za vsak w in $P(X < w) < P(Y < w)$ za neki w). V tem primeru postavimo $K_\alpha := (z_{1-\alpha}, \infty)$. Podobno, če H_1 trdi, da je X stohastično strogo manjša od Y, postavimo $K_\alpha := (-\infty, -z_{1-\alpha})$.

19. Tekmovalci dveh ekip, "rdečih" in "modrih", so se pomerili v teku. Vrstni red tekmovalcev je naslednji:

$$R,\ R,\ M,\ R,\ R,\ M,\ R,\ R,\ M,\ R,\ M,\ M,\ M,\ R,\ M,\ M,\ M,\ M,\ R,\ M$$

(t. j. prvi, ki je prispel na cilj, je bil član "rdečih", drugi prav tako, tretji je bil član "modrih" itd.). Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte hipotezo, da so rdeči enako dobri od modrih, proti alternativni hipotezi, da je med njimi razlika.

12. Povezanost dveh številskih spremenljivk

Interval zaupanja za korelacijski koeficient. Testiranje nekoreliranosti. Linearna regresija.

Interval zaupanja za korelacijski koeficient $\rho = r(X,Y)$. *Najprej izračunamo vzorčni korelacijski koeficient R:*

$$\begin{aligned} C_x^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\ C_y^2 &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 \\ C_{xy} &= \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} \\ R &= \frac{C_{xy}}{C_x C_y} \end{aligned}$$

in ga normaliziramo:

$$Z := \operatorname{Arth} R = \frac{1}{2}\ln\frac{1+R}{1-R}$$

Približen interval zaupanja za ρ*:*

$$\operatorname{th}\left(Z - \frac{c}{\sqrt{n-3}}\right) \le \rho \le \operatorname{th}\left(Z + \frac{c}{\sqrt{n-3}}\right)$$

kjer je $c = z_{(1+\beta)/2}$ *in* $\operatorname{th}(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$.

1. Meritve krvnega pritiska so zbrane v naslednji tabeli:

sistolični	130	120	120	125	125	125	105	130	130	135
diastolični	80	80	85	80	75	75	75	80	85	70

sistolični	130	125	140	130	120
diastolični	75	80	90	80	85

Poiščite 95% interval zaupanja za korelacijski koeficient med sistoličnim in diastoličnim pritiskom.

Testiranje nekoreliranosti. *Če velja ničelna hipoteza H_0, da sta X in Y nekorelirani, je približno:*

$$T := \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \sim \mathrm{S}(n-2)$$

kjer je R vzorčni korelacijski koeficient. Tako glede na H_1 postavimo:

$$\begin{aligned} H_1\colon\ X \text{ in } Y \text{ sta korelirani}:&\quad K_\alpha = (-\infty, -t_{1-\alpha/2}) \cup (t_{1-\alpha/2}, \infty) \\ H_1\colon\ X \text{ in } Y \text{ sta negativno korelirani}:&\quad K_\alpha = (-\infty, -t_{1-\alpha}) \\ H_1\colon\ X \text{ in } Y \text{ sta pozitivno korelirani}:&\quad K_\alpha = (t_{1-\alpha}, \infty) \end{aligned}$$

pri $df = n-2$.

2. Za meritve krvnega pritiska iz 1. naloge pri stopnji značilnosti $\alpha = 0{\cdot}1$ testirajte ničelno hipotezo, da sta sistolični in diastolični pritisk nekorelirana, proti alternativni hipotezi, da sta korelirana.

Linearna regresija. *Spremenljivki X in Y zadoščata zvezi:*

$$Y = a + bX + R$$

kjer je $R \sim \mathrm{N}(0,\sigma)$ neodvisna od X in kjer so a, b in σ neznani parametri. Želeli bi točkasto oceniti a in b (t. j. potegniti premico skozi podatke), poleg tega pa še točkasto in intervalsko oceniti vrednost spremenljivke Y za dano realizacijo, pri kateri poznamo X. Cenilki za a in b sta:

$$\hat{b} = \frac{C_{\mathrm{xy}}}{C_{\mathrm{xx}}}\,, \quad \hat{a} = \bar{Y} - \hat{b}\bar{X}$$

Cenilka za Y pri danem X pa je:

$$\hat{Y} = \hat{a} + \hat{b}X$$

Interval zaupanja:

$$\hat{Y} - \Delta \le Y \le \hat{Y} + \Delta$$

kjer je:

$$\Delta = t_{(1+\beta)/2}\, S\sqrt{1 + \frac{1}{n} + \frac{(X-\bar{X})^2}{C_x^2}}$$

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \hat{a} - \hat{b}X_i)^2$$

pri $df = n-2$.

3. Meritve dajo naslednje vrednosti:

X_i	1	2	3	4	5
Y_i	2	6	7	10	10

Določite regresijsko premico, napovejte Y pri $X = 10$ in poiščite 95% interval zaupanja.

REŠITVE

1. Osnove kombinatorike

1. $(4+3)\cdot 5 = 35$.

2. a) 900, b) 450, c) 400, d) 9, e) 648, f) 90.

3. Označimo s k število kroglic, ki jih vzamemo iz posode. Tedaj so rezultati zbrani v naslednji tabeli:

k	vrstni red vlečenja			
	pomemben		ni pomemben	
	vračamo	ne vračamo	vračamo	ne vračamo
1	5	5	5	5
2	25	20	15	10
3	125	60	35	10

Pri različici, ko kroglic ne vračamo in vrstni red ni pomemben, sta rezultata za $k=2$ in $k=3$ enaka, ker lahko namesto tega, katere kroglice smo vzeli, gledamo, katere kroglice so ostale v posodi.

Različico, ko vrstni red ni pomemben, kroglice pa vračamo, v splošnem primeru, ko iz posode z n kroglicami vzamemo k kroglic, izračunamo tako, da jemanja ponazorimo z razporeditvijo $n-1$ rdečih in k modrih puščic v vrsto, pri čemer puščice ločimo le po barvi. Vsaka rdeča puščica predstavlja pregrajo med dvema kroglicama v škatli in vsaka modra puščica predstavlja jemanje ustrezne kroglice. Tako npr. razporeditev $MRRMMRR$ pomeni, da iz posode s 5 kroglicami vzamemo tri kroglice, in sicer prvo kroglico enkrat, tretjo dvakrat, ostalih pa ne vzamemo. Vsaka razporeditev puščic ustreza natanko enemu jemanju kroglic. Tako dobimo, da je vseh jemanj $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$ (glej naslednjo nalogo).

4. Če vse razločujemo: $(3+4)! = 5040$.
Če cvetlic iste vrste ne razločujemo: $\displaystyle\binom{7}{3} = 35$.

5. a) $\displaystyle\binom{6}{4}\binom{4}{2} = 90$,
b) $\displaystyle\binom{6}{1}\binom{4}{3} + \binom{6}{2}\binom{4}{2} + \binom{6}{3}\binom{4}{1} = \binom{10}{4} - \binom{6}{4} - \binom{4}{4} = 194$.

6. a) $9! = 362880$, b) $3!\,2!\,4! = 288$, c) $288\cdot 3! = 1728$, d) $\dfrac{9!}{3!\,2!\,4!} = 1260$.

7. $5! = 120$; $9\cdot 8\cdot 7\cdot 6\cdot 5 = 15120$.

2. Elementarna verjetnost

1. $\dfrac{5}{36} \doteq 0\dot{}139.$

2. Verjetneje je, da so trije enega, eden pa nasprotnega spola: verjetnost tega dogodka je $8/16 = 0\dot{}5$, verjetnost enake zastopanosti pa je $6/16 = 0\dot{}375$.

3. $1 - \dfrac{5^5}{6^5} \doteq 0\dot{}598.$

4. a) $\dfrac{5}{12} \cdot \dfrac{4}{12} \cdot \dfrac{3}{12} \cdot 6 \doteq 0\dot{}208,$ b) $\dfrac{5 \cdot 4 \cdot 3}{\binom{12}{3}} \doteq 0\dot{}273.$

5. $\dfrac{3}{10} + \dfrac{2}{10} - \dfrac{3}{10} \cdot \dfrac{2}{9} \doteq 0\dot{}433.$

6. $1 - \dfrac{364 \cdot 363 \cdot \dots \cdot (366-n)}{365^{n-1}}$. Tabela prvih nekaj verjetnosti:

n	P	n	P
1	0	36	0˙832
2	0˙00274	37	0˙849
3	0˙0082	38	0˙864
4	0˙0164	39	0˙878
5	0˙0271	40	0˙891
6	0˙0405	41	0˙903
7	0˙0562	42	0˙914
8	0˙0743	43	0˙924
9	0˙0946	44	0˙933
10	0˙117	45	0˙941
11	0˙141	46	0˙948
12	0˙167	47	0˙955
13	0˙194	48	0˙961
14	0˙223	49	0˙966
15	0˙253	50	0˙97
16	0˙284	51	0˙974
17	0˙315	52	0˙978
18	0˙347	53	0˙981
19	0˙379	54	0˙984
20	0˙411	55	0˙986
21	0˙444	56	0˙988
22	0˙476	57	0˙99
23	0˙507	58	0˙992
24	0˙538	59	0˙993
25	0˙569	60	0˙994
26	0˙598	61	0˙995
27	0˙627	62	0˙996
28	0˙654	63	0˙997
29	0˙681	64	0˙997
30	0˙706	65	0˙998
31	0˙73	66	0˙998
32	0˙753	67	0˙998
33	0˙775	68	0˙999
34	0˙795	69	0˙999
35	0˙814	70	0˙999

7. $\dfrac{\binom{4}{2}\binom{12}{6}}{\binom{16}{8}} = \dfrac{\left[\binom{8}{2}\right]^2}{\binom{16}{4}} \doteq 0\text{˙}431.$

8. $1 - \dfrac{\binom{90}{10} + \binom{10}{1}\binom{90}{9}}{\binom{100}{10}} \doteq 0\text{˙}262.$

9. $\dfrac{\binom{8}{4}\binom{4}{2}\binom{2}{1}}{\binom{14}{7}} \doteq 0\text{˙}245.$

10. $B \cap C$

11. a) $P(\bar{A} \cap \bar{B} \cap \bar{C}) = 0\text{˙}25$,
b) $P\big((A \cap \bar{B} \cap \bar{C}) \cup (\bar{A} \cap B \cap \bar{C}) \cup (\bar{A} \cap \bar{B} \cap C)\big) = 0\text{˙}25$,
c) $P\big((A \cap B) \cup (A \cap C) \cup (B \cap C)\big) = 0\text{˙}5$.

12. 76/120. V splošnem, ko je kuvert n, je rezultat:

$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + \frac{(-1)^{n-1}}{n!} = 1 - \frac{1}{n!}\left\langle \frac{n!}{e} \right\rangle = \frac{1}{n!}\left\langle n!\left(1 - \frac{1}{e}\right)\right\rangle$$

13. $\dfrac{3}{4} = 0\text{˙}75, \quad \dfrac{9}{16} - \dfrac{1}{32} \doteq 0\text{˙}531.$

14. a) $\dfrac{5^2}{2 \cdot 10 \cdot 5} \doteq 0\text{˙}179$, b) $\dfrac{2^2}{2 \cdot 10 \cdot 5} \doteq 0\text{˙}0286.$

15. Označimo s h razdaljo med točko na najbolj spodnjem delu igle in prvo črto, ki je nad to točko ($0 \le h < a$). Nadalje naj bo φ kot med iglo in pravokotnico na črte ($0 \le \varphi \le \pi/2$). Tedaj igla seka katero izmed črt natanko tedaj, ko je $h \le b\cos\varphi$, zato je:

$$P = \frac{2}{\pi a}\int_0^{\pi/2} \min\{a, b\cos\varphi\}\,\mathrm{d}\varphi = \begin{cases} \dfrac{2}{\pi}\left[\arccos\dfrac{a}{b} + \dfrac{b - \sqrt{b^2 - a^2}}{a}\right] & ;\ a \le b \\[2ex] \dfrac{2b}{\pi a} & ;\ a \ge b \end{cases}$$

3. Pogojna verjetnost

1. Če je kocka poštena, velja $P(A \mid L) = 1/3$ in $P(B \mid L) = 0$. Če pa je nepoštena na način, opisan v nalogi, pride $P(A \mid L) = 0{\dot{}}25$ in še vedno $P(B \mid L) = 0$.

2. Označimo z A dogodek, da je prva karta pik, z B pa dogodek, da sta med izvlečenimi kartami natanko dva pika.

Prvi način. S preprostim premislekom dobimo brezpogojni verjetnosti obeh dogodkov pa tudi pogojno verjetnost $P(B \mid A)$:

$$\begin{aligned}
P(A) &= \frac{1}{4}\,,\\
P(B) &= \frac{\binom{4}{2}\binom{12}{2}}{\binom{16}{4}} = \binom{4}{2} \cdot \frac{4}{16} \cdot \frac{3}{15} \cdot \frac{12}{14} \cdot \frac{11}{13} = \frac{9 \cdot 11}{5 \cdot 7 \cdot 13} = \frac{99}{455}\,,\\
P(B \mid A) &= \frac{\binom{3}{1}\binom{12}{2}}{\binom{15}{3}} = 3 \cdot \frac{3}{15} \cdot \frac{12}{14} \cdot \frac{11}{13} = \frac{2 \cdot 9 \cdot 11}{5 \cdot 7 \cdot 13} = \frac{198}{455}\,.
\end{aligned}$$

Sledi:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)\,P(B \mid A)}{P(B)} = \frac{1}{2}\,.$$

Drugi način. Za izide v našem verjetnostnem prostoru postavimo kar konfiguracije, pri katerih opišemo, katere izvlečene karte so piki in katere ne. Tedaj vsi izidi sicer *niso* enako verjetni, pač pa so enako verjetni izidi, ki sestavljajo dogodek B. Teh je $\binom{4}{2} = 6$, trije med njimi so v dogodku A. Zato je $P(A \mid B) = 1/2$.

3. Razmislek je napačen: brezpogojni verjetnosti sta enaki, ne pa tudi pogojni verjetnosti glede na dogodek, da smo izvlekli zlat kovanec.

4. Če je naš namen dobiti porscheja, je v drugo bolje izbrati vrata, ki jih v prvo *nismo* izbrali. Če s črko P označimo porscheja, s črko K pa kozo in s polkrepko črko vrata, ki smo jih prvotno izbrali, bi na prvi pogled lahko sklepali, da sta možnosti **P**K in **K**P enako verjetni (odprta vrata smo izpustili). Vendar pa temu ni tako. Če pišemo tudi odprta vrata, ki jih označimo z malo črko, imamo štiri možnosti: **P**kK, **P**Kk, **K**Pk in **K**kP. Le-te niso enako verjetne: prvi dve imata verjetnost 1/6, zadnji dve pa 1/3, saj moramo najprej gledati, kje je porsche, šele nato pa tudi, katera vrata je voditelj odprl.

5. Končni izid je lahko $6 : 2$, $6 : 3$, $6 : 4$ ali $6 : 5$, zato je verjetnost enaka $13/16 \doteq 0{\dot{}}822$. Še hitreje pridemo do rezultata, če vzamemo nasprotni dogodek.

6. Označimo s p verjetnost, da Janez zmaga. Po dveh igrah je rezultat lahko $2 : 0$, $1 : 1$ ali $0 : 2$ za Janeza. Od tod dobimo zvezo $p = \frac{1}{9} + \frac{4p}{9}$, kar da $p = 0{\dot{}}2$.

7. $\dfrac{5}{10} \cdot \dfrac{4 \cdot 3}{\binom{7}{2}} + \dfrac{3}{10} \cdot \dfrac{3 \cdot 3}{\binom{7}{2}} + \dfrac{2}{10} \cdot \dfrac{4 \cdot 3}{\binom{7}{2}} \doteq 0{\dot{}}529.$

8. $\dfrac{0{\cdot}6 \cdot 0{\cdot}2}{0{\cdot}6 \cdot 0{\cdot}2 + 0{\cdot}4 \cdot 0{\cdot}1} = 0{\cdot}75.$

9. $\dfrac{0{\cdot}2 \cdot 0{\cdot}05}{0{\cdot}2 \cdot 0{\cdot}05 + 0{\cdot}5 \cdot 0{\cdot}01 + 0{\cdot}3 \cdot 0{\cdot}02} \doteq 0{\cdot}476.$

10. $$\dfrac{\dfrac{\binom{4}{3}}{\binom{10}{3}} \cdot \dfrac{4}{5} \cdot \dfrac{3}{4}}{\dfrac{\binom{4}{3}}{\binom{10}{3}} \cdot \dfrac{4}{5} \cdot \dfrac{3}{4} + \dfrac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} \cdot \dfrac{3}{5} \cdot \dfrac{2}{4} + \dfrac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} \cdot \dfrac{2}{5} \cdot \dfrac{1}{4}} = 0{\cdot}125.$$

11. Da, da, ne, ne.

12. Ne (čeprav je $P(A \cap B \cap C) = P(A)\ P(B)\ P(C)$).

13. a) Ne. b) Pri $p = 0$ in $p = 2/3$.

14. Pri ženskah je bilo prvo zdravilo uspešno v $200/1000 = 10\%$, drugo pa v $10/100 = 5\%$ primerov.

Pri moških je bilo prvo zdravilo uspešno v $190/200 = 95\%$, drugo pa v $1000/2000 = 50\%$ primerov.

Kaže torej, da je bilo prvo zdravilo uspešnejše od drugega tako pri moških kot tudi pri ženskah. Toda če pogledamo oboje skupaj, je bilo prvo zdravilo uspešno v $390/2200 \doteq 17{\cdot}7\%$, drugo zdravilo pa v $1010/2200 \doteq 45{\cdot}9\%$ primerov. Ko skupini združimo, kaže, da je uspešnejše drugo zdravilo.

Kako naj si to razložimo? Zapišimo vse skupaj v jeziku verjetnostnega računa. Za verjetnostni prostor vzamemo množico testirancev, pri čemer so vsi enako verjetni. Označimo:

$$\begin{aligned} M &:= \text{množica moških} \\ Z &:= \text{množica žensk} \\ S_1 &:= \text{množica tistih, ki so prejemeli prvo zdravilo} \\ S_2 &:= \text{množica tistih, ki so prejemeli prvo zdravilo} \\ Z &:= \text{množica tistih, pri katerih je bilo zdravljenje uspešno} \end{aligned}$$

Tedaj velja:

$$\begin{aligned} P(U \mid S_1 \cap Z) &> P(U \mid S_2 \cap Z) \\ P(U \mid S_1 \cap M) &> P(U \mid S_2 \cap M) \\ P(U \mid S_1) &< P(U \mid S_2)\,. \end{aligned}$$

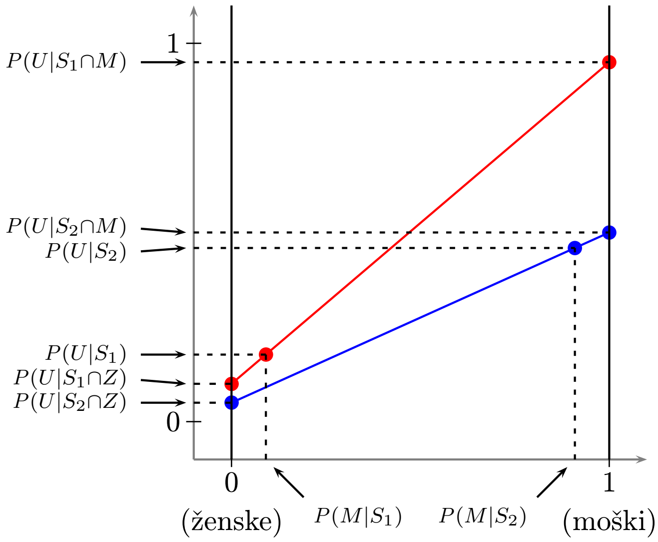
Oglejmo si zdaj, kako se $P(U \mid S_i)$ izraža z $P(U \mid S_i \cap Z)$ in $P(U \mid S_i \cap M)$. Velja:

$$\begin{aligned}
P(U \mid S_1) &= \frac{P(U \cap S_1)}{P(S_1)} = \\
&= \frac{P(U \cap S_1 \cap Z)}{P(S_1)} + \frac{P(U \cap S_1 \cap M)}{P(S_1)} = \\
&= \frac{P(S_1 \cap Z)\, P(U \mid S_1 \cap Z)}{P(S_1)} + \frac{P(S_1 \cap M)\, P(U \mid S_1 \cap M)}{P(S_1)} = \\
&= P(Z \mid S_1)\, P(U \mid S_1 \cap Z) + P(M \mid S_1)\, P(U \mid S_1 \cap M)\,.
\end{aligned}$$

in podobno:

$$P(U \mid S_2) = P(Z \mid S_2)\, P(U \mid S_2 \cap Z) + P(M \mid S_2)\, P(U \mid S_2 \cap M)\,.$$

Za $i = 1, 2$ sta $P(Z \mid S_i)$ in $P(M \mid S_i)$ uteži z vsoto 1, zato lahko verjetnosti oz. deleže takole prikažemo:



Do paradoksa ne bi prišlo, če bi bili obe vmesni bunkici navpično poravnani, t. j. če bi veljalo $P(M \mid S_1) = P(M \mid S_2)$ ali ekvivalentno $P(Z \mid S_1) = P(Z \mid S_2)$. Ker sta dogodka S_1 in S_2 nasprotna, je to ekvivalentno zahtevi, da sta dogodka Z in S_1 neodvisna, oziroma zahtevi, da sta spol testiranca in zdravilo, ki ga prejme, neodvisna. Z drugimi besedami, do paradoksa ne bi prišlo, če bi pri raziskavi pazili na delež žensk, ki so jim dali posamezno zdravilo – natančneje, če bi imeli skupini testirancev, ki so prejemali posamezno zdravilo, enako razporeditev spolov.

Posploševanje (statistično sklepanje) z vzorca na populacijo se obnese, če je vzorec *reprezentativen* glede na populacijo, torej če so lastnosti (v našem primeru spol), ki vplivajo na eksperimentalne spremenljivke (v našem primeru uspešnost zdravljenja), vsaj približno enako zastopane kot v populaciji. Do paradoksa torej ne bi prišlo, če bi bili obe skupini reprezentativni glede na spol.

V končni fazi bi lahko izrekli sklep, da je učinkovitejše *prvo* zdravilo, če bi bila različna zastopanost spolov edina šibka točka te raziskave (t. j. če bi bile vse ostale lastnosti, ki pomembno vplivajo na delovanje zdravila, reprezentativno zastopane, to pa je dostikrat težko doseči, ker jih ne poznamo).

15. Najprej pogledamo, ali srednje stikalo prepušča tok. Od tod dobimo, da je verjetnost, da vezje prepušča tok, enaka:

$$\frac{1}{3}\left[1-\left(\frac{2}{3}\right)^2\right]^2+\frac{2}{3}\left[1-\left(1-\left(\frac{1}{3}\right)^2\right)^2\right]=\frac{59}{243}\doteq 0\text{˙}243\,.$$

16. a) $\dfrac{0\text{˙}1}{1-0\text{˙}9\cdot 0\text{˙}8\cdot 0\text{˙}7}\doteq 0\text{˙}202,$

b) $\dfrac{0\text{˙}1\cdot 0\text{˙}8\cdot 0\text{˙}7}{0\text{˙}1\cdot 0\text{˙}8\cdot 0\text{˙}7+0\text{˙}9\cdot 0\text{˙}2\cdot 0\text{˙}7+0\text{˙}9\cdot 0\text{˙}8\cdot 0\text{˙}3}\doteq 0\text{˙}141.$

17. $\dfrac{0\text{˙}4\cdot 0\text{˙}7\cdot 0\text{˙}5\cdot 0\text{˙}9}{(0\text{˙}6\cdot 0\text{˙}3+0\text{˙}4\cdot 0\text{˙}7)\cdot(0\text{˙}5\cdot 0\text{˙}1+0\text{˙}5\cdot 0\text{˙}9)}\doteq 0\text{˙}548.$

18.
$$\frac{\binom{20}{3}}{\binom{50}{3}}\big((3\cdot 0\text{˙}1^2\cdot 0\text{˙}9+0\text{˙}1^3\big)+\frac{\binom{30}{1}\binom{20}{2}}{\binom{50}{3}}\big(0\text{˙}1^2+2\cdot 0\text{˙}1\cdot 0\text{˙}9\cdot 0\text{˙}7\big)+$$
$$+\frac{\binom{30}{2}\binom{20}{1}}{\binom{50}{3}}\big(0\text{˙}7^2+2\cdot 0\text{˙}7\cdot 0\text{˙}3\cdot 0\text{˙}1\big)+\frac{\binom{30}{3}}{\binom{50}{3}}\big(3\cdot 0\text{˙}7^2\cdot 0\text{˙}3+0\text{˙}7^3\big)\doteq 0\text{˙}440.$$

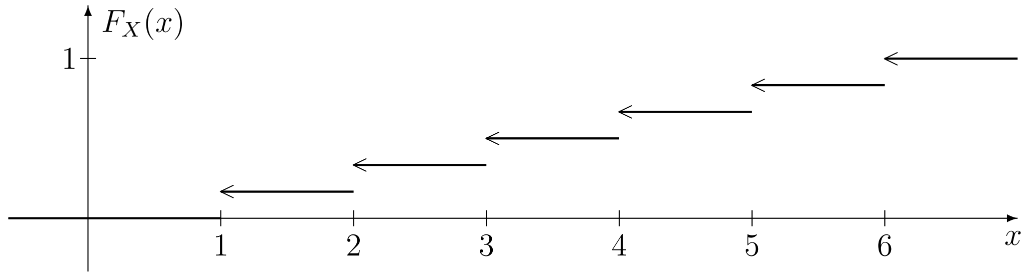
19. $\dfrac{1}{g}\big(0\text{˙}6\cdot 0\text{˙}4\cdot 0\text{˙}9\cdot 0\text{˙}7+0\text{˙}6\cdot 0\text{˙}4\cdot 0\text{˙}1\cdot 1\big)\doteq 0\text{˙}407$, kjer je:

$$g=0\text{˙}4\cdot 0\text{˙}6\cdot 0\text{˙}9\cdot 0\text{˙}1+\big(0\text{˙}6\cdot 0\text{˙}6\cdot 0\text{˙}9+0\text{˙}4\cdot 0\text{˙}4\cdot 0\text{˙}9+0\text{˙}4\cdot 0\text{˙}6\cdot 0\text{˙}1\big)\cdot 0\text{˙}4+$$
$$+\big(0\text{˙}6\cdot 0\text{˙}4\cdot 0\text{˙}9+0\text{˙}6\cdot 0\text{˙}6\cdot 0\text{˙}1+0\text{˙}4\cdot 0\text{˙}4\cdot 0\text{˙}1\big)\cdot 0\text{˙}7+0\text{˙}6\cdot 0\text{˙}4\cdot 0\text{˙}1\cdot 1.$$

4. Slučajne spremenljivke

1. $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$

Porazdelitev je diskretna enakomerna $\mathrm{E}\{1,2,3,4,5,6\}$.



2. $S \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ \frac{1}{24} & \frac{1}{24} & \frac{1}{8} & \frac{1}{8} & \frac{1}{6} & \frac{1}{6} & \frac{1}{8} & \frac{1}{8} & \frac{1}{24} & \frac{1}{24} \end{pmatrix}$

3. Gre za zvezno enakomerno porazdelitev: $X \sim \mathrm{E}(0, 2\pi)$,

$$F_X(x) = \begin{cases} 0 & ; x \le 0 \\ \dfrac{x}{2\pi} & ; 0 \le x \le 2\pi \\ 1 & ; x \ge 2\pi \end{cases}, \qquad p_X(x) = \begin{cases} \dfrac{1}{2\pi} & ; 0 < x < 2\pi \\ 0 & ; \text{sicer} \end{cases}$$

4. $$F_M(t) = \begin{cases} 0 & ; t \le 0 \\ \dfrac{t^2}{100} & ; 0 \le t \le 10 \\ 1 & ; t \ge 10 \end{cases}, \qquad p_M(t) = \begin{cases} \dfrac{t}{50} & ; 0 < t < 10 \\ 0 & ; \text{sicer} \end{cases}$$

5. Gre za binomsko porazdelitev: $X \sim \mathrm{B}(10, 1/6)$, t. j.

$$P(X = k) = \binom{10}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{10-k}.$$

Približno velja:

k	$P(X = k)$
0	0˙162
1	0˙323
2	0˙291
3	0˙155
4	0˙0543
5	0˙0130
6	0˙00217
7	$2˙48 \cdot 10^{-4}$
8	$1˙86 \cdot 10^{-5}$
9	$8˙27 \cdot 10^{-7}$
10	$1˙65 \cdot 10^{-8}$

6. $1-\left(\frac{2}{3}\right)^6-6\cdot\frac{1}{3}\left(\frac{2}{3}\right)^5-\binom{6}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4 \doteq 0\text{'}320.$

7. Poisson: 0'315, Laplace: 0'289.
Točen rezultat: 0'323.

8. Poisson: 0'0888, Laplace: 0'1152.
Točen rezultat: 0'1146.

9. Poisson: 0'12511, točen rezultat: 0'12574.

10. Označimo z X število okvarjenih izdelkov.

a) Laplaceova aproksimacija za $P(X=160)$: 0'033245
Točen rezultat za $P(X=160)$: 0'033228.
Laplaceova aproksimacija za $P(X=175)$: 0'015221.
Točen rezultat za $P(X=175)$: 0'014929.

b) Laplaceova aproksimacija za $P(175<X<\infty)$: 0'10565.
Laplaceova aproksimacija za $P(176\le X<\infty)$: 0'09121.
Laplaceova aproksimacija za $P(175\text{'}5<X<\infty)$: 0'09824
Točen rezultat za $P(X>175)$: 0'09944.
Laplaceova aproksimacija za $P(-\infty<X<150)$: 0'20232.
Laplaceova aproksimacija za $P(-\infty<X\le 149)$: 0'17966.
Laplaceova aproksimacija za $P(-\infty<X<149\text{'}5)$: 0'19079.
Točen rezultat za $P(X<150)$: 0'19147.

c) Označimo z x potrebno velikost skladišča. Po Laplaceovi integralski formuli je x približno najmanjše celo število, za katerega velja:

$$\Phi\left(\frac{x-159\text{'}5}{12}\right)\ge 0\text{'}45\,.$$

Torej približno velja $x=\lceil y\rceil$, kjer je y rešitev enačbe:

$$\Phi\left(\frac{y-159\text{'}5}{12}\right)=0\text{'}45$$

in iz $y\doteq 179.24$ dobimo $x=180$.

Dejansko je verjetnost, da bo pokvarjenih izdelkov (strogo) več kot 179, enaka 0'0539, da jih bo več kot 180, pa 0'0457.

11. Označimo z X število ponesrečenih.

Poissonov obrazec za $P(X=0)$: 0'22313.
Laplaceova lokalna formula za $P(X=0)$: 0'15381,
Točen rezultat za $P(X=0)$: 0'22288.
Poissonov obrazec za $P(X>2)$: 0'19115.
Laplaceova integralska formula za $P(2<X<\infty)$: 0'34143.
Laplaceova integralska formula za $P(3<X<\infty)$: 0'11016.
Laplaceova integralska formula za $P(2\text{'}5<X<\infty)$: 0'20693.
Točen rezultat za $P(X>2)$: 0'19106.

12. Po Laplaceovi integralski formuli (brez uporabe polovičk) dobimo, da moramo naročiti najmanj 13030 izdelkov.
V resnici je najmanjše možno število naročenih izdelkov, ki ustrezajo zahtevi, že 12922. Ne ustreza pa *vsako* število izdelkov, ki je večje ali enako 12922: prvo število, od katerega naprej vsako število naročenih ustreza, je 13096. Nekaj točnih verjetnosti, kjer z n označimo število izdelkov, z S pa število prvovrstnih:

$$\begin{aligned} n = 12921 :& \quad P(S \geq 7624) \doteq 0^{\cdot}9897146436 \\ n = 12922 :& \quad P(S \geq 7624) \doteq 0^{\cdot}9900021378 \\ n = 13095 :& \quad P(S \geq 7727) \doteq 0^{\cdot}9899715692 \\ n = 13096 :& \quad P(S \geq 7727) \doteq 0^{\cdot}9902509306 \end{aligned}$$

13. Po Laplaceovi integralski formuli dobimo, da je treba naročiti najmanj 1159 izdelkov. V resnici je potrebno naročiti vsaj 1161 izdelkov: verjetnost, da je prvovrstnih vsaj 100, pride pri 1160 izdelkih 0˙9493, pri 1161 izdelkih pa 0˙9502.

14. Geometrijska G(1/6).

15. Pascalova (negativna binomska) P(10, 1/6).

16. $P(X = k) = (k-1)\,2^{-k}$ za $k = 2, 3, \ldots$. Porazdelitev je Pascalova P(2, 1/2).
Verjetnostna razlaga: če takoj za prvo cifro vse cifre zamenjamo z grbi, grbe pa s ciframi, se problem iz naloge prevede na čakanje, dokler ne padeta dve cifri.

17. $P(X = k) = 2^{-(k-1)}$ za $k = 2, 3, \ldots$, $X - 1 \sim$ G(1/2).
Verjetnostna razlaga: če prvič pade cifra, pri nadaljnjih metih cifre zamenjamo z grbi, grbe pa s ciframi. Tako se problem prevede na čakanje na cifro (od drugega meta dalje).

18. Slučajna spremenljivka X je porazdeljena hipergeometrijsko H(7, 4, 16) = = H(4, 7, 16). Velja:

$$P(X = k) = \frac{\binom{4}{k}\binom{12}{7-k}}{\binom{16}{7}} = \frac{\binom{7}{k}\binom{9}{4-k}}{\binom{16}{4}}$$

Približno velja $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0^{\cdot}0692 & 0^{\cdot}3231 & 0^{\cdot}4154 & 0^{\cdot}1731 & 0^{\cdot}0192 \end{pmatrix}$.

19. $X \sim H(4, 3, 12) = H(3, 4, 12)$.

$$P(X = k) = \frac{\binom{3}{k}\binom{9}{4-k}}{\binom{12}{4}} = \frac{\binom{4}{k}\binom{8}{3-k}}{\binom{12}{3}}$$

ali približno $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0^{\cdot}2545 & 0^{\cdot}5091 & 0^{\cdot}2182 & 0^{\cdot}0182 \end{pmatrix}$.

20. $c = 1/55$, $P(X > 3) = 49/55$.

21. $c = \lambda$, X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Velja še:

$$F_X(x) = \begin{cases} 0 & ; x \geq 0 \\ 1 - e^{-\lambda x} & ; x \geq 0 \end{cases},$$
$$P(1 < X < 2) = e^{-\lambda} - e^{-2\lambda}.$$

22. $Y \sim \begin{pmatrix} 0 & 1 & 4 \\ 0\text{'}3 & 0\text{'}3 & 0\text{'}4 \end{pmatrix}$

23. $Y \sim \begin{pmatrix} -1 & 0 & 1 \\ \dfrac{(1-p)^2}{(2-p)(2-2p+p^2)} & \dfrac{1-p}{2-p} & \dfrac{1}{(2-p)(2-2p+p^2)} \end{pmatrix}$

24. Slučajna spremenljivka lahko s pozitivno verjetnostjo zavzame vrednosti $1, 2, \ldots b-1$. Za d iz te množice velja:

$$P(D = d) = P(d \leq b^U < d+1) = P(\log_b d \leq U \log_b(d+1)) =$$
$$= \log_b(d+1) - \log_b d = \log_b\left(1 + \frac{1}{d}\right)$$

Tej porazdelitvi pravimo *Benfordova porazdelitev* in predstavlja idealizirano porazdelitev prve števke oz. neničelne decimalke zapisa slučajnega podatka v številskem sistemu z osnovo b. Prva števka oz. neničelna decimalka števila X je namreč enaka $\lfloor b^{\log_b X - \lfloor \log_b X \rfloor} \rfloor$ in za veliko porazdelitev slučajne spremenljivke X je slučajna spremenljivka $\log_b X - \lfloor \log_b X \rfloor$ porazdeljena približno enakomerno $\mathrm{E}(0, 1)$. Porazdelitvena shema za Benfordovo porazdelitev pri $b = 10$:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 0\text{'}3010 & 0\text{'}1761 & 0\text{'}1249 & 0\text{'}0969 & 0\text{'}0792 & 0\text{'}0669 & 0\text{'}0580 & 0\text{'}0512 & 0\text{'}0458 \end{pmatrix}$$

25. Naj bo:

$$X := \begin{cases} -1 & ; U < 0\text{'}1 \\ 0 & ; 0\text{'}1 \leq U < 0\text{'}4 \\ 1 & ; 0\text{'}4 \leq U < 0\text{'}6 \\ 2 & ; U \geq 0\text{'}6 \end{cases} \quad \text{in} \quad Y := -\frac{\ln(1-U)}{\lambda}$$

Tedaj je $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0\text{'}1 & 0\text{'}3 & 0\text{'}2 & 0\text{'}4 \end{pmatrix}$ in $Y \sim \mathrm{Exp}(\lambda)$.

26. $0\text{'}93319$.

27. Normalno $\mathrm{N}(a\mu + b, |a|\sigma)$.

28. $1/2 - \Phi(9/5) \doteq 0\text{'}03593$.

29. $p_X(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi}}\, e^{-(\ln x)^2/2} & ; x > 0 \\ 0 & ; \text{sicer} \end{cases}$, $p_Y(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi y}}\, e^{-y/2} & ; y > 0 \\ 0 & ; \text{sicer} \end{cases}$

Slučajna spremenljivka Y ima porazdelitev $\chi^2(1)$.

30. $$P(R = r, M = m) = \frac{\binom{3}{r}\binom{2}{m}\binom{5}{3-r-m}}{\binom{10}{3}},$$

$$P(R = r) = \frac{\binom{3}{r}\binom{7}{3-r}}{\binom{10}{3}}, \quad P(M = m) = \frac{\binom{2}{m}\binom{8}{3-m}}{\binom{10}{3}},$$

$$P(R + M = k) = \frac{\binom{5}{k}\binom{5}{3-k}}{\binom{10}{3}}.$$

Vse te slučajne spremenljivke so porazdeljene hipergeometrijsko: $R \sim \mathrm{H}(3, 3, 10)$, $M \sim \mathrm{H}(3, 2, 10)$, $R + M \sim \mathrm{H}(3, 5, 10)$.

Slučajni spremenljivki R in M sta odvisni.

31. Porazdelitev slučajnega vektorja (X, Y) z robnima porazdelitvama:

	$Y = 0$	$Y = 1$	$Y = 2$	
$X = -1$	0˙05	0˙1	0˙1	0˙25
$X = 0$	0˙1	0˙2	0˙2	0˙5
$X = 1$	0˙05	0˙1	0˙1	0˙25
	0˙2	0˙4	0˙4	1

Velja še $Y - X \sim \begin{pmatrix} -1 & 0 & 1 & 2 & 3 \\ 0\dot{}05 & 0\dot{}2 & 0\dot{}35 & 0\dot{}3 & 0\dot{}1 \end{pmatrix}$

32. Velja $S \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0\dot{}7^3 & 3 \cdot 0\dot{}7^2 \cdot 0\dot{}3 & 3 \cdot 0\dot{}7 \cdot 0\dot{}3^2 & 0\dot{}3^3 \end{pmatrix} = \mathrm{B}(3, 0\dot{}3)$.

Če so $X_1, \ldots X_n$ neodvisne in imajo vse Bernoullijevo porazdelitev $\mathrm{Be}(p)$, je njihova vsota porazdeljena binomsko $\mathrm{B}(n, p)$.

33. $S \sim \mathrm{P}(\lambda_1 + \lambda_2)$.

34. $S \sim \mathrm{P}(n, p)$.

35. $1 - e^{-2\lambda}\left(1 + 2\lambda + 2\lambda^2 + \dfrac{7\lambda^3}{6}\right)$.

36. $\left(\frac{29}{30}\right)^{30} \cdot \left[1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2}\right)\right] +$
$+ \binom{30}{1} \cdot \frac{1}{30} \cdot \left(\frac{29}{30}\right)^{29} \cdot \left[1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6}\right)\right] +$
$+ \binom{30}{2} \cdot \left(\frac{1}{30}\right)^{2} \cdot \left(\frac{29}{30}\right)^{28} \cdot \left[1 - \frac{1}{e}\left(1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24}\right)\right] \doteq$
$\doteq 0^{\cdot}036831.$
Točen rezultat (brez zanemarjanja): $0^{\cdot}036867$.

37. $c = 1, \quad p_X(x) = \begin{cases} x\,e^{-x} & ; x > 0 \\ 0 & ; \text{sicer} \end{cases}, \quad p_Y(y) = \begin{cases} e^{-y} & ; y > 0 \\ 0 & ; \text{sicer} \end{cases}.$
Slučajni spremenljivki X in Y sta odvisni.
$P(2 < Y < 3) = e^{-2} - e^{-3}, \quad P(Y < X) = 1, \quad P(2Y < X) = 1/2.$

38. $p_Z(z) = \begin{cases} e^{-z/2} - e^{-z} & ; z > 0 \\ 0 & ; \text{sicer} \end{cases}$

39. $c = 2, \quad p_Z(z) = \begin{cases} (1+z)^{-2} & ; z > 0 \\ 0 & ; \text{sicer} \end{cases},$
$P(X < 2Y) = P(Z > 1/2) = 2/3.$

40. $F_Z(z) = \begin{cases} 0 & ; z \le 0 \\ z(1 - \ln z) & ; 0 \le z \le 1 \\ 1 & ; z \ge 1 \end{cases}$

41. $p_Z(z) = \begin{cases} \dfrac{1}{\pi(1+z)\sqrt{z}} & ; z > 0 \\ 0 & ; \text{sicer} \end{cases}$

42. $X \sim \mathrm{N}\big(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\big).$

43. $1/2 - \Phi(1/5) \doteq 0^{\cdot}42074.$

44. $T_n \sim \mathrm{Gama}(n, \lambda); \quad \mathrm{Gama}(1, \lambda) = \mathrm{Exp}(\lambda).$

45. $S \sim \mathrm{Gama}(a_1 + a_2, \lambda)$ (za $a_1, a_2 \in \mathbb{N}$ to sledi iz prejšnje naloge, sicer pa je potrebno izpeljati).

46. Če s p_X in p_Y označimo ustrezni gostoti:

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/(2\sigma^2)}, \quad p_Y(y) = \frac{a^\lambda}{\Gamma(\lambda)}\, y^{\lambda-1}\, e^{-ay}\,;\ y > 0\,,$$

velja:

$$\begin{aligned}
p_T(t) &= \int_0^\infty p_X\big(t\sqrt{y}\big) p_Y(y)\sqrt{y}\,\mathrm{d}y = \\
&= \frac{a^\lambda}{\sigma\sqrt{2\pi}\,\Gamma(\lambda)} \int_0^\infty y^{\lambda-1/2}\, e^{-(a+t^2/(2\sigma^2))}\,\mathrm{d}y = \\
&= \frac{\Gamma\big(\lambda+\frac{1}{2}\big)}{\Gamma(\lambda)} \frac{1}{\sqrt{2\pi a\sigma^2}\left(1+\dfrac{t^2}{2a\sigma^2}\right)^{\lambda+1/2}} = \\
&= \frac{1}{\mathrm{B}\big(\lambda,\frac{1}{2}\big)\sqrt{2a\sigma^2}\left(1+\dfrac{t^2}{2a\sigma^2}\right)^{\lambda+1/2}}\,.
\end{aligned}$$

Za $a\sigma^2 = \lambda$ dobimo *Studentovo porazdelitev* z 2λ prostostnimi stopnjami, ki je pomembna v statistiki (porazdelitev Gama$(n/2, 1/2)$ se tudi ujema s porazdelitvijo $\chi^2(n)$).

47. Če ustrezni gostoti označimo s p_1 in p_2, za $y > 0$ velja:

$$\begin{aligned}
p_Y(y) &= \int_0^\infty p_1(xy)p_2(x)\,x\,\mathrm{d}x = \\
&= \frac{a_1{}^{\lambda_1} a_2^{\lambda_2}}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)} y^{\lambda_1-1} \int_0^\infty x^{\lambda_1+\lambda_2-1} e^{-(a_1 y+a_2)x}\,\mathrm{d}x = \\
&= \frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\,\Gamma(\lambda_2)} \frac{a_1^{\lambda_1} a_2^{\lambda_2}}{(a_1 y + a_2)^{\lambda_1+\lambda_2}}\, y^{\lambda_1-1} = \\
&= \frac{1}{\mathrm{B}(\lambda_1,\lambda_2)\left(1+\dfrac{a_2}{a_1 y}\right)^{\lambda_1}\left(1+\dfrac{a_1 y}{a_2}\right)^{\lambda_2} y}\,.
\end{aligned}$$

Torej je:

$$p_Y(y) = \begin{cases} \Big[\mathrm{B}(\lambda_1,\lambda_2)\big(1+\frac{a_2}{a_1 y}\big)^{\lambda_1}\big(1+\frac{a_1 y}{a_2}\big)^{\lambda_2}\, y\Big]^{-1} & ;\ y > 0 \\ 0 & ;\ \text{sicer} \end{cases}\,.$$

Za $\lambda_1 = a_1 = m/2$ in $\lambda_2 = a_2 = n/2$ dobimo *Snedecorjevo* porazdelitev $\mathrm{F}(m,n)$, ki je prav tako pomembna v statistiki.

48. Pogojno na $Y = 0$ je $X \sim \begin{pmatrix} -1 & 0 \\ 1/4 & 3/4 \end{pmatrix}$.

Pogojno na $X = -1$ je $Y \sim \begin{pmatrix} 0 & 2 \\ 1/3 & 2/3 \end{pmatrix}$.

49. $\mathrm{B}\left(Z, \dfrac{\lambda}{\lambda+\mu}\right)$.

50. Za $y > 0$ je $p_{X|Y}(x \mid y) = \begin{cases} e^{-(x-y)} & ; x > y \\ 0 & ; \text{sicer} \end{cases}$.

Za $x > 0$ je $p_{Y|X}(y \mid x) = \begin{cases} 1/x & ; 0 < y < x \\ 0 & ; \text{sicer} \end{cases}$.

Pogojno glede na Y ima torej X eksponentno porazdelitev $\mathrm{Exp}(1)$, pomaknjeno za Y v desno.

Pogojno glede na X pa ima Y eksponentno porazdelitev $\mathrm{E}(0, X)$.

5. Številske karakteristike

1. $E(X) = 2, \quad D(X) = 3{\cdot}8, \quad \sigma(X) \doteq 1{\cdot}949.$

2. $E(X) = \dfrac{1}{\lambda}, \quad D(X) = \dfrac{1}{\lambda^2}, \quad E(e^{-X}) = \dfrac{\lambda}{\lambda+1}.$

3. $E(1+X^2) = 2, \quad D(X) = 1.$

5. $1/2.$

6. $c = \dfrac{2}{\pi}, \quad E(X) = E(Y) = 0, \quad E(X^2) = E(Y^2) = \dfrac{1}{2}, \quad E(X^2+Y^2) = 1.$

7. Če je n lih, je $E(Z^n) = 0$.
Če je n sod, je $E(Z^n) = 1 \cdot 3 \cdot \cdots \cdot (n-1)$.

8. Če je $S \sim \mathrm{B}(n,p)$, je $E(S) = np$ in $D(S) = np(1-p)$.

9. Če je $S \sim \mathrm{Gama}(n,\lambda)$, je $E(S) = n/\lambda$ in $D(S) = n/\lambda^2$.

10. $D(3X-Y) = 64$, $E\big((X+2Y)^2\big) = 215$.

11. Za $i = 1, \ldots 8$ definirajmo slučajne spremenljivke X_i, kjer naj bo X_i enaka 1, če i-ti igralec stavi svojo ženo, sicer pa 0. Tedaj je očitno:

$$E(X_i) = p_0 := P(i\text{-ti igralec stavi svojo ženo}) = \frac{4}{52} \cdot \frac{48}{51} \cdot \frac{47}{50}$$

Ker je $S = X_1 + \cdots + X_8$, je tudi $E(S) = 8p_0 \doteq 0{\cdot}5444$. Izračunajmo še disperzijo: $D(S) = E(S^2) - \big(E(S)\big)^2$. Velja:

$$\begin{aligned} E(S^2) &= \sum_{i=1}^{8}\sum_{j=1}^{8} E(X_i X_j) = \\ &= \sum_{i=1}^{8}\sum_{j=1}^{8} P\big(i\text{-ti in } j\text{-ti igralec oba stavita svojo ženo}\big) = \\ &= 8(p_0 + 2p_1 + 2p_2 + 2p_3 + p_4) \end{aligned}$$

kjer je p_k verjetnost, da stavita svoji ženi posamezna igralca, ki sta oddaljena za k (t. j. med njima je $k-1$ sedežev). Verjetnost p_0 smo že izračunali, velja pa še:

$$\begin{aligned} p_1 &= 0 \\ p_2 &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{48}{50} \cdot \frac{47}{49} \cdot \frac{46}{48} \\ p_3 = p_4 &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{48}{50} \cdot \frac{47}{49} \cdot \frac{46}{48} \cdot \frac{45}{47} \end{aligned}$$

Torej je:

$$D(S) = 8(p_0 + 2p_2 + 3p_3) - (8p_0)^2 \doteq 0{\cdot}4037\,.$$

12. Če je $X \sim \mathrm{G}(p)$, je $E(X) = 1/p$ in $D(X) = q/p^2$.
Če je $S \sim \mathrm{P}(n,p)$, je $E(X) = n/p$ in $D(X) = nq/p^2$.

13. Če je $S \sim \mathrm{H}(s;r,n)$, je $E(S) = rs/n$.

14. $E(XY^2) = 3\text{˙}55$, $\quad E(Y \mid X = 4) = 5/7 \doteq 0\text{˙}7143$, $\quad E(Y^2 \mid X = 4) = 11/7 \doteq 1\text{˙}571$, $E(XY^2 \mid X = 4) = 4\,E(Y^2 \mid X = 4) = 44/7 \doteq 6\text{˙}286$.

15. $p_X(x) = \frac{3}{2}x^2 + x$,
$E(Y \mid X = x) = \dfrac{4x+3}{6x+4}$,
$E(XY \mid X = x) = x\,E(Y \mid X = x) = \dfrac{4x^2+3x}{6x+4}$.

16. $K(X,Y) = -0\text{˙}93$, $\quad r(X,Y) \doteq -0\text{˙}3952$.

17. Slučajni spremenljivki X in Y sta odvisni, čeprav sta nekorelirani. To lahko vidimo iz poljubne navzkrižne verjetnosti.

18. $K(X,Y) = -1/576 \doteq -0\text{˙}001736$, $\quad r(X,Y) = -5/139 \doteq -0\text{˙}03597$.

19. $K(U,V) = a - a^2$, U in V sta nekorelirani pri $a = 0$ in $a = 1$.

20. $A(X) = 2$, $K(X) = 6$.

21. Iz eksponentne vrste dobimo:

$$\begin{aligned}
E(X) &= \lambda \\
E\big(X(X-1)\big) &= E(X^2 - X) = \lambda^2 \\
E\big(X(X-1)(X-2)\big) &= E(X^3 - 3X^2 + 2X) = \lambda^3 \\
E\big(X(X-1)(X-2)(X-3)\big) &= E(X^4 - 6X^3 + 11X^2 - 6X) = \lambda^4
\end{aligned}$$

od koder najprej izračunamo:

$$\begin{aligned}
E(X^2) &= \lambda^2 + \lambda \\
E(X^3) &= \lambda^3 + 3\lambda^2 + \lambda \\
E(X^4) &= \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda
\end{aligned}$$

in nato še:

$$\begin{aligned}
E\big((X-\lambda)^2\big) &= \lambda \\
E\big((X-\lambda)^3\big) &= \lambda \\
E\big((X-\lambda)^4\big) &= 3\lambda^2 + \lambda
\end{aligned}$$

Sledi $A(X) = \dfrac{1}{\sqrt{\lambda}}$ in $K(X) = \dfrac{1}{\lambda}$.

22. $m = 1$, $s = 4/3$.

6. Rodovne in karakteristične funkcije

1. $G_X(s) = e^{\lambda(s-1)}$, $\quad E(X) = D(X) = \lambda$.

2. Neposredno ali z uporabo dejstva, da se binomska porazdelitev ujema s porazdelitvijo vsote neodvisnih Bernoullijevih slučajnih spremenljivk, dobimo $G_X(s) = (1 - p + ps)^n$. Nadalje iz zveze:

$$\frac{1}{1+X} = \int_0^1 s^X \, \mathrm{d}s$$

dobimo:

$$E\left(\frac{1}{1+X}\right) = \int_0^1 G_X(s) \, \mathrm{d}s = \frac{1-(1-p)^{n+1}}{(n+1)p} \, .$$

3. Rodovna funkcija:

$$G(s) = \left(\frac{1}{2} + \frac{1}{4}s + \frac{1}{4}s^2\right)^2 = \frac{1}{4} + \frac{1}{4}s + \frac{5}{16}s^2 + \frac{1}{8}s^3 + \frac{1}{16}s^4 \, ,$$

porazdelitev:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1/4 & 1/4 & 5/16 & 1/8 & 1/16 \end{pmatrix} .$$

4. Rodovno funkcijo števila Nikinih otrok označimo z:

$$G_1(s) = \frac{1}{4} + \frac{1}{4}s + \frac{1}{4}s^2 + \frac{1}{4}s^3 \, ,$$

rodovno funkcijo števila otrok posameznega Nikinega otroka pa z:

$$G_2(s) = \frac{1}{2} + \frac{1}{4}s + \frac{1}{4}s^2 \, ,$$

Tedaj je rodovna funkcija števila Nikinih vnukov enaka:

$$G(s) = G_1\big(G_2(s)\big) \, ,$$

verjetnost, da Nika ostane brez vnukov, je $G_1(G_2(0)) = 15/32$, matematično upanje pa je:

$$G'(1) = G_1'\big(G_2(1)\big) G_2'(1) = \frac{9}{8} \, .$$

5. Označimo z $G(s) = \frac{1}{6} + \frac{1}{3}s + \frac{1}{2}s^2$ rodovno funkcijo števila otrok. Verjetnost, da Maks v n-ti generaciji ne bo imel potomcev, je enaka $G_n(0)$, kjer je:

$$G_n(s) = G(G(G(\ldots G(s) \ldots))) \qquad (n \text{ znakov } G).$$

Verjetnost, da bo Maksovo potomstvo izumrlo, pa je enaka $p = \lim_{n\to\infty} G_n(0)$. Velja $G(p) = p$, kar ima rešitvi $p = 1/3$ in $p = 1$. Z indukcijo po n dokažemo, da je $G_n(0) < 1/3$ za vse n, torej je $p = 1/3$.

6. $\phi_X(t) = \dfrac{p\,e^{it}}{1-(1-p)e^{it}}$.

7. $\phi_X(t) = \dfrac{\lambda}{\lambda - it}$.

8. $\phi_X(t) = \dfrac{1}{1+t^2}$.

9. Če je $X \sim \mathrm{B}(n,p)$, je $\phi_X(t) = \big(1-p+p\,e^{it}\big)^n$.

10. Če je $X \sim \mathrm{P}(n,p)$, je $\phi_X(t) = \left(\dfrac{p\,e^{it}}{1-(1-p)e^{it}}\right)^n$.

11. Če je $X \sim \mathrm{Gama}(n,\lambda)$, je $\phi_X(t) = \left(\dfrac{\lambda}{\lambda - it}\right)^n$.

12. $p_X(x) = \begin{cases} \dfrac{1-\cos x}{\pi x^2} & ;\ |x| \neq 0 \\ \dfrac{1}{2\pi} & ;\ x = 0 \end{cases}$

13. Slučajna spremenljivka X ima *Cauchyjevo porazdelitev*, t. j.:

$$p_X(x) = \frac{1}{\pi(1+x^2)}$$

14. $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1/6 & 1/3 & 1/6 & 1/3 \end{pmatrix}$.

15. Slučajna spremenljivka $\bar{X}$ ima prav tako Cauchyjevo porazdelitev.

7. Limitni izreki

1. Po krepkem zakonu velikih števil Kolmogorova zaporedje konvergira tako v verjetnosti kot tudi skoraj gotovo.

2. Za $n \in \mathbb{N}$ označimo s p_n verjetnost, da slučajni sprehod še kdaj pride v izhodišče iz točke n oziroma $-n$ (zaradi simetrije sta verjetnosti očitno enaki). Pokazati moramo, da je $p_1 = 1$.

Če označimo še $p_0 = 1$, iz dinamike slučajnega sprehoda razberemo, da mora za vsak $n \in \mathbb{N}$ veljati $p_n = (p_{n-1} + p_{n+1})/2$ oziroma $p_{n+1} = 2p_n - p_{n-1}$. Tako so vse verjetnosti p_n enolično določene s p_1. Če označimo $\delta := 1 - p_1$, ni težko videti, da za vsak $n \in \mathbb{N}$ velja $p_n = 1 - n\delta$. Toda ker so p_n verjetnosti in zato $0 \le p_n \le 1$, ni druge možnosti, kot da je $\delta = 0$. Torej je $p_1 = 1$ (in $p_n = 1$ za vse n), to pa je bilo potrebno dokazati.

3. Označimo iskano verjetnost s π_k (za π_0 pa se dogovorimo, da je to verjetnost, da se slučajni sprehod *še* kdaj vrne v izhodišče). Iz neodvisnosti in enake porazdeljenosti sledijo naslednje rekurzivne zveze:

$$\pi_k = p\pi_{k-1} + (1-p)\pi_{k+1}\,; \quad |k| > 1 \tag{1}$$
$$\pi_1 = p + (1-p)\pi_2 \tag{2}$$
$$\pi_{-1} = p\pi_{-2} + (1-p) \tag{3}$$
$$\pi_0 = p\pi_{-1} + (1-p)\pi_1\,. \tag{4}$$

Enačba (1) je diferenčna enačba, katere rešitve so linearne kombinacije členov $k^r\lambda^k$, kjer je λ rešitev karakteristične enačbe:

$$(1-p)\lambda^2 - \lambda + p = 0\,, \tag{5}$$

ki ima rešitvi $\lambda_1 = 1$ in $\lambda_2 = p/(1-p)$.

Oglejmo si najprej primer, ko je $p < 1/2$. V tem primeru sta rešitvi različni. Ker se zveza (1) pri izhodišču prekine, rešitev nastavimo v obliki:

$$\pi_k = C_1^+ + C_2^+\left(\frac{p}{1-p}\right)^k \quad ; \quad k \ge 1$$
$$\pi_k = C_1^- + C_2^-\left(\frac{p}{1-p}\right)^k \quad ; \quad k \le -1\,.$$

Najprej za negativne k opazimo, da mora biti $C_2^- = 0$, sicer verjetnosti uidejo izven intervala $[0,1]$. Torej je $\pi_k = C_1^-$ za vse negativne k. Ko to vstavimo v enačbo (3), dobimo, da je $C_1^- = 1$.

Za pozitivne k pa si pomagamo z dejstvom, da je $\lim_{k\to\infty} \pi_k = 0$, kar je intuitivno jasno in bomo tudi eksaktno dokazali malo kasneje. Iz tega dejstva dobimo, da mora biti $C_1^+ = 0$. Ko vse skupaj vstavimo v enačbo (2) in poračunamo, dobimo, da mora veljati $C_2^+ = 1$. Iz enačbe (4) zdaj dobimo še $\pi_0 = 2p$. Sledi:

$$\pi_k = \begin{cases} \left(\frac{p}{1-p}\right)^k & ; k > 1 \\ 2p & ; k = 0 \\ 1 & ; k < 0 \end{cases} . \tag{$*$}$$

Preostane nam le še dokazati, da je $\pi := \lim_{k\to\infty} \pi_k = 0$. Za ta namen definirajmo dogodke:

$$A_k = \{\text{slučajni sprehod obišče stanje } k\} .$$

Najprej opazimo, da je $A_1 \supseteq A_2 \supseteq A_3 \supseteq \ldots$. Torej je π verjetnost njihovega preseka, to pa je dogodek, da slučajni sprehod obišče vsa pozitivna stanja. Pokazati je torej treba, da ima ta dogodek verjetnost nič.

Po krepkem zakonu velikih števil Kolmogorova z verjetnostjo ena velja:

$$\lim_{n\to\infty} \frac{S_n}{n} = 2p - 1 < 0 .$$

Torej so z verjetnostjo ena vse vrednosti $S_1, S_2, S_3, \ldots$ od nekod naprej negativne ali nič, se pravi, da z verjetnostjo nič velja, da je neskončno mnogo vrednosti $S_1, S_2, S_3, \ldots$ strogo pozitivnih. Dogodek, da slučajni sprehod obišče vsa pozitivna stanja, je način tega dogodka, torej ima tudi sam verjetnost nič. Zveza $(*)$ je tako dokazana.

Za $p > 1/2$ je obnašanje zrcalno simetrično okrog $1/2$ – velja torej:

$$\pi_k = \begin{cases} 1 & ; k > 1 \\ 2(1-p) & ; k = 0 \\ \left(\frac{p}{1-p}\right)^k & ; k < 0 \end{cases} .$$

Končno si oglejmo še primer, ko je $p = 1/2$. Le-tega smo sicer že obravnavali v 2. nalogi, a ga lahko vseeno pogledamo v luči karakteristične enačbe (5). Le-ta ima sedaj dvojno ničlo $\lambda = 1$, torej bo imela rešitev rekurzivnih enačb obliko:

$$\begin{aligned} \pi_k &= C_1^+ + C_2^+\, k ; \quad k \geq 1 \\ \pi_k &= C_1^- + C_2^-\, k ; \quad k \leq -1 . \end{aligned}$$

Veljati mora $C_2^+ = C_2^- = 0$, sicer verjetnosti π_k spet uidejo izven intervala $[0, 1]$. Iz zvez (2) in (3) dobimo $C_1^+ = C_1^- = 1$. Še iz zveze (4) dobimo, da mora biti $\pi_k = 1$ za vse k.

4. Slučajna spremenljivka T_n se razlikuje od S_n le v primeru, ko je $S_n = 0$. Toda po Laplaceovi lokalni formuli je $P(S_n = 0) \sim \sqrt{2/n\pi}$, če je n sod, in $P(S_n = 0) = 0$, če je n lih. Torej je $\lim_{n\to\infty} P(S_n = 0) = 0$. To pa pomeni, da se mora zaporedje

T_n glede konvergence v verjetnosti obnašati enako kot S_n/n. Natančneje, za vsak $\varepsilon > 0$ velja:

$$P(|T_n| < \varepsilon) = P\left(\left|\frac{S_n}{n}\right| < \varepsilon, S_n \neq 0\right) \geq P\left(\left|\frac{S_n}{n}\right| < \varepsilon\right) - P(S_n = 0)$$

od koder sledi $\lim_{n\to\infty} P(|T_n| < \varepsilon) = 1$.

Po drugi strani pa smo v prejšnji nalogi videli, da se standardni slučajni sprehod skoraj gotovo vrne v izhodišče. To pomeni, da se skoraj gotovo tudi *neskončno mnogokrat* vrne v izhodišče, torej je skoraj gotovo neskončno mnogo slučajnih spremenljivk T_n enakih 2. Toda če je $T_n = 2$, je $S_n = 0$ torej $|S_{n+1}| = 1$ in zato $|T_{n+1}| = 1/(n+1)$. Zaporedje, v katerem se neskončno mnogokrat ponovi ta vzorec, pa ne more nikamor konvergirati.

5. Rezultat, dobljen s pomočjo centralnega limitnega izreka:
$\Phi\left(\frac{9˙5}{\sqrt{120}}\right) + \Phi\left(\frac{29˙5}{\sqrt{120}}\right) \doteq 0˙80355$.
Točen rezultat: 0˙80493.

6. Ker ima X enako porazdelitev kot vsota 100 neodvisnih slučajnih spremenljivk, porazdeljenih po hi kvadrat z eno prostostno stopnjo, so izpolnjeni pogoji centralnega limitnega izreka in lahko porazdelitev aproksimiramo z $\mathrm{N}\big(100, \sqrt{200}\big)$.
$P(X > 110)$: CLI: 0˙23975, točen rezultat: 0˙23220.
$P(90 < X < 110)$: CLI: 0˙52050, točen rezultat: 0˙52099.

7. a) 400

b) *Prvi način.* Dogodek, da se bo moral potopiti več kot 450-krat, lahko zapišemo kot dogodek, da bo imel po 450 potopih manj kot 80 biserov. Tako, če binomsko porazdelitev $\mathrm{B}(450, 0˙2)$ aproksimiramo z normalno $\mathrm{N}\big(90, \sqrt{72}\big)$, dobimo približen rezultat 0˙10796.

Drugi način. Število potopov je porazdeljeno po Pascalovi porazdelitvi $\mathrm{P}(80, 0˙2)$, ki jo, ker jo dobimo iz vsote 80 neodvisnih slučajnih spremenljivk, lahko aproksimiramo z normalno $\mathrm{N}(400, 40)$. Tako dobimo približen rezultat 0˙10338.

Točen rezultat: 0˙10669.

8. Najprej izračunamo, da je $E(X_n) = 0$ in $D(X_n) = 1$ za vse n. Da njihove delne vsote, ki imajo matematična upanja 0 in disperzije n, izpolnjujejo pogoje centralnega limitnega izreka, pomeni, da slučajne spremenljivke:

$$\frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}$$

šibko konvergirajo proti standardni normalni porazdelitvi, t. j. da za poljubna $a, b \in \mathbb{R}$ velja:

$$\lim_{n\to\infty} P\left(a \leq \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a) \qquad (*)$$

brž ko je $a \leq b$. Toda:

$$P\left(\frac{X_2 + X_3 + \ldots + X_n}{\sqrt{n}} = 0\right) \geq P(X_2 = \cdots = X_n = 0) =$$
$$= \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\ldots\left(1 - \frac{1}{n^2}\right) = \frac{n+1}{2n} \geq \frac{1}{2}$$

in zato:

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}}\right| \leq 1\right) \geq \frac{1}{2}$$

Če zdaj v $(*)$ vzamemo $a = -1$ in $b = 1$, bi moralo veljati:

$$\lim_{n\to\infty} P\left(-1 \leq \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \leq 1\right) = 2\Phi(1) \doteq 0\dot{}317 < \frac{1}{2}\,,$$

kar je v nasprotju s prej dobljenim.

8. Zadostne in postranske statistike

2. a) dokažemo s popolno indukcijo. Dokazati moramo, da za vsak k velja $P_\theta(X_k = 1) = \theta$, in to smo že privzeli za $k = 1$. Indukcijski korak s k na $k+1$ izpeljemo tako, da najprej opazimo, da je tudi:

$$P_\theta(X_{k+1} = x_{k+1} \mid X_k = x_k) = p_{x_k x_{k+1}}$$

(sledi iz izreka o popolni verjetnosti). Sledi:

$$\begin{aligned} P_\theta(X_{k+1} = 1) &= P_\theta(X_k = 0)\, P_\theta(X_{k+1} = 1 \mid X_k = 0) + \\ &\qquad + P_\theta(X_k = 1)\, P_\theta(X_{k+1} = 1 \mid X_k = 1) = \\ &= (1-\theta)\,\frac{\theta}{2} + \theta\frac{1+\theta}{2} = \theta\,. \end{aligned}$$

b) Označimo z S število uspešnih poskusov in pri $n = 3$ izračunajmo pogojno porazdelitev našega opažanja (X_1, X_2, X_3) glede na $S = 1$. Velja:

$$\begin{aligned} P_\theta(X_1 = 0, X_2 = 0, X_3 = 1) &= \frac{\theta(1-\theta)(2-\theta)}{4} \\ P_\theta(X_1 = 0, X_2 = 1, X_3 = 0) &= \frac{\theta(1-\theta)^2}{4} \\ P_\theta(X_1 = 1, X_2 = 0, X_3 = 0) &= \frac{\theta(1-\theta)(2-\theta)}{4}\,. \end{aligned}$$

Ko seštejemo, dobimo $P_\theta(S = 1) = \theta(1-\theta)(5-3\theta)/4$, torej je pogojna porazdelitev enaka:

$$\begin{pmatrix} (0,0,1) & (0,1,0) & (1,0,0) \\ \frac{2-\theta}{5-3\theta} & \frac{1-\theta}{5-3\theta} & \frac{2-\theta}{5-3\theta} \end{pmatrix},$$

kar je odvisno od θ, zato S ni zadostna.

c) Ustrezna zadostna statistika je npr. $(X_1, N_{00}, N_{01}, N_{10}, N_{11})$, kjer je N_{ij} število pojavljanj sosledij 00, 01, 10, 11 v našem vzorcu. Če so $n_{00}, n_{01}, n_{10}, n_{11}$ ustrezna števila za zaporedje $(x_1, x_2, \ldots x_n)$, namreč velja:

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = p_{x_1} p_{00}^{n_{00}} p_{01}^{n_{01}} p_{10}^{n_{10}} p_{11}^{n_{11}}\,,$$

kjer je $p_0 = 1-\theta$ in $p_1 = \theta$. Pogojna porazdelitev našega vzorca $(X_1, X_2, \ldots X_n)$ glede na $(X_1 = x_1, N_{00} = n_{00}, N_{01} = n_{01}, N_{10} = n_{10}, N_{11} = n_{11})$ je torej enakomerna na množici vseh zaporedij $(x_1, x_2, \ldots x_n)$, kjer je x_1 že določen, preostale komponente pa morajo imeti predpisano število ustreznih sosledij.

3. V diskretnem primeru zadostnost pomeni, da so pogojne verjetnosti $P_\theta(X = x \mid T = \tau)$ neodvisne od θ, torej da obstaja taka funkcija q, da je

$$P_\theta(X = x, T = \tau) = P_\theta(T = \tau) q(x, \tau)$$

za vse x in τ. Toda opisani dogodek se lahko zgodi le, če je $\tau = t(x)$. Torej je T zadostna natanko tedaj, ko obstaja taka funkcija f, da je:

$$P_\theta(X = x) = P_\theta(T = t(x))\, f(x)$$

(velja $f(x) = q(x, t(x))$ oziroma $q(x, \tau) = f(x)$, če je $\tau = t(x)$, sicer pa $q(x, \tau) = 0$). Brž ko je statistika zadostna, lahko torej definiramo $g(\tau, \theta) := P_\theta(T = \tau)$ in imamo zahtevano izražavo verjetnosti. Dokažimo še obratno: če velja $(*)$, je potrebno le še izračunati:

$$P_\theta(T = \tau) = \sum_{x;t(x)=\tau} P_\theta(X = x) = \sum_{x;t(x)=\tau} f(x)\, g(\tau, \theta) = g(\tau, \theta) \sum_{x;t(x)=\tau} f(x)\,.$$

Torej je $g(\tau, \theta) = P_\theta(T = \tau) / \sum_{x;t(x)=\tau} f(x)$ in zato:

$$P_\theta(X = x) = P_\theta(T = t(x)) \frac{f(x)}{\sum_{y;t(y)=t(x)} f(y)}\,.$$

4. Iz:

$$P_\lambda(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = \frac{\lambda^{x_1+x_2+\cdots+x_n} e^{-n\lambda}}{x_1!\, x_2! \cdots x_n!}$$

razberemo, da je statistika $X_1 + X_2 + \cdots + X_n$ zadostna.

5. Gre za eksponentno družino, saj lahko verjetnostno funkcijo zapišemo v obliki:

$$P_\lambda(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = \frac{e^{-n\lambda}}{x_1!\, x_2! \cdots x_n!}\, e^{(x_1+x_2+\cdots+x_n)\ln\lambda}\,,$$

funkciji $\lambda \mapsto 1$ in $\lambda \mapsto \ln\lambda$ pa sta zagotovo linearno neodvisni kot funkciji iz $(0, \infty)$ v $\mathbb{R}$.

6. Vzorec ima navzkrižno gostoto:

$$L := p_{X_1,\ldots X_n}(x_1, \ldots x_n; \mu, \sigma) =$$
$$= \frac{1}{\sigma^n (2\pi)^{n/2}} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i\right)\,,$$

torej gre za eksponentno družino.

a) Vzorčna vsota $X_1 + X_2 + \cdots + X_n$ je minimalna zadostna statistika.

b) Minimalna zadostna statistika je $\sum_{i=1}^n X_i^2 - 2\mu \sum_{i=1}^n X_i$ ali pa tudi $\sum_{i=1}^n (X_i - \mu)^2$.

c) Minimalna zadostna statistika je par $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$.

d) Iz zapisa:

$$L = \frac{e^{-na/2}}{(2a\pi)^{n/2}} \exp\left(\sum_{i=1}^n X_i\right) \exp\left(\frac{1}{2a} \sum_{i=1}^n x_i^2\right)$$

sledi, da je $\sum_{i=1}^n X_i^2$ minimalna zadostna statistika.

e) V tem primeru pa je:

$$L = \frac{e^{-n/2}}{a^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2a^2} \sum_{i=1}^n x_i^2 + \frac{1}{a} \sum_{i=1}^n x_i\right)$$

in ker so $a \mapsto 1$, $a \mapsto 1/a$ in $a \mapsto -1/(2a^2)$ linearno neodvisne funkcije, je minimalna zadostna statistika spet par $\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$, tako kot če bi bila oba parametra neznana.

7. Pri porazdelitvi Gama(λ, a) je navzkrižna gostota za vzorec enaka:

$$\begin{aligned} p(x_1, \ldots x_n; \lambda, a) &= \frac{a^{n\lambda}}{\big(\Gamma(\lambda)\big)^n}\, x_1^{\lambda-1} e^{-ax_1}\, x_2^{\lambda-1} e^{-ax_2} \cdots x_n^{\lambda-1} e^{-ax_n} = \\ &= \frac{a^{n\lambda}}{\big(\Gamma(\lambda)\big)^n}\, e^{(\lambda-1)(\ln x_1 + \ln x_2 + \cdots + \ln x_n) - a(x_1 + x_2 + \ldots + x_n)}\,, \end{aligned}$$

torej gre za eksponentno družino in minimalna zadostna statistika je recimo:

$$\big(\ln X_1 + \ln X_2 + \cdots + \ln X_n,\ X_1 + X_2 + \cdots + X_n\big)$$

ali pa tudi:

$$\big(X_1 X_2 \cdots X_n,\ X_1 + X_2 + \cdots + X_n\big)\,.$$

8. To je recimo $|X_1 + X_2 + \cdots + X_n|$.

9. Pišemo lahko $X_i = \mu + \sigma Z_i$, kjer so $Z_1, \ldots Z_n$ neodvisne in porazdeljene standardno normalno (torej poznamo porazdelitev slučajnega vektorja $(Z_1, Z_2, \ldots Z_n)$).
a) Vzemimo:

$$\big(X_1 - \bar{X}, X_2 - \bar{X}, \ldots X_n - \bar{X}\big) = \sigma\big(Z_1 - \bar{Z}, Z_2 - \bar{Z}, \ldots Z_n - \bar{Z}\big)\,,$$

kjer je $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$ vzorčno povprečje, na enak način pa je definiran tudi $\bar{Z}$. Iz zapisa z Z-ji se vidi, da je to res postranska statistika, saj (kot funkcija) ni odvisna od μ.

Dimenzijo zaloge vrednosti preslikave, ki vektor $(x_1, x_2, \ldots x_n)$ preslika v $(x_1 - \bar{x}, x_2 - \bar{x}, \ldots x_n - \bar{x})$, kjer je $\bar{x}$ definiran na enak način kot $\bar{X}$, lahko izračunamo tako, da izračunamo rang njene matrike:

$$\begin{bmatrix} \frac{n-1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & \frac{n-1}{n} \end{bmatrix},$$

ki je enak $n-1$, ali pa dokažemo, da je ta preslikava projektor na prostor vseh vektorjev $(x_1, x_2, \ldots x_n)$, za katere je $\bar{x} = 0$.
b) Vzemimo:

$$\begin{aligned} &\left(\frac{X_1 - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}, \frac{X_2 - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}, \ldots \frac{X_n - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}\right) = \\ &= \left(\frac{Z_1}{\sqrt{\sum_{i=1}^n Z_i^2}}, \frac{Z_2}{\sqrt{\sum_{i=1}^n Z_i^2}} \cdots \frac{Z_n}{\sqrt{\sum_{i=1}^n Z_i^2}}\right). \end{aligned}$$

Spet vidimo, da je porazdelitev tega slučajnega vektorja neodvisna od σ in je zato statistika postranska. Njena zaloga vrednosti je enotska sfera v $\mathbb{R}^n$, torej $(n-1)$-dimenzionalen prostor. Še več: iz radialne simetrije gostote slučajnega vektorja

$(Z_1, \ldots Z_n)$ se vidi, da je naša statistika porazdeljena enakomerno na enotski sferi.
c) Vzemimo:

$$\left(\frac{X_1 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \frac{X_2 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \ldots \frac{X_n - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}\right) =$$
$$= \left(\frac{Z_1 - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}, \frac{Z_2 - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}} \ldots \frac{Z_n - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}\right).$$

Zaloga vrednosti tega slučajnega vektorja je presek enotske sfere in množice vseh vektorjev x z $\bar{x} = 0$, kar je $(n-2)$-dimenzionalen prostor. Da se dokazati, da je porazdelitev tudi tega slučajnega vektorja enakomerna na tej množici.

10. Če so $X_1, \ldots X_n$ neodvisne in porazdeljene normalno $\mathrm{N}(\mu, \sigma)$,

a) sta $\left(\frac{X_1 - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}, \ldots \frac{X_n - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}\right)$ in $\sum_{i=1}^n (X_i - \mu)^2$ neodvisna;

b) so $\left(\frac{X_1 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \ldots \frac{X_n - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}\right)$, $\sum_{i=1}^n (X_i - \bar{X})^2$ in $\bar{X}$ neodvisni.

11. Velja:

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 .$$

Iz 10. naloge vemo, da sta statistiki $\sum_{i=1}^n (X_i - \bar{X})^2$ in $n(\bar{X} - \mu)^2$ neodvisni. Poleg tega iz 29. in 45. naloge iz 4. razdelka sledi, da je:

$$\sum_{i=1}^n (X_i - \mu)^2 \sim \text{Gama}\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right) \quad \text{in} \quad n(\bar{X} - \mu)^2 \sim \text{Gama}\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$$

ali, ekvivalentno,

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n) \quad \text{in} \quad \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi^2(1) .$$

Vemo, da, če sta $U \sim \chi^2(n-1)$ in $V \sim \chi^2(1)$ neodvisni, mora veljati $U + V \sim \chi^2(n)$. Iz teorije momentno-rodovnih ali karakterističnih funkcij pa sledi, da velja tudi obratno: če sta U in V neodvisni ter je $V \sim \chi^2(1)$ in $U + V \sim \chi^2(n)$, mora biti $U \sim \chi^2(n-1)$. Sledi:

$$\sum_{i=1}^n (X_i - \bar{X})^2 \sim \text{Gama}\left(\frac{n-1}{2}, \frac{1}{2\sigma^2}\right) \quad \text{oziroma} \quad \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1) .$$

12. Računajmo:

$$U = \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}} = \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2}} = \frac{T}{\sqrt{n(n-1) + nT^2}} ,$$

kar nam da zahtevano bijektivno korespondenco na dogodku, da niso vse slučajne spremenljivke X_i enake, le-ta pa ima verjetnost ena. Za izračun porazdelitve uporabimo neodvisnost slučajnih spremenljivk $\bar{X} - \mu$ in $\sum_{i=1}^{n}(X_i - \bar{X})^2$ ter rezultat iz 11. naloge, ki pravi, da je $\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \text{Gama}((n-1)/2, 1/(2\sigma^2))$, nakar se skličemo na 46. nalogo iz 4. razdelka, ki pravi, da mora imeti zato T Studentovo porazdelitev z $n-1$ prostostnimi stopnjami.

9. Točkasto ocenjevanje parametrov

1. V obeh primerih so slučajne spremenljivke X_i porazdeljene enako kot X, zato je:

$$E(\bar{X}) = \frac{n\,E(X)}{n} = E(X) = \mu\,,$$

torej gre res za nepristransko cenilko. Srednja kvadratična napaka je torej enaka kar disperziji. Če gre za vzorec s ponavljanjem, so slučajne spremenljivke $X_1, \ldots X_n$ neodvisne, zato je:

$$D(\bar{X}) = \frac{D(X_1) + \cdots + D(X_n)}{n^2} = \frac{\sigma^2}{n}\,,$$

kjer je:

$$\begin{aligned}\sigma^2 = D(X) &= \frac{(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_N-\mu)^2}{N} = \\ = E(X^2) - \mu^2 &= \frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N} - \mu^2\end{aligned}$$

populacijska disperzija.

Če gre za vzorec brez ponavljanja, pa nastavimo:

$$\begin{aligned}D(\bar{X}) &= \frac{1}{n^2}\,D\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n K(X_i, X_j) = \\ &= \frac{1}{n^2}\left[\sum_{i=1}^n D(X_i) - \sum_{i,j;i\neq j} K(X_i, X_j)\right].\end{aligned}$$

Velja $D(X_i) = D(X) = \sigma^2$, za $i \neq j$ pa je:

$$\begin{aligned}E(X_iX_j) &= \frac{1}{N(N-1)}\sum_{k,l;k\neq l} x_k x_l = \\ &= \frac{1}{N(N-1)}\left[\sum_{k=1}^N\sum_{l=1}^N x_k x_l - \sum_{k=1}^N x_k^2\right] = \\ &= \frac{1}{N(N-1)}\big(N^2\mu^2 - N(\mu^2+\sigma^2)\big) = \\ &= \mu^2 - \frac{\sigma^2}{N-1}\,,\end{aligned}$$

torej je $K(X_i, X_j) = -\sigma^2/(N-1)$ in končno:

$$D(\bar{X}) = \frac{N-n}{N-1}\,\frac{\sigma^2}{n}\,.$$

Tako pri vzorcu s ponavljanjem kot brez ponavljanja je cenilka dosledna, ker gre srednja kvadratična napaka proti nič. Pri vzorcu brez ponavljanja gre za končno zaporedje cenilk, pri čemer pri zadnji zajamemo vso populacijo, zato je kar $\bar{X} = \mu$ in seveda $D(\bar{X}) = 0$.

2. a) $D(\bar{X}) = \dfrac{p_1^2\sigma_1^2}{n_1} + \dfrac{p_2^2\sigma_2^2}{n_2} + \cdots + \dfrac{p_r^2\sigma_r^2}{n_r}$.
b) Ko vstavimo $n_i = np_i$, dobimo:

$$D(\bar{X}) = \frac{\sigma_{\mathrm{W}}^2}{n}\,.$$

Če z X_i označimo našo statistično spremenljivko, zoženo na i-ti stratum, iz zveze:

$$D(X_i) = E(X_i - \mu)^2 - (\mu_i - \mu)^2$$

dobimo, da populacijska disperzija $\sigma^2 = D(X)$ zadošča pomembni relaciji:

$$\sigma^2 = \sigma_{\mathrm{W}}^2 + \sigma_{\mathrm{B}}^2\,,$$

torej je $\sigma_{\mathrm{W}}^2 \le \sigma^2$.
c) Poiskati moramo ekstrem disperzije $D(\bar{X})$, ki smo jo izračunali v točki a), kot funkcije spremenljivk $n_1, n_2, \ldots n_r$, pri pogoju $n_1 + n_2 + \cdots + n_r = n$. Nastavimo Lagrangeovo funkcijo:

$$L = \frac{p_1^2\sigma_1^2}{n_1} + \frac{p_2^2\sigma_2^2}{n_2} + \cdots + \frac{p_r^2\sigma_r^2}{n_r} - \lambda(n_1 + n_2 + \cdots + n_r)$$

in njeni parcialni odvodi so enaki:

$$\frac{\partial L}{\partial n_i} = -\frac{p_i^2\sigma_i^2}{n_i^2} - \lambda\,.$$

Ko jih postavimo na nič, še iz pogoja po nekaj računanja dobimo:

$$n_i = \frac{p_i\sigma_i}{p_1\sigma_1 + p_2\sigma_2 + \cdots + p_r\sigma_r}\, n\,.$$

Pri tej izbiri dobimo:

$$D(\bar{X}) = \frac{\big(p_1\sigma_1 + p_2\sigma_2 + \cdots + p_r\sigma_r\big)^2}{n}$$

in iz Jensenove neenakosti sledi, da je to res manjše ali enako σ_{W}^2/n.

3. Ne, ker se porazdelitev vzorčnega povprečja ujema s porazdelitvijo populacije (glej 15. nalogo iz 6. razdelka) in je zato verjetnost v limiti pri pogoju za doslednost konstantna.

4. $\hat{\alpha} = \bar{X} = (X_1 + X_2 + \cdots + X_n)/n$. Cenilka je nepristranska. Je tudi dosledna, ker je dobljena po metodi momentov.

5. $\hat{a} = (\hat{z}_2 - 1)/2, \quad \hat{b} = (3 + \hat{z}_1 - 2\hat{z}_2)/2$.
Na našem konkretnem vzorcu iz $\hat{z}_1 = 1/2$ in $\hat{z}_2 = 13/10$ dobimo $\hat{a} = 3/20$ in $\hat{b} = 9/20$.

6. Velja $E(X) = 0$ (neodvisno od α), zato iz prvega momenta ne dobimo ničesar. Iz drugega momenta dobimo $\hat{\alpha} = \sqrt{\hat{z}_2/2}$.

7. Po metodi momentov dobimo $\bar{X}$ kot cenilko za $\mu := E(X)$ in:

$$\hat{z}_2 = \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}$$

kot cenilko za $E(Y^2) = \mu^2 + \sigma^2$. Cenilka za σ^2 bo torej $\hat{\sigma}^2 = \hat{z}_2 - \bar{X}^2$. Z nekaj računanja dobimo, da lahko to zapišemo tudi v obliki:

$$\hat{\sigma}^2 = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

Označimo $Y := X - \mu$. Tedaj velja:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 = \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Y_i Y_j \\
&= \frac{n-1}{n^2}\sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2}\sum\sum_{\substack{1 \le i,j, \le n \\ i \ne j}} Y_i Y_j
\end{aligned}$$

Ker je $E(Y) = 0$, je očitno:

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\,\sigma^2\,,$$

torej je $\hat{\sigma}^2$ pristranska cenilka za σ^2 (je pa *asimptotično nepristranska*). Če postavimo $k' := n/(n-1)$, dobimo, da je:

$$S^2 = \frac{n}{n-1}\,\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

nepristranska cenilka za σ^2.

Za izračun srednje kvadratne napake pišimo:

$$\hat{\sigma}^4 = \frac{(n-1)^2}{n^4} \sum_{i=1}^{n} \sum_{j=1}^{n} Y_i^2 Y_j^2 - \frac{2(n-1)}{n^4} \sum_{i=1}^{n} \sum\sum_{\substack{1 \le j,k, \le n \\ j \neq k}} Y_i^2 Y_j Y_k +$$

$$+ \frac{1}{n^4} \sum\sum_{\substack{1 \le j,k, \le n \\ i \neq j}} \sum\sum_{\substack{1 \le k,l, \le n \\ k \neq l}} Y_i Y_j Y_k Y_l =$$

$$= \frac{(n-1)^2}{n^4} \sum_{i=1}^{n} Y_i^4 + \frac{(n-1)^2}{n^4} \sum\sum_{\substack{1 \le i,j, \le n \\ i \neq j}} Y_i^2 Y_j^2 - \frac{2(n-1)}{n^4} \sum_{i=1}^{n} \sum\sum_{\substack{1 \le j,k, \le n \\ j \neq k}} Y_i^2 Y_j Y_k +$$

$$+ \frac{1}{n^4} \sum\sum_{\substack{1 \le i,j, \le n \\ i \neq j}} \sum\sum_{\substack{1 \le k,l, \le n \\ k \neq l}} Y_i Y_j Y_k Y_l \,.$$

Za prva dva člena v zadnjem izrazu je izražava očitna. Pri tretjem členu opazimo, da je, če je $j \neq k$, $E(Y_i^2 Y_j Y_k)$ enako bodisi $E(Y^3)\,E(Y)$ bodisi $E(Y^2)\big(E(Y)\big)^2$, kar je v vsakem primeru enako nič. Za četrti člen, ko je $i \neq j$ in $k \neq l$, pa dobimo, da je $E(Y_i Y_j Y_k Y_l)$ različno od nič, kvečjemu če je $i = k$ in $j = l$ ali pa $i = l$ in $j = k$. Takih členov je $2n(n-1)$ in njihova matematična upanja so enaka $\big(E(Y^2)\big)^2 = \sigma^4$. Sledi:

$$\begin{aligned} E(\hat{\sigma}^4) &= \frac{(n-1)^2}{n^3}\,\kappa^4 + \frac{(n-1)^3}{n^3}\,\sigma^4 + \frac{2(n-1)}{n^3}\,\sigma^4 = \\ &= \frac{(n-1)^2}{n^3}\,\kappa^4 + \frac{(n-1)(n^2-2n+3)}{n^3}\,\sigma^4 \,. \end{aligned}$$

od koder dobimo:

$$q(k\hat{\sigma}^2) = \left(\frac{(n-1)^2\kappa^4}{n^3} + \frac{(n-1)(n^2-2n+3)\sigma^4}{n^3} \right) k^2 - \frac{2(n-1)\sigma^4}{n}\, k + \sigma^4$$

Za $k = 1$ dobimo:

$$q(\hat{\sigma}^2) = \frac{(n-1)^2\kappa^4 + (-n^2+5n-3)\sigma^4}{n^3}$$

(to je vedno nenegativno, ker po Jensenovi neenakosti velja $\kappa \ge \sigma$). Za $k = n/(n-1)$ pa dobimo:

$$q(S^2) = \frac{1}{n}\,\kappa^4 + \frac{(3-n)}{n(n-1)}\,\sigma^4 \,.$$

Če je X porazdeljena normalno, je $\kappa = 3\sigma$ in velja:

$$q(k\hat{\sigma}^2) = \left(\frac{n^2-1}{n^2}\, k^2 - \frac{2(n-1)}{n}\, k + 1 \right) \sigma^4$$

kar je minimalno pri $k = n/(n+1)$. Najučinkovitejša izmed cenilk oblike $k\hat{\sigma}^2$ je torej:

$$(\hat{\sigma}^*)^2 = \frac{1}{n+1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \,.$$

Srednje kvadratične napake cenilk pa so:

$$q(\hat{\sigma}^2) = \frac{2n-1}{n^2}\,\sigma^4\,, \quad q(S^2) = \frac{2}{n-1}\,\sigma^4\,, \quad q\big((\hat{\sigma}^*)^2\big) = \frac{2}{n+1}\,\sigma^4\,.$$

Ni težko preveriti, da je $\hat{\sigma}^2$ vselej učinkovitejša od S^2 in da je $(\hat{\sigma}^*)^2$ vselej učinkovitejša od $\hat{\sigma}^2$.

8. Ker vemo, da je $E(X) = \lambda$ in da je $\bar{X}$ nepristranska cenilka za $E(X)$, je $\bar{X}$ res nepristranska cenilka za λ. Poleg tega je tudi $D(X) = \lambda$ (glej npr. 21. nalogo iz 5. razdelka) in iz prejšnje naloge vemo, da je S^2 nepristranska cenilka za $D(X) = \lambda$.

Izračun disperzije prve cenilke je preprost:

$$D(\bar{X}) = \frac{D(X_1) + \cdots + D(X_n)}{n^2} = \frac{\lambda}{n}\,.$$

Pri izračunu disperzije druge cenilke pa se lahko spet opremo na prejšnjo nalogo: ker je cenilka nepristranska, velja:

$$D(S^2) = q(S^2) = \frac{1}{n}\,\kappa^4 + \frac{(3-n)}{n(n-1)}\,\sigma^4\,,$$

kjer je σ^2 drugi centralni moment (t. j. disperzija), κ^4 pa je četrti centralni moment. Spet iz 21. naloge iz 5. razdelka poberemo $\kappa^4 = 3\lambda^2 + \lambda$. Sledi:

$$D(S^2) = \frac{\lambda}{n} + \frac{2\lambda^2}{n-1}\,.$$

Torej ima S^2 v vsakem primeru večjo disperzijo kot $\bar{X}$.

9. Tu dobimo isto cenilko kot pri metodi momentov, t. j. $\hat{\alpha} = \bar{X}$.

10. $\hat{a} = 1/4$, $\hat{b} = 3/10$, kar se ne ujema z ocenama po metodi momentov.

11. Po metodi momentov dobimo $\bar{X}$ kot nepristransko cenilko za $E(X) = a/2$, torej je tudi $a = 2\bar{X}$ nepristranska cenilka za a.
Iz metode največjega verjetja dobimo cenilko $M := \max\{X_1, \ldots X_n\}$. Za $x \in [0, a]$ je porazdelitvena funkcija te slučajne spremenljivke enaka:

$$F_M(x) = P(M < x) = P(X_1 < x, X_2 < x, \ldots X_n < x) = \left(\frac{x}{a}\right)^n\,,$$

verjetnostna gostota pa je potemtakem enaka:

$$p_M(x) = \begin{cases} nx^{n-1}/a^n & ;\ 0 \le x \le a \\ 0 & ;\ \text{sicer} \end{cases}\,.$$

Od tod izračunamo $E(M) = \dfrac{n}{n+1}\,a$, torej je cenilka M pristranska (je pa *asimptotično nepristranska*). Cenilka M' bo torej nepristranska za $k' = (n+1)/n$.

Za srednjo kvadratično napako izračunamo:

$$q(kM) = \left[\frac{n}{n+2}\,k^2 - \frac{2n}{n+1}\,k + 1\right] a^2\,,$$

kar bo minimalno pri $k^* = (n+2)/(n+1)$, neodvisno od a. Srednje kvadratične napake vseh omenjenih cenilk so prikazane v naslednji tabeli:

C	A	M	M'	M^*
$q(C)$	$\dfrac{a^2}{3n}$	$\dfrac{2a^2}{(n+1)(n+2)}$	$\dfrac{a^2}{n(n+2)}$	$\dfrac{a^2}{(n+1)^2}$

Pri $n = 1$ so torej vse cenilke A, M in M' enako učinkovite.
Pri $n = 2$ sta A in M enako učinkoviti, M' pa je učinkovitejša.
Pri $n \geq 3$ je M učinkovitejša od A in M' učinkovitejša od M.
Vselej pa je M^* učinkovitejša od vseh ostalih cenilk.

10. Intervali zaupanja

1. $\Delta \doteq 0\dot{}078, \quad 0\dot{}122 \leq p \leq 0\dot{}278.$

2. $\Delta \doteq 0\dot{}0082, \quad 0\dot{}4966 \leq p \leq 0\dot{}5130.$

3. $\bar{X} = 97, \quad \Delta \doteq 3\dot{}27, \quad 93\dot{}73 \leq \mu \leq 100\dot{}27.$

4. $S = 5, \quad df = 8, \quad t_{0.975} = 2\dot{}31, \quad \Delta \doteq 3\dot{}85, \quad 93\dot{}15 \leq \mu \leq 100\dot{}85.$

5. $\bar{X} = 125, \quad S \doteq 2\dot{}646, \quad df = 4, \quad \chi^2_{0.05} \doteq 0\dot{}711, \quad \chi^2_{0.95} \doteq 9\dot{}49, \quad 1\dot{}72 \leq \sigma \leq 6\dot{}28.$

6. $\bar{X} \doteq 45\dot{}51, \quad S \doteq 3\dot{}71, \quad df = 74, \quad \chi^2_{0.025} \doteq 52\dot{}1, \quad \chi^2_{0.975} \doteq 99\dot{}7, \quad 3\dot{}20 \leq \sigma \leq 4\dot{}42.$

11. Testi značilnosti

1. Če je več kot pol srečk dobitnih, je verjetnost dogodka, da sta med osmimi srečkami dobitni *največ* dve, navzgor omejena z 0˙145, kar s stališča statistike ni zadosten razlog, da bi loterijo sumili goljufije (če kupimo 8 srečk in loterijo osumimo goljufije, brž ko sta dobitni dve srečki ali manj, lahko verjetnost, da jo sumimo po krivici, doseže 0˙145, kar je preveč; verjetnost krivičnega suma sme biti največ 0˙05, če smo še previdnejši, pa največ 0˙01).

2. Merilo za to, kateri vzorec bolj "ustreza" ničelni hipotezi, bo odstopanje števila grbov od 5000 (v eno ali drugi stran, ker je alternativna hipoteza obojestranska). Če število grbov označimo z X, po centralnem limitnem izreku približno izračunamo:

$$P(X < 4875 \text{ ali } X > 5125) \approx 1 - 2\Phi(2\dot{}51) \doteq 0\dot{}0121\,.$$

V resnici so vse decimalke natančne. Pri $\alpha = 0\dot{}05$ bomo torej našo hipotezo zavrnili, pri $\alpha = 0\dot{}01$ pa ne.

3. Če se omejimo na prvih 100 iger, je glede na izjave prireditelja in naše sume najmanj "ustrezen" prav dogodek, da po 100 igrah glavni dobitek še ni bil izžreban. Verjetnost, da po 100 igrah glavni dobitek še ni izžreban, je $(49/50)^{100} \doteq 0\dot{}133$, kar pomeni, da ničelne hipoteze ne moremo zavrniti.

4. $n = 100, k = 23$: $Z \doteq 0\dot{}75$, $K_\alpha \doteq (1\dot{}645, \infty)$, hipoteze ne moremo zavrniti.
$n = 1000, k = 230$: $Z \doteq 2\dot{}37$, hipotezo zavrnemo.

5. $\bar{X} = 97$, $Z = -1\dot{}8$.
Pri alternativni hipotezi $\mu \neq 100$ je $K_\alpha \doteq (-\infty, -1\dot{}96) \cup (1\dot{}96, \infty)$ in hipoteze ne moremo zavrniti.
Pri alternativni hipotezi $\mu < 100$ je $K_\alpha \doteq (-\infty, -1\dot{}645)$ in hipotezo zavrnemo.
Pri alternativni hipotezi $\mu > 100$ pa je $K_\alpha \doteq (1\dot{}645, \infty)$ in hipoteze ne moremo zavrniti.

6. $\bar{X} = 107$, $S = 10$, $T = 2\dot{}1$, $df = 8$, $K_\alpha \doteq (-\infty, -2\dot{}31) \cup (2\dot{}31, \infty)$.
Hipoteze ne moremo zavrniti.
Če bi vedeli, da je $\sigma = 10$, pa bi bilo $K_\alpha \doteq (-\infty, -1\dot{}96) \cup (1\dot{}96, \infty)$ in hipotezo bi zavrnili.

7. $\bar{X} = 48$, $S \doteq 2\dot{}58$, $T = -2\dot{}45$, $df = 9$.
Če za H_1 vzamemo, da je $\mu \neq 50$, je $K_\alpha \doteq (-\infty, -2\dot{}26) \cup (2\dot{}26, \infty)$ in hipotezo zavrnemo.
Če za H_1 vzamemo, da je $\mu < 50$, pa je $K_\alpha \doteq (-\infty, -1\dot{}83)$ in hipotezo prav tako zavrnemo.

8. Označimo z Δ razliko v teži posamezne osebe (teža po dieti minus teža pred dieto). Tedaj je $\Delta \sim \mathrm{N}(\mu, \sigma)$ in testiramo ničelno hipotezo, da je $\mu = 0$, proti alternativni hipotezi, da je $\mu < 0$.
Velja $\bar{\Delta} \doteq -5\dot{}6$, $S \doteq 6\dot{}433$, $df = 9$, $T \doteq -2\dot{}75$, $c \doteq (-\infty, -1\dot{}83]$.
Ničelno hipotezo zavrnemo, torej sprejmemo hipotezo, da dieta deluje.

9. $\bar{X} = 20, \quad \bar{Y} = 22, \quad S = 2{\cdot}65, \quad T = -1{\cdot}5, \quad df = 14,$
$K_\alpha \doteq (-2{\cdot}14, -\infty) \cup (2{\cdot}14, \infty)$. Hipoteze ne moremo zavrniti.

10. $\bar{X} = 100, \quad \bar{Y} = 96, \quad S = 2{\cdot}5, \quad T = 3{\cdot}2, \quad df = 16, \quad K_\alpha \doteq (2{\cdot}58, \infty)$.
Hipotezo zavrnemo.

11. $\bar{X}_1 = 4, \quad \bar{X}_2 = 3, \quad \bar{X}_3 \doteq 1{\cdot}667,$
$S_{\mathrm{B}}^2 \doteq 9{\cdot}333, \quad S_{\mathrm{W}}^2 \doteq 8{\cdot}667, \quad F \doteq 4{\cdot}31,$
$df_1 = 2, \quad df_2 = 8, \quad K_\alpha \doteq (4{\cdot}46, \infty)$.
Hipoteze ne moremo zavrniti.

12. $S = 7{\cdot}45$.

a) $\chi^2 = 20, \quad K_\alpha \doteq (0, 2{\cdot}70) \cup (19{\cdot}0, \infty)$, hipotezo zavrnemo.

b) $\chi^2 = 5, \quad K_\alpha \doteq (0, 3{\cdot}33)$, hipoteze ne moremo zavrniti.

13. Če so frekvence 20, 45 in 35, je: $\chi^2 = 5{\cdot}5, \quad df = 2, \quad K_\alpha \doteq (5{\cdot}99, \infty)$.
Hipoteze ne moremo zavrniti.
Pri frekvencah 200, 450 in 350 pa je $\chi^2 = 55$. Ker je kritično območje isto, hipotezo zdaj zavrnemo.

14. Pri frekvencah 21, 42, 77 in 116 je: $\chi^2 \doteq 81{\cdot}3 \quad df = 3, \quad K_\alpha \doteq (11{\cdot}3, \infty)$.
Hipotezo zavrnemo.
Če so frekvence 21, 37, 53 in 25, pa je $\chi^2 \doteq 18{\cdot}2$ in hipotezo še vedno zavrnemo.

15. Diskretizirana porazdelitev iz ničelne hipoteze:

pod 10	10–20	20–30	30–40	nad 40
0˙0668	0˙2417	0˙3829	0˙2417	0˙0668

$\chi^2 \doteq 11{\cdot}8, \quad df = 4, \quad K_\alpha \doteq (9{\cdot}49, \infty)$.
Hipotezo zavrnemo.

16. $\chi^2 \doteq 42{\cdot}6, \quad df = 6, \quad K_\alpha \doteq (16{\cdot}8, \infty)$.
Hipotezo zavrnemo.

17. $\chi^2 \doteq 169, \quad df = 8, \quad K_\alpha \doteq (15{\cdot}5, \infty)$.
Hipotezo zavrnemo.

18. $S^+ = 34$, $S^- = 16$, $Z \doteq 2{\cdot}55$. Hipotezo zavrnemo.

19. $m = 9$, $n = 11$, $\sum_{i=1}^{9} R_i = 70, \quad Z \doteq -1{\cdot}86, \quad K_\alpha \doteq (-\infty, -1{\cdot}96) \cup (1{\cdot}96, \infty)$.
Hipoteze ne moremo zavrniti.

12. Povezanost dveh številskih spremenljivk

1. Označimo z X sistolični, z Y pa diastolični pritisk. Izračunajmo:
$\bar{X} = 126$, $\bar{Y} \doteq 79{\cdot}67$, $C_x \doteq 30{\cdot}17$, $C_y \doteq 19{\cdot}32$, $C_{xy} = 130$, $R \doteq 0{\cdot}223$, $Z \doteq 0{\cdot}227$, $c \doteq 1{\cdot}96$.
Interval zaupanja: $-0.327 < \rho < 0.656$.

2. $T \doteq 0{\cdot}825$, $df = 13$, $K_\alpha = (-\infty, 1{\cdot}77) \cup (1{\cdot}77, \infty)$.
Hipoteze ne moremo zavrniti.

3. $\bar{X} = 3$, $\bar{Y} = 7$, $C_{\mathrm{xx}} = 10$, $C_{\mathrm{xy}} = 20$.
Regresijska premica: $y = 2x + 1$.
Pri $X = 10$ je $\hat{Y} = 21$.
$df = 3$, $t_{0.975} \doteq 3{\cdot}18$, $S \doteq 1{\cdot}155$, $\Delta \doteq 7{\cdot}06$.
Interval zaupanja: $13{\cdot}94 < Y < 28{\cdot}06$.