VAJE IZ VERJETNOSTI IN STATISTIKE

Martin Raič

Datum zadnje spremembe: 15. junij 2014

Kazalo

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	$V_n^k = n(n-1)\ldots(n-k+1)$	$C_n^k = \binom{n}{k} = \frac{V_n^k}{k!}$

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

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.

8. [1] 7 moških in 5 žensk se odpravi na taborjenje. Na voljo imajo dva šotora za tri osebe in tri šotore za dve osebi. Vse šotore med seboj ločimo. Na koliko načinov se lahko razporedijo v šotore, če:

   a) ni omejitev?

   b) smejo biti v posameznem šotoru le osebe istega spola?

   c) mora biti v posameznem šotoru najmanj en moški in najmanj ena ženska?

[1] Avtor naloge: Gregor Šega

2. Elementarna verjetnost

Klasična verjetnost, klasična geometrijska verjetnost. Računanje z dogodki.

Klasična verjetnost

Če so vsi izidi enako verjetni, za dogodek A velja:

$$P(A) = \frac{\text{število izidov, ki so v } A}{\text{število vseh izidov}}.$$

Temu, da so vse možnosti enako verjetne, pravimo **slepa izbira**. *Izbirati dve (splošneje n) stvari na slepo in* **neodvisno** *pa pomeni, da so vse kombinacije možnosti (kjer stvari ločimo) enako verjetne. Z drugimi besedami, to pomeni slepo izbiro ustreznega urejenega para oz. n-terice.*

1. Vržemo dve neodvisni standardni kocki. Kolikšna je verjetnost, da bo 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? Privzamemo, da sta oba spola pri posameznem rojstvu enako verjetna in da so spoli pri posameznih rojstvih neodvisni.

$$\boxed{P(\overline{A}) = 1 - P(A)}$$

3. Vržemo pet neodvisnih standardnih kock. Kolikšna je verjetnost, da bo 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?

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

Dokaz. Ker sta A in $B \setminus A$ nezdružljiva, velja $P(A \cup B) = P(A) + P(B \setminus A)$. Podobno, ker sta $A \cap B$ in $B \setminus A$ nezdružljiva, velja $P(B) = P(A \cap B) + P(B \setminus A)$. Ko dobljeni enakosti odštejemo in preoblikujemo, dobimo natančno želeno formulo.

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. Najmanj koliko ljudi mora biti, da je ta verjetnost enaka vsaj $1/2$? Zapišite rezultat še za splošno število dni v letu in raziščite asimptotično obnašanje, ko gre le-to proti neskončno.

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?

10. Pri igri Loto na kombinacijskem listku prekrižamo 7 številk izmed 39. Izžreba se 7 rednih številk in še ena dodatna. Možni so naslednji dobitki:

    - sedmica: vse prekrižane številke so redno izžrebane;
    - šest in dodatna: med prekrižanimi številkami je šest redno izžrebanih in ena dodatna;
    - šestica: natanko šest prekrižanih številk je redno izžrebanih, dodatna ni prekrižana;
    - petica: natanko pet prekrižanih številk je redno izžrebanih (dodatna pa je lahko prekrižana ali ne);
    - štirica: natanko štiri prekrižane številke so redno izžrebane (dodatna pa je lahko prekrižana ali ne);
    - tri in dodatna: natanko tri prekrižane številke so redno izžrebane, prekrižana pa je tudi dodatna številka.

    Izračunajte verjetnosti posameznih dobitkov.

11. V kupu je 10 kart, od tega dve rdeči, tri zelene in pet belih. Kup dobro premešamo in drugo za drugo brez vračanja vlečemo karte. Izračunajte verjetnosti dogodkov:

    a) da bo prva rdeča karta izvlečena pred prvo zeleno;

    b) da bo prva rdeča karta izvlečena pred zadnjo zeleno.

12. Študenti, ki bodo pisali izpit, se posedejo v tri vrste in tri kolone, tako kot je prikazano spodaj:

Aljaž	Brigita	Cveto
Dragica	Edo	Fani
Gregor	Hana	Iztok

    Asistent na slepo izbere tri študente in jih zamenja: prvega premesti na mesto drugega, drugega na mesto tretjega in tretjega na mesto prvega. Kolikšna je verjetnost, da Aljaž in Brigita po premestitvi še vedno sedita skupaj v isti vrsti?

Računanje z dogodki

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

13. Poenostavite naslednji izraz z dogodki:

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

14. 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 $P(A \cap B \cap C) = 0\dot{}1$.

Načelo vključitev in izključitev

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

Dokaz z indukcijo. Za $n = 1$ in $n = 2$ velja. Naredimo indukcijski korak z n na $n + 1$. Najprej opazimo:

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

Po indukcijski predpostavki je:

$$\begin{aligned}
P(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) = P(A_1) &+ P(A_2) + \cdots + P(A_n) - \\
&- P(A_1 \cap A_2) - P(A_1 \cap A_3) - \cdots - P(A_{n-1} \cap A_n) + \\
&+ \mathbf{1}_{A_1 \cap A_2 \cap A_3} + \mathbf{1}_{A_1 \cap A_2 \cap A_4} + \cdots + \mathbf{1}_{A_{n-2} \cap A_{n-1} \cap A_n} - \\
&\qquad \cdots \\
&+ (-1)^{n+1} P(A_1 \cap \cdots \cap A_n) + \\
&+ P(A_{n+1}) - \\
&- P(A_1 \cap A_{n+1}) - P(A_2 \cap A_{n+1}) - \cdots + P(A_n \cap A_{n+1}) - \\
&+ P(A_1 \cap A_2 \cap A_{n+1}) + P(A_1 \cap A_3 \cap A_{n+1}) + \cdots + P(A_{n-1} \cap A_n \cap A_{n+1}) + \\
&\qquad \cdots \\
&+ (-1)^{n+2} P(A_1 \cap A_2 \cap \cdots \cap A_n \cap A_{n+1}) ,
\end{aligned}$$

kar je natančno ustrezna desna stran: iz razvoja verjetnosti $P(A_1 \cup A_2 \cup \cdots \cup A_n)$ smo dobili tiste člene, ki ne vsebujejo dogodka A_{n+1}, iz $P(A_{n+1})$ in razvoja verjetnosti $P\big((A_1 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1})\big)$ pa tiste člene, ki dogodek A_{n+1} vsebujejo.

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

*σ-**algebra** na množici Ω je družina $\mathcal{F}$ njenih podmnožic, ki izpolnjuje naslednje pogoje:*

- *$\emptyset \in \mathcal{F}$.*
- *Če je $A \in \mathcal{F}$, je tudi $\overline{A} \in \mathcal{F}$.*
- *Za poljubno zaporedje množic $A_1, A_2, \ldots \in \mathcal{F}$ je tudi $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.*

16. Za vsak $n \in \mathbb{N}$ naj bo $\mathcal{A}_n$ najmanjša σ-algebra na $\mathbb{N}$, ki vsebuje množice $\{1\}, \{2\}, \ldots$ $\ldots \{n\}$.

    a) Opišite družine množic $\mathcal{A}_n$.
    b) Pokažite, da njihova unija $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{A}_n$ ni σ-algebra.
    c) Določite najmanjšo σ-algebro na $\mathbb{N}$, ki vsebije $\mathcal{A}$.

Klasična geometrijska verjetnost

Točka je izbrana **na slepo** *iz množice G, ki je lahko interval, lik, telo ipd., če za vsako merljivo podmnožico $A \subseteq G$ velja:*

$$P(\text{točka pripada } A) = \frac{\text{mera množice } A}{\text{mera množice } G} \,.$$

Pri tem je mera lahko dolžina, ploščina itd.
Na slepo in **neodvisno** *izbrati dve točki (splošneje, n točk) pomeni slepo izbiro njunega urejenega para (oz. n-terice) v ustreznem kartezijskem produktu.*

17. 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? Seveda privzamemo, da je faza semaforja izbrana na slepo (oz. da sta fazi semaforjev izbrani na slepo in neodvisno).

18. Avtobus se ustavi na postaji na slepo med 6:55 in 7:05. Študent pa je nagnjen k zamujanju in pride na postajo na slepo med 7:00 in 7:07, neodvisno od avtobusa.

a) Kolikšna je verjetnost, da ujame ta avtobus?
b) Če želi študent še pravočasno priti na predavanje, mora biti na tem avtobusu najkasneje ob 7:02. Kolikšna je verjetnost, da se to zgodi?

19. Kolikšna je verjetnost, da je na slepo izbrana točka v kvadratu bližje robu kot središču kvadrata?

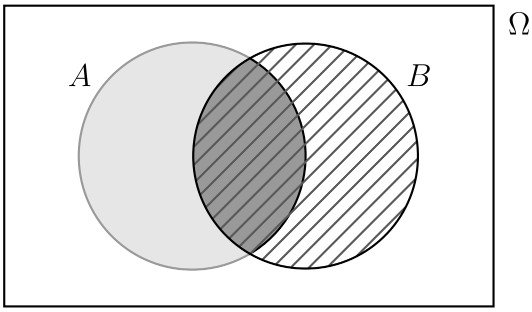
20. *Buffonova igla.* Na list papirja z ravnimi vzporednimi črtami, razmaknjenimi za a, na slepo vržemo iglo dolžine b. Kolikšna je verjetnost, da igla seka katero od črt?

3. Pogojna verjetnost

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

Definicija pogojne verjetnosti

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



Če je dogodek B sestavljen iz samih enako verjetnih izidov, pa je tudi:

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

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 vse v redu?

4. *Monty–Hallov paradoks.* Dana so tri vrata, za enimi je skrit avto, za dvojimi pa buča. Najprej izberemo ena vrata, ne da bi jih odprli, nakar vodja igre odpre ena

izmed vrat, za katerima je buča in ki jih nismo izbrali. Nato nam ponovno ponudi, da izberemo ena izmed še zaprtih vrat. Tisto, kar se skriva za njimi, dobimo.

Kako naj ravnamo, če želimo dobiti avto? Privzamemo, da so vse možnosti za vrata, za katerimi stoji avto, enako verjetne. Prav tako privzamemo, da vodja igre v primeru, ko ima možnost izbire, izbere na slepo.

Razmislek: Recimo, da smo najprej pokazali na prva vrata, vodja igre pa je nato odprl tretja vrata. Prva in druga vrata so še zaprta. Ker so vsa vrata enako verjetna, je pri obojih verjetnost, da bo zadaj avto, enaka 1/2. Torej je čisto vseeno, kaj storimo.

Je s tem razmislekom vse v redu?

Kaj pa, če v prvo izberemo prva vrata in vemo, da vodja igre *favorizira* tretja vrata, torej ta vrata odpre vedno, kadar jih ima na voljo?

5. Janez in Peter igrata namizni tenis. V vsaki rundi nekdo zmaga in oba sta enakovredna, ne glede na zgodovino, 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, a tokrat Janez dobi posamezno rundo z verjetnostjo $1/3$ (ne glede na zgodovino), 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

7. Pri določenem slučajnem poskusu lahko med drugim pride do opažanja A in do opažanja B. Lahko pride tudi do obeh opažanj, to označimo z $A \cap B$. Poskus ponavljamo in pri vsaki izvedbi pride do opažanja A z verjetnostjo $P(A)$, do opažanja B z verjetnostjo $P(B)$ in do obeh opažanj z verjetnostjo $P(A \cap B)$, ne glede na zgodovino. Privzemimo, da je $P(B) > 0$.

   Poskus ponavljamo, dokler ne pride do opažanja B. Kolikšna je verjetnost, da pri zadnjem poskusu pride tudi do opažanja A?

8. Mečemo pošten kovanec, pri čemer privzamemo, da je verjetnost, da v posameznem metu pade grb, enaka $1/2$ ne glede na prejšnje mete. Kolikšna je verjetnost, da v prvih n metih *nista* padli dve zaporedni cifri?

Izrek o polni verjetnosti

Če $H_1, H_2, H_3, \ldots$ tvorijo **popoln sistem dogodkov** *(t. j. vedno se zgodi natanko eden izmed njih), velja:*

$$P(A) = P(H_1)\,P(A \mid H_1) + P(H_2)\,P(A \mid H_2) + P(H_3)\,P(A \mid H_3) + \cdots$$

Dogodkom H_i često pravimo **hipoteze** *in jih je lahko končno ali pa števno neskončno.*

9. 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 $H_1, H_2, H_3, \ldots$ tvorijo popoln sistem dogodkov, velja:

$$P(H_i \mid A) = \frac{P(H_i)\, P(A \mid H_i)}{P(H_1)\, P(A \mid H_1) + P(H_2)\, P(A \mid H_2) + P(H_3)\, P(A \mid H_3) + \cdots}.$$

Brezpogojnim verjetnostim $P(H_i)$ pravimo **apriorne**, *pogojnim verjetnostim $P(H_i \mid A)$ pa* **aposteriorne** *verjetnosti hipotez.*

10. Žena pošlja moža na trg po solato, ki jo prodajata dve branjevki, Francka in Micka. Verjetnost, da mož kupi solato pri Francki, je 60%, verjetnost, da kupi pri Micki, pa 40%. Francka ima 10%, Micka pa 20% nagnite solate. Mož prinese domov nagnito glavo solate. Katero branjevko lahko žena bolj upravičeno osumi, da mu je prodala nagnito solato? Privzamemo, da branjevki solato izbirata na slepo.

11. 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?

12. 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 (na slepo in 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. $P(A \mid B) = P(A)$).
Če je $P(B) > 0$, je to ekvivalentno pogoju, da je $P(A \mid B) = P(A)$.
Če je $0 < P(B) < 1$, pa je to ekvivalentno tudi pogoju, da je $P(A \mid B) = P(A \mid \overline{B})$.
Dogodki $A_1, A_2, A_3 \ldots$ so neodvisni, če za poljubne različne indekse $i_1, i_2, \ldots i_k$ velja:

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

13. 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?

14. Danih je osem kart: as, kralj, dama, fant, 10, 9, 8 in 7. Na slepo izvlečemo eno karto. Definirajmo naslednje dogodke:

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

So dogodki F, G in H neodvisni?

15. 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?

16. *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

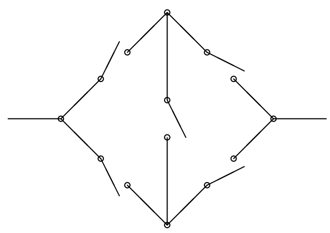
Naj bo $\mathcal{F}$ družina dogodkov. S $\sigma(\mathcal{F})$ označimo najmanjšo σ-algebro, ki vsebuje $\mathcal{F}$, t. j. družino vseh dogodkov, ki jih dobimo iz dogodkov iz $\mathcal{F}$ s števnimi unijami in komplementi.

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.

17. 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?



18. 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?

19. 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?

20. 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?

21. Miha se odpravi na obisk k vinogradnikom Janezu, Lojzu in Štefanu. Vsak mu ponudi kozarec vina, ki je lahko cviček ali pa šmarnica. 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.

22. Pustolovec Albert pride v tujo deželo, kjer ga takoj primejo in vtaknejo v ječo. Po prvi noči, prebiti v ječi, ga obišče kralj in mu ponudi posodo, v kateri je ena rdeča in ena zelena kroglica. Albert na slepo izvleče eno kroglico. Če izvleče zeleno, je izpuščen, če izvleče rdečo, pa mora prebiti v ječi še eno noč. Naslednji dan ga spet obišče kralj in spet mu ponudi posodo, le da sta tokrat notri dve rdeči in ena zelena kroglica. Spet je Albert izpuščen, če izvleče zeleno kroglico, sicer pa mora ponovno prespati v ječi. Tako se nadaljuje: vsak dan je v posodi ena rdeča kroglica več.

    a) Dokažite, da Albert z verjetnostjo ena nekoč pride iz ječe.

    b) Ko Alberta izpustijo, mu kralj izroči posodo s kroglicami (n rdečimi in eno zeleno, če je Albert v ječi prespal n-krat). Albert nato sam takoj izvleče eno kroglico. Če je zelena, takoj zapusti deželo, sicer pa izvlečeno rdečo kroglico odvrže in tam prespi (tokrat na svobodi). Nato spet vleče kroglice (tokrat z eno rdečo manj) in če izvleče zeleno, deželo zapusti, sicer pa ponovno prespi. Tako nadaljuje, vsakič z eno rdečo kroglico manj.

       Recimo, da je Albert v tej deželi prespal natanko petkrat. Kolikšna je pogojna verjetnost, da je v ječi prespal trikrat?

23. Miranda je na nočni zabavi spoznala Ferdinanda. V dneh po zabavi čaka na njegov klic. Verjetnost, da jo Ferdinand prvič pokliče k-ti dan po zabavi, je enaka 3^{-k}. Vsako noč, ki sledi dnevu, ko Ferdinand Mirande ne pokliče, Miranda spozna novega fanta z verjetnostjo $1/10$.

    Recimo, da je Ferdinand poklical Mirando. Kolikšna je pogojna verjetnost, da je, preden jo je prvič poklical, že spoznala novega fanta (privzamemo, da jih spoznava le ponoči)?

4. Slučajne spremenljivke

Pojem porazdelitve, kumulativna porazdelitvena funkcija, porazdelitvena shema diskretno porazdeljene slučajne spremenljivke, porazdelitvena gostota zvezno porazdeljene slučajne spremenljivke. Ugotavljanje in prepoznavanje porazdelitev. Približni obrazci za binomsko porazdelitev. Vrstilne karakteristike. Transformacije (funkcije) slučajnih spremenljivk. Generiranje slučajnih spremenljivk.

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 (če so vse vrednosti a_i različne) pomeni $P(X = a_1) = p_1$, $P(X = a_2) = p_2$ itd. Velja:

$$p_1 + p_2 + p_3 + \cdots = 1\,.$$

Slučajna spremenljivka X je diskretna natanko tedaj, ko njena **verjetnostna funkcija**:

$$f_X(x) = P(X = x)$$

zadošča $\sum_x f_X(x) = 1$.

V splošnem porazdelitev opišemo z verjetnostmi $P(X \in C)$ za vse merljive množice C. Pri **realnih** *slučajnih spremenljivkah pa zadostuje za C vzeti poltrake $(-\infty, x]$. Tako dobimo* **kumulativno porazdelitveno funkcijo**:

$$F_X(x) = P(X \le x)\,.$$

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

Diskretna enakomerna porazdelitev *na n-elementni množici $\{a_1, a_2, \ldots a_n\}$ je porazdelitev na slepo izbranega elementa te množice, t. j. porazdelitev s shemo:*

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$$

Označevali jo bomo z $\mathrm{E_d}\{a_1, a_2, \ldots a_n\}$.

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.

Realna slučajna spremenljivka X je porazdeljena **zvezno**, *če za poljubna $a \le b$ velja:*

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

To je natanko tedaj, ko za vsak $x \in \mathbb{R}$ velja:

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

Poleg tega je to natanko tedaj, ko je kumulativna porazdelitvena funkcija F_X absolutno zvezna – veljata implikaciji:

zvezna, odsekoma zvezno odvedljiva $\Longrightarrow$ *absolutno zvezna* $\Longrightarrow$ *zvezna*.

Za skoraj vsak x velja:

$$\begin{aligned} p_X(x) = F_X'(x) &= \lim_{h \downarrow 0} \frac{F_X(x+h) - F_X(x)}{h} = \lim_{h \downarrow 0} \frac{P(x < X \le x+h)}{h} = \\ &= \lim_{h \downarrow 0} \frac{F_X(x-h) - F_X(x)}{-h} = \lim_{h \downarrow 0} \frac{P(x-h < X \le x)}{h}\,. \end{aligned}$$

Za vsak x velja $P(X = x) = 0$. Velja tudi $\int_{-\infty}^{\infty} p_X(x)\,\mathrm{d}x = 1$.

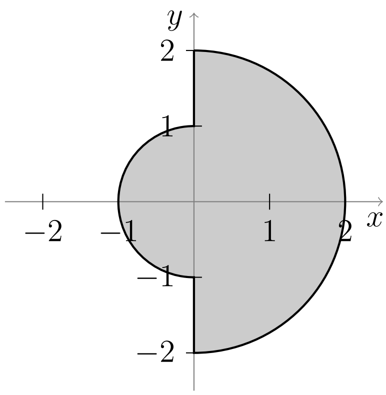
3. Avtobus vozi na 10 minut, na postajo pa pridemo na slepo. Slučajna spremenljivka X naj predstavlja čas čakanja na avtobus v minutah. Zapišite kumulativno porazdelitveno funkcijo te slučajne spremenljivke. Nadalje dokažite, da je porazdelitev zvezna, in zapišite še njeno gostoto.

Zvezna enakomerna porazdelitev *na intervalu (a, b) $(a < b)$ je porazdelitev na slepo izbrane točke iz tega intervala. To je porazdelitev z gostoto:*

$$p(x) = \begin{cases} \dfrac{1}{b-a} & ;\ a < x < b \\ 0 & ;\ \textit{sicer} \end{cases}\,.$$

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 kumulativno porazdelitveno funkcijo in gostoto te slučajne spremenljivke.

5. Na slepo izberemo točko iz lika, ki ga sestavljajo točke, ki ležijo levo od ordinatne osi in so od izhodišča oddaljene največ 1, in točke, ki ležijo desno od ordinatne osi in so od izhodišča oddaljene največ 2 (glej sliko). Naj bo Z oddaljenost izbrane točke od izhodišča. Zapišite kumulativno porazdelitveno funkcijo in gostoto porazdelitve te slučajne spremenljivke.



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

Bernoullijevo zaporedje poskusov *je zaporedje neodvisnih slučajnih poskusov, od katerih lahko vsak uspe ali ne uspe, in sicer vsak poskus uspe z isto verjetnostjo.*
Binomska porazdelitev $\mathrm{Bin}(n, p)$ *je porazdelitev števila uspelih poskusov v Bernoullijevem zaporedju n poskusov, od katerih vsak uspe z verjetnostjo p. Če je $X \sim \mathrm{Bin}(n, p)$, velja:*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}\,; \quad k = 0, 1, \ldots n\,.$$

7. 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?

Aproksimacija točkastih verjetnosti pri binomski porazdelitvi

Naj bo $X \sim \mathrm{Bin}(n, p)$ in $n \to \infty$ ter še $k \in \mathbb{N}_0$. Če gre $p \to 0$ in je $|k - np| \ll \sqrt{n}$, 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 za velike n približno pri $p = 0{\cdot}6/\sqrt[3]{n}$.

V okviru dometa aproksimacij lahko relativne napake pri aproksimaciji točkastih verjetnosti $P(X = k)$ navzgor omejimo s količinami naslednjih velikostnih redov:

- pri Poissonovi aproksimaciji: $p + \frac{(k-np)^2}{n}$;
- pri Laplaceovi lokalni formuli: $\dfrac{1}{\sigma} + \dfrac{|k-np|^3}{\sigma^4}$;
- pri Laplaceovi lokalni formuli, če je $k \in \mathbb{Z} + 1/2$: $\dfrac{1 + |k-np|}{\sigma^2} + \dfrac{|k-np|^3}{\sigma^4}$.

Izboljšave aproksimacij (asimptotski razvoj 1. reda):

- pri Poissonovem obrazcu: $P(X=k) \approx \dfrac{(np)^k}{k!} \exp\left(-np + \dfrac{k-(k-np)^2}{2n}\right) \approx \dfrac{(np)^k}{k!} \exp\left(-np + \dfrac{p-(k-np)^2}{2n}\right)$;
- pri Laplaceovi lokalni formuli: $P(X=k) \approx \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{x^2}{2} + \dfrac{2p-1}{6\sigma}(3x - x^3)\right)$;
- pri Laplaceovi integralski formuli: $P(X<k) \approx \dfrac{1}{2} + \Phi\left(x + \dfrac{2p-1}{6\sigma}(x^2-1)\right)$ (za $k \in \mathbb{Z} + 1/2$);

Označili smo $x = (k-np)/\sigma$. Iz zgornjih izboljšanih aproksimacij lahko izpeljemo asimptotično obnašanje napake pri Poissonovi aproksimaciji in pri Laplaceovi lokalni formuli, če je $1/n \ll p \ll 1$:

- $\max\limits_k \left|P(X=k) - \dfrac{(np)^k e^{-np}}{k!}\right| \sim \dfrac{p}{2\sigma\sqrt{2\pi}} \max\limits_x \left|1-x^2\right| e^{-x^2/2} = \dfrac{p}{2\sigma\sqrt{2\pi}} \doteq 0{\cdot}199\,\dfrac{p}{\sigma}$;
- $\sum\limits_k \left|P(X=k) - \dfrac{(np)^k e^{-np}}{k!}\right| \sim \dfrac{p}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} \left|1-x^2\right| e^{-x^2/2}\,\mathrm{d}x = \dfrac{2p}{\sqrt{2\pi e}} \doteq 0{\cdot}484\,p$;
- $\max\limits_k \left|P(X=k) - \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(k-np)^2}{2\sigma^2}\right)\right| \sim \dfrac{1}{6\sigma^2\sqrt{2\pi}} \max\limits_x \left|3x - x^3\right| e^{-x^2/2} = \dfrac{1}{\sigma^2\sqrt{2\pi}} \sqrt{\dfrac{3-\sqrt{6}}{6}}\, e^{-(3-\sqrt{6})/2} \doteq \dfrac{0{\cdot}0918}{\sigma^2}$;
- $\sum\limits_k \left|P(X=k) - \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left(-\dfrac{(k-np)^2}{2\sigma^2}\right)\right| \sim \dfrac{1}{6\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \left|3x - x^3\right| e^{-x^2/2}\,\mathrm{d}x = \dfrac{1}{\sigma\sqrt{2\pi}}\, \dfrac{1 + 4\,e^{-3/2}}{3} \doteq \dfrac{0{\cdot}252}{\sigma}$.

Asimptotična meja med smotrnostjo uporabe Poissonove in Laplaceove aproksimacije bo torej:

- če gledamo maksimalno absolutno napako: pri $p = \left(\frac{2(3-\sqrt{6})}{3}\, e^{-(3-\sqrt{6})}\right)^{1/3} n^{-1/3} \doteq 0{\cdot}596\, n^{-1/3}$;
- če gledamo vsoto absolutnih napak: pri $p = \left(\frac{e^{1/2} + 4\,e^{-1}}{6}\right)^{2/3} n^{-1/3} \doteq 0{\cdot}647\, n^{-1/3}$.

8. 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 $P(X=1)$.

9. 50-krat vržemo nepošten kovanec, pri katerem je verjetnost, da pade grb, enaka $0{\cdot}4$. Meti so neodvisni. Kolikšna je verjetnost, da bo padlo natanko 20 grbov? Točen rezultat primerjajte z rezultatoma, dobljenima po Poissonovem obrazcu in po Laplaceovi lokalni formuli.

10. Verjetnost, da uporabnik stranišča potegne vodo, je $0{\cdot}99$. Kolikšna je verjetnost, da se pri 1000 uporabah voda potegne natanko 990-krat?

Aproksimacija intervalskih verjetnosti pri binomski porazdelitvi

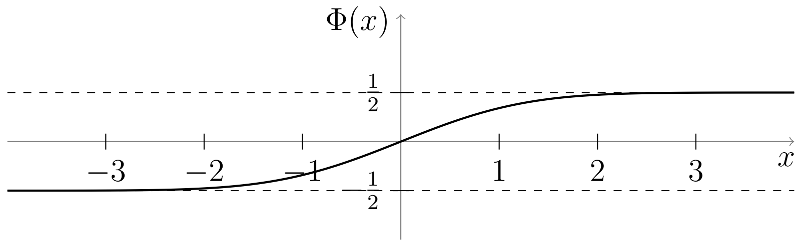
Če je $X \sim \mathrm{Bin}(n,p)$, $a \le b$, $\sigma := \sqrt{np(1-p)} \to \infty$ *in je* $|a - np| \ll \sigma^{4/3}$ **ali** $|b - np| \ll \sigma^{4/3}$ *ter še* $a, b \in \mathbb{Z} + \frac{1}{2}$ *ali* $b - a \gg 1$, *velja* **Laplaceova integralska formula***:*

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

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. Graf:



V literaturi so definicije funkcije Φ *različne, zato je treba paziti!*

11. V tovarni vsak dan proizvedejo 1600 izdelkov. Za vsakega je verjetnost, da bo okvarjen, enaka 10%. Izdelki so med seboj neodvisni.

    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˙05?

12. Zavarovalnica je proti nezgodi zavarovala 1000 oseb. Vsako od njih doleti nezgoda z verjetnostjo 0˙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.

13. Verjetnost, da je izdelek prvovrsten, je 60%. Najmanj koliko izdelkov približno moramo naročiti, če naj bo med naročenimi izdelki z verjetnostjo najmanj 0˙99 vsaj 59% izdelkov prvovrstnih? Seveda privzamemo, da so posamezni izdelki med seboj neodvisni.

14. Verjetnost, da je izdelek prvovrsten, je 10%. Najmanj koliko izdelkov približno moramo naročiti, če naj bo med naročenimi izdelki z verjetnostjo najmanj 0˙95 vsaj 100 izdelkov prvovrstnih?

15. 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.

Geometrijska porazdelitev *je porazdelitev na* $\mathbb{N}$, *pri kateri točkaste verjetnosti tvorijo geometrijsko zaporedje. Natančneje, zapis* $X \sim \mathrm{Geo}(p)$, *kjer je* $0 < p \leq 1$, *pomeni:*

$$P(X = k) = p(1-p)^{k-1}\,; \quad k \in \mathbb{N}\,.$$

Geometrijska porazdelitev je tudi porazdelitev števila poskusov do vključno prvega uspelega, če izvajamo Bernoullijevo zaporedje poskusov, pri katerih vsak uspe z verjetnostjo p.

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

Pascalova (negativna binomska) porazdelitev $\mathrm{NegBin}(n,p)$ *je porazdelitev števila poskusov do vključno* n*-tega uspelega, če izvajamo Bernoullijevo zaporedje poskusov, pri katerih vsak uspe z verjetnostjo* p. *Če je* $X \sim \mathrm{NegBin}(n,p)$, *velja:*

$$P(X = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}\,.$$

17. 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.

    Kaj pa, če kovanec ni pošten?

18. 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?

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

Hipergeometrijska porazdelitev

Iz posode, v kateri je n kroglic, od tega r rdečih, na slepo in brez vračanja izvlečemo s kroglic. Če z X označimo število rdečih med izvlečenimi, ima ta slučajna spremenljivka hipergeometrijsko porazdelitev: $X \sim \mathrm{Hip}(s, r, n) = \mathrm{Hip}(r, s, n)$. *Velja:*

$$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{s-k}}{\binom{n}{s}} = \frac{\binom{s}{k}\binom{n-s}{r-k}}{\binom{n}{r}} \; ; \quad k = 0, 1, 2, \ldots$$

20. Danih je 12 praznih škatel. Mimo pride Janezek, na slepo izbere tri škatle in v vsako vrže po eno kroglico. Mimo pride še Marička, na slepo (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. Kaj pa porazdelitev števila škatel, v katerih ni nobene kroglice?

21. Na nekem izpitu dobi študent dve na slepo izbrani vprašanji izmed 10 možnih. Študent se je učil le polovico vseh vprašanj. Vendar pa na vsako vprašanje, ki se ga ni učil, z verjetnostjo 20% ugane odgovor. Glede tega so vprašanja neodvisna, prav tako je študentova zmožnost ugibanja odgovorov neodvisna od izbire izpitnih vprašanj.

    Slučajna spremenljivka U naj pove število vprašanj, ki se jih študent ni učil, je pa uganil odgovor. Zapišite njeno porazdelitev numerično na 4 decimalke natančno.

22. Slučajna spremenljivka X ima diskretno porazdelitev z vrednostmi v množici $\{1, 2, \ldots 10\}$. Verjetnost, da je X enaka vrednosti iz te množice, je premo sorazmerna s to vrednostjo. Izračunajte $P(X > 3)$.

23. 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 in določite kumulativno porazdelitveno funkcijo $F_X(x)$. Izračunajte še $P(1 < X < 2)$.

Eksponentna porazdelitev *je zvezna porazdelitev, skoncentrirana na intervalu* $[0, \infty)$ *in katere gostota na tem intervalu je eksponentna funkcija. Natančneje, porazdelitev* $\mathrm{Exp}(\lambda)$ *ima gostoto:*

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & ; \; x > 0 \\ 0 & ; \; \textit{sicer} \end{cases}$$

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

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

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

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

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

27. Naj bo $a \in \mathbb{R}$ in $\lambda > 0$. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Dokažite, da je slučajna spremenljivka $Y = (X + a)^2$ zvezno porazdeljena, in zapišite njeno porazdelitveno gostoto.

28. Na razpolago imamo generator slučajnih števil, ki generira enakomerno porazdelitev $\mathrm{E_z}(0,1)$. Kako bi generirali porazdelitev slučajne spremenljivke X iz 24. naloge? Kaj pa eksponentno porazdelitev $\mathrm{Exp}(\lambda)$?

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

$$P(X < q_\alpha) \leq \alpha, \;\; P(X \leq q_\alpha) \geq \alpha\,.$$

Kvantilu za verjetnost 1/2 pravimo **mediana**, *kvantiloma za verjetnosti 1/3 in 2/3 pravimo prvi in drugi* **tercil**, *kvantili za verjetnosti 1/4, 2/4 in 3/4 so* **kvartili**, *kvantili za verjetnosti $0{\cdot}1, 0{\cdot}2, \ldots 0{\cdot}9$ so* **decili**, *kvantili za verjetnosti $0{\cdot}01, 0{\cdot}02, \ldots 0{\cdot}09$ pa so* **centili** *ali* **percentili**.

29. Slučajna spremenljivka X je porazdeljena diskretno po shemi:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0{\cdot}05 & 0{\cdot}15 & 0{\cdot}15 & 0{\cdot}15 & 0{\cdot}15 & 0{\cdot}35 \end{pmatrix}.$$

Določite $q_{0.3}$ in $q_{0.5}$.

Če je X zvezno porazdeljena in je q_α kvantil za verjetnost α, velja kar:

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

Če ima X v okolici točke q_α strogo pozitivno gostoto, je q_α edini kvantil za verjetnost α. Brž ko je torej gostota na nekem intervalu strogo pozitivna, izven tega intervala pa enaka nič, so kvantili za vse verjetnosti iz $(0,1)$ natančno določeni.

Mediana $m = q_{1/2}$ je mera centralne tendence, **semiinterkvartilni razmik**:

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

pa je ena od mer razpršenosti. 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.

30. Določite vse kvantile slučajne spremenljivke, porazdeljene zvezno z gostoto:

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

Posebej izračunajte še njeno mediano in semiinterkvartilni razmik.

31. Naj bo X slučajna spremenljivka s kumulativno porazdelitveno funkcijo F.

a) Naj bo q kvantil slučajne spremenljivke X za verjetnost p. Dokažite, da za vsak x veljajo naslednje implikacije:

$$\begin{aligned} x < q &\Longrightarrow F(x) \le p\,, & \qquad F(x) < p &\Longrightarrow x \ge p\,, \\ x > q &\Longrightarrow F(x) \ge p\,, & \qquad F(x) > p &\Longrightarrow x \le p\,. \end{aligned}$$

Ali lahko katero od teh implikacij še okrepimo, tako da na levi strani dodamo enačaj in/ali ga na desni odvzamemo?

b) Naj bo $Q\colon (0,1) \to \mathbb{R}$ *kvantilna funkcija* slučajne spremenljivke X, t. j. za vsak $p \in (0,1)$ naj bo $Q(p)$ kvantil slučajne spremenljivke X za verjetnost p. Dokažite, da je Q (ne nujno strogo) naraščajoča.

c) Naj bo spet $Q\colon (0,1) \to \mathbb{R}$ kvantilna funkcija slučajne spremenljivke X. Nadalje naj bo slučajna spremenljivka U porazdeljena zvezno enakomerno na intervalu $(0,1)$. Dokažite, da ima slučajna spremenljivka $Q(U)$ enako porazdelitev kot X. To je torej splošni recept za generiranje porazdelitev.

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* $P(X = \mu) = 1$*).*
Parametru μ *pravimo* **pričakovana vrednost**, *parametru* σ *pa* **standardni odklon**.
Standardna normalna porazdelitev $\mathrm{N}(0, 1)$ *ima potemtakem gostoto:*

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

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

32. Slučajna spremenljivka Z je porazdeljena standardno normalno. Izračunajte $P(Z < 1{\cdot}5)$.

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

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

33. 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:*

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

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

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

kjer je $q = 1 - p$*.*

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

Porazdelitev gama, *ki jo bomo označevali z* $\mathrm{Gama}(a, \lambda)$, *je zvezna porazdelitev z gostoto:*

$$p(x) = \begin{cases} \dfrac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} & ;\ x > 0 \\ 0 & ;\ \textit{sicer} \end{cases} .$$

Poseben primer te porazdelitve je eksponentna porazdelitev $\mathrm{Exp}(\lambda) = \mathrm{Gama}(1, \lambda)$.

35. Če je $X \sim \mathrm{Gama}(a, \lambda)$ in $k > 0$, določite porazdelitev slučajne spremenljivke $Y = kX$.

36. Slučajna spremenljivka X je porazdeljena standardno normalno. Določite porazdelitev slučajne spremenljivke $Y = e^X$.

Naj bo X zvezno porazdeljena slučajna spremenljivka z gostoto p_X, skoncentrirana na dovolj lepi množici A. Če je $h\colon A \to \mathbb{R}$ dovolj lepa funkcija in:

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

je slučajna spremenljivka Y porazdeljena zvezno z gostoto:

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

37. Slučajna spremenljivka X je porazdeljena standardno normalno. Določite porazdelitvi slučajnih spremenljivk $Y = X^2$ in $Z = (X-1)^2$.

5. Slučajni vektorji

Skupne (navzkrižne) in robne porazdelitve. Neodvisnost slučajnih spremenljivk. Transformacije slučajnih vektorjev.

Diskretni slučajni vektorji

- *Porazdelitev slučajnega vektorja* (X, Y) *podamo s* $P(X = x, Y = y)$ *(***skupna** *ali* **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* $P(X = x, Y = y) = P(X = x)\,P(Y = y)$.

1. 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 $R + M$.

2. Slučajni spremenljivki R in M sta neodvisni in porazdeljeni hipergeometrijsko: $R \sim \mathrm{Hip}(3, 3, 10)$, $M \sim \mathrm{Hip}(3, 2, 10)$. Zapišite porazdelitev slučajne spremenljivke $R+M$.

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

	$Y = 1$	$Y = 2$
$X = 1$	$\frac{5}{6} - \frac{1}{6}a - \frac{1}{3}a^2$	$\frac{1}{3}a^2$
$X = 2$	$\frac{1}{6}a$	$\frac{1}{6}$

   a) Za katere vrednosti parametra a je z zgornjo shemo določena porazdelitev slučajnega vektorja?
   b) Pri katerih vrednostih parametra a sta X in Y enako porazdeljeni?
   c) Pri katerih vrednostih parametra a sta X in Y skoraj gotovo enaki?
   d) Pri katerih vrednostih parametra a sta X in Y neodvisni?

4. 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 $Y - X$.

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

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

Posledica. Če sta $S \sim \mathrm{Bin}(m, p)$ in $T \sim \mathrm{Bin}(n, p)$ neodvisni slučajni spremenljivki, je $U := S + T \sim \mathrm{Bin}(m + n, p)$.

Poissonova porazdelitev

Slučajna spremenljivka X ima Poissonovo porazdelitev, kar označimo z $X \sim \mathrm{Poi}(\lambda)$, če velja:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}\,; \quad k = 0, 1, 2, \ldots$$

Poissonova porazdelitev je torej limita binomske porazdelitve $\mathrm{Bin}(n, p)$, ko gre $n \to \infty$ in $np \to \lambda$.

6. Naj bosta $S \sim \mathrm{Poi}(\lambda)$ in $T \sim \mathrm{Poi}(\mu)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $U := S + T$?

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

Posledica. Če sta $S \sim \mathrm{NegBin}(m, p)$ in $T \sim \mathrm{NegBin}(n, p)$ neodvisni slučajni spremenljivki, je $U := S + T \sim \mathrm{NegBin}(m + n, p)$.

8. Ministrstvo dobi vsak dan na mizo slučajno število prošenj, ki je porazdeljeno po Poissonovem zakonu $\mathrm{Poi}(\lambda)$. 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?

9. 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 $\mathrm{Poi}(1)$ 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 (za $a \le b$ in $c \le d$):*

$$P\big(a \le X \le b, c \le Y \le d\big) = \int_a^b \int_c^d p_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x$$

*Splošneje, porazdelitev zveznega slučajnega vektorja X z vrednostmi v $\mathbb{R}^n$ lahko opišemo z n-***razsežno gostoto** *p_X, ki ima to lastnost, da za vsako merljivo množico $A \subseteq \mathbb{R}^n$ velja:*

$$P(X \in A) = \int_A p_X(x)\,\mathrm{d}x\,.$$

Seveda velja:

$$\int_{\mathbb{R}^n} p_X(x)\,\mathrm{d}x = 1\,.$$

Če sta X in Y slučajna vektorja z vrednostmi v $\mathbb{R}^m$ in $\mathbb{R}^n$ in je slučajni vektor (X, Y) porazdeljen zvezno z $(m+n)$-razsežno **skupno (navzkrižno)** *gostoto $p_{X,Y}$, sta tudi njegovi komponenti X in Y porazdeljeni zvezno, in sicer z* **robnima gostotama***:*

$$p_X(x) = \int_{\mathbb{R}^n} p_{X,Y}(x,y)\,\mathrm{d}y\,, \qquad p_Y(y) = \int_{\mathbb{R}^m} p_{X,Y}(x,y)\,\mathrm{d}x\,.$$

Zvezno porazdeljena slučajna vektorja X in Y sta neodvisna natanko tedaj, ko je tudi slučajni vektor (X, Y) porazdeljen zvezno z gostoto:

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

10. 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 $P(2 < Y < 3)$, $P(Y < 3X)$ in $P(X < 3Y)$.

Večrazsežna normalna porazdelitev *s* **pričakovano vrednostjo** $\boldsymbol{\mu}$ *in* **kovariančno matriko** $\boldsymbol{\Sigma}$, *kjer je* $\boldsymbol{\mu}$ *vektor v* $\mathbb{R}^n$, $\boldsymbol{\Sigma}$ *pa je pozitivno semidefinitna matrika v* $\mathbb{R}^{n\times n}$, *je določena z naslednjimi lastnostmi:*

- *Komponente* **standardne n-razsežne normalne porazdelitve**, *kjer je* $\boldsymbol{\mu} = \mathbf{0}$ *in* $\boldsymbol{\Sigma} = \mathbf{I}_n$ *(identiteta na* $\mathbb{R}^n$*), so neodvisne in porazdeljene (enorazsežno) standardno normalno.*
- *Če je* $\boldsymbol{X}$ *porazdeljen večrazsežno normalno s pričakovano vrednostjo* $\boldsymbol{\mu}$ *in kovariančno matriko* $\boldsymbol{\Sigma}$, $\mathbf{A}$ *matrika v* $\mathbb{R}^{m\times n}$ *in* $\boldsymbol{b} \in \mathbb{R}^m$, *je tudi* $\mathbf{A}\boldsymbol{X}+\boldsymbol{b}$ *porazdeljena večrazsežno normalno, in sicer s pričakovano vrednostjo* $\mathbf{A}\boldsymbol{\mu}+\boldsymbol{b}$ *in kovariančno matriko* $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}}$. *Med drugim to pomeni, da je v primeru, ko je bločni vektor* $\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}$ *porazdeljen večrazsežno normalno s pričakovano vrednostjo* $\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}$ *in kovariančno matriko* $\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$, *njegova komponenta* $\boldsymbol{X}_1$ *porazdeljena večrazsežno normalno s pričakovano vrednostjo* $\boldsymbol{\mu}$ *in kovariančno matriko* $\boldsymbol{\Sigma}_{11}$.

Če je $\boldsymbol{\Sigma}$ *pozitivno definitna, ima večrazsežna normalna porazdelitev s pričakovano vrednostjo* $\boldsymbol{\mu}$ *in kovariančno matriko* $\boldsymbol{\Sigma}$ *gostoto:*

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}}\, e^{-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2}\,.$$

V enorazsežnem primeru je torej to porazdelitev $\mathrm{N}\big(\mu, \sqrt{\Sigma}\big)$ *(vektor* $\boldsymbol{\mu}$ *in matrika* $\boldsymbol{\Sigma}$ *sta skalarja* μ *in* Σ*).*

Če je $\begin{bmatrix}\boldsymbol{X}_1\\ \vdots\\ \boldsymbol{X}_n\end{bmatrix}$ *večrazsežni normalni bločni slučajni vektor s pričakovano vrednostjo* $\begin{bmatrix}\boldsymbol{\mu}_1\\ \vdots\\ \boldsymbol{\mu}_n\end{bmatrix}$ *in kovariančno matriko* $\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \cdots & \boldsymbol{\Sigma}_{1n}\\ \vdots & \ddots & \vdots\\ \boldsymbol{\Sigma}_{n1} & \cdots & \boldsymbol{\Sigma}_{nn}\end{bmatrix}$, *so komponente* $\boldsymbol{X}_1, \ldots \boldsymbol{X}_n$ *neodvisne natanko tedaj, ko za poljubna* $i \neq j$ *velja* $\boldsymbol{\Sigma}_{ij} = 0$.

11. Dana naj bosta neodvisna n-razsežna slučajna vektorja $\boldsymbol{X}_1$ in $\boldsymbol{X}_2$, pri čemer naj ima $\boldsymbol{X}_1$ pričakovano vrednost $\boldsymbol{\mu}_1$ in kovariančno matriko $\boldsymbol{\Sigma}_1$, $\boldsymbol{X}_2$ pa naj ima pričakovano vrednost $\boldsymbol{\mu}_2$ in kovariančno matriko $\boldsymbol{\Sigma}_2$. Kako je porazdeljena njuna vsota?

12. 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 $\mathrm{N}(7{:}01, 4\,\mathrm{min})$, neodvisno od avtobusa. Kolikšna je verjetnost, da še ujamem avtobus?

13. Naj bo $\begin{bmatrix}X_1\\ X_2\end{bmatrix}$ slučajni vektor, porazdeljen dvorazsežno normalno s pričakovano vrednostjo $\begin{bmatrix}0\\ 0\end{bmatrix}$ in kovariančno matriko $\begin{bmatrix}8 & 2\\ 2 & 5\end{bmatrix}$. Izračunajte $P(X_1 > 0, X_2 > 0)$.

14. Naj bosta $\boldsymbol{X}_1$ in $\boldsymbol{X}_2$ slučajna vektorja z vrednostmi v $\mathbb{R}^m$ in $\mathbb{R}^n$, pri čemer naj bo $\begin{bmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{bmatrix}$ porazdeljen $(m+n)$-razsežno normalno s pričakovano vrednostjo $\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}$ in kovariančno matriko $\begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{bmatrix}$. Dokažite, da obstaja taka matrika $\mathbf{C} \in \mathbb{R}^{n\times m}$, da je $\boldsymbol{X}_2 + \mathbf{C}\boldsymbol{X}_1$ neodvisen od $\boldsymbol{X}_1$.

Jacobijeva matrika *diferencialne preslikave* $f = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix} : \mathbb{R}^n \to \mathbb{R}^m$ *je matrika iz njenih parcialnih odvodov:*

$$\mathbf{D}f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}.$$

Jacobijeva determinanta *je determinanta Jacobijeve matrike:* $Jf = \det \mathbf{D}f$ *(v tem primeru mora biti seveda* $m = n$*).*

Naj bo $\boldsymbol{X}$ *zvezno porazdeljen slučajni vektor z zalogo vrednosti v odprti množici* $A \subseteq \mathbb{R}^n$ *in gostoto* $p_{\boldsymbol{X}}$*. Nadalje naj bo dana bijektivna zvezno diferenciabilna preslikava* $h\colon A \to B$ *z lastnostjo, da njena Jacobijeva matrika ni nikjer izrojena. Tedaj ima slučajni vektor* $\boldsymbol{Y} := h(\boldsymbol{X})$ *gostoto:*

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \begin{cases} p_{\boldsymbol{X}}\big(h^{-1}(\boldsymbol{y})\big)\big|J(h^{-1})(\boldsymbol{y})\big| & ;\ \boldsymbol{y} \in B \\ 0 & ;\ \textit{sicer} \end{cases}.$$

15. Naj bosta X in Y neodvisni standardni normalni slučajni spremenljivki in naj bodo (R, Θ) polarne koordinate slučajnega vektorja (X, Y), t. j. $R = \sqrt{X^2 + Y^2}$ in $\Theta = \arg(X, Y) \in (-\pi, \pi]$. Določite porazdelitev slučajnega vektorja (R, Θ).

Naj bo $\boldsymbol{X}$ *zvezno porazdeljen* n*-razsežen slučajni vektor z gostoto* p_X*, skoncentriran na merljivi množici* A*. Če je* $h\colon A \to \mathbb{R}^n$ *dovolj lepa vektorska funkcija in:*

$$P\big(h \text{ v } \boldsymbol{X} \text{ ni diferenciabilna ali } Jh(\boldsymbol{X}) = 0\big) = 0\,,$$

je slučajni vektor $\boldsymbol{Y}$ *porazdeljena zvezno z gostoto:*

$$p_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}\in A;\, h(\boldsymbol{x})=\boldsymbol{y}} \frac{p_{\boldsymbol{X}}(\boldsymbol{x})}{\big|Jh(\boldsymbol{x})\big|}\,.$$

16. Naj bosta X in Y neodvisni slučajni spremenljivki, pri čemer naj bo $X \sim \mathrm{N}(0,1)$ in $Y \sim \mathrm{N}(1,1)$. Določite porazdelitev slučajnega vektorja $\left(XY, \frac{Y}{X}\right)$.

Naj bosta $\boldsymbol{X}$ in $\boldsymbol{Y}$ slučajna vektorja z vrednostmi v $\mathbb{R}^m$ in $\mathbb{R}^n$. Bločni slučajni vektor $(\boldsymbol{X}, \boldsymbol{Y})$ naj bo skoncentriran na merljivi množici $A \subseteq \mathbb{R}^{m+n}$ in naj ima navzkrižno gostoto $p_{\boldsymbol{X},\boldsymbol{Y}}$. Nadalje naj bo $h\colon A \to \mathbb{R}^n$ dovolj lepa vektorska funkcija. Označimo z $J_y h$ Jacobijevo determinanto iz parcialnih odvodov vektorske funkcije $h(x,y)$ po komponentah argumenta $y \in \mathbb{R}^n$. Če je:

$$P\big(h \text{ ni diferenciabilna v } (\boldsymbol{X}, \boldsymbol{Y}) \text{ ali } J_{\boldsymbol{y}} h(\boldsymbol{X}, \boldsymbol{Y}) = 0\big) = 0\,,$$

je slučajni vektor $\boldsymbol{Z} = h(\boldsymbol{X}, \boldsymbol{Y})$ porazdeljen zvezno z gostoto:

$$p_{\boldsymbol{Z}}(\boldsymbol{z}) = \int_{\mathbb{R}^m} \sum_{\boldsymbol{y};(\boldsymbol{x},\boldsymbol{y})\in A, h(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{z}} \frac{p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})}{|J_{\boldsymbol{y}} h(\boldsymbol{x},\boldsymbol{y})|}\, \mathrm{d}\boldsymbol{x}\,.$$

17. Slučajni spremenljivki X in Y sta neodvisni in porazdeljeni eksponentno $\mathrm{Exp}(\lambda)$. Zapišite porazdelitev njune razlike $Z := X - Y$.

18. 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 $P(X < 2Y)$.

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

20. 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 $Z := X^2 + Y^2$ (v celoti, konstante C pa ni potrebno izračunati).

21. Določene pojave, ki se pojavljajo v času (npr. telefonski klici, nesreče, radioaktivni razpadi), lahko modeliramo s t. i. *homogenim Poissonovim tokom* oz. *Poissonovim procesom štetja* z intenzivnostjo λ, ki je karakteriziran z lastnostma, da je število pojavov, ki se zgodijo v katerem koli časovnem intervalu dolžine t, porazdeljeno po Poissonu $\mathrm{Poi}(\lambda t)$, in da 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.

Privzemimo le, da je za vsak $t \geq 0$ število pojavov, ki se zgodijo do vključno časa t, porazdeljeno po Poissonu $\mathrm{Poi}(\lambda t)$. Naj bo T_n slučajna spremenljivka, ki pove čas, ob katerem se zgodi n-ti pojav (čas štejemo od 0 naprej). Določite njeno porazdelitev. Kaj pride pri $n = 1$?

22. Naj bosta $S \sim \mathrm{Gama}(a, \lambda)$ in $T \sim \mathrm{Gama}(b, \lambda)$ neodvisni slučajni spremenljivki. Kako je porazdeljena slučajna spremenljivka $U := S + T$?

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

Porazdelitev hi hvadrat z n prostostnimi stopnjami, *ki jo označujemo s* $\chi^2(n)$, *je porazdelitev vsote* $Z_1^2 + \cdots + Z_n^2$, *kjer so* $Z_1, \ldots Z_n$ *neodvisne slučajne spremenljivke, porazdeljene standardno normalno. Ta porazdelitev je pomembna v statistiki, saj med drugim igra ključno vlogo pri ocenjevanju disperzije. Velja* $\chi^2(n) = \mathrm{Gama}\left(\frac{n}{2}, \frac{1}{2}\right)$.

23. Na zabavo, ki se začne ob določeni uri, je povabljenih n gostov. Vsak malo zamudi, zamuda vsakega je porazdeljena enakomerno na intervalu $[0, 1]$ in zamude so med seboj neodvisne. Slučajna spremenljivka T_k naj predstavlja čas, ko je na zabavo prišel k-ti gost po vrsti (glede na čas prihoda). Določite njeno porazdelitev.

24. Slučajni spremenljivki X in Y sta neodvisni, pri čemer je X porazdeljena normalno $\mathrm{N}(0, \sigma)$, Y pa ima porazdelitev $\mathrm{Gama}(a, \lambda)$.

    a) Določite porazdelitev slučajne spremenljivke $T = \dfrac{X}{\sqrt{Y}}$.

    b) *Studentova porazdelitev* z n prostostnimi stopnjami je porazdelitev slučajne spremenljivke:
    $$T = \frac{X}{\sqrt{\dfrac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}}}\,, \qquad (*)$$
    kjer so $X, X_1, X_2, \ldots X_n \sim \mathrm{N}(0, \sigma)$ neodvisne slučajne spremenljivke. Dokažite, da je ta porazdelitev neodvisna od σ, in zapišite njeno gostoto. Kam konvergira ta gostota, ko gre n proti neskončno?

Opomba. Studentova porazdelitev je pomembna v statistiki. Zaenkrat si lahko predstavljamo, da želimo standardizirati opažanje $X \sim \mathrm{N}(0, \sigma)$, pri čemer pa parametra σ ne poznamo; pač pa lahko σ ocenimo tako, da opažanje n-krat neodvisno ponovimo. Imenovalec v $(*)$ je *cenilka* za σ.

25. Slučajni spremenljivki X in Y sta neodvisni s porazdelitvijo gama, in sicer $X \sim \mathrm{Gama}(a, \lambda)$ in $Y \sim \mathrm{Gama}(b, \lambda)$. Določite porazdelitev slučajne spremenljivke $Z = X/(X + Y)$.

26. Slučajni spremenljivki X in Y sta neodvisni s porazdelitvijo gama: $X \sim \mathrm{Gama}(a, \lambda)$ in $Y \sim \mathrm{Gama}(b, \mu)$. Določite porazdelitev slučajne spremenljivke $Q = X/Y$.

Snedecorjeva (Fisherjeva) *porazdelitev z m in n prostostnimi stopnjami,* $\mathrm{F}(m, n)$, *je porazdelitev, dobljena na enega od naslednjih dveh ekvivalentnih načinov:*

- *kot porazdelitev kvocienta X/Y, kjer sta X in Y neodvisni ter $X \sim \mathrm{Gama}(m/2, m/2)$ in $Y \sim \mathrm{Gama}(n/2, n/2)$;*
- *kot porazdelitev kvocienta $\dfrac{U/m}{V/n}$, kjer sta U in V neodvisni ter $U \sim \chi^2(m)$ in $V \sim \chi^2(n)$.*

27. Konstruirajte funkcijsko zvezo, za katero obstajata slučajni spremenljivki $F \sim \mathrm{F}(m, n)$ in $B \sim \mathrm{Beta}(m/2, n/2)$, ki sta v tej zvezi.

6. Matematično upanje in sorodne karakteristike

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

Matematično upanje:

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

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

Disperzija (varianca): $D(X) = E\big((X - E(X))^2\big)$.
Standardni odklon: $\sigma(X) = \sqrt{D(X)}$.

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

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

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

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

***U*-metoda**

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

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

$$\begin{pmatrix} 490 & 500 & 520 \\ 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}6 \end{pmatrix}.$$

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

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

$$\begin{pmatrix} 1 & 2 & 4 & 8 & \cdots \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \cdots \end{pmatrix}.$$

Izračunajte $E(X)$ in $E\big(\sqrt{X}\big)$.

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

5. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

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

   Pokažite, da $E(X)$ ne obstaja.

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

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

$$\boxed{E(aX + b) = a\,E(X) + b}$$

Posledica. Če je $X \sim \mathrm{N}(\mu, \sigma)$, je $E(X) = \mu$ in $\sigma(X) = \sigma$. Tako se pričakovana vrednost in standardni odklon iz definicije normalne porazdelitve ujemata s pojmoma, definiranima tukaj.

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

$$p_X(x) = \frac{4}{\pi(1+4x^2)^2}$$

   Izračunajte $E(1 + 4X^2)$ in $D(X)$.

$$\boxed{\begin{aligned} E\big(h(X,Y)\big) &= \sum_x \sum_y h(x,y)\ P(X = x, Y = y) \\ E\big(h(X,Y)\big) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y)\, p_{X,Y}(x,y)\, \mathrm{d}x\, \mathrm{d}y \end{aligned}}$$

8. 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 $E(XY^2)$.

9. 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 $E\big(e^{Y-X}\big)$.

$$\boxed{E(X+Y) = E(X) + E(Y)}$$

10. 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 $E(X^2+Y^2)$.

11. Izračunajte pričakovani kot, pod katerim meteorit pade na planet. Privzemite, da ima planet obliko krogle, da so vse smeri v vesolju, iz katerih prileti meteorit, enako zastopane, pri posamezni smeri pa privzemite tudi enakomerno porazdelitev trajektorij (premic), ki sekajo planet. Prav tako zanemarite ukrivljenje trajektorij zaradi gravitacije.

Indikatorji dogodkov

Indikator *dogodka je slučajna spremenljivka, ki je na danem dogodku enaka* 1*, zunaj njega pa* 0*.*

Indikator dogodka A bomo označevali z $\mathbf{1}_A$.

Indikator dogodka, da je izjava $\mathcal{A}$ pravilna, bomo označevali z $\mathbf{1}(\mathcal{A})$.

Velja $E(\mathbf{1}_A) = P(A)$.

12. Danih je šest ploščic z naslednjimi oznakami:

$$\boxed{1} \quad \boxed{1} \quad \boxed{1} \quad \boxed{2} \quad \boxed{3} \quad \boxed{\text{S}}$$

Ploščice se naključno premešajo in obrnejo. Nato jih odkrivamo, dokler ne odkrijemo ploščice $\boxed{\text{S}}$. Označimo z X vsoto vseh odkritih števil. Izračunajte $E(X)$.

13. Urška kupuje čevlje. Obiskati namerava tri trgovine. Če v prvi ne dobi čevljev, ki so ji všeč, po ceni največ 50 evrov, gre naprej v drugo trgovino in če tam ne dobi čevljev, ki so ji všeč, po ceni največ 50 evrov, gre še v tretjo trgovino, kjer kupi čevlje v vsakem primeru.

Cena najugodnejših čevljev, ki so Urški všeč, je v prvi trgovini porazdeljena diskretno enakomerno na množici $\{36, 45, 60\}$, v drugi zvezno z gostoto:

$$p_2(x) = \begin{cases} \frac{60-x}{200} & ; 40 < x < 60 \\ 0 & ; \text{sicer} \end{cases},$$

in v tretji normalno $\mathrm{N}(54, 10)$. Te tri cene so neodvisne.

Naj bo C cena, po kateri Urška kupi čevlje. Izračunajte $E(C)$.

14. Izračunajte matematično upanje hipergeometrijske porazdelitve.

15. Izrazite indikator nasprotnega dogodka in indikator preseka dogodkov z indikatorji izvirnih dogodkov. Nato s pomočjo indikatorjev izpeljite načelo vključitev in izključitev.

16. Dokažite, da za poljubno slučajno spremenljivko N z vrednostmi v $\{0, 1, 2, \ldots\}$ velja:

$$E(N) = \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=0}^{\infty} P(N > n)$$

17. Pošten kovanec mečemo, dokler ne padeta dve cifri zapored. Meti so med seboj neodvisni. Izračunajte pričakovano število vseh metov.

18. 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 $D(S)$.

Nekoreliranost

Slučajni spremenljivki X in Y sta nekorelirani, če velja:

$$E(XY) = E(X)\, E(Y)\,.$$

Poljubni neodvisni slučajni spremenljivki sta nekorelirani, obratno pa ni nujno res: nekoreliranost še ne pomeni neodvisnosti.

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

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

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 g in h velja, da sta slučajni spremenljivki $g(X)$ in $h(Y)$ nekorelirani.*

20. Diskretno porazdeljen slučajni vektor (X, Y) ima naslednjo porazdelitveno tabelo:

	$Y=-1$	$Y=0$	$Y=1$
$X=-1$	$\frac{1}{6}-\frac{t^2}{18}$	$\frac{t}{9}$	$\frac{t^2}{18}-\frac{t}{9}+\frac{1}{6}$
$X=0$	$\frac{t}{9}$	$\frac{1}{3}-\frac{2t}{9}$	$\frac{t}{9}$
$X=1$	$\frac{t^2}{18}-\frac{t}{9}+\frac{1}{6}$	$\frac{t}{9}$	$\frac{1}{6}-\frac{t^2}{18}$

a) Pri katerih t je z zgornjo tabelo res določena porazdelitev slučajnega vektorja?
b) Za katere vrednosti t sta X in Y nekorelirani?
c) Za katere vrednosti t sta X in Y neodvisni?

Kovarianca:

$$K(X,Y) := E\big((X - E(X))\,(Y - E(Y))\big) = \\ = E(XY) - E(X)\,E(Y)$$

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 $K(aX+bY, Z) = a\,K(X,Z) + b\,K(Y,Z)$. Slučajni spremenljivki X in Y sta nekorelirani natanko tedaj, ko je $K(X,Y) = 0$.

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 $r(aX+b, cY+d) = r(X,Y)$.

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

	$Y=0$	$Y=10$	$Y=30$
$X=0$	0˙05	0˙1	0˙2
$X=10$	0˙15	0	0˙15
$X=40$	0˙2	0˙1	0˙05

Izračunajte $K(X,Y)$ in $r(X,Y)$.

22. 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 $r(X,Y)$.

23. Za okroglo mizo izmenoma sedijo štiri ženske in štirje moški. Vsak vrže kocko, meti so neodvisni.

    a) Za i-to žensko naj bo A_i dogodek, da vrže šestico, obenem pa nobeden od njenih sosedov ne vrže šestice. Nadalje naj bo za j-tega moškega B_j dogodek, da vrže enojko, obenem pa nobena od njegovih sosed ne vrže enojke. Za vse i in j izračunajte verjetnosti $P(A_i \cap A_j)$, $P(B_i \cap B_j)$ in $P(A_i \cap B_j)$.

    b) Naj bo X število žensk, za katere velja, da vržejo šestico, obenem pa nobeden od njenih sosedov ne vrže šestice. Nadalje naj bo Y število moških, ki vržejo enojko, obenem pa velja, da nobena od njenih sosed ne vrže enojke. Izračunajte $r(X, Y)$.

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

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

Brž ko sta X in Y nekorelirani in imata disperzijo, velja:

$$D(X + Y) = D(X) + D(Y)\,.$$

Pozor! *Zakaj ne velja:*

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

24. Izračunajte matematično upanje in disperzijo binomske porazdelitve.

25. Izračunajte matematično upanje in disperzijo Poissonove porazdelitve.

26. Izračunajte matematično upanje in disperzijo porazdelitve gama.

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

28. Izračunajte matematično upanje in disperzijo geometrijske in Pascalove porazdelitve.

Matematično upanje slučajnega vektorja $\boldsymbol{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ *definiramo po komponentah:*

$$\boldsymbol{E}(\boldsymbol{X}) := \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix},$$

disperziji pa ustreza **kovariančna matrika***:*

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

Za poljubno deterministično matriko $\mathbf{A}$ *velja* $\mathbf{K}(\mathbf{A}\boldsymbol{X}) = \mathbf{A}\,\mathbf{K}(\boldsymbol{X})\,\mathbf{A}^T$.
Za poljubna deterministična vektorja $\boldsymbol{u}$ *in* $\boldsymbol{v}$ *velja*
$K\big(\langle \boldsymbol{X}, \boldsymbol{u}\rangle, \langle \boldsymbol{X}, \boldsymbol{b}\rangle\big) = K\big(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{X}, \boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}\big) = \boldsymbol{u}^{\mathrm{T}}\,\mathbf{K}(\boldsymbol{X})\boldsymbol{v}$.

29. 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.

30. Dokažite, da se pričakovana vrednost in kovariančna matrika iz definicije večrazsežne normalne porazdelitve ujemata s pričakovano vrednostjo in kovariančno matriko, definirano v tem razdelku.

31. Slučajni vektor $\boldsymbol{X}$ ima naslednjo kovariančno matriko:

$$\mathbf{K} = \begin{bmatrix} 5 & 2 & 4 \\ 2 & 2 & 2 \\ 4 & 2 & 5 \end{bmatrix}.$$

V katerih smereh je disperzija projekcije največja in v katerih najmanjša in koliko znaša? Z drugimi besedami, za katere enotske vektorje $\boldsymbol{u}$ je disperzija slučajne spremenljivke $\langle \boldsymbol{X}, \boldsymbol{u}\rangle$ največja in za katere najmanjša?

7. Pogojne porazdelitve in pogojevanje na slučajne spremenljivke

Pogojna porazdelitev diskretne slučajne spremenljivke glede na dogodek in glede na diskretno slučajno spremenljivko. Pogojna verjetnost dogodka glede na poljubno slučajno spremenljivko. Pogojna gostota. Pogojno matematično upanje.

Pogojno porazdelitev *slučajne spremenljivke X glede na dogodek B opišemo s pogojnimi verjetnostmi $P(X \in C \mid B)$, kjer C preteče vse merljive množice. Če je X diskretna, lahko njeno pogojno porazdelitev opišemo s* **pogojno porazdelitveno shemo**:

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ P(X = a_1 \mid B) & P(X = a_2 \mid B) & \cdots \end{pmatrix}.$$

1. Vržemo tri poštene in neodvisne kovance in jih razporedimo v vrsto. Naj bo X število grbov pri prvih dveh kovancih. Nato pride Pepček, na slepo izbere dva različna kovanca in ju zamenja. Zdaj je na obeh prvih kovancih grb. Zapišite pogojno porazdelitev slučajne spremenljivke X glede na omenjeno opažanje.

2. Slučajna spremenljivka X je porazdeljena geometrijsko $\mathrm{Geo}(1/3)$. Določite njeno pogojno porazdelitev glede na dogodek $\{X < 5\}$.

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

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

Določite pogojno porazdelitev slučajne spremenljivke X glede na $Y = 2$ ter Y glede na $X = 0$ in glede na $Y \geq X$.

Za vsako realno slučajno spremenljivko in vsak dogodek B s pozitivno verjetnostjo lahko pogojno porazdelitev slučajne spremenljivke X glede na B opišemo s **pogojno kumulativno porazdelitveno funkcijo**:

$$F_{X|B}(x) = P(X \le x \mid B)\,.$$

Če je pogojna porazdelitev zvezna, obstaja tudi **pogojna porazdelitvena gostota***:*

$$p_{X|B}(x) = F'_{X|B}(x)\,.$$

Brž ko je X zvezno porazdeljena, je tudi njena pogojna porazdelitev zvezna – glede na vsak dogodek s pozitivno verjetnostjo.
Če je X porazdeljena zvezna z gostoto p_X in $P(X \in C) > 0$, je:

$$p_{X|X\in C}(x) = \begin{cases} \dfrac{p_X(x)}{P(X \in C)} & ;\ x \in C \\ 0 & ;\ \textit{sicer} \end{cases}\,.$$

Podobno velja tudi za zvezne slučajne vektorje.

4. Slučajna spremenljivka X je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Za vsak $a \ge 0$ določite pogojno porazdelitev slučajne spremenljivke X glede na dogodek $\{X \ge a\}$.

5. Slučajni spremenljivki X in Y sta neodvisni in porazdeljeni enakomerno na intervalu $[0,1]$. Določite pogojno porazdelitev slučajne spremenljivke X glede na dogodek $\{X < Y\}$.

6. Računalnik na slepo izbere realno število X med 0 in 1. Tine ga pogleda in Tone ga z verjetnostjo 1/2 vpraša, ali je to število manjše od 2/3, z verjetnostjo 1/2 pa, ali je večje od 1/3 (izbira vprašanja je neodvisna od izbranega števila). Odgovor je pritrdilen. Zapišite pogojno porazdelitveno gostoto izbranega števila X glede na dani odgovor (kaj točno je Tone vprašal Tineta, ne vemo).

Pogojno matematično upanje *slučajne spremenljivke X glede na dogodek B je matematično upanje, ki pripada ustrezni pogojni porazdelitvi, in ga označimo z $E(X \mid B)$. Tako velja:*

$$E(X \mid B) = \sum_x x\, P(X = x \mid B)$$

in splošneje:

$$E\big(h(X) \mid B\big) = \sum_x h(x)\, P(X = x \mid B)\,.$$

Pogojno matematično upanje ima vse lastnosti običajnega matematičnega upanja, npr. linearnost.
Podobno definiramo tudi **pogojno disperzijo**. *Velja:*

$$D(X \mid B) = E\Big[\big(X - E(X \mid B)\big)^2 \;\Big|\; B\Big] = E\big(X^2 \mid B\big) - \big(E(X \mid B)\big)^2\,.$$

7. Za slučajno spremenljivko X iz 1. naloge in dogodek B, da je po Pepčkovi zamenjavi na obeh prvih kovancih grb, izračunajte $E(X \mid B)$ in $D(X \mid B)$. Prav tako izračunajte $E(X \mid D)$ in $D(X \mid D)$ za slučajno spremenljivko X iz 6. naloge in dogodek D, da Tine odgovori pritrdilno.

Za vsako slučajno spremenljivko X in vsak dogodek B velja:

$$E(X \mid B) = \frac{E(XZ)}{P(B)} = \frac{E(XZ)}{E(Z)}\,,$$

kjer je slučajna spremenljivka Z **indikator** *dogodka B, t. j. enaka 1 na dogodku B in 0 zunaj njega.*

8. Naj bosta X in Y neodvisni slučajni spremenljivki, pri čemer naj bo X porazdeljena enakomerno na $[0, 1]$, Y pa enakomerno na $\{0, 1, 2\}$. Izračunajte $E(XY \mid 2X > Y)$.

Za vsako slučajno spremenljivko X z matematičnim upanjem in vsak popoln sistem dogodkov $H_1, H_2, H_3, \ldots$ velja **izrek o polnem matematičnem upanju**:

$$E(X) = P(H_1)\, E(X \mid H_1) + P(H_2)\, E(X \mid H_2) + P(H_3)\, E(X \mid H_3) + \cdots$$

9. V prvi posodi je 6 belih in 4 rdeče kroglice, v drugi pa 3 bele in 6 rdečih. Najprej na slepo premestimo eno kroglico iz prve posode v drugo, nato pa iz druge posode na slepo in brez vračanja izvlečemo tri kroglice. Izračunajte pričakovano število belih med njimi.

10. Pošten kovanec mečemo, dokler ne padeta dve cifri zapored. Meti so med seboj neodvisni. Izračunajte pričakovano število vseh metov.

Pogojno verjetnost *dogodka* A **glede na diskretno slučajno spremenljivko** Y *lahko definiramo bodisi kot funkcijo, ki* y *slika v* $P(A \mid Y = y)$*, bodisi kot slučajno spremenljivko, ki jo označimo s* $P(A \mid Y)$*: na dogodku* $\{Y = y\}$ *s pozitivno verjetnostjo definiramo:*

$$P(A \mid Y) := P(A \mid Y = y)\,,$$

na dogodku $\{Y = y\}$ *z verjetnostjo nič pa vrednost izberemo poljubno, a konstantno na celem dogodku. Tako dobimo slučajno spremenljivko, ki je funkcija slučajne spremenljivke* Y*, določena pa je* **skoraj gotovo**: *poljubni izbiri se z verjetnostjo ena ujemata.*
Tako definirana slučajna spremenljivka je odvisna samo od informacije, ki jo nudi Y *(natančneje* σ*-algebre, generirane z* Y*): če je* g *merljiva bijektivna funkcija, je* $P(A \mid g(Y)) = P(A \mid Y)$ *(natančneje, vsaka izbira, ki je dobra za levo stran, je dobra tudi za desno stran).*

11. Pošteno kocko mečemo, dokler ne pade šestica, meti so neodvisni. Izračunajte pogojno verjetnost, da pri tem ne pade nobena enojka, glede na število vseh metov.

12. Na kupu je 16 dobro premešanih kart, in sicer po štirje asi, kralji, dame in fanti. Naj bo A dogodek, da je prva karta as, z Y pa označimo število kraljev med prvimi štirimi kartami.

    a) Določite pogojno verjetnost dogodka A glede na Y.
    b) Izračunajte $E\big(P(A \mid Y)\big)$. Ali kaj opazite?
    c) Dokažite, da za *vsako* diskretno slučajno spremenljivko Y in *vsak* dogodek A velja:
    $$E\big(P(A \mid Y)\big) = P(A)\,.$$

13. Pošteno kocko spet mečemo, dokler ne pade šestica, meti so neodvisni. Kolikšna je verjetnost, da pri tem ne pade nobena enojka?

Pogojno porazdelitev *slučajne spremenljivke X glede na diskretno slučajno spremenljivko Y opišemo s pogojnimi verjetnostmi $P(X \in C \mid Y)$, kjer C preteče vse merljive množice. Tako dobimo slučajno verjetnostno mero. Če je X diskretna, lahko seveda njeno pogojno porazdelitev opišemo s pogojno porazdelitveno shemo:*

$$\begin{pmatrix} a_1 & a_2 & \cdots \\ P(X = a_1 \mid Y) & P(X = a_2 \mid Y) & \cdots \end{pmatrix}.$$

Za poljubno realno slučajno spremenljivko X lahko definiramo pogojno kumulativno porazdelitveno funkcijo:

$$F_{X|Y}(x) = P(X \le x \mid Y) \quad \text{ali tudi} \quad F_{X|Y}(x \mid y) = P(X \le x \mid Y = y)\,,$$

za zvezno porazdeljene pa pogojno gostoto: $p_{X|Y}(x)$ je slučajna funkcija spremenljivke x, $p_{X|Y}(x \mid y)$ pa je (deterministična) funkcija spremenljivk x in y. Lahko definiramo tudi pogojno matematično upanje:

$$E(X \mid Y) = g(Y)\,, \quad \text{kjer je} \quad g(y) = E(X \mid Y = y)\,.$$

Funkciji g pravimo **regresijska funkcija**. *Prav tako lahko definiramo pogojno disperzijo $D(X \mid Y)$ in podobno.*

14. Iz posode, v kateri so najprej ena rdeča, dve zeleni in sedem belih kroglic, na slepo in brez vračanja vlečemo kroglice, dokler ne izvlečemo rdeče. Naj bo X število zelenih, Y pa število vseh izvlečenih kroglic.

    a) Zapišite pogojno porazdelitev slučajne spremenljivke X glede na Y.

    b) Izračunajte $E(X \mid Y)$ in $E\big(E(X \mid Y)\big)$. Ali kaj opazite?

    c) Dokažite, da za vsako slučajno spremenljivko X in diskretno slučajno spremenljivko Y velja $E\big(E(X \mid Y)\big) = E(X)$.

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

16. Naj bosta X in Y slučajni spremenljivki, pri čemer naj bo Y porazdeljena binomsko $\mathrm{Bin}(2, 1/2)$ ter še $E(X \mid Y) = Y$ in $D(X \mid Y) = Y + 1$. Izračunajte $E(X)$ in $D(X)$.

17. Pošteno kocko spet mečemo, dokler ne pade šestica, meti so neodvisni.

    a) Določite *pogojno* porazdelitev števila vseh enojk glede na število vseh metov.

    b) Določite *brezpogojno* porazdelitev števila vseh enojk.

18. Naj bodo $X_1, X_2, \ldots$ neodvisne enako porazdeljene Bernoullijeve slučajne spremenljivke, t. j.:

$$X_i \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

in naj bo $N \sim \mathrm{Poi}(\lambda)$ neodvisna od slučajnih spremenljivk X_i. Označimo:

$$S = X_1 + X_2 + \cdots + X_N \,.$$

a) Določite *pogojno* porazdelitev slučajne spremenljivke S glede na N.
b) Določite *brezpogojno* porazdelitev slučajne spremenljivke S.
c) Dokažite, da sta slučajni spremenljivki S in $T := N - S$ neodvisni.
d) Dokažite, da, če spremenimo porazdelitev slučajne spremenljivke N, ni več nujno, da sta S in T neodvisni.

Opomba. Transformaciji, pri katerih iz slučajne spremenlivke N nastane slučajna spremenljivka S, pravimo *redčenje* (angl. *thinning*).

Če je Y porazdeljena diskretno, X pa pogojno na Y porazdeljena zvezno z gostoto $p_{X|Y}$, je tudi brezpogojna porazdelitev slučajne spremenljivke X zvezna z gostoto:

$$p_X(x) = E\big[p_{X|Y}(x)\big] \,.$$

19. Naj bodo $X_1, X_2, \ldots$ neodvisne slučajne spremenljivke, porazdeljene eksponentno $\mathrm{Exp}(\lambda)$, N pa naj bo neodvisna od prej omenjenih slučajnih spremenljivk in porazdeljena geometrijsko $\mathrm{Geo}(p)$. Zapišite porazdelitveno gostoto slučajne spremenljivke:

$$S = X_1 + X_2 + \cdots + X_N \,.$$

Če sta X in Y slučajni spremenljivki in $P(Y = y) > 0$, se pogojna porazdelitev slučajne spremenljivke $h(X, Y)$ glede na $Y = y$ ujema s pogojno porazdelitvijo slučajne spremenljivke $h(X, y)$ glede na ta dogodek. Če je X diskretna, torej velja:

$$E\big[h(X, Y) \mid Y = y\big] = \sum_x h(x, y)\, P(X = x \mid Y = y) \,.$$

20. Naj bo slučajni vektor (X, Y) porazdeljen tako kot v 3. nalogi. Izračunajte $E(X \mid Y = 2)$, $E(X^2 \mid Y = 2)$ in $E(X^2Y^2 \mid Y = 2)$.

Če je X slučajna spremenljivka z matematičnim upanjem in Y diskretna slučajna spremenljivka, velja:

$$E\big[X\,g(Y)\;\big|\;Y\big] = E(X \mid Y)\,g(Y)$$

in posledično:

$$E\big[X\,g(Y)\big] = E\big[E(X \mid Y)\,g(Y)\big]\,.$$

21. Pošteno kocko spet mečemo, dokler ne pade šestica, meti so neodvisni. Izračunajte pričakovani delež enojk med vsemi meti.

Za vsako slučajno spremenljivko X z matematičnim upanjem in **vsako** *slučajno spremenljivko Y obstaja slučajna spremenljivka Z, ki je funkcija slučajne spremenljivke Y in za katero je $E\big[Z\,g(Y)\big] = E\big[X\,g(Y)\big]$ za vsako merljivo funkcijo g, za katero desna stran obstaja. Slučajna spremenljivka Z je določena do skoraj gotovega ujemanja natančno. Slučajni spremenljivki Z pravimo* **pogojno matematično upanje glede na slučajno spremenljivko** *Y in pišemo $Z = E(X \mid Y)$. Velja tudi:*

$$E\big[X\,g(Y)\;\big|\;Y\big] = E(X \mid Y)\,g(Y)\,.$$

Regresijska funkcija *je definirana kot $E(X \mid Y = y) = h(y)$, kjer je $h(Y) = E(X \mid Y)$. Ni pa nujno natančno določena.*
Pogojna verjetnost *dogodka A glede na slučajno spremenljivko Y je definirana kot pogojno matematično upanje njegovega indikatorja Z: $P(A \mid Y) = E(Z \mid Y)$.*
Pogojna porazdelitev *slučajne spremenljivke X glede na Y je nabor pogojnih verjetnosti $P(X \in C \mid Y)$, kjer C preteče vse merljive množice. Lahko jo gledamo tudi kot slučajno preslikavo $C \mapsto P(X \in C \mid Y)$, ki pa mora biti povsod verjetnostna mera. V splošnem ni nujno, da pogojna porazdelitev obstaja, a pogojne porazdelitve realnih (in mnogih drugih) slučajnih spremenljivk vedno obstajajo.*
Na podlagi pogojne porazdelitve lahko definiramo pogojno kumulativno porazdelitveno funkcijo, pogojno gostoto, pogojno matematično upanje, pogojno disperzijo itd. Definicija pogojnega matematičnega upanja poljubne funkcije slučajne spremenljivke na podlagi pogojne porazdelitve se ujema s prvotno definicijo.

22. Naj bodo $X_1, X_2, \ldots$ in N tako kot v 19. nalogi. Zapišite pogojno porazdelitev slučajne spremenljivke N glede na S.

23. Slučajna spremenljivka U naj bo porazdeljena zvezno enakomerno na $(0,1)$ in pogojno na U naj bo slučajna spremenljivka X porazdeljena binomsko $\mathrm{Bin}(n, U)$.

Določite porazdelitev slučajne spremenljivke X in še pogojno porazdelitev slučajne spremenljivke U glede na X.

Namig: Ker je X porazdeljena diskretno, je dovolj izračunati pogojne porazdelitve slučajne spremenljivke U glede na dogodke $\{X = k\}$, kjer je $k = 0, 1, \ldots n$; le-te pa lahko dobimo na podlagi pogojnih matematičnih upanj $E\big[h(U) \;\big|\; X = k\big]$, kjer je h merljiva *testna funkcija*.

Opomba. Ta naloga sodi v *Bayesovo statistiko*: enakomerna porazdelitev je *apriorna* porazdelitev slučajne spremenljivke U, želimo pa izračunati njeno *aposteriorno* porazdelitev glede na opažanje X.

24. Za slučajni vektor (X, Y) iz 3. naloge zapišite porazdelitev slučajnih spremenljivk $E(X \mid Y)$ in $D(X \mid Y)$, nato pa izračunajte še $D\big(E(X \mid Y)\big)$ in $E\big(D(X \mid Y)\big)$. Rezultata primerjajte z $D(X)$. Kaj opazite?

25. Naj obstaja $E(X^2)$. Dokažite zvezo:

$$D(X) = D\big(E(X \mid Y)\big) + E\big(D(X \mid Y)\big) .$$

    Opomba: prvemu členu pravimo *pojasnjena*, drugemu členu pa *nepojasnjena* ali *rezidualna* disperzija. Z izrazom pojasnjena je mišljena pojasnjenost z odvisnostjo slučajne spremenljivke X od Y (t. j. slučajna spremenljivka X ima za različne vrednosti Y različne pogojne porazdelitve, z njimi pa lahko različna pogojna matematična upanja).

26. Naj bo $Y \sim \mathrm{Exp}(\lambda)$, pogojno na Y pa naj bo $X \sim \mathrm{Poi}(Y)$. Izračunajte $E(X^2)$.

27. Naj bo spet $\lambda > 0$, $Y \sim \mathrm{Exp}(\lambda)$ in pogojno na Y naj bo $X \sim \mathrm{Poi}(Y)$. Za katere $a \in \mathbb{R}$ obstaja $E(a^X)$ in koliko je to enako?

*Če je Y **poljubna** slučajna spremenljivka, X pa pogojno na Y porazdeljena zvezno z gostoto $p_{X|Y}$, je tudi brezpogojna porazdelitev slučajne spremenljivke X zvezna z gostoto $p_X(x) = E\big[p_{X|Y}(x)\big]$.*

28. Dani sta slučajni spremenljivki N in T. Slučajna spremenljivka N je porazdeljena geometrijsko $\mathrm{Geo}(p)$ in pogojno na N naj ima T porazdelitev $\mathrm{Gama}(N, \lambda)$. Določite brezpogojno porazdelitev slučajne spremenljivke T.

Slučajni vektor (X, Y) je zvezno porazdeljen natanko tedaj, ko je hkrati zvezno porazdeljena slučajna spremenljivka Y in pogojna porazdelitev slučajne spremenljivke X glede Y skoraj gotovo zvezna. Za ustrezne (pogojne) gostote tedaj velja zveza:

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

29. 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. Poiščite še kakšno netrivialno slučajno spremenljivko, neodvisno od X, in še kakšno neodvisno od Y.

30. Slučajna spremenljivka X je porazdeljena standardno normalno $\mathrm{N}(0, 1)$, pogojna porazdelitev slučajne spremenljivke Y glede na X pa je normalna $\mathrm{N}(0, 1/|X|)$. Določite porazdelitev slučajne spremenljivke Y.

31. 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 $E(X \mid Y)$ in $E(XY \mid Y)$.

Dvorazsežna normalna porazdelitev $\mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, *kjer je* $\sigma_1, \sigma_2 > 0$ *in* $-1 < \rho < 1$, *je porazdelitev z gostoto:*

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{(x-\mu_1)^2}{2(1-\rho^2)\sigma_1^2} + \frac{\rho(x-\mu_1)(y-\mu_2)}{(1-\rho^2)\sigma_1\sigma_2} - \frac{(y-\mu_2)^2}{2(1-\rho^2)\sigma_2^2}}\,.$$

Če ima slučajni vektor (X, Y) to porazdelitev, je $X \sim \mathrm{N}(\mu_1, \sigma_1)$ in $Y \sim \mathrm{N}(\mu_2, \sigma_2)$.

32. a) Naj ima slučajni vektor (X_1, X_2) dvorazsežno normalno porazdelitev $\mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Določite pogojno porazdelitev slučajne spremenljivke X_2 glede na X_1.

b) Če za rezultate prvega in drugega kolokvija iz verjetnosti in statistike za računalničarje privzamemo model z dvorazsežno normalno porazdelitvijo, iz rezultatov 963 parov kolokvijev iz let od 1997 do 2010 dobimo naslednje ocene parametrov:

$$\mu_1 \doteq 59{\cdot}2\,, \mu_2 \doteq 58{\cdot}0\,, \sigma_1 \doteq 20{\cdot}1\,, \sigma_2 \doteq 25{\cdot}3\,, \rho \doteq 0{\cdot}291\,.$$

Na podlagi privzetega modela ocenite verjetnost, da bo kandidat, ki na prvem kolokviju zbere 25 točk, kolokvije naredil, t. j. na obeh zbral skupaj vsaj 100 točk.

8. Rodovne, momentno-rodovne in karakteristične funkcije

Rodovne funkcije: osnove, konvolucijski izrek, procesi razvejanja. Momentno-rodovne funkcije, kumulante, asimetrija, sploščenost. Neenakosti. Karakteristične funkcije: osnove, povezava z rodovnimi funkcijami, konvolucijski izrek, inverzna formula.

Rodovna funkcija

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

Če je $X \ge 0$, je $G_X(s)$ definirana za $0 \le s \le 1$ ali še splošneje za vse kompleksne s z $|s| \le 1$, 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 $|s| \le 1$.
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 $\mathrm{Poi}(\lambda)$. Izračunajte še $E(X)$ in $D(X)$.

2. Izračunajte rodovno funkcijo geometrijske porazdelitve $\mathrm{Geo}(p)$ ter še matematično upanje in disperzijo.

3. Rodovna funkcija slučajne spremenljivke X je podana s predpisom:

$$G_X(s) = a\,e^{s+s^2}\,.$$

Določite konstanto a in izračunajte $P(X = 2)$.

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

4. Naj bosta $X \sim \mathrm{Geo}(1/2)$ in $Y \sim \mathrm{Geo}(1/3)$ neodvisni slučajni spremenljivki. Določite porazdelitev njune vsote.

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

6. 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.

7. 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 spremenljivke 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 $G(s) = G_1\big(G_2(s)\big)$.

8. Naj bodo N in $X_1, X_2, X_3, \ldots$ neodvisne slučajne spremenljivke, pri čemer naj bo $N \sim \mathrm{Geo}(a)$ in $X_i \sim \mathrm{Geo}(b)$ za vse i. Določite porazdelitev vsote $X_1+X_2+\ldots+X_N$.

9. Slučajne spremenljivke $X_1, X_2, \ldots$ so porazdeljene geometrijsko Geo(2/3), slučajna spremenljivka N pa geometrijsko Geo(3/4). Vse omenjene slučajne spremenljivke so neodvisne. Zapišite porazdelitev slučajne spremenljivke $S = X_1+X_2+\cdots+X_{2N}$.

10. Maks še nima otrok. Porazdelitev število njegovih otrok in števila otrok posameznega njegovega potomca je enaka:
$$\begin{pmatrix} 0 & 1 & 2 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}.$$
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.

Moment *reda r je matematično upanje r-te potence: $z_r := E(X^r)$.*
Centralni moment *reda r je matematično upanje r-te potence centrirane slučajne spremenljivke: $m_r := E\left[\big((X - E(X)\big)^r\right]$. Drugi centralni moment je torej ravno disperzija.*
Momentno-rodovna funkcija *je eksponentna rodovna funkcija momentov:*

$$M_X(t) = E\big(e^{tX}\big) = 1 + z_1\, t + \frac{z_2}{2!}\, t^2 + \frac{z_3}{3!}\, t^3 + \cdots$$

Tudi če so vsi momenti definirani, momentno-rodovna funkcija ni nujno definirana za vse t (lahko se zgodi, da je definirana le za $t = 0$).
Če je X celoštevilska, je tudi:

$$M_X(t) = G_X(e^t)\,.$$

Če sta X in Y neodvisni, je $M_{X+Y} = M_X\, M_Y$.

11. Naj bo $a < b$. Določite momentno-rodovno funkcijo enakomerne porazdelitve na intervalu $[a, b]$. Kje je definirana?

12. Naj bo $\lambda > 0$. Določite momentno-rodovno funkcijo Poissonove porazdelitve $\mathrm{Poi}(\lambda)$. Kje je definirana?

13. Naj bo $\lambda > 0$. Določite vse momente in momentno-rodovno funkcijo eksponentne porazdelitve na intervalu $\mathrm{Exp}(\lambda)$. Kje je definirana? Izračunajte še tretji centralni moment.

14. Izračunajte momentno-rodovno funkcijo in vse momente standardne normalne porazdelitve.

15. Izračunajte momentno-rodovno funkcijo splošne normalne porazdelitve. Na podlagi tega sklepajte o porazdelitvi vsote neodvisnih normalno porazdeljenih slučajnih spremenljivk.

16. Izračunajte momentno-rodovno funkcijo porazdelitve gama. Kje je le-ta definirana? Na podlagi tega sklepajte o porazdelitvi vsote določenih neodvisnih slučajnih spremenljivk s to porazdelitvijo.

Kumulante *so odvodi logaritma momentno-rodovne funkcije v izhodišču:*

$$\kappa_r(X) := (\ln M_X)^{(r)}(0)\,.$$

17. Izračunajte vse kumulante normalne porazdelitve $\mathrm{N}(\mu, \sigma)$.

18. a) Koliko je enaka prva kumulanta?

b) Dokažite, da so višje kumulante invariantne za translacije: za $r \geq 2$ velja $\kappa_r(X + a) = \kappa_r(X)$.

c) Izrazite $\kappa_r(aX)$ s $\kappa_r(X)$.

Asimetrija:

$$A(X) = \frac{\kappa_3(X)}{\big(D(X)\big)^{3/2}}$$

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

Sploščenost (kurtozis):

$$K(X) = \frac{\kappa_4(X)}{\big(D(X)\big)^2}$$

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

19. Izrazite drugo, tretjo in četrto kumulanto ter še asimetrijo in sploščenost s centralnimi momenti. *Namig*: pomagajte si z razvojem v potenčne vrste.

20. Naj bo $a < b$. Izračunajte prve štiri kumulante ter asimetrijo in sploščenost enakomerne porazdelitve na intervalu $[a, b]$.

21. Naj bo $\lambda > 0$. Izračunajte drugi, tretji in četrti centralni moment Poissonove porazdelitve $\mathrm{Poi}(\lambda)$.

Neenačba Markova: *če je $X \geq 0$, za vsak $x > 0$ velja ocena:*

$$P(X \geq x) \leq \frac{E(X)}{x}\,.$$

Dostikrat se namesto slučajne spremenljivke X splača vzeti kakšno njeno funkcijo. Tako lahko dobimo **neenačbo Čebiševa**:

$$P\big(|X - E(X)| \geq t\big) \leq \frac{D(X)}{t^2}\,.$$

Dobre ocene pa često dobimo tudi iz momentno-rodovne funkcije:

$$P(X \geq x) \leq e^{-tx}\, M_X(t)\,.$$

22. V tovarni proizvedejo 1600 izdelkov, a je vsak okvarjen z verjetnostjo 0˙1. Posamezni izdelki so med seboj neodvisni. Navzgor ocenite verjetnost, da bo pokvarjenih več kot 250 izdelkov:

a) s pomočjo neenačbe Markova;

b) s pomočjo neenačbe Čebiševa z upoštevanjem simetrije;

c) s pomočjo momentno-rodovne funkcije.

Primerjajte še z aproksimacijo po Laplaceovi integralski formuli!

23. Slučajne spremenljivke $X_1, X_2, \ldots X_n$ naj bodo neodvisne in naj imajo standardno *Laplaceovo porazdelitev*, t. j. porazdelitev z gostoto:

$$p(x) = \tfrac{1}{2}\, e^{-|x|}\,.$$

Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Ocenite $P(S_5 > 5)$ in $P(S_{20} > 20)$, in sicer z uporabo neenakosti Čebiševa in simetrije ter z uporabo momentno-rodovne funkcije.

Karakteristična funkcija

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

Definirana je za vsak $t \in \mathbb{R}$.

Velja tudi $\phi_X(t) = M_X(it)$. Brž ko je momentno-rodovna funkcija definirana za kakšno neničelno realno število, lahko karakteristično funkcijo izračunamo že iz vrednosti momentno-rodovne funkcije na realnih številih, in sicer tako, da jo ustrezno analitično razširimo.

Če je X celoštevilska, je tudi:

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

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

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

26. 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 $\phi_{X+Y} = \phi_X\, \phi_Y$.

27. Karakteristična funkcija binomske porazdelitve.

28. Karakteristična funkcija Pascalove porazdelitve.

29. Karakteristična funkcija porazdelitve gama.

30. Rekonstruirajte porazdelitev slučajne spremenljivke s karakteristično funkcijo:

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

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).

31. 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.

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

33. 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:

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

9. Limitni izreki

Konvergenca slučajnih spremenljivk. Šibki in krepki zakon velikih števil. Centralni limitni izrek.

Prva Borel–Cantellijeva lema

Če za dogodke $A_1, A_2, \ldots$ velja $\sum_{n=1}^{\infty} P(A_n) < \infty$, se skoraj gotovo zgodi kvečjemu končno mnogo teh dogodkov.

1. [2] Mečemo pošten kovanec, meti so neodvisni. Naj bo L_n dolžina najdaljšega sosledja samih cifer do vključno n-tega meta. Dokažite, da skoraj gotovo velja:

$$\lim_{n\to\infty} \frac{L_n}{\log_2 n} = 1\,.$$

Konvergenca slučajnih spremenljivk

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 $\lim_{n\to\infty} P(|X_n - X| < \varepsilon) = 1$.*
- **v porazdelitvi** *ali tudi* **šibko** *($X_n \xrightarrow[n\to\infty]{\text{d}} X$), če velja ena izmed naslednjih dveh ekvivalentnih trditev:*
  - *Za vsak x, kjer je F_X zvezna, velja $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.*
  - *Za vsako zvezno in omejeno funkcijo h velja $\lim_{n\to\infty} E\big(h(X_n)\big) = E\big(h(X)\big)$.*

Iz skoraj gotove konvergence sledi konvergenca v verjetnosti, iz nje pa šibka konvergenca.
Šibko konvergenco lahko definiramo zgolj za porazdelitve.

2. *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 šibko? Pa skoraj gotovo?

[2] Vir: R. Durrett: Probability: Theory and Examples, druga izdaja, Duxbury Press, 1995.

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:*

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

Druga Borel–Cantellijeva lema

Če za **neodvisne** *dogodke $A_1, A_2, \ldots$ velja $\sum_{n=1}^{\infty} P(A_n) = \infty$, se skoraj gotovo zgodi neskončno mnogo teh dogodkov.*

3. [3] Naj bodo $X_2, X_3, X_4, \ldots$ neodvisne slučajne spremenljivke z:

$$X_n \sim \begin{pmatrix} -n & 0 & n \\ \frac{1}{2n\ln n} & 1 - \frac{1}{n\ln n} & \frac{1}{2n\ln n} \end{pmatrix}.$$

   Dokažite, da za to zaporedje velja šibki, ne pa tudi krepki zakon velikih števil: če označimo $S_n = X_1 + X_2 + \cdots + X_n$, zaporedje S_n/n v verjetnosti konvergira proti nič, vendar skoraj gotovo divergira.

4. 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 $S_n = 0$.

5. 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.

6. 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.

7. Dokažite, da iz konvergence v porazdelitvi ne sledi nujno konvergenca v verjetnosti, pač pa to sledi, če gre za konvergenco proti konstantni slučajni spremenljivki.

[3] Vir: G. R. Grimmett, D. R. Stirzacker: Probability and Random Processes. Problems and Solutions. Clarenson Press, Oxford, 1997.

8. Dokažite, da za vsako zaporedje porazdelitev, ki šibko konvergira proti dani porazdelitvi, obstaja verjetnostni prostor, na njem pa zaporedje slučajnih spremenljivk z ustreznimi porazdelitvami, ki skoraj gotovo konvergira proti slučajni spremenljivki z limitno porazdelitvijo.

9. Dan je naslednji prototip trditve:

   Če X_n konvergira proti X in Y_n konvergira proti Y, tudi $X_n + Y_n$ konvergira proti $X + Y$.

   Raziščite, kako je z njegovo veljavnostjo pri skoraj gotovi konvergenci ter konvergenci v verjetnosti in porazdelitvi.

10. Dokažite *izrek Sluckega*: če X_n konvergira proti X in Y_n konvergira proti konstanti c (oboje v porazdelitvi), tudi $X_n + Y_n$ konvergira proti $X + c$.

Centralni limitni izrek
(klasična formulacija)

Naj bodo $X_1, X_2, \ldots$ neodvisne in enako porazdeljene slučajne spremenljivke z $E(X_i^2) < \infty$ ter $E(X_i) = \mu_1$ in $D(X_i) = \sigma_1^2$. Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Tedaj standardizirane vsote:

$$\frac{S_n - n\mu_1}{\sigma_1\sqrt{n}}$$

šibko konvergirajo proti standardni normalni porazdelitvi. Bolj ohlapno povedano, vsota S_n je, če je n velik, porazdeljena približno normalno $\mathrm{N}\big(n\mu_1, \sigma_1\sqrt{n}\big)$. Spet natančneje, velja:

$$P(a < S_n < b),\ P(a \le S_n \le b) \approx \Phi\left(\frac{b - n\mu_1}{\sigma_1\sqrt{n}}\right) - \Phi\left(\frac{a - n\mu_1}{\sigma_1\sqrt{n}}\right),$$

kjer napaka v zgornji aproksimaciji pri fiksni porazdelitvi seštevancev X_i konvergira proti nič, ko gre n proti neskončno, in sicer enakomerno po a in b.

11. Standardno kocko vržemo 105-krat, meti so neodvisni. Ocenite verjetnost, da bo skupaj padlo manj kot 350 pik.

12. Slučajne spremenljivke $X_1, \ldots X_{100}$ so neodvisne in porazdeljene diskretno po naslednji shemi:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0{\cdot}5 & 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}1 \end{pmatrix}$$

   Ocenite $P(170 < X_1 + \cdots + X_{100} < 210)$ in utemeljite odgovor.

13. Loterija izda srečko, ki stane 1 evro. Srečka z verjetnostjo 5% zadene 10 evrov, z verjetnostjo 45% zadene 1 evro (v tem primeru torej kupec dobi povrnjen vložek), z verjetnostjo 50% pa ne zadene nič. Kupimo 50 srečk. Kolikšna je verjetnost, da imamo dobiček? Privzamemo, da so srečke neodvisne.

Naj bodo $X_1, \dots X_n$ neodvisne in enako porazdeljene slučajne spremenljivke s $\sigma_1 := \sigma(X_i)$ in $\gamma_1 := \left[E\left(|X_i - E(X_i)|^3\right)\right]^{1/3}$. Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Naj gre n proti neskončno, pri čemer pa se lahko porazdelitev slučajnih spremenljivk X_i tudi spreminja z n. Zadosten pogoj, da standardizirane vsote S_n šibko konvergirajo proti standardni normalni porazdelitvi, t. j. da gre absolutna napaka pri normalni aproksimaciji intervalskih verjetnosti proti nič, je:

$$n \gg \frac{\gamma_1^6}{\sigma_1^6} .$$

Da se tudi oceniti napaka, ki jo naredimo pri normalni aproksimaciji. Ocen je veliko, med najenostavnejšimi pa je ocena na podlagi tretjega absolutnega momenta, znana kot **Berry–Esséenov izrek**. Če so $X_1, X_2, \dots X_n$ neodvisne in enako porazdeljene z $E(X_i) = \mu_1$, $\sigma_1 = \sigma(X_i)$ in $\gamma_1 = \left[E(|X_i - E(X_i)|^3)\right]^{1/3}$, vsota $S_n = X_1 + X_2 + \cdots + X_n$ zadošča oceni:

$$\sup_{x \in \mathbb{R}} \left| P(S_n \le x) - \Phi\left(\frac{x - n\mu_1}{\sigma_1 \sqrt{n}}\right) - \frac{1}{2} \right| \le \frac{0{\cdot}4774}{\sqrt{n}} \frac{\gamma_1^3}{\sigma_1^3} .$$

(Berry, 1941; Esséen, 1942; Tjurin, 2012).

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

15. Slučajne spremenljivke $X_1, X_2, \dots X_n$ naj bodo neodvisne in naj imajo standardno *Laplaceovo porazdelitev*, t. j. porazdelitev z gostoto:

$$p(x) = \tfrac{1}{2}\, e^{-|x|} .$$

Označimo $S_n := X_1 + X_2 + \cdots + X_n$. Približno izračunajte $P(S_5 > 5)$ in $P(S_{20} > 20)$.

Naj bodo spet $X_1, \dots X_n$ neodvisne in enako porazdeljene slučajne spremenljivke s $\sigma_1 := \sigma(X_i)$ in $\gamma_1 := \left[E\left(|X_i - \mu_1|^3\right)\right]^{1/3}$ ter naj bo $S_n := X_1 + X_2 + \cdots + X_n$. Naj gre n proti neskončno, pri čemer pa se lahko porazdelitev slučajnih spremenljivk X_i tudi spreminja z n. Zadosten pogoj, da gre **relativna** *napaka pri normalni aproksimaciji intervalskih verjetnosti $P(a_n \le S_n \le b_n)$ proti nič, je:*

$$n \gg \frac{\gamma_1^6}{\sigma_1^6}, \qquad \min\left\{|a_n - n\mu_1|, |b_n - n\mu_1|\right\} \ll \frac{n^{2/3}\sigma_1^2}{\gamma_1} .$$

16. Slučajne spremenljivke $U_1, \dots U_{100}$ so porazdeljene enakomerno na intervalu $[0, 1]$, slučajne spremenljivke $V_1, \dots V_{100}$ pa diskretno po shemi:

$$\begin{pmatrix} 1 & 2 \\ 2/3 & 1/3 \end{pmatrix} .$$

Vse omenjene slučajne spremenljivke so med seboj neodvisne.

a) Definirajmo $X_i := U_i V_i$. Izračunajte $E(X_i)$ in $D(X_i)$.

b) Naj bo $S := X_1 + X_2 + \cdots + X_{100}$. Približno izračunajte $P(S < 60)$.

17. Slučajne spremenljivke $X_1, X_2, \ldots X_{100}$ so neodvisne in porazdeljene diskretno po shemi:
$$\begin{pmatrix} 0 & 1 & 2 \\ p & 1-2p & p \end{pmatrix}.$$
Označimo $S := X_1 + X_2 + \cdots + X_{100}$. Približno določite vrednost parametra p, pri kateri je $P(S < 90) = 0{\cdot}05$.

18. 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)!

Centralni limitni izrek ima veliko posplošitev. Pomembna posplošitev je na primer, ko seštevanci niso nujno enako porazdeljeni. Površno povedano, 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:

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

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

$$\mu = E(X_1) + E(X_2) + \cdots + E(X_n)$$
$$\sigma^2 = D(X_1) + D(X_2) + \cdots + D(X_n).$$

Tudi v tem primeru se da napaka pri normalni aproksimaciji eksplicitno oceniti. Velja:

$$\sup_{x\in\mathbb{R}} \left| P(S \le x) - \Phi\left(\frac{x-\mu}{\sigma}\right) - \frac{1}{2} \right| \le \frac{0{\cdot}5591}{\sigma^3} \sum_{k=1}^{n} E\left(|X_k - E(X_k)|^3\right).$$

(Tjurin, 2012). Količini $\frac{1}{\sigma^3} \sum_{k=1}^{n} E(|X_k - E(X_k)|^3)$ pravimo **razmerje Ljapunova**.

19. Naj bodo $X_1, X_2, \ldots$ neodvisne in naj bo $X_k \sim \mathrm{Exp}(1/k)$. Označimo $S_n := X_1 + X_2 + \cdots + X_n$.

a) Dokažite, da standardizirane vsote $(S_n - E(S_n))/\sigma(S_n)$ šibko konvergirajo proti standardni normalni porazdelitvi.

b) Ocenite $P(S_{100} < 6000)$.

20. Slučajne spremenljivke $X_1, X_2, \dots X_{100}$ so neodvisne in porazdeljene diskretno po shemi:
$$\begin{pmatrix} -1 & 1 \\ 1/3 & 2/3 \end{pmatrix}.$$
Označimo $S = 10X_1 + X_2 + X_3 + \cdots + X_{100}$. Približno izračunajte $P(S < 41)$.

21. Naj bodo $X_1, X_2, X_3, \dots$ 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.

Centralni limitni izrek ima posplošitve celo na situacije, ko med sumandi obstaja določena vrsta odvisnosti. Med drugim ga lahko uporabimo za hipergeometrijsko porazdelitev.

22. Naj bo $N \sim \mathrm{Hip}(s, r, n)$.

    a) Izračunajte $E(N)$ in $D(N)$.

    b) Za $n = 500$, $r = 200$ in $s = 100$ približno izračunajte $P(N > 45)$.

23. Slučajne spremenljivke $X, Y_1, Y_2, \dots Y_{100}$ so neodvisne, pri čemer je $P(X = 1) = 2/3$, $P(X = 2) = 1/3$, $E(Y_i) = 3$ in $D(Y_i) = 100$. Slučajne spremenljivke $Y_1, \dots Y_{100}$ so tudi enako porazdeljene. Označimo $S = X(Y_1 + Y_2 + \cdots + Y_{100})$. Približno izračunajte $P(400 < S < 500)$.

10. Zadostne in postranske statistike

Zadostnost statistik. Fisher–Neymanov faktorizacijski izrek. Minimalne zadostne statistike. Eksponentne družine porazdelitev. 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:

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

   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. Naj bo X opažanje v statističnem modelu, kjer je verjetnost odvisna od parametra $\theta \in \Theta$. Statistični model naj bo *diskreten*, kar pomeni, da opažanje skoraj gotovo zavzame vrednosti le v neki števni množici, neodvisni od θ. Naj bo $T = t(X)$ statistika. Dokažite naslednjo ekvivalenco:

   (1) T je zadostna.
   (2) Obstajata taki funkciji ρ in g, da je:
   $$P_\theta(X = x) = \rho(x)\, g\big(t(x), \theta\big) \qquad (*)$$
   za vse x in Θ.
   (3) Brž ko je $t(x_1) = t(x_2)$, sta verjetnosti $P_\theta(X = x_1)$ in $P_\theta(X = x_2)$ *sorazmerni*, kar pomeni, da je bodisi $P_\theta(X = x_1) = 0$ za vse $\theta \in \Theta$, bodisi obstaja tak k, neodvisen od θ, da je $P_\theta(X = x_2) = k\, P_\theta(X = x_1)$ za vse $\theta \in \Theta$.

   Opomba: ekvivalenci (1) $\Leftrightarrow$ (2) pravimo *Fisher–Neymanov faktorizacijski izrek.*

Fisher–Neymanov faktorizacijski izrek v splošnem

Če je $f_X(x, \theta)$ verjetnostna funkcija ali pa gostota opažanja X, je statistika $T = t(X)$ zadostna za parameter θ natanko tedaj, ko obstajata taki funkciji ρ in g, da je:

$$f_X(x, \theta) = \rho(x)\, g\big(t(x), \theta\big)$$

za vse x in Θ.

4. Poiščite kakšno enorazsežno zadostno statistiko za Poissonovo porazdelitev $\mathrm{Poi}(\lambda)$, kjer imamo na voljo vzorec neodvisnih opažanj $X_1, \dots X_n$.

Zadostna statistika S je **minimalna**, *če je skoraj gotovo funkcija vsake zadostne statistike. Natančneje, za vsako zadostno statistiko T morata obstajati taka merljiva funkcija h in tak dogodek A, ki ima pri vseh vrednostih parametrov verjetnost ena, da na A velja $S = h(M)$.*

5. Opažanje X je porazdeljeno diskretno po shemi:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ a & a^2 & 2a^2 & 4a & 3a^2 & b \end{pmatrix}$$

   a) Izrazite b z a.
   b) Pri katerih vrednostih parametra a je z zgornjo shemo določena porazdelitev slučajne spremenljivke?
   c) Poiščite minimalno zadostno statistiko za model, pri katerem a zavzame vse vrednosti iz prejšnje točke.
   d) Poiščite minimalno zadostno statistiko za model, pri katerem a zavzame le robni vrednosti iz prejšnje točke.

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

$$f_X(x) = \rho(x)\, g(\gamma_1, \dots \gamma_m)\, e^{h_1(x)\,\gamma_1 + h_2(x)\,\gamma_2 + \cdots + h_m(x)\,\gamma_m}\,.$$

Pri zgoraj opisani družini je $\big(h_1(X), \dots h_m(X)\big)$ zadostna statistika za to družino. Pravimo ji **pripadajoča (zadostna) statistika**.
Naboru $\gamma_1, \dots \gamma_m$ pravimo **naravni nabor parametrov** *za ta zapis, množico v $\mathbb{R}^m$, ki jo preteče, pa imenujemo* **parametrični prostor**.

6. Dookažite, da je družina Bernoullijevih porazdelitev:

$$\begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}$$

eksponentna. Določite parametrični prostor. Poskusite to storiti tako, da bo število naravnih parametrov čim manjše. Nato enako storite še za binomsko porazdelitev $\mathrm{Bin}(n, p)$.

Če parametrični prostor eksponentne družine vsebuje afino neodvisno množico (denimo oglišča neizrojenega simpleksa) ali, ekvivalentno, če so parametri $1, \gamma_1, \gamma_2, \ldots \gamma_m$ linearno neodvisni, je pripadajoča zadostna statistika minimalna.

7. Pokažite, da je vzorčna vsota minimalna zadostna statistika za parameter pri Poissonovi porazdelitvi, če imamo na voljo vzorec neodvisnih opažanj $X_1, \ldots X_n$.

Če porazdelitev statistične spremenljivke X pripada eksponentni družini, ki ima za določen naraven nabor parametrov pripadajočo zadostno statistiko $\boldsymbol{h}(X) = \big(h_1(X), \ldots h_m(X)\big)$, eksponentno družino tvori tudi porazdelitev slučajnega vektorja $(X_1, \ldots X_n)$ neodvisnih kopij spremenljivke X (torej porazdelitev vzorca n neodvisnih opažanj). Parametrični prostor je isti, pripadajoča zadostna statistika pa je $\boldsymbol{h}(X_1) + \boldsymbol{h}(X_2) + \cdots + \boldsymbol{h}(X_n)$. Tako lahko tudi iskanje minimalne zadostne statistike pri vzorcu več opažanj prevedemo na iskanje le-te pri enem samem opažanju.

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.

8. 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 $\mathrm{N}(a, a)$.

9. 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 13. razdelku).

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

11. 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.

> *Brž ko naravni parametrični prostor eksponentne družine vsebuje kakšno odprto množico v $\mathbb{R}^m$, je pripadajoča zadostna statistika $\big(h_1(X), \ldots h_m(X)\big)$ neodvisna od katere koli postranske statistike pri vseh vrednostih parametrov.*

12. Iz postranskih statistik iz 11. naloge potegnite zaključke o neodvisnosti pomembnih statistik pri Gaussovih modelih.

13. 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.

14. 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.

11. Točkasto ocenjevanje in vzorčenje

Doslednost in nepristranskost. Srednja kvadratična napaka. Pridobivanje cenilk: metoda momentov in metoda največjega verjetja, nepristranske cenilke z enakomerno najmanjšo disperzijo.

Cenilka $\hat{\zeta}$ karakteristike 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 $x_1, x_2, \dots x_N$, na vzorcu pa z $X_1, X_2, \dots X_n$. 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. Če vrednosti $X_1, \dots X_n$ različnih opažanj določene statistične spremenljivke uredimo po velikosti, dobimo *vrstilne statistike*:

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}\,.$$

a) Privzemimo, da so $X_1, \dots X_n$ neodvisna opažanja statistične spremenljivke, porazdeljene zvezno enakomerno na $(0, 1)$. Izračunajte matematična upanja vseh vrstilnih statistik.

b) Naj bodo $X_1, \dots X_n$ neodvisna opažanja statistične spremenljivke, porazdeljene zvezno enakomerno na poljubnem intervalu. Poiščite nepristransko cenilko za kvantil porazdelitve za verjetnost p, ki temelji na danih dveh vrstilnih statistikah $X_{(i)}$ in $X_{(j)}$, kjer sta i in j različna. Seveda mora biti funkcija statistik $X_{(i)}$ in $X_{(j)}$ neodvisna od intervala.

c) Če predpišemo i in j, za katera p je cenilka enaka kar $X_{(i)}$ oz. $X_{(j)}$? Se to ujema z ustreznima *vzorčnima* kvantiloma (ki ju definiramo kot kvantila porazdelitve, ki jo dobimo, če na slepo izberemo enoto vzorca)? Kako bi s pomočjo tega poljudno opisali cenilko za kvantil za poljubno verjetnost?

d) Pri konstrukciji cenilke za q_p pri danem p lahko i in j poljubno izberemo. Kateri vrstilni statistiki pa je smiselno izbrati, če p ni preblizu 0 ali 1?

e) Izračunajte ustrezno oceno prvega kvartila populacije, za katero privzamemo, da je porazdeljena zvezno enakomerno, pri vzorčnih vrednostih:

$$3,\ 6,\ 14,\ 16,\ 17,\ 17,\ 18,\ 20\,.$$

3. Pri *stratificiranem vzorčenju* populacijo razdelimo na več podpopulacij – *stratumov* – in iz vsakega vzamemo vzorec predpisane velikosti. Recimo, da je populacija velika in da so deleži posameznih stratumov v celotni populaciji enaki $p_1, p_2, \ldots p_r$.

   Na populaciji imamo spet definirano populacijsko spremenljivko X. Na i-tem stratumu naj ima X povprečje μ_i in disperzijo σ_i^2.

   a) Označimo z $X_1, X_2, \ldots X_r$ zožitve populacijske spremenljivke X na posamezne stratume. Dokažite, da za poljubno funkcijo f velja:

   $$E\big[f(X)\big] = p_1\, E\big[f(X_1)\big] + p_2\, E\big[f(X_2)\big] + \cdots + p_r\, E\big[f(X_r)\big]\,.$$

   Pravimo, da je porazdelitev statistične spremenljivke X *mešanica* porazdelitev statističnih spremenljivk $X_1, X_2, \ldots X_r$.

   b) Naj bo μ povprečje, σ^2 pa disperzija statistične spremenljivke na celi populaciji. Izrazite μ in σ^2 z μ_i in σ_i^2.

   c) Iz vsakega stratuma vzamemo enostavni slučajni vzorec – iz i-tega stratuma vzorec velikosti n_i. Označimo z $\bar{X}_1, \ldots \bar{X}_r$ vzorčna povprečja na teh stratumih. Zapišite nepristransko cenilko za μ, ki temelji na teh vzorčnih povprečjih, in izračunajte njeno disperzijo za primer, ko so delni vzorci majhni v primerjavi s stratumi.

   d) 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.

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

4. Niso vedno vse enote v vzorcu z enako verjetnostjo. Vzorčni načrt je lahko zasnovan celo tako, da niso nujno vsi vzorci enako veliki. Predstavimo vzorec s slučajno množico S in za različne enote $i_1, i_2, \ldots i_k$ označimo:

$$\pi_{i_1 i_2 \cdots i_k} := P(i_1 \in S, i_2 \in S, \ldots i_k \in S)\,.$$

Homogene linearne statistike spremenljivke X na tako predstavljenih vzorcih so oblike:

$$\sum_{i \in S} a_i x_i \,,$$

kjer je x_i vrednost spremenljivke X na i-ti enoti.

a) Recimo, da želimo oceniti populacijsko povprečje dane spremenljivke s homogeno linearno statistiko. Pokažite, da obstaja največ ena izbira koeficientov a_i, pri katerih je ta statistika nepristranska (kdaj obstaja?). Pravimo ji *Horvitz–Thompsonova cenilka.*

b) Kaj je Horvitz–Thompsonova cenilka na enostavnih slučajnih vzorcih brez ponavljanja?

c) Denimo, da iz populacije velikosti 10 izberemo enostavni slučajni vzorec velikosti 3 s ponavljanjem (in ga nato zapišemo kot množico). To pomeni, da izvedemo tri runde in v vsaki izberemo eno enoto, vsakič na slepo in neodvisno, pri čemer enote pustimo notri; na koncu vzamemo v vzorec enote, ki so bile vsaj enkrat izbrane v vzorec. Izračunajte vrednost Horvitz–Thompsonove cenilke za naslednje primere:
   - izbrane so tri različne enote, vrednosti statistične spremenljivke na njih pa so 1, 2 in 3;
   - v prvih dveh rundah je izbrana ista enota, na kateri je vrednost spremenljivke enaka 1, v tretji rundi pa je izbrana enota, nakateri je vrednost spremenljivke enaka 4;
   - v prvih dveh rundah sta izbrani različni enoti in na obeh je vrednost spremenljivke enaka 1, v tretji rundi pa je izbrana enota, na kateri je vrednost spremenljivke enaka 4;
   - izbrane so tri različne enote, vrednosti statistične spremenljivke na njih pa so 21, 22 in 23.

   Komentirajte!

d) Poiščite potreben in zadosten pogoj, da Horvitz–Thompsonova cenilka komutira s translacijami.

e) Pri *stratificiranem vzorčenju* populacijo razdelimo na K delov – *stratumov* velikosti $N_1, N_2, \dots N_K$. Nato najprej vzamemo enostavni slučajni vzorec iz k stratumov, na vsakem izbranem stratumu pa pogojno na to vzemem enostaven slučajni vzorec: če je i-ti stratum izbran, na njem vzamemo enostavni slučajni vzorec velikosti n_i. Privzamemo, da so ti delni vzorci pogojno na izbrane stratume med seboj neodvisni. Nato vse delne vzorce združimo v enoten vzorec. Poiščite potreben in zadosten pogoj za to, da Horvitz–Thompsonova cenilka komutira s translacijami.

f) Izračunajte disperzijo Horvitz–Thompsonove cenilke.

V vseh nadaljnjih nalogah tega razdelka je, če ni določeno drugače, 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.

5. Je vzorčno povprečje dosledna cenilka parametra a pri zvezni porazdelitvi z gostoto:
$$p(x) = \frac{1}{\pi\bigl(1 + (x-a)^2\bigr)}\ ?$$

> **Metoda momentov** *temelji na tem, da za cenilke populacijskih momentov* $z_k = E(X^k)$ *vzamemo:*
> $$\hat{z}_k = \frac{X_1^k + X_2^k + \cdots + X_n^k}{n}$$
> *Če želimo oceniti karakteristiko* ζ*, jo najprej izrazimo kot funkcijo momentov:* $\zeta = g(z_1, \ldots z_r)$*. Njena cenilka po metodi momentov je ista funkcija vzorčnih momentov:* $\hat{\zeta} = g(\hat{z}_1, \ldots \hat{z}_r)$*. 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.*

6. Statistična spremenljivka je porazdeljena enakomerno na $[0, a]$.
   a) Poiščite cenilko za a, ki temelji na prvem momentu.
   b) Na podlagi cenilke iz prejšnje točke ocenite a iz vzorca:
   $$1,\ 2,\ 1,\ 3,\ 10\,.$$
   Kaj opazite?
   c) Dokažite, da za prejšnji vzorec dobimo smiseln rezultat, če vzamemo kakšen višji moment.

7. 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?

8. 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:
$$-1,\ 0,\ 1,\ 1,\ 0,\ 0,\ -1,\ 2,\ 1,\ 2$$

9. 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.

10. 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 z s_0^2. Pokažite, da obstaja tak faktor k', odvisen le od velikosti vzorca, da je $s^2 := k' s_0^2$ nepristranska cenilka za σ^2.

    c) Privzemimo, da ima X končen četrti moment. Izračunajte srednji kvadratični napaki cenilk s_0^2 in s^2.

    d) Naj bo X porazdeljena normalno. Pokažite, da obstaja tak faktor k^*, odvisen le od velikosti vzorca, da je $k^* s_0^2$ najučinkovitejša izmed cenilk oblike $k s_0^2$, $k > 0$. Izračunajte še $q(k^* s_0^2)$.

11. Statistična spremenljivka X je porazdeljena po Poissonu $\mathrm{Poi}(\lambda)$. 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?

Metoda največjega verjetja

*Naj bo $f_{\mathbf{X}}(\mathbf{x} \mid \theta)$ verjetnostna funkcija ali gostota (verjetje) porazdelitve opažanja $\mathbf{X}$, ki je odvisna še od parametra θ. Cenilka za θ po metodi največjega verjetja (maksimalne zanesljivosti) pri opažanju $\mathbf{X}$ je tisti θ, pri katerem je vrednost $L := f_{\mathbf{X}}(\mathbf{X} \mid \theta)$ (**verjetje**) maksimalna.*
Za iskanje maksimuma navadno rešujemo enačbo:
$$\frac{\mathrm{d} \ln L}{\mathrm{d}\theta} = 0\,,$$
če je parameter θ enorazsežen, in ustrezen sistem enačb s parcialnimi odvodi, če je večrazsežen.

12. Denimo, da zaporedje poskusov tako kot v 2. nalogi iz 10. razdelka tvori markovsko verigo z začetno verjetnostjo uspeha θ in prehodno matriko:
$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\theta}{2} & \frac{\theta}{2} \\ \frac{1-\theta}{2} & \frac{1+\theta}{2} \end{bmatrix}.$$
Recimo, da smo opazili, da je prvi poskus uspel, drugi ni uspel, tretji pa je spet uspel. Ocenite θ po metodi največjega verjetja.

13. 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?

Če imamo na voljo vzorec z vrednostmi $X_1, \ldots X_n$, ki so neodvisne in enako porazdeljene z verjetnostno funkcijo ali gostoto $f(x \mid \theta)$, je verjetje kar produkt verjetij za posamezne spremenljivke:

$$L = L_1 L_2 \cdots L_n \,,$$

kjer je $L_i = f(X_i \mid \theta)$.

14. Statistična spremenljivka X ima *Paretovo porazdelitev*, ki je zvezna z gostoto:

$$p_X(x \mid a) = \begin{cases} \dfrac{c}{x^a} & ; x > 1 \\ 0 & ; \text{sicer} \end{cases} .$$

   a) Določite tiste vrednosti parametra a, pri katerih ima porazdelitev smisel, in izrazite c z a.
   b) Ocenite a po metodi največjega verjetja.

15. 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?

16. Statistična spremenljivka X je porazdeljena zvezno z gostoto:

$$p_X(x) = \begin{cases} e^{-\lambda x} & ; x \geq -\frac{\ln \lambda}{\lambda} \\ 0 & ; \text{sicer} \end{cases} ,$$

kjer je $\lambda > 0$ neznan parameter. Po metodi največjega verjetja poiščite cenilko za λ, ki temelji na enem samem opažanju.

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

   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^*.

Nepristranska cenilka z enakomerno najmanjšo možno disperzijo

V eksponentni družini, kjer parametrični prostor vsebuje odprto množico, ima nepristranska cenilka za dano karakteristiko ζ enakomerno najmanjšo možno disperzijo natanko tedaj, ko se izraža s pripadajočo zadostno statistiko $T = \big(h_1(X), \ldots h_m(X)\big)$.
Taka cenilka je (do skoraj gotove enakosti) enolična in obstaja, brž ko obstaja sploh kakšna nepristranska cenilka Z za ζ. Nepristransko cenilko z enakomerno najmanjšo možno disperzijo v tem primeru dobimo kot $E(Z \mid T)$ (ki je opazljiva, ker je T zadostna).

18. Vsak poskus določene vrste uspe z verjetnostjo $p \in (0, 1)$. Poiščite nepristransko cenilko za p^2 z enakomerno najmanjšo možno disperzijo, ki temelji na izvedbi treh neodvisnih poskusov te vrste.

    Kaj pa, če izvedemo n poskusov?

19. Statistična spremenljivka X je porazdeljena diskretno po shemi:

$$\begin{pmatrix} 0 & 1 & 2 \\ \dfrac{1}{1+3\theta+2\theta^2} & \dfrac{3\theta}{1+3\theta+2\theta^2} & \dfrac{2\theta^2}{1+3\theta+2\theta^2} \end{pmatrix},$$

    kjer je $\theta > 0$ neznan parameter.

    a) Zapišite to kot naravno enoparametrično eksponentno družino porazdelitev. *Namig*: pomagajte si s funkcijami $\rho_i(x) = \begin{cases} 1 & ; x = i \\ 0 & ; \text{sicer} \end{cases}$.
    b) Poiščite nepristransko cenilko za $1/(1 + \theta)$ z enakomerno najmanjšo možno disperzijo.

20. Statistična spremenljivka je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$, na voljo pa imamo vzorec iz n neodvisnih opažanj.

    a) Poiščite nepristransko cenilko za λ z enakomerno najmanjšo možno disperzijo.
    b) Poiščite večkratnik cenilke iz prejšnje točke z najmanjšo srednjo kvadratično napako.

21. Statistična spremenljivka je porazdeljena geometrijsko $\mathrm{Geo}(p)$, na voljo pa imamo vzorec iz n neodvisnih opažanj. Poiščite nepristransko cenilko za p z enakomerno najmanjšo možno disperzijo.

    Namig: najprej poiščite nepristransko cenilko za p, ki se izraža le z enim opažanjem.

22. Statistična spremenljivka X je porazdeljena po Poissonu $\mathrm{Poi}(\lambda)$. Želeli bi oceniti λ^2, na voljo pa imamo eno samo opažanje (X).

    a) Poiščite cenilko po metodi največjega verjetja. Je le-ta nepristranska?

b) Poiščite nepristransko cenilko z enakomerno najmanjšo možno disperzijo.

c) Recimo, da nastavimo cenilko v obliki $X^2 - aX$. Če natančnost cenilke merimo s srednjo kvadratično napako, ali obstaja najnatančnejša cenilka te oblike? Katere izbire pa so smiselne?

Pomoč: prvi štirje momenti Poissonove porazdelitve so:

$$E(X) = \lambda\,,\ E(X^2) = \lambda^2 + \lambda\,,\ E(X^3) = \lambda^3 + 3\lambda^2 + \lambda\,,\ E(X^4) = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda\,.$$

23. Statistična spremenljivka je porazdeljena po Poissonu $\mathrm{Poi}(\lambda)$, na voljo pa imamo eno samo opažanje (X).

a) Poiščite nepristransko cenilko za $e^{-3\lambda}$ z enakomerno najmanjšo možno disperzijo. Je videti smiselna?

b) Dokažite, da ima cenilka po metodi največjega verjetja enakomerno manjšo srednjo kvadratično napako.

c) Kaj pa, če opazimo vzorec neodvisnih opažanj $X_1, X_2, \ldots X_n$? Kako se nepristranska cenilka asimptotično obnaša?

12. Intervali zaupanja

Normalna porazdelitev z znanim σ. Normalna porazdelitev z obema neznanima parametroma, iščemo μ ali σ. Asimptotični intervali zaupanja za ne-Gaussove porazdelitve, Bernoullijevo zaporedje poskusov.

Interval zaupanja $(\zeta_{\min}, \zeta_{\max})$ *za karakteristiko* ζ *pri* **stopnji zaupanja** β *je določen z neenačbo:*

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

ki mora veljati za vse verjetnostne mere P *iz našega statističnega modela,* $\zeta_{\min}$ *in* $\zeta_{\max}$ *pa morata biti opazljivi. Če je res* $\zeta_{\min} < \zeta < \zeta_{\max}$*, pravimo, da pride do* **pokritosti**. *Če se da, interval izberemo tako, da je* β *natančna spodnja meja za verjetnost pokritosti. Tipično vzamemo* $\beta = 0{\cdot}90$, $0{\cdot}95$ *ali* $0{\cdot}99$.
Konstrukcija intervalov zaupanja navadno temelji na pojmu **pivotne funkcije** $T(\mathbf{X}, \zeta)$, *kjer je* $\mathbf{X}$ *opažanje (recimo vzorec): interval zaupanja za* ζ *je* $\{\zeta \,;\, t_{\min} < T(\mathbf{X}, \zeta) < t_{\max}\}$ *ali kaj podobnega.* T *je pivotna funkcija, če je porazdelitev slučajne spremenljivke* $T(\mathbf{X}, \zeta)$ *konstantna (ko spreminjamo parametre modela in se s tem spreminja* ζ*, prav tako pa tudi porazdelitev opažanja* $\mathbf{X}$*). Tega sicer ni možno vedno doseči, a je dovolj, če dosežemo primerno približno. Obenem pa se mora pri fiksni vrednosti opažanja* $\mathbf{X}$ *statistika* T *čimbolj spreminjati z drugim argumentom.*
Če zaokrožujemo, spodnjo mejo vedno zaokrožimo navzdol, zgornjo pa navzgor.

1. Za $n = 1$ in $n = 2$ ter poljuben $\beta \in (0, 1)$ konstruirajte interval zaupanja za parameter θ (z verjetnostjo pokritosti vsaj β, kjer opazimo $S \sim \mathrm{Bin}(n, \theta)$, z naslednjimi lastnostmi:
   - *monotonost*: krajišči sta naraščajoči funkciji opažanja X;
   - *simetrija*: interval zaupanja za $S = n - k$ je interval zaupanja za $S = k$, prezrcaljen okoli $1/2$;
   - *minimalnost*: interval je minimalen med vsemi intervali zaupanja z zgornjima lastnostma.

2. Statistična spremenljivka ima diskretno porazdelitev, podano s shemo:
$$\begin{pmatrix} 0 & 1 & 2 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & \dfrac{\theta(1-\theta)}{\theta^2 - \theta + 1} & \dfrac{(1-\theta)^2}{\theta^2 - \theta + 1} \end{pmatrix}.$$
   Poiščite minimalni *enostranski* 95% interval zaupanja, ki bo temeljil na enem samem opažanju te spremenljivke. Natančneje, če opazimo k, naj bo ta interval oblike $I = [0, b_k)$ ali $[0, 1]$ (za $b_k = 1$).

3. Konstruirajte minimalni enostranski interval zaupanja za parameter θ na podlagi opažanja $S \sim \mathrm{Bin}(n, \theta)$, t. j. za $S = k$ naj bo interval oblike $[0, b_k)$ ali $[0, b_k]$, kjer je $b_0 < b_1 < \cdots < b_n$.

Če je I_1 množica zaupanja pri stopnji zaupanja $1 - \alpha_1$, I_2 pa množica zaupanja pri stopnji zaupanja $1 - \alpha_2$, je $I_1 \cap I_2$ množica zaupanja pri stopnji zaupanja $1 - \alpha_1 - \alpha_2$. Na podlagi te ugotovitve lahko iz enostranskih intervalov zaupanja dobimo dvostranske. Tako iz prejšnje konstrukcije dobimo naslednji dvostranski interval zaupanja:

Clopper–Pearsonov interval zaupanja za neznano verjetnost

Naj se določen dogodek zgodi z verjetnostjo θ. Izvedemo Bernoullijevo zaporedje n poskusov in opazimo, da se dani dogodek zgodi pri natanko k poskusih. Tedaj za interval zaupanja postavimo:

$$\begin{array}{cl} \big[0,\ B_{(1+\beta)/2}(1,n)\big) & ;\ k=0 \\ \big(B_{(1-\beta)/2}(k,n-k+1),\ B_{(1+\beta)/2}(k+1,n-k)\big) & ;\ k=1,2,\ldots n-1 \\ \big(B_{(1-\beta)/2}(n,1),\ 1\big] & ;\ k=n \end{array} ,$$

kjer je $B_p(r,s)$ kvantil porazdelitve $\mathrm{Beta}(r,s)$ za verjetnost p. Velja (glej 27. nalogo iz 5. razdelka):

$$B_p(r,s) = \frac{r\,F_p(2r,2s)}{r\,F_p(2r,2s)+s} = \frac{r}{r+s\,F_{1-p}(2s,2r)} ,$$

kjer je $F_p(a,b)$ kvantil Snedecorjeve (Fisherjeve) porazdelitve z a in b prostostnimi stopnjami za verjetnost p.

Iz 3. naloge sledi, da Clopper–Pearsonov interval zagotavlja verjetnost pokritosti vsaj β, vendar pa ni vedno minimalen. Zadošča pa monotonosti in simetriji.

4. Za primer, ko izvedemo dva poskusa, primerjajte Clopper–Pearsonov interval zaupanja za neznano verjetnost s tistim iz 1. naloge. Privzemite, da je $\beta \ge 4/9$.

5. Slučajne spremenljivke $X_1, \ldots X_n$ so neodvisne in porazdeljene diskretno po shemi:

$$\begin{pmatrix} 0 & a & 3a \\ 1/3 & 1/2 & 1/6 \end{pmatrix} ,$$

kjer je $a > 0$ neznan parameter. Za velike n poiščite asimptotični 95% interval zaupanja za a oblike $[0, a_{\max}]$, kjer mora biti meja $a_{\max}$ opazljiva.

Sredina pri normalni porazdelitvi z znanim standardnim odklonom

Ocenjujemo parameter μ *pri porazdelitvi* $\mathrm{N}(\mu, \sigma)$*, kjer je* σ *znan, na voljo pa imamo vzorec neodvisnih opažanj* $X_1, \ldots X_n$*. Tedaj ne glede na* μ *velja:*

$$\frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim \mathrm{N}(0, 1)\,,$$

kjer je $\bar{X}$ *vzorčno povprečje, definirano spodaj. Izračunamo torej:*

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

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

6. 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.

Sredina pri normalni porazdelitvi z neznanim standardnim odklonom

Ocenjujemo parameter μ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, kjer σ ni znan, na voljo pa imamo spet vzorec neodvisnih opažanj $X_1, \ldots X_n$. Tedaj ne glede na μ in σ velja:

$$\frac{\bar{X} - \mu}{s} \sqrt{n} \sim \mathrm{Student}(n-1)\,,$$

kjer je s popravljeni vzorčni standardni odklon, definiran spodaj. Izračunamo torej:

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c &= t_{(1+\beta)/2}(n-1) \\
s &= \sqrt{\frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}} \\
\Delta &= \frac{cs}{\sqrt{n}}\,,
\end{aligned}$$

kjer je $t_p(df)$ kvantil Studentove porazdelitve z df prostostnimi stopnjami za verjetnost p.

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

Za velike vzorce konstrukcija asimptotično deluje tudi pri ne-Gaussovih spremenljivkah. V tem primeru lahko kvantil Studentove porazdelitve $t_{(1+\beta)/2}(n-1)$ zamenjamo s kvantilom normalne porazdelitve $z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$.

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

Asimptotični interval zaupanja za matematično upanje pri ne-Gaussovih spremenljivkah

Naj bo X nekonstantna statistična spremenljivka z matematičnim upanjem μ in disperzijo σ, na voljo pa imamo vzorec neodvisnih opažanj $X_1, \ldots X_n$. Tedaj vemo, da sta $s_0 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ in s dosledni cenilki za σ (glej 10. nalogo iz 11. razdelka), torej po izreku Sluckega za deljenje velja:

$$\frac{\bar{X} - \mu}{S}\sqrt{n}\,,\ \frac{\bar{X} - \mu}{s_0}\sqrt{n} \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)\,.$$

Konstrukcija asimptotičnega intervala zaupanja torej sovpada s konstrukcijo eksaktnega intervala zaupanja za Gaussove spremenljivke pri neznanem σ, ki ga lahko ocenimo z s_0 ali s.

8. Za 860 žensk poizvemo, koliko otrok imajo. Dobimo naslednjo frekvenčno porazdelitev:

število otrok	0	1	2	3	4	5	6	7	8	9	10
število žensk	227	168	320	96	29	11	4	2	2	0	1

Poiščite 95% interval zaupanja za povprečno število otrok na žensko na celotni populaciji.

Opomba. Podatki so sicer izmišljeni, so pa ukrojeni po popisu Slovenije iz leta 2002 (števila žensk so deljena s 1000 in zaokrožena, izmišljen je tudi konec tabele).

9. Izvedemo n neodvisnih poskusov, vsak uspe z verjetnostjo θ. Opazimo, da je uspelo S poskusov. Konstruirajte asimptotični interval zaupanja za θ, pri čemer za oceno standardnega odklona uporabite s_0.

Wilsonov interval zaupanja za neznano verjetnost

To je interval, ki ga dobimo neposredno iz Laplaceove integralske formule brez popravka s polovico. Ob oznakah iz prejšnje naloge dobimo:

$$\hat{\theta} - c\sqrt{\frac{\theta(1-\theta)}{n}} < \theta < \hat{\theta} + c\sqrt{\frac{\theta(1-\theta)}{n}}\,.$$

V zgornji formuli interval zaupanja za θ še ni eksplicitno izražen z opažanjem S (oz. $\hat{\theta}$). Za ta namen je potrebno rešiti kvadratno enačbo. Dobimo:

$$\tilde{\theta} - \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}(1-\hat{\theta}) + \frac{c^2}{4}} < \theta < \tilde{\theta} + \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}(1-\hat{\theta}) + \frac{c^2}{4}}\,,$$

kjer je $\tilde{n} = n + c^2$ in $\tilde{\theta} = \dfrac{S + \frac{c^2}{2}}{\tilde{n}}$. Če kot spodnje krajišče dobimo 0, ga vključimo v interval, prav tako vključimo zgornje krajišče 1.

Ta interval ima tipično boljšo pokritost kot Waldov. Ima dobro pokritost **v povprečju** *in podobno kot pri Waldovem intervalu velja, da se, ko θ preteče interval $[a, b]$ $(0 < a \le b < 1)$, minimalna verjetnost pokritosti bliža nominalni β, ko gre n proti neskončno. Za $n \ge 20$ in $0{\cdot}1 \le p \le 0{\cdot}9$ je pri $\beta = 0{\cdot}95$ verjetnost pokritosti vsaj $0{\cdot}92$.*

Wilsonov interval zaupanja
s popravkom za zveznost

Ta interval dobimo neposredno iz Laplaceove integralske formule s popravkom s polovico. Ob oznakah iz prejšnje naloge dobimo:

$$\tilde{\theta}_- - \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}_-(1-\hat{\theta}_-) + \frac{c^2}{4}} < \theta < \tilde{\theta}_+ + \frac{c}{\tilde{n}}\sqrt{n\hat{\theta}_+(1-\hat{\theta}_+) + \frac{c^2}{4}}\,,$$

kjer je $\hat{\theta}_- = \dfrac{S-\frac{1}{2}}{\hat{n}}$, $\tilde{\theta}_- = \dfrac{S-\frac{1}{2}}{\tilde{n}}$, $\hat{\theta}_+ = \dfrac{S+\frac{1}{2}}{\hat{n}}$ *in* $\tilde{\theta}_+ = \dfrac{S+\frac{1}{2}}{\tilde{n}}$.
Če je $S = 0$*, za spodnje krajišče postavimo* 0 *in ga vključimo.*
Če je $S = n$*, za zgornje krajišče postavimo* 1 *in ga vključimo.*
Numerični izračuni kažejo, da različica s popravkom vedno doseže nominalno pokritost. V povprečju pa pride do prepokritosti (ki je višjega velikostnega reda kot povprečna prepokritost pri različici brez popravka).

Agresti–Coullov interval zaupanja
za neznano verjetnost

Je zelo blizu Wilsonovemu, a je lažje izračunljiv. Če upoštevamo oznake iz Wilsonovega intervala, je različica brez popravka za zveznost:

$$\tilde{\theta} - c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}} < \theta < \tilde{\theta} + c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}}\,.$$

Tu že za $n \ge 10$ *velja, da je pri* $0\text{˙}1 \le p \le 0\text{˙}9$ *in* $\beta = 0\text{˙}95$ *verjetnost pokritosti vsaj* $0\text{˙}92$.
Različica s popravkom za zveznost:

$$\tilde{\theta} - c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}} - \frac{1}{2n} < \theta < \tilde{\theta} + c\sqrt{\frac{\tilde{\theta}(1-\tilde{\theta})}{\tilde{n}}} + \frac{1}{2n}\,.$$

Tudi tu numerični izračuni kažejo, da različica s popravkom vedno doseže nominalno pokritost.

10. Zanima nas, kolikšnemu deležu študentov je izmed zvrsti filma najbolj všeč romantični film. Anketiramo 20 študentov in romantični film je najbolj všeč štirim. Določite vse prej omenjene intervale zaupanja za delež vseh študentov, ki jim je najbolj všeč romantični film, pri 90% stopnji zaupanja.

11. V 100 metih kocke je 20-krat padla šestica. Določite vse prej omenjene intervale zaupanja za verjetnost padca šestice pri 95% stopnji zaupanja.

12. Pri 10000 metih kovanca je padlo 5048 grbov. Določite vse prej omenjene 90% intervale zaupanja za verjetnost, da pade grb.

Standardni odklon pri normalni porazdelitvi

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

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)\,.$$

Izračunamo torej:

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

kjer je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost p.

Interval zaupanja:

$$s\sqrt{\frac{n-1}{c_2}} < \sigma < s\sqrt{\frac{n-1}{c_1}}$$

oziroma:

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

13. Isti podatki kot pri 6. nalogi, le da ocenjujemo σ.

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

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

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

15. 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 $\mathrm{N}(\mu, \sigma)$. Poiščite 99% interval zaupanja za μ in za σ (za vsakega posebej, pri čemer privzemite, da drugi parameter ni znan).

16. Pri statistični spremenljivki X s končnim četrtim momentom bi želeli oceniti standardni odklon σ, na voljo pa imamo vzorec neodvisnih opažanj $X_1, \ldots X_n$. Modificirajte neopazljivo spremenljivko:

$$\sum_{i=1}^{n} (X_i - \mu)^2$$

v opazljivo funkcijo parametra σ, ki ima, če je v argumentu prava vrednost parametra σ, asimptotično normalno porazdelitev. Na podlagi tega konstruirajte asimptotični interval zaupanja za σ.

Asimptotični interval zaupanja za populacijski standardni odklon pri ne-Gaussovih spremenljivkah

$$\hat{\sigma}^2 - c\,\frac{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}{\sqrt{n}} < \sigma^2 < \hat{\sigma}^2 + c\,\frac{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}{\sqrt{n}}\,,$$

kjer je $c = z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$.

17. Poiščite 95% interval zaupanja za standardni odklon števila otrok na žensko za podatke iz 8. naloge:

število otrok	0	1	2	3	4	5	6	7	8	9	10
število žensk	227	168	320	96	29	11	4	2	2	0	1

13. Testi značilnosti

p-vrednosti. Test na podlagi razmerja verjetij. Nerandomizirani test neznane verjetnosti. Neyman–Pearsonova lema. Wilksov izrek. Z-test neznane verjetnosti. Z- in T-test sredine. T-test razlike sredin (za parne in neparne vzorce). Analiza variance z enojno klasifikacijo. Testiranje disperzije. Test skladnosti s fiksno porazdelitvijo in z družino porazdelitev. Test z znaki. Inverzijski test.

Želeli bi testirati, ali so opaženi podatki v vzorcu v skladu z ničelno hipotezo H_0 o porazdelitvi 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:*

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

Če se da, test načrtujemo tako, da je α natančna zgornja meja, z drugimi besedami, da je stopnja značilnosti **eksaktna**.
Če ničelno hipotezo zavrnemo pri $\alpha = 0\text{˙}05$, *pravimo, da so odstopanja statistično* **značilna**. *Če se to zgodi pri pragu* $0\text{˙}01$, *pa pravimo, da so statistično* **zelo značilna**.
Moč testa *je natančna spodnja meja za verjetnost, da H_0 zavrnemo v primeru, ko velja H_1.*

Odločanje o tem, ali bomo hipotezo zavrnili ali ne, navadno temelji na **testni statistiki**, *t. j. opazljivi spremenljivki z vrednostmi v* $\mathbb{R}$. *Hipotezo zavrnemo, če testna statistika pade v* **kritično območje**, *ki ga navadno označimo s* K_α. *Lahko si predstavljamo, da testna statistika* T *meri, koliko opažanje "ustreza" ničelni hipotezi, npr. večja kot je vrednost* T, *"ustreznejše" je opažanje. Tako hipotezo zavrnemo, če je* $T \le c$ *ali pa* $T < c$. *Pragu* c *pravimo* **kritična vrednost**. *Da bi dosegli eksaktno stopnjo značilnosti, lahko test tudi* **randomiziramo**: *če je* $T < c$, *hipotezo zavrnemo, če je* $T > c$, *je ne zavrnemo, če je* $T = c$, *pa jo zavrnemo z določeno verjetnostjo.*
Lahko si pomagamo tudi s p-vrednostmi: definirajmo $p(t) := \bar{P}(T' \le t)$, *kjer smo s* $\bar{P}(A)$ *označili supremum vseh verjetnosti dogodka* A *pri ničelni hipotezi,* T' *pa je kopija testne statistike* T *(pomožna slučajna spremenljivka, porazdeljena tako kot* T*). Hipotezo zavrnemo, brž ko je* $p(T) \le \alpha$. *Če to ni res, lahko ravnamo na dva načina: ali je ne zavrnemo ali pa* **randomiziramo**: *definiramo še* $p^-(t) := \bar{P}(T' < t)$. *Če je* $p^-(T) > \alpha$, *hipoteze nikakor ne zavrnemo, sicer pa jo zavrnemo s pogojno verjetnostjo:*

$$\frac{\alpha - p^-(T)}{p(T) - p^-(T)}\,.$$

Test na podlagi razmerja verjetij

Naj bo statistični model parametriziran s $\theta \in \Theta$. *Testiramo ničelno hipotezo, da je* $\theta \in \Theta_{H_0}$, *proti alternativni hipotezi, da je* $\theta \in \Theta_{H_1} := \Theta \setminus \Theta_{H_0}$. *Test na podlagi razmerja verjetij temelji na testni statistiki:*

$$\Lambda = \frac{\sup_{\Theta_{H_0}} L}{\sup_{\Theta} L}\,,$$

kjer je L *verjetje, t. j. verjetnostna funkcija ali gostota, evaluirana na opažanju. Večje kot je razmerje verjetij, ustreznejši je vzorec. Z drugimi besedami, ničelno hipotezo zavrnemo, če je razmerje verjetij* **premajhno**. *Lahko vzamemo tudi:*

$$\Lambda' = \frac{\sup_{\Theta_{H_0}} L}{\sup_{\Theta_{H_1}} L}\,.$$

1. Loterija za neko vrsto srečke trdi, da jih je vsaj pol dobitnih. Kupili smo osem srečk in le dve sta bili dobitni. Lahko pri stopnji značilnosti $\alpha = 0{\cdot}05$ začnemo sumiti, da loterija laže? Seveda privzamemo, da so posamezne kupljene srečke med seboj neodvisne.

2. Pri 20 metih kovanca je padlo 5 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 $\alpha = 0{\cdot}01$? Uporabite test na podlagi razmerja verjetij.

3. Na neki fakulteti študira 70% žensk in 30% moških. Posebno priznanje za izjemne študijske dosežke je bilo podeljeno 5 ženskam in 7 moškim. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte ničelno hipotezo, da so izjemni študijski dosežki (po merilih komisije) neodvisni od spola, proti alternativni hipotezi, da so od spola odvisni. Privzemite, da se spoli posameznikov, ki dobijo nagrado, obnašajo kot Bernoullijevo zaporedje poskusov.

Neznana verjetnost (enostavnejši, nerandomiziran test)

Naj bo θ verjetnost, da se zgodi določen dogodek. Pri stopnji značilnosti α testiramo ničelno hipotezo H_0, ki pravi, da je $\theta = \theta_0$. Izvedemo n poskusov in dani dogodek se zgodi pri natanko S poskusih. Ničelno hipotezo zavrnemo, če je $\rho(S) \le \alpha$, kjer je funkcija ρ odvisna od alternativne hipoteze:

- *pri* $H_1\colon\ \theta > \theta_0$ *je* $\rho(k) = P(S' \ge k)$;
- *pri* $H_1\colon\ \theta < \theta_0$ *je* $\rho(k) = P(S' \le k)$;
- *pri* $H_1\colon\ \theta \ne \theta_0$ *je* $\rho(k) = 2\min\{P(S' \le k), P(S' \ge k)\}$.

Tu je $S' \sim \mathrm{Bin}(n, \theta_0)$.

4. Uporabite zgornjo različico testa na podatkih iz prejšnje naloge. Izvedite test še proti alternativni hipotezi, da so izjemni študijski dosežki pristranski v korist moških. Komentirajte!

5. Prireditelj neke igre na srečo navaja, da je verjetnost, da bo v posamezni igri izžreban dobitek, enaka 1/50. To želimo testirati na osnovi opažanja, po kolikšnem številu iger je dobitek prvič izžreban. Natančno opišite izvedbo ustreznega randomiziranega testa za stopnjo značilnosti $\alpha = 0{\cdot}1$. Alternativna hipoteza naj bo, da je verjetnost, da bo dobitek izžreban, manjša od 1/50.

Neyman–Pearsonova lema

Če ima parameter le dve možni vrednosti (t. j. statistični model obsega le dve verjetnostni meri), je test na podlagi razmerja verjetij najmočnejši.

6. Življenjska doba originalne žarnice je porazdeljena eksponentno s pričakovano vrednostjo 500 ur, življenjska doba ponaredka pa je porazdeljena eksponentno s pričakovano vrednostjo 100 ur. Na podlagi opažene življenjske dobe ene žarnice testiramo ničelno hipotezo, da je originalna, proti alternativni hipotezi, da je ponaredek. Konstruirajte najmočnejši test pri stopnji značilnosti $\alpha = 0{\cdot}05$. Kolikšna je njegova moč?

7. Tine je zgeneriral 100 slučajnih števil, ki so neodvisna in porazdeljena normalno $\mathrm{N}(0,1)$, Tone pa je zgeneriral 10.000 takih števil. Oba povesta povprečje števil, ki sta jih dobila.

   a) Zapišite porazdelitev Tinetovega in Tonetovega povprečja.

   b) Zapomnimo si enega izmed povprečij, ne pa tudi, čigavo je. Konstruirajte najmočnejši test, ki na podlagi tega opažanja pri stopnji značilnosti $\alpha = 0{\cdot}05$ testira ničelno hipotezo, da je opaženo povprečje Tonetovo, proti alternativni hipotezi, da je Tinetovo. Kolikšna je njegova moč?

Asimptotično obnašanje testa na podlagi razmerja verjetij (Wilksov izrek)

Denimo, da opazimo vzorec n neodvisnih opažanj iz eksponentne družine z naravnimi parametri $\gamma_1, \ldots \gamma_m$ in odprtim naravnim parametričnim prostorom. Testiramo ničelno hipotezo H_0: $\zeta_1 = z_1, \ldots \zeta_r = z_r$ za karakteristike $\zeta_1, \ldots \zeta_r$ ($r \le m$), ki naj bodo funkcije naravnih parametrov $\gamma_1, \ldots \gamma_m$. Pri veljavnosti H_0 naj imajo komponente $h_1(X), \ldots h_m(X)$ pripadajoče zadostne statistike končne druge momente in neizrojeno kovariančno matriko. Nadalje, če H_0 gledamo kot ustrezno podmnožico parametričnega prostora, naj bo v kakšni njeni okolici funkcija, ki naravnim parametrom priredi vektor karakteristik, dvakrat parcialno zvezno odvedljiva in matrika njenih prvih parcialnih odvodov naj ima na H_0 maksimalni rang. Označimo z Λ razmerje verjetij. Tedaj pri H_0 velja:

$$-2 \ln \Lambda \xrightarrow[n\to\infty]{\mathrm{d}} \chi^2(r)\,.$$

8. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Na voljo imamo vzorec $X_1, X_2, \ldots X_n$, kjer so vse spremenljivke v vzorcu neodvisne in imajo predpisano porazdelitev.

   Opazimo $n = 100$, $\sum_{i=1}^n X_i = 37$, $\sum_{i=1}^n X_i^2 = 75$. Pri stopnji značilnosti $\alpha = 0{\cdot}01$ na podlagi razmerja verjetij testirajte hipotezo, da je $\mu = \sigma^2$, proti alternativni hipotezi, da to ni res.

Neznana verjetnost pri veliko poskusih

Naj bo θ verjetnost, da se zgodi določen dogodek uspeha. Testiramo ničelno hipotezo H_0, ki pravi, da je $\theta = \theta_0$. Izvedemo n poskusov in pri dani dogodek se zgodi pri natanko S poskusih. Tokrat privzamemo, da je število poskusov dovolj veliko: za minimalno sprejemljivo natančnost testa je potrebno, da je $n\theta_0 \ge 5$ in $n(1-\theta_0) \ge 5$. Glede na alternativno hipotezo H_1 ničelno hipotezo zavrnemo:

- *pri* $H_1\colon\ \theta \ne \theta_0$: *če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\left|\dfrac{S}{n} - \theta_0\right| - \dfrac{1}{2n}\right) \ge z_{(1-\alpha)/2}$;
- *pri* $H_1\colon\ \theta > \theta_0$: *če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\dfrac{S}{n} - \theta_0 - \dfrac{1}{2n}\right) \ge z_{1-\alpha}$;
- *pri* $H_1\colon\ \theta < \theta_0$: *če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\dfrac{S}{n} - \theta_0 + \dfrac{1}{2n}\right) \le -z_{1-\alpha}$.

Z $z_p = \Phi^{-1}\big(p - \frac{1}{2}\big)$ *smo označili kvantil standardne normalne porazdelitve za verjetnost* p.

Z-test

Z-test temelji na standardni normalni porazdelitvi. Natančneje, Z-test na testni statistiki Z s popravkom δ ima tri različice:

- **Dvostranska različica** *zavrne ničelno hipotezo, če je* $|Z| - \delta \ge z_{1-\alpha/2}$;
- **Enostranska različica v desno** *zavrne ničelno hipotezo, če je* $Z - \delta \ge z_{1-\alpha}$;
- **Enostranska različica v levo** *zavrne ničelno hipotezo, če je* $Z + \delta \le -z_{1-\alpha}$.

9. Tovarna jamči, da je delež izdelkov z napako enak 20%. V vzorcu 100 izdelkov pa jih je 24 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 120 izdelkov izmed 500? In kaj, če bi v slednjem primeru vzeli stopnjo značilnosti $0\text{˙}01$?

10. 10000-krat vržemo kovanec in 5090-krat je padel grb. Pri stopnji značilnosti $\alpha = 0\text{˙}05$ testirajte hipotezo, da je kovanec pošten, proti alternativni hipotezi, da ni pošten. Kaj pa, če bi za alternativno hipotezo vzeli, da grb pade z večjo verjetnostjo kot cifra?

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

Sredina pri normalni porazdelitvi z znanim standardnim odklonom

Testiramo ničelno hipotezo $H_0\colon\ \mu = \mu_0$. *Ker pri* H_0 *velja:*

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

izvedemo Z*-test brez popravka na testni statistiki* Z*, in sicer:*

- **dvostransko različico**, *če* H_1 *trdi, da je* $\mu \neq \mu_0$*;*
- **enostransko različico v desno**, *če* H_1 *trdi, da je* $\mu > \mu_0$*;*
- **enostransko različico v levo**, *če* H_1 *trdi, da je* $\mu < \mu_0$*;*

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

$$101,\ 91,\ 93,\ 103,\ 91,\ 101,\ 103,\ 95,\ 95$$

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ 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 $\mu > 100$?

Sredina pri normalni porazdelitvi z neznanim standardnim odklonom

Testiramo ničelno hipotezo $H_0\colon\ \mu = \mu_0$. *Ker pri* H_0 *velja:*

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

kjer je s *definiran tako kot pri konstrukciji intervala zaupanja za ta primer, ničelno hipotezo zavrnemo:*

- *proti* $H_1\colon\ \mu \neq \mu_0$*, če je* $|T| \geq t_{1-\alpha/2}(n-1)$*;*
- *proti* $H_1\colon\ \mu > \mu_0$*, če je* $T \geq t_{1-\alpha}(n-1)$*;*
- *proti* $H_1\colon\ \mu < \mu_0$*, če je* $T \leq -t_{1-\alpha}(n-1)$*.*

Tu je $t_p(df)$ *kvantil Studentove porazdelitve z* df *prostostnimi stopnjami za verjetnost* p*.*

T-test

Prej opisani test je primer T*-testa. Ta temelji na Studentovi porazdelitvi. Natančneje,* T*-test na testni statistiki* T *z* df *prostostnimi stopnjami ima tri različice:*

- **Dvostranska različica** *zavrne ničelno hipotezo, če je* $|T| \geq t_{1-\alpha/2}(df)$*;*
- **Enostranska različica v desno** *zavrne ničelno hipotezo, če je* $T \geq t_{1-\alpha}(df)$*;*
- **Enostranska različica v levo** *zavrne ničelno hipotezo, če je* $T \leq -t_{1-\alpha}(df)$*.*

Tu je $t_p(df)$ *kvantil Studentove porazdelitve z* df *prostostnimi stopnjami za verjetnost* p*.*

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

$$109,\ 107,\ 95,\ 121,\ 107,\ 97,\ 111,\ 121,\ 95$$

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

13. 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 $\alpha = 0{\cdot}05$ 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 $\mu < 50$?

Enakost sredin dveh statističnih spremenljivk na istih opažanjih (Gaussov model)

Večkrat hkrati izmerimo statistični spremenljivki X in Y. Privzamemo, da je $(X, Y) \sim \mathrm{N}(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, pri čemer pa se μ_X in μ_Y lahko spreminjata od meritve do meritve. Če meritve izhajajo iz enostavnega slučajnega vzorca iz velike populacije, to velja, če je porazdelitev vektorja (X, Y) na populaciji mešanica porazdelitev $\mathrm{N}(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$, kjer se μ_X in μ_Y lahko spreminjata. Testiramo ničelno hipotezo H_0, da je ves čas $\mu_X = \mu_Y$, alternativna hipoteza pa je lahko:
- *da ni ves čas $\mu_X = \mu_Y$;*
- *da je ves čas $\mu_X \geq \mu_Y$ in vsaj kdaj $\mu_X > \mu_Y$;*
- *da je ves čas $\mu_X \leq \mu_Y$ in vsaj kdaj $\mu_X < \mu_Y$.*

Testiramo tako, da testiramo sredino razlike $X - Y$ glede na 0 (pri normalni porazdelitvi z neznanim standardnim odklonom).

14. 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 $\alpha = 0{\cdot}05$ 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.

Enakost sredin dveh statističnih spremenljivk na različnih opažanjih (homoskedastični Gaussov model)

Naj bo $X \sim \mathrm{N}(\mu_X, \sigma)$ *in* $Y \sim \mathrm{N}(\mu_Y, \sigma)$. *Pri statistični spremenljivki* X *opazimo* $X_1, \ldots X_m$, *pri* Y *pa* $Y_1, \ldots Y_n$. *Privzamemo, da so vsa opažanja med seboj neodvisna. Če opažanja temeljijo na jemanju vzorcev, sta lahko* X *in* Y *definirani na različnih populacijah.*
Testiramo hipotezo $H_0\colon\ \mu_X = \mu_Y$. *Ker pri* H_0 *velja:*

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

kjer je:

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

izvedemo T*-test na testni statistiki* T *z* $m + n - 2$ *prostostnimi stopnjami, in sicer:*

- **dvostransko različico**, *če* H_1 *trdi, da je* $\mu_X \neq \mu_Y$;
- **enostransko različico v desno**, *če* H_1 *trdi, da je* $\mu_X > \mu_Y$;
- **enostransko različico v levo**, *če* H_1 *trdi, da je* $\mu_X < \mu_Y$;

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

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

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

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

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je $\mu_X = \mu_Y$, proti alternativni hipotezi, da je $\mu_X \neq \mu_Y$.

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

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

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

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

Pri stopnji značilnosti $\alpha = 0{\cdot}01$ testirajte hipotezo, da je $\mu_X = \mu_Y$, proti alternativni hipotezi, da je $\mu_X > \mu_Y$.

Enakost sredin več normalnih statističnih spremenljivk: analiza variance (ANOVA) z enojno klasifikacijo

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. V skladu s tem ničelno hipotezo zavrnemo, če je $F \ge F_{1-\alpha/2}(k-1, n-k)$.

F-test

Prej opisani test je primer F-testa. Ta temelji na Fisher–Snedecorjevi porazdelitvi. Natančneje, F-test na testni statistiki F z (df_1, df_2) prostostnimi stopnjami ima tri različice:

- **Dvostranska različica**: *H_0 zavrnemo, če je $F \le F_{\alpha/2}(df_1, df_2)$ ali $F \ge F_{1-\alpha/2}(df_1, df_2)$;*
- **Enostranska različica v desno**: *H_0 zavrnemo, če je $F \ge F_{1-\alpha}(df_1, df_2)$;*
- **Enostranska različica v levo**: *H_0 zavrnemo, če je $F \le F_{\alpha}(df_1, df_2)$.*

Tu je $F_p(df_1, df_2)$ kvantil Snedecorjeve porazdelitve z (df_1, df_2) prostostnimi stopnjami za verjetnost p.

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

Standardni odklon pri normalni porazdelitvi

Testiramo ničelno hipotezo $H_0\colon\ \sigma = \sigma_0$. Ker pri H_0 velja:

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

ničelno hipotezo zavrnemo:

- *proti $H_1\colon\ \sigma \neq \sigma_0$, če je $\chi^2 \le \chi^2_{\alpha/2}(n-1)$ ali $\chi^2 \ge \chi^2_{1-\alpha/2}(n-1)$;*
- *proti $H_1\colon\ \sigma > \sigma_0$, če je $\chi^2 \ge \chi^2_{1-\alpha}(n-1)$;*
- *proti $H_1\colon\ \sigma < \sigma_0$, če je $\chi^2 \le \chi^2_{\alpha}(n-1)$.*

Tu je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost p.

Test hi kvadrat

Prej opisani test je primer testa hi kvadrat. Ta temelji na porazdelitvi hi kvadrat. Natančneje, test hi kvadrat na testni statistiki χ^2 z df prostostnimi stopnjami ima tri različice:

- **Enostranska različica v desno** *zavrne ničelno hipotezo, če je $\chi^2 \ge \chi^2_{1-\alpha}(df)$;*
- **Enostranska različica v levo** *zavrne ničelno hipotezo, če je $\chi^2 \le \chi^2_{\alpha}(df)$;*
- **Dvostranska različica** *zavrne ničelno hipotezo, če je $\chi^2 \le \chi^2_{\alpha/2}(df)$ ali $\chi^2 \ge \chi^2_{1-\alpha/2}(df)$.*

Tu je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost p.

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

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

Pri $\alpha = 0{\cdot}05$ 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 $\sigma < 10$.

Pearsonov test skladnosti s fiksno porazdelitvijo

Testiramo, ali je porazdelitev dane statistične spremenljivke enaka:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & \cdots & p_r \end{pmatrix} \qquad (r \geq 3)\,.$$

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** $\tilde{N}_i := np_i$. *Ker tedaj pri veljavnosti ničelne hipoteze približno velja:*

$$\chi^2 := \sum_{i=1}^{r} \frac{(N_i - \tilde{N}_i)^2}{\tilde{N}_i} \sim \chi^2(r-1)$$

hipotezo o porazdelitvi testiramo s testom hi kvadrat na testni statistiki χ^2 z $r-1$ prostostnimi stopnjami, in sicer enostransko v desno.
Za naše potrebe je test dovolj natančen, če je $r \geq 3$ in $\tilde{N}_i \geq 5$ za vse i. Če dobimo $\tilde{N}_i < 5$, lahko razrede združimo. Za $r = 2$ pa lahko uporabimo kar dvostranski test uspeha poskusa.

19. 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.

20. V vzorcu so 2 osebka tipa RR, 5 tipa Rr in 4 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 20 osebkov tipa RR, 50 tipa Rr in 50 tipa rr?

21. Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte hipotezo, da ima statistična spremenljivka X *standardno* Laplaceovo porazdelitev, t. j. zvezno z gostoto:

$$p_X(x) = \frac{1}{2}\, e^{-|x|}\,,$$

    če ima vzorec naslednjo frekvenčno porazdelitev:

pod -1	od -1 do 0	od 0 do 1	nad 1
32	20	21	27

22. Podatki imajo naslednjo frekvenčno porazdelitev:

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

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

Test skladnosti z družino porazdelitev Pearsonovega tipa

Testiramo, ali porazdelitev dane statistične spremenljivke pripada družini:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1(\theta_1, \ldots \theta_k) & p_2(\theta_1, \ldots \theta_k) & \cdots & p_r(\theta_1, \ldots \theta_k) \end{pmatrix} \qquad (r \ge 3)\,,$$

kjer nabor $(\theta_1, \ldots \theta_k)$ preteče odprto množico v $\mathbb{R}^k$, vektorska funkcija $(p_1, \ldots p_r)$ pa je dvakrat zvezno odvedljiva vložitev. Vzorec velikosti n naj ima frekvenčno porazdelitev:

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

Izračunamo **teoretične frekvence** *$\tilde{N}_i := np_i(\hat{\theta}_1, \ldots \hat{\theta}_k)$, kjer so $\hat{\theta}_1, \ldots \hat{\theta}_k$ ocene za $\theta_1, \ldots \theta_k$ po metodi največjega verjetja. Tedaj pri veljavnosti ničelne hipoteze približno velja:*

$$\chi^2 := \sum_{i=1}^{r} \frac{(N_i - \tilde{N}_i)^2}{\tilde{N}_i} \sim \chi^2(r - 1 - k)\,.$$

Torej izvedemo test hi kvadrat z $df = r - 1 - k$, in sicer enostransko različico v desno. Test je dovolj natančen, če je $\tilde{N}_i \ge 5$ za vse i.

23. V vzorcu je 20 osebkov tipa RR, 30 tipa Rr in 50 tipa rr. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da populacija ustreza *Hardy–Weinbergovemu modelu*, t. j. da so deleži osebkov tipa RR, Rr in rr enaki $(1-\theta)^2$, $2\theta(1-\theta)$ in θ^2.

24. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je statistična spremenljivka X porazdeljena diskretno po naslednji shemi:

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

če je v vzorcu 20 enot z vrednostjo -1, 30 enot z vrednostjo 0, 30 enot z vrednostjo 1 in 20 enot z vrednostjo 2.

25. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da ima statistična spremenljivka X *splošno usredinjeno* Laplaceovo porazdelitev, t. j. zvezno z gostoto:

$$p_X(x) = \frac{\lambda}{2}\, e^{-\lambda|x|}\,; \quad \lambda > 0\,,$$

če ima vzorec isto frekvenčno porazdelitev kot v 21. nalogi.

Test z znaki

Večkrat hkrati izmerimo urejenostni statistični spremenljivki X in Y. Privzamemo, da sta X in Y ob vsaki meritvi neodvisni, sicer pa se lahko porazdelitvi od meritve do meritve spreminjata. Če meritve izhajajo iz enostavnega slučajnega vzorca iz velike populacije, to velja, če je porazdelitev vektorja (X, Y) na populaciji mešanica porazdelitev parov neodvisnih slučajnih spremenljivk. Testiramo ničelno hipotezo H_0, da sta X in Y ob vsaki meritvi enako porazdeljeni, alternativna hipoteza pa je lahko:

- $H_1^{\pm}$, *da X in Y nista ves čas enako porazdeljeni;*
- H_1^{+}, *da je X vsaj kdaj stohastično strogo večja od Y, t. j. $F_X \le F_Y$, za določen x pa tudi $F_X(x) < F_Y(x)$;*
- H_1^{-}, *da je X vsaj kdaj stohastično strogo manjša od Y, t. j. $F_X \ge F_Y$, za določen x pa tudi $F_X(x) > F_Y(x)$.*

Naj bo n število meritev, z S^+ označimo število meritev, pri katerih je $X > Y$, z S^- pa število meritev, pri katerih je $X < Y$. Naj bo še $\tilde{n} = S^+ + S^-$ število meritev, pri katerih je $X \ne Y$ (primere, kjer pride enako, torej preprosto izločimo). Ničelno hipotezo zavrnemo, če je $p(S^+, S^-) \le \alpha$. Funkcija p (p-vrednost) je odvisna od ničelne hipoteze in ustreza testu uspeha poskusa pri ničelni hipotezi, da je le-ta enaka $1/2$:

- *pri H_1^+ postavimo $p(k^+, k^-) = P(S' \ge k^+) = P(S' \le k^-)$;*
- *pri H_1^- postavimo $p(k^+, k^-) = P(S' \ge k^-) = P(S' \le k^+)$;*
- *pri $H_1^{\pm}$ postavimo $p(k^+, k^-) = 2\min\{P(S' \ge k^+), P(S' \ge k^-)\} = 2\min\{P(S' \le k^+), P(S' \le k^-)\}$.*

Tu je $S' \sim \mathrm{Bin}(\tilde{n}, 1/2)$.

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

X_i	28	14	16	16	31	17	13	14	12	13
Y_i	26	29	31	18	37	10	19	33	23	45

Pri stopnji značilnosti $\alpha = 0{,}05$ testirajte hipotezo, da sta X in Y vsakič enako porazdeljeni, proti alternativni, da je X ves čas stohastično manjša od Y in vsaj kdaj stohastično strogo manjša od Y. Nato naredite ustrezni enostranski test povprečij.

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

X_i	25	28	30	23	28	26	29	23	33	21	33	28
Y_i	35	27	29	21	18	25	28	27	31	19	32	26

Pri stopnji značilnosti $\alpha = 0{,}05$ testirajte hipotezo, da sta X in Y vsakič enako porazdeljeni, proti alternativni, da je X ves čas stohastično večja od Y in vsaj kdaj stohastično strogo večja od Y. Nato naredite ustrezni enostranski test povprečij.

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

X_i	30	24	22	28	26	19	25	31	36	21	25	26	29	29	19	18
Y_i	28	121	21	25	25	17	122	129	34	20	22	23	126	26	18	17

Pri stopnji značilnosti $\alpha = 0'05$ testirajte hipotezo, da sta X in Y vsakič enako porazdeljeni, proti obema enostranskima alternativnima hipotezama. Nato naredite še ustrezna enostranska testa povprečij.

Test z znaki za veliko število meritev

Če je število meritev dovolj veliko (za naše potrebe vsaj 10), lahko test z znaki nadomestimo s približnim testom, pri katerem ničelno hipotezo zavrnemo:

- *proti* H_1^+, *če je* $\dfrac{S_+ - S_- - 1}{\sqrt{S^+ + S^-}} \geq z_{1-\alpha}$;
- *proti* H_1^-, *če je* $\dfrac{S_+ - S_- + 1}{\sqrt{S^+ + S^-}} \leq -z_{1-\alpha}$.
- *proti* $H_1^\pm$, *če je* $\dfrac{|S_+ - S_-| - 1}{\sqrt{S^+ + S^-}} \geq z_{1-\alpha/2}$;

Z drugimi besedami, izvedemo Z-test na testni statistiki $\dfrac{S_+ - S_-}{\sqrt{S_+ + S_-}}$ *s popravkom* $\dfrac{1}{\sqrt{S_+ + S_-}}$.

29. 50 ljudi so pred ogledom in po ogledu filma povprašali, kako se počutijo: zelo slabo, slabo, srednje, dobro ali zelo dobro. Rezultati so naslednji:

pred	po
srednje	srednje
dobro	zelo dobro
srednje	zelo dobro
dobro	srednje
srednje	zelo dobro
dobro	dobro
srednje	dobro
dobro	dobro
dobro	zelo dobro
zelo dobro	zelo dobro
dobro	zelo dobro
zelo dobro	dobro
dobro	srednje
zelo dobro	srednje
srednje	dobro
srednje	dobro
dobro	zelo dobro
srednje	dobro
dobro	zelo dobro
zelo dobro	dobro
zelo dobro	zelo dobro
zelo dobro	dobro
slabo	dobro
dobro	srednje
srednje	zelo dobro

pred	po
dobro	zelo dobro
dobro	dobro
zelo dobro	zelo dobro
dobro	dobro
srednje	zelo slabo
srednje	zelo dobro
zelo dobro	srednje
dobro	dobro
dobro	dobro
srednje	slabo
slabo	srednje
srednje	srednje
zelo slabo	slabo
slabo	srednje
slabo	srednje
slabo	zelo dobro
zelo slabo	srednje
srednje	slabo
srednje	slabo
zelo slabo	srednje
srednje	dobro
slabo	zelo dobro
slabo	slabo
slabo	slabo
zelo slabo	srednje

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte ničelno hipotezo, da ogled filma ne spremeni počutja, proti alternativni hipotezi, da ga spremeni.

Inverzijski (Wilcoxon–Mann–Whitneyjev) test

Testiramo, ali sta **urejenostni** *statistični spremenljivki X in Y enako porazdeljeni. Pri statistični spremenljivki X opazimo $X_1, \ldots X_m$, pri Y pa $Y_1, \ldots Y_n$. Privzamemo, da so vsa opažanja med seboj neodvisna. Če opažanja temeljijo na jemanju vzorcev, sta lahko X in Y definirani na različnih populacijah.*

Opažene vrednosti združimo in jih uredimo po velikosti, recimo od namanjše do največje. Naj bodo $R_1, \ldots R_m$ mesta (rangi), ki pripadajo opažanjem spremenljivke X. Privzemimo, da sta vzorca dovolj velika.

Tudi tu ločimo enostransko in dvostransko različico testa. Obravnavali bomo torej tri alternativne hipoteze:

- *$H_1^\pm$, da X in Y nista enako porazdeljeni;*
- *H_1^+, da je X stohastično strogo večja od Y, t. j. $F_X \le F_Y$, za določen x pa tudi $F_X(x) < F_Y(x)$;*
- *H_1^-, da je X stohastično strogo manjša od Y, t. j. $F_X \ge F_Y$, za določen x pa tudi $F_X(x) > F_Y(x)$.*

Če je število meritev dovolj veliko, ničelno hipotezo zavrnemo:

- *proti $H_1^\pm$, če je* $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(\left|2\sum_{i=1}^m R_i - m(m+n+1)\right| - 1\right) \ge z_{1-\alpha/2}$;
- *proti H_1^+, če je* $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i - m(m+n+1) - 1\right) \ge z_{1-\alpha}$;
- *proti H_1^-, če je* $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i - m(m+n+1) + 1\right) \le -z_{1-\alpha}$.

Z drugimi besedami, izvedemo Z-test na testni statistiki

$$\sqrt{\frac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i - m(m+n+1)\right) \text{ s popravkom } \sqrt{\frac{3}{mn(m+n+1)}}.$$

30. Tekmovalci dveh ekip, "zelenih" in "oranžnih", so se pomerili v teku. Vrstni red tekmovalcev je naslednji:

$$Z,\ Z,\ O,\ Z,\ Z,\ O,\ Z,\ Z,\ O,\ Z,\ O,\ O,\ O,\ Z,\ O,\ O,\ O,\ O,\ Z,\ O$$

(t. j. prvi, ki je prispel na cilj, je bil član "zelenih", drugi prav tako, tretji je bil član "oranžnih" itd.). Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da so zeleni enako dobri kot oranžni, proti alternativni hipotezi, da je med njimi razlika.

31. Dijaki so se pomerili v teku na 60 metrov. Določeni so izjavili, da so se prej pripravljali, določeni pa, da ne. Rezultati tistih, ki so se pripravljali, so:

$$7{\cdot}6\,,\ 7{\cdot}6\,,\ 7{\cdot}7\,,\ 7{\cdot}8\,,\ 7{\cdot}8\,,\ 8{\cdot}0\,,\ 8{\cdot}1\,,\ 8{\cdot}2\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 9{\cdot}3\,,$$

rezultati tistih, ki se niso pripravljali, pa so:

$$7{\cdot}9\,,\ 8{\cdot}2\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}4\,,\ 8{\cdot}7\,,\ 8{\cdot}8\,.$$

Z inverzijskim testom testirajte ničelno hipotezo, da tisti, ki se pripravljajo, tečejo enako kot tisti, ki se ne pripravljajo, proti alternativni hipotezi, da tisti, ki se pripravljajo, tečejo bolje od tistih, ki se ne pripravljajo. Kaj pa pravi T-test? Stopnja značilnosti naj bo obakrat $\alpha = 0{\text{˙}}05$.

32. Med 17 študenti so izvedli anketo z naslednjima vprašanjema:

    1. Ocenite stopnjo stresa pri vas v zadnjih dveh tednih.
       (1 – zelo majhna, 2 – majhna, 3 – srednja, 4 – velika, 5 – zelo velika)
    2. Ali ste se v zadnjih dveh tednih posvečali študiju bolj kot ponavadi?
       (da/ne)

    Rezultati ankete so:

	1	2	3	4	5
da	0	2	1	5	2
ne	0	5	2	0	0

    Pri stopnji značilnosti $\alpha = 0{\text{˙}}01$ testirajte hipotezo, ali so študenti, ki se v zadnjih dveh tednih bolj posvečajo študiju, v tem času enako pod stresom kot tisti, ki se ne, proti alternativni hipotezi, da so študenti, ki se v zadnjih dveh tednih bolj posvečajo študiju, v tem času bolj pod stresom kot tisti, ki se ne.

14. Povezanost dveh številskih spremenljivk

Interval zaupanja za korelacijski koeficient. Testiranje nekoreliranosti. Kontingenčni test. Enostavna linearna regresija.

Interval zaupanja za korelacijski koeficient

Naj bo $\rho = r(X,Y)$. *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 (Gaussov model)

Če velja ničelna hipoteza H_0, da sta X in Y nekorelirani, je približno:

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

kjer je R vzorčni korelacijski koeficient. Tako nekoreliranost testiramo s T-testom na testni statistiki T z $n-2$ prostostnimi stopnjami, in sicer z:

- *dvostransko različico, če H_1 trdi, da sta X in Y korelirani;*
- *enostransko različico v desno, če H_1 trdi, da sta X in Y pozitivno korelirani;*
- *enostransko različico v levo, če H_1 trdi, da sta X in Y negativno korelirani.*

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.

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 ($i = 1, \dots r$, $j = 1, \dots s$). 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:

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

neodvisnost testiramo s testom hi kvadrat z $(r-1)(s-1)$ prostostnimi stopnjami na testni statistiki χ^2, in sicer uporabimo enostransko različico v desno.

3. Na vzorcu 62 oseb dobimo naslednjo navzkrižno frekvenčno porazdelitev barve oči in las:

oči\lasje	rdeči, blond	rjavi, črni	Skupaj
modre	12	1	13
zelene	14	9	23
rjave	4	22	26
Skupaj	30	32	62

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

Kontingenčni test neodvisnosti za dihotomni spremenljivki

Če sta spremenljivki X in Y dihotomni in so navzkrižne frekvence podane s tabelo:

A	B
C	D

velja:

$$\chi^2 = \frac{(A+B+C+D)(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)}\,.$$

4. Rezultati ankete z dvema vprašanjema 'Ali verjamete v horoskop?' in 'Ali verjamete v NLP-je?' so zbrani v naslednji tabeli:

Horoskop\NLP	vsaj malo	ne	Skupaj
vsaj malo	5	7	12
ne	6	9	15
Skupaj	11	16	27

Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte neodvisnost verovanja ljudi v horoskop in v NLP-je.

5. Na neki spletni 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 $0\text{'}05$.

Enostavna 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:

$$\hat{b} = \frac{C_{xy}}{C_x^2}\,, \quad \hat{a} = \bar{Y} - \hat{b}\bar{X}$$

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:

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

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

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

8. a) $\dfrac{12!}{3!\,3!\,2!\,2!\,2!} = 1633200$,

b) $2 \cdot 3 \cdot \dfrac{7!}{3!\,2!\,2!} \cdot \dfrac{5!}{3!\,2!} = 12600$,

c) $\dfrac{7!}{2!\,2!} \cdot 5! = 151200$.

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. Označimo iskani dogodek z A. Tedaj velja:

$$P(A) = 1 - \frac{364 \cdot 363 \cdot \cdots \cdot (366 - n)}{365^{n-1}} .$$

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

Že pri 23 ljudeh verjetnost prvič preseže 1/2.

Splošneje, če je d dni v letu, velja:

$$P(A) = 1 - \left(1 - \frac{1}{d}\right)\left(1 - \frac{2}{d}\right)\cdots\left(1 - \frac{n-1}{d}\right).$$

Oglejmo si asimptotično obnašanje te verjetnosti, ko gre $d \to \infty$ (n pa se lahko spreminja z d). Najprej opazimo, da je dogodek A unija dogodkov A_{ij}, da imata i-ti in j-ti človek oba rojstni dan na isti dan ($i \neq j$). Vsi ti dogodki imajo verjetnosti $1/d$. Dogodka A_{ij} in A_{ji} sovpadata, sicer pa je plavzibilno, da so ti dogodki, če je n majhen v primerjavi z d, približno paroma nezdružljivi, zato je $P(A) \approx n(n-1)/(2d)$. Natančnejši argument pa nam da analiza. Z aproksimacijo $1 - x \approx e^{-x}$ in oceno logaritma ostankov se da izpeljati:

$$\begin{aligned}
&P(\overline{A}) \sim e^{-n(n-1)/(2d)} &&, \text{ brž ko je } n \ll d^{2/3};\\
&P(A) \sim 1 - e^{-n(n-1)/(2d)} &&, \text{ brž ko gre } d \to \infty;\\
&P(A) \sim \frac{n(n-1)}{2d} &&, \text{ brž ko je } n \ll \sqrt{d}.
\end{aligned}$$

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

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

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

10. Verjetnosti dobitkov lahko računamo na dva načina. Pri prvem načinu si predstavljamo, da so prekrižane številke fiksne, nakar gledamo vsa možna žrebanja (*pogled igralca*). Lahko pa si predstavljamo tudi, da so fiksne izžrebane številke, nakar gledamo vsa možna križanja (pogled Loterije). Dobimo:

$$\begin{aligned}
P(\text{sedmica}) &= \frac{1}{\binom{39}{7}} = \frac{1}{15.380.937} \doteq 6{\dot{}}50 \cdot 10^{-8}\\
P(\text{šest in dodatna}) &= \frac{\binom{7}{6}\binom{32}{1}\cdot 1}{\binom{39}{7}\cdot 32} = \frac{\binom{7}{6}\binom{1}{1}\binom{31}{0}}{\binom{39}{7}} \approx \frac{1}{2.197.277} \doteq 4{\dot{}}55 \cdot 10^{-7}\\
P(\text{šestica}) &= \frac{\binom{7}{6}\binom{32}{1}\cdot 31}{\binom{39}{7}\cdot 32} = \frac{\binom{7}{6}\binom{1}{0}\binom{31}{1}}{\binom{39}{7}} \approx \frac{1}{70.880} \doteq 1{\dot{}}41 \cdot 10^{-5}\\
P(\text{petica}) &= \frac{\binom{7}{5}\binom{32}{2}}{\binom{39}{7}} \approx \frac{1}{1477} \doteq 6{\dot{}}77 \cdot 10^{-4}\\
P(\text{štirica}) &= \frac{\binom{7}{4}\binom{32}{3}}{\binom{39}{7}} \approx \frac{1}{88{\dot{}}6} \doteq 0{\dot{}}0113\\
P(\text{tri in dodatna}) &= \frac{\binom{7}{3}\binom{32}{4}\cdot 4}{\binom{39}{7}\cdot 32} = \frac{\binom{7}{3}\binom{1}{1}\binom{31}{4}}{\binom{39}{7}} \approx \frac{1}{97{\dot{}}8} \doteq 0{\dot{}}0102
\end{aligned}$$

11. Mešanje kupa si lahko predstavljamo tako, da najprej na slepo v vrsto razporedimo vse rdeče in zelene karte (t. j. vseh 5! možnih permutacij, kjer vse karte ločimo, je enako verjetnih), nakar začnemo razporejati bele: najprej prvo belo karto razporedimo na 6 možnih pozicij, od skrajne leve do skrajne desne, vse z enako verjetnostjo pogojno na razporeditev prejšnjih kart. Nato na 7 možnih pozicij razporedimo drugo belo karto, spet na vse z enako pogojno verjetnostjo. Tako nadaljujemo, dokler niso razporejene vse karte. Na koncu z desne proti levi poberemo vse karte, tako da se najbolj desna karta znajde na dnu, najbolj leva pa na vrhu kupa. Tako so, če vse karte ločimo, vse permutacije enako verjetne.

a) Dogodek, da je prva rdeča karta izvlečena pred prvo zeleno, se ujema z dogodkom, da je v trenutku, ko so razporejene vse rdeče in zelene karte, prva karta rdeča. Ker si lahko predstavljamo, da rdeče in zelene karte razporejamo tako, da najprej na slepo razporedimo eno karto na skrajno levo pozicijo, nato eno od štirih preostalih na naslednjo pozicijo desno in tako naprej, je verjetnost tega dogodka enaka $2/5$.

b) Dogodek, da bo prva rdeča karta izvlečena pred zadnjo zeleno, se ujema s komplementom dogodka, da so v trenutku, ko so razporejene vse rdeče in zelene karte, z leve najprej tri zelene in nato dve rdeči. Lahko si predstavljamo, da na označenih pet pozicij najprej razporedimo obe rdeči karti, nato pa na preostale tri razporedimo zelene karte. Verjetnost dogodka, da sta obe rdeči karti skrajno desno, je enaka $\dfrac{1}{\binom{5}{2}} = \dfrac{1}{10}$, verjetnost iskanega dogodka pa je enaka $\dfrac{9}{10}$.

12. Dogodek, čigar verjetnost iščemo, je disjunktna unija naslednjih treh dogodkov:

- Asistent ne premesti niti Aljaža niti Brigite. Naj bo to dogodek S_0.
- Asistent premesti Aljaža na Cvetovo mesto, Brigite pa ne premesti. Naj bo to dogodek S_1.
- Asistent premesti Aljaža na Brigitino mesto, Brigito na Cvetovo mesto in Cveta na Aljaževo mesto. Naj bo to dogodek S_2.

Verjetnostni prostor lahko razdelimo na $9 \cdot 8 \cdot 7 = 504$ enako verjetne izide – glede na to, katere študente asistent izbere za premestitev in v katerem vrstnem redu. Vsak tak izid lahko torej ponazorimo z zaporedjem črk oblike xyz. Tako zaporedje si lahko predstavljamo kot povelje, naj gre x tja, kjer je bil prej y, y tja, kjer je bil prej z in z tja, kjer je bil prej x. Naj črka A označuje Aljaža, črka B Brigito, črka C Cveta, črke X, Y in Z katerega koli, ki ni Aljaž ali Brigita, črka W pa katerega koli, ki ni Aljaž, Brigita ali Cveto.

- Dogodek S_0 sestavljajo izidi oblike XYZ, ki jih je $7 \cdot 6 \cdot 5 = 210$.
- Dogodek S_1 sestavljajo vsi izidi oblike ACW, WAC in CWA. Teh je $6 \cdot 3 = 18$.
- Dogodek S_2 pa sestavljajo izidi oblike ABC, BCA in CAB, ki so trije.

Iskana verjetnost je tako enaka:

$$P(S_0) + P(S_1) + P(S_2) = \frac{210 + 18 + 3}{504} = \frac{231}{504} = \frac{11}{24} \doteq 0{\cdot}458\,.$$

Opomba. Po trije in trije izidi, kot smo jih definirali, določajo enako premestitev. Tako bi lahko za izide vzeli tudi kar premestitve. Teh je $\dfrac{9 \cdot 8 \cdot 7}{3} = 2 \cdot \dbinom{9}{3} = 168$.

13. $B \cap C$

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

15. 76/120. V splošnem, ko je kuvert n, je rezultat:

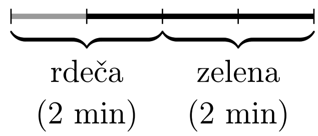
$$1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + \frac{(-1)^{n-1}}{n!} = 1 - \frac{1}{n!}\left\langle \frac{n!}{e} \right\rangle = \frac{1}{n!}\left\langle n!\left(1 - \frac{1}{e}\right)\right\rangle$$

16. a) Družina $\mathcal{A}_n$ vsebuje natanko množice K in $K \cup \{n+1, n+2, \ldots\}$, kjer je $K \subseteq \{1, 2, \ldots n\}$.

b) Družina $\mathcal{A}$ vsebuje vse enoelementne množice $\{2\}$, $\{4\}$, $\{6\}, \ldots$, ne vsebuje pa njihove unije, množice sodih naravnih števil.

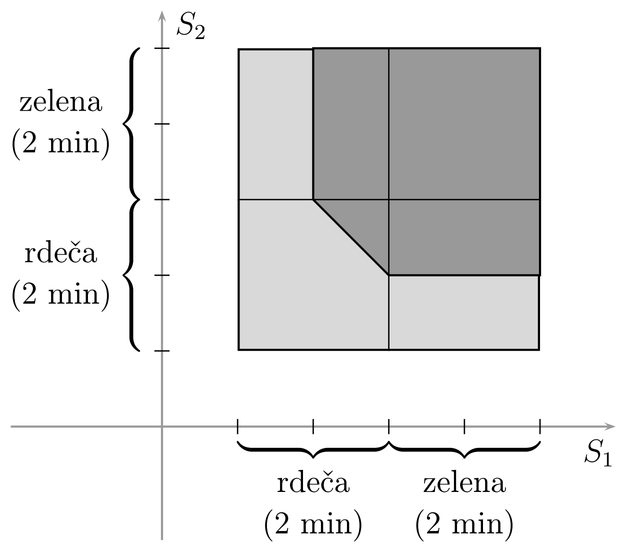
c) Najmanjša σ-algebra, ki vsebuje $\mathcal{A}$, je kar potenčna množica množice naravnih števil, saj vsaka σ-algebra, ki vsebuje $\mathcal{A}$, vsebuje vse podmnožice $\mathbb{N}$. Vsako podmnožico namreč lahko zapišemo v obliki $\{a_1, a_2, a_3, \ldots\} = \{a_1\} \cup \{a_2\} \cup \{a_3\} \cup \cdots$.

17. Za en semafor lahko verjetnostni prostor ponazorimo z naslednjo sliko:



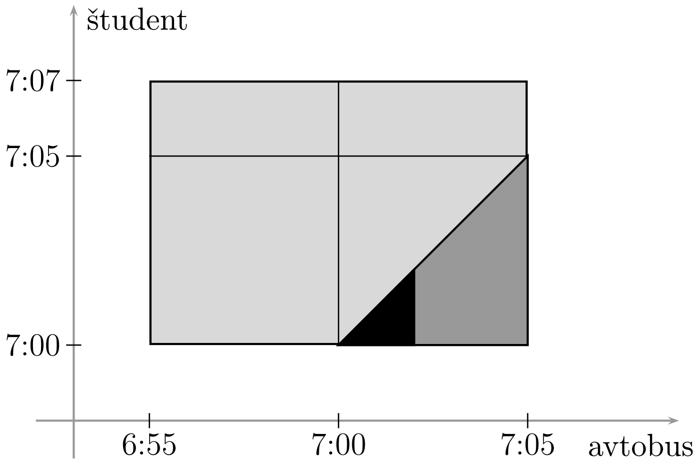
in verjetnost iskanega dogodka je $\dfrac{3}{4} = 0{\cdot}75$.

Za dva semaforja pa je slika naslednja:

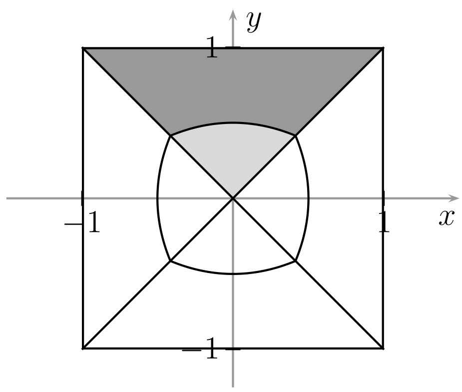


in verjetnost iskanega dogodka je $\dfrac{17}{32} \doteq 0{\cdot}531$.

18. a) $\dfrac{5}{28} \doteq 0{\cdot}179$, b) $\dfrac{1}{35} \doteq 0{\cdot}0286$. Slika:



19. Kvadrat razdelimo na štiri trikotnike z vrhovi v središču kvadrata in osnovnicami, ki se ujemajo s stranicami kvadrata (glej sliko). Zaradi simetrije je dovolj, če se omejimo na en sam trikotnik. Kvadrat postavimo v koordinatni sistem (x, y), tako da se središče kvadrata ujema z izhodiščem koordinatnega sistema:



V zgornjem trikotniku bo tedaj točka bližje robu kot središču kvadrata (sivo) natanko tedaj, ko bo $1 - y < \sqrt{x^2 + y^2}$. Po ureditvi (in ob upoštevanju lege točke) dobimo, da je to natanko tedaj, ko je $y > (1 - x^2)/2$. Mejna črta gre od točke $(-a, a)$ do točke (a, a), kjer je $a = 1/(1 + \sqrt{2}) = \sqrt{2} - 1$.

Ploščina celega trikotnika je 1, ploščina ugodnega območja (temnosivo) in tudi verjetnost našega dogodka pa je enaka:

$$(1-a)^2 + \int_{-a}^{a} \left(1 - \frac{1 - x^2}{2}\right) \mathrm{d}x = (1-a)^2 + \int_0^a (1 + x^2)\,\mathrm{d}x =$$

$$= 1 - a + a^2 + \frac{a^3}{3} = \frac{8 - 4\sqrt{2}}{3} \doteq 0{\cdot}781\,.$$

20. 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 $h \le b\cos\varphi$.

Slepo metanje igle je smiselno interpretirati tako, da sta razdalja h in kot φ izbrana na slepo in neodvisno iz ustreznih intervalov. V tem primeru dobimo:

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

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 $P(A \mid L) = 0\text{'}25$ 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. Ker so (na začetku) vse možnosti za vrata, za katerimi se skriva avto, enako verjetne, lahko brez škode za splošnost privzamemo, da najprej pokažemo na prva vrata. Igro lahko modeliramo z verjetnostnim prostorom iz naslednjih štirih izidov:

Izid	Verjetnost
AOB	1/6
ABO	1/6
BAO	1/3
BOA	1/3

Pri tem črka A pomeni avto za (še zaprtimi) vrati, črka B bučo za zaprtimi vrati, črka O pa odprta vrata. Iz tega se takoj vidi, da se je pri drugem izbiranju *bolje premisliti*, saj v tem primeru dobimo avto pri izidih BAO in BOA, torej z verjetnostjo 2/3 (če se ne premislimo, pa ga dobimo na nasprotnem dogodku, sestavljenem iz izidov AOB in ABO, ki ima verjetnost 1/3).

Zakaj je razmislek iz besedila naloge napačen? Spet zato, ker so v njem pomešane pogojne in brezpogojne verjetnosti. Brezpogojni verjetnosti sta res enaki:

$$P(\text{avto za prvimi vrati}) = P(\text{avto za drugimi vrati}) = \frac{1}{3},$$

to pa ne drži za pogojni verjetnosti:

$$P\big(\text{avto za prvimi vrati} \,\big|\, \text{tretja vrata odprta}\big) = \frac{P(\{\text{ABO}\})}{P(\{\text{ABO}, \text{BAO}\})} = \frac{1}{3},$$
$$P\big(\text{avto za drugimi vrati} \,\big|\, \text{tretja vrata odprta}\big) = \frac{P(\{\text{BAO}\})}{P(\{\text{ABO}, \text{BAO}\})} = \frac{2}{3}.$$

Pogojni verjetnosti sta drugačni od brezpogojnih: brž ko je vodja igre odprl tretja vrata, nam je s tem dal dodatno informacijo.

Opazimo tudi, da se z odprtjem vrat ni prav nič spremenila verjetnost, da je avto za prvimi vrati (spremenili sta se le verjetnosti za druga in tretja vrata). To pa se lahko spremeni, če se spremeni protokol igre. Ključno je, da vodja igre takrat, ko ima izbiro, oboja možna vrata odpre z enako verjetnostjo. Prav tako bi bila, če bi vodjo igre vprašali, kaj je za tretjimi vrati, ob odgovoru 'buča' pogojna verjetnost, da je za prvimi vrati avto, enaka $1/2$, ob odgovoru 'avto' pa seveda 0. *Dogodek, da vodja igre odpre tretja vrata, namreč ni enak dogodku, da je za tretjimi vrati buča.*

Monty–Hallov paradoks je podoben *zaporniškemu paradoksu*, pri katerem tri raziskovalce eksotične dežele, recimo Allana, Billyja in Charlieja, zaprejo (vsakega posebej) in se poglavar odloči enega ubiti, preostala dva pa izpustiti. Vsak zapornik razmišlja, kaj bo z njim, stražarju pa je prepovedano povedati zapornikom, kaj bo z njimi. Če zapornik Allan od stražarja izprosi, da mu pove za enega izmed preostalih dveh zapornikov, ki bo izpuščen, in mu stražar recimo pove za Billyja, se verjetnost, da bo Allan izpuščen, prav nič ne spremeni, če privzamemo, da poglavar na slepo izbere zapornika, ki ga bo dal ubiti, stražar pa v primeru, če ima dve možnosti, prav tako izbere na slepo. Drugače pa bi bilo, če bi Allan lahko dobil odgovor, ali bo Billy izpuščen.

Če vemo, da vodja igre favorizira tretja vrata, je v primeru, ko jih odpre, enako verjetno, da je avto za prvimi in drugimi vrati. Če pa vodja igre odpre druga vrata, zagotovo vemo, da je avto za tretjimi.

5. Končni izid je lahko:

- $6 : 2$, če v nadaljnjih dveh rundah zmaga Janez;
- $6 : 3$, če v nadaljnjih dveh rundah enkrat zmaga Janez in enkrat Peter, v tretji rundi pa zmaga Janez;
- $6 : 4$, če v nadaljnjih treh rundah enkrat zmaga Janez in dvakrat Peter, v četrti rundi pa zmaga Janez;
- $6 : 5$, če v nadaljnjih štirih rundah enkrat zmaga Janez in trikrat Peter, v peti rundi pa zmaga Janez.

Verjetnost je zato enaka:

$$\left(\frac{1}{2}\right)^2 + 2 \cdot \left(\frac{1}{2}\right)^3 + 3 \cdot \left(\frac{1}{2}\right)^4 + 4 \cdot \left(\frac{1}{2}\right)^5 = \frac{13}{16} = 0\text{'}8125 \,.$$

Še hitreje pridemo do rezultata, če vzamemo nasprotni dogodek (izida $4:6$ in $5:6$).

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 $p = 0\text{'}2$.

Opomba: izpeljava je intuitivno zelo jasna, njena teoretična utemeljitev pa ni tako preprosta. Ujemanje pogojnih verjetnosti ne glede na zgodovino je namreč potrebno z dogodka, da Janez dobi na posamezno rundo, posplošiti na dogodek, da Janez nasploh zmaga. To naredimo tako, da najprej posplošimo na dogodke, ki se nanašajo na naslednjih nekaj rund, nato pa še na dogodke, ki se nanašajo vse nadaljnje runde. Za slednje pa potrebujemo netrivialne teoretične rezultate: glej naslednjo nalogo.

7. Tu imamo opravka z dvema verjetnostnima prostoroma: prvi z verjetnostno mero P se nanaša na posamezen poskus. Drugi pa se nanaša na celotno zaporedje poskusov; verjetnostno mero na njem označimo s $\tilde{P}$. Na prvem verjetnostnem prostoru imamo definirana dogodka A in B, na drugem pa dve celi seriji dogodkov: $A_1, A_2, \ldots$ in $B_1, B_2, \ldots$, kjer A_n oziroma B_n označuje dogodek, da je v n-tem poskusu prišlo do opažanja A oz. B. Tedaj velja:

$$\begin{aligned} \tilde{P}(A_n \mid Z_1, Z_2, \ldots Z_{n-1}) &= P(A) \,, \\ \tilde{P}(B_n \mid Z_1, Z_2, \ldots Z_{n-1}) &= P(B) \,, \\ \tilde{P}(A_n \cap B_n \mid Z_1, Z_2, \ldots Z_{n-1}) &= P(A \cap B) \,, \end{aligned}$$

kjer je lahko za vsak $k = 1, \ldots n-1$ dogodek Z_k kateri koli dogodek s strogo pozitivno verjetnostjo, ki ga lahko dobimo iz dogodkov A_n in B_n (torej iz σ-*algebre*, ki jo generirata dogodka A_n in B_n).

Označimo s C_0 dogodek, da pri prvem poskusu, pri katerem pride do opažanja B, pride tudi do opažanja A. Do verjetnosti tega dogodka lahko pridemo na vsaj dva načina.

Prvi način: z uporabo zasnove izreka o polni verjetnosti, pri čemer za hipoteze postavimo dogodke $H_n := \bar{B}_1 \cap \bar{B}_2 \cap \cdots \cap \bar{B}_{n-1} \cap B_n$. Velja:

$$\begin{aligned} \tilde{P}(C_0) &= \sum_{n=1}^{\infty} \tilde{P}(H_n \cap A_n) = \\ &= \sum_{n=1}^{\infty} \tilde{P}(\bar{B}_1)\tilde{P}(\bar{B}_2 \mid \bar{B}_1)\tilde{P}(\bar{B}_3 \mid \bar{B}_1 \cap \bar{B}_2) \cdots \tilde{P}(\bar{B}_{n-1} \mid \bar{B}_1 \cap \cdots \cap \bar{B}_{n-2}) \times \\ &\qquad \times \tilde{P}(A_n \cap B_n \mid \bar{B}_1 \cap \cdots \cap \bar{B}_{n-1}) = \\ &= \sum_{n=1}^{\infty} (1 - P(B))^{n-1} P(A \cap B) = \frac{P(A \cap B)}{P(B)} = P(A \mid B) \,. \end{aligned}$$

Opomba. Če je $P(B) = 1$, določene vmesne pogojne verjetnosti sicer nimajo smisla, rezultat pa še vedno velja, saj je $\tilde{P}(H_1) = 1$ in $\tilde{P}(H_m) = 0$ za $m > 1$; posledično je $\tilde{P}(H_1 \cap A_1) = \tilde{P}(A_1) = \tilde{P}(A_1)$ in $\tilde{P}(H_m \cap A_m) = 0$, zato je $\tilde{P}(C_0) = \tilde{P}(A_1) = P(A) = P(A \mid B)$.

Drugi način: z uporabo intuicije, katere rigorozna utemeljitev pa zahteva nekaj več teorije. Naj bo D_0 poljuben dogodek, ki ga dobimo iz dogodkov $A_1, B_1, A_2, B_2, \ldots$ (natančneje, dogodek iz σ-algebre, generirane s temi dogodki). Nadalje naj bo $D_0^{\to}$ dogodek, generiran tako kot dogodek D_0, le da poskuse zakasnimo za enega: če je npr. $D = A_1 \setminus B_2$, je $D_0^{\to} = A_2 \setminus B_3$. Natančneje, če je verjetnostni prostor, ki ponazarja celotno zaporedje poskusov, sestavljen iz zaporedij $(\omega_1, \omega_2, \ldots)$, je dogodek $D_0^{\to}$ sestavljen iz izidov oblike $(\omega, \omega_1, \omega_2, \ldots)$, kjer $(\omega_1, \omega_2, \ldots) \in D_0$. Tedaj iz pomembnih teoretičnih rezultatov (natančneje, *Dynkinove leme*, znane tudi pod imenom *izrek* π–λ) sledi sicer intuitivno zelo očitna zveza:

$$\tilde{P}(D_0^{\to} \mid Z_1) = \tilde{P}(D_0)\,,$$

kjer je Z_1 spet kateri koli dogodek s pozitivno verjetnostjo, ki generiran z A_1 in B_1.

Dogodek C_0 je disjunktna unija dogodkov $A_1 \cap B_1$ in $\bar{B}_1 \cap C_0^{\to}$. Iz prej povedanega sledi:

$$\begin{aligned}\tilde{P}(C_0) &= \tilde{P}(A_1 \cap B_1) + \tilde{P}(\bar{B}_1 \cap C_0^{\to}) = \tilde{P}(A_1 \cap B_1) + \tilde{P}(\bar{B}_1)\tilde{P}(C_0) = \\ &= P(A \cap B) + (1 - P(B))\tilde{P}(C_0)\,,\end{aligned}$$

od koder dobimo:

$$\tilde{P}(C_0) = \frac{P(A \cap B)}{P(B)} = P(A \mid B)\,.$$

8. Označimo iskano verjetnost s p_n. Za $n \geq 2$ se dogodek, da v prvih n metih še ni bilo dveh zaporednih cifer, zgodi v naslednjih dveh primerih:

- v prvem metu je padel grb in kasneje ni bilo dveh zaporednih cifer;
- v prvem metu je padla cifra, v drugem metu grb, kasneje pa ni bilo dveh zaporednih cifer. prej pa ni bilo dveh zaporednih cifer.

Od tod dobimo, da verjetnosti p_n za $n \geq 2$ zadoščajo rekurzivni zvezi $p_n = \frac{1}{2}p_{n-1} + \frac{1}{4}p_{n-2}$ z začetnima pogojema $p_0 = p_1 = 1$. Od tod jih lahko eksplicitno izračunamo – velja:

$$p_n = \frac{4}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{4}\right)^{n+2} - \left(\frac{1-\sqrt{5}}{4}\right)^{n+2}\right] = \frac{F_{n+2}}{2^n}\,,$$

kjer je $F_1, F_2, F_3, F_4, F_5, F_6 \ldots = 1, 1, 2, 3, 5, 8, \ldots$ znano Fibonaccijevo zaporedje.

9. $\displaystyle \frac{5}{10} \cdot \frac{4 \cdot 3}{\binom{7}{2}} + \frac{3}{10} \cdot \frac{3 \cdot 3}{\binom{7}{2}} + \frac{2}{10} \cdot \frac{4 \cdot 3}{\binom{7}{2}} \doteq 0{\cdot}529.$

10. Če s H_F označimo dogodek, da mož kupi solato pri Francki, s H_M dogodek, da kupi pri Micki, z G pa dogodek, da je kupil nagnito solato, velja:

$$\begin{aligned} P(H_F \mid G) &= \frac{0'6 \cdot 0'1}{0'6 \cdot 0'1 + 0'4 \cdot 0'2} \doteq 0'429\,, \\ P(H_M \mid G) &= \frac{0'4 \cdot 0'2}{0'6 \cdot 0'1 + 0'4 \cdot 0'2} \doteq 0'571\,. \end{aligned}$$

Torej žena bolj upravičeno sumi Micko.

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

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

13. Da, da, da, ne.

14. Ne. Velja $P(F \cap G \cap H) = P(F)\,P(G)\,P(H)$, $P(F \cap G) = P(F)\,P(G)$ in $P(F \cap H) = P(F)\,P(H)$, ne pa tudi $P(G \cap H) = P(G)\,P(H)$.

15. a) Ne: $P(A) = \dfrac{1}{2}$, $P(B) = \dfrac{3}{8}$, $P(A \cap B) = \dfrac{1}{4} \neq P(A)\ P(B)$.

b) Iz:

$$\begin{gathered} P(A) = \frac{1}{2}\,, \quad P(B) = 2 \cdot \frac{1}{2} \cdot p(1-p) + \frac{1}{2}\,p^2 = \frac{p(2-p)}{2}\,, \\ P(A \cap B) = 2 \cdot \frac{1}{2} \cdot p(1-p) = p(1-p) \end{gathered}$$

dobimo, da sta dogodka A in B neodvisna pri $p = 0$ in $p = 2/3$.

16. Pri ženskah je bilo prvo zdravilo uspešno v $200/2000 = 10\%$, drugo pa v $10/200 = 5\%$ 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 $390/2200 \doteq 17'7\%$, drugo zdravilo pa v $1010/2200 \doteq 45'9\%$ 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:

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

Tedaj velja:

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

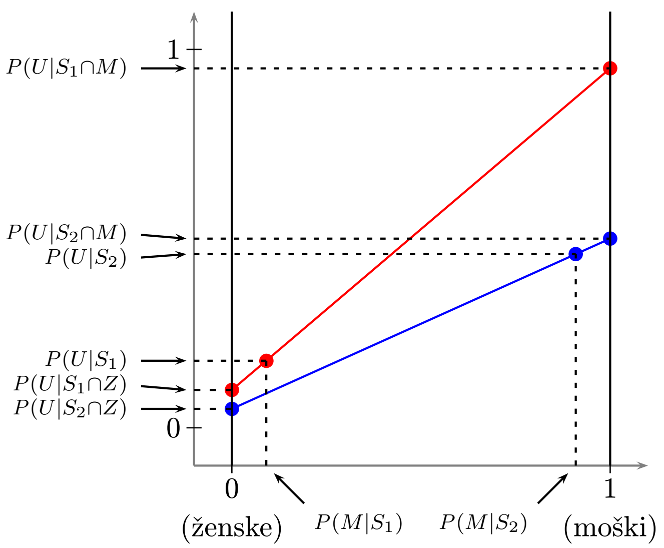
Oglejmo si zdaj, kako se $P(U \mid S_i)$ izraža z $P(U \mid S_i \cap Z)$ in $P(U \mid S_i \cap M)$. Velja:

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

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).

17. Naj bodo A, B, C, D in E dogodki, da posamezna stikala prepuščajo tok: A in B za stikali na levi, C za stikalo v sredini, D in E pa za stikalo na desni. Naj bo še T dogodek, da vezje prepušča tok. Tedaj najprej po izreku o polni verjetnosti velja:

$$P(T) = P(C)\,P(T \mid C) + P(\overline{C})\,P(T \mid \overline{C}) = \frac{1}{3}\,P(T \mid C) + \frac{2}{3}\,P(T \mid \overline{C})\,.$$

Nadalje je:

$$P(T \mid C) = P\big((A \cup B) \cap (D \cup E) \bigm| C\big)\,.$$

Dogodek $(A \cup B) \cap (D \cup E)$ pripada $\sigma(A, B, D, E)$, dogodek C pa $\sigma(C)$. Iz izreka o neodvisnosti izpeljanih dogodkov sledi, da sta dogodka $(A \cup B) \cap (D \cup E)$ in C neodvisna, torej je tudi:

$$P(T \mid C) = P\big((A \cup B) \cap (D \cup E)\big)\,.$$

Spet iz izreka o neodvisnosti izpeljanih dogodkov sledi neodvisnost dogodkov $A \cup B$ in $D \cup E$, torej je:

$$\begin{aligned} P(T \mid C) &= P(A \cup B)\,P(D \cup E) = \\ &= \big(P(A) + P(B) - P(A \cap B)\big)\big(P(D) + P(E) - P(D \cap E)\big) = \\ &= \left(\frac{1}{3} + \frac{1}{3} - \frac{1}{9}\right)^2 = \\ &= \frac{25}{81}\,. \end{aligned}$$

Podobno dobimo:

$$\begin{aligned} P(T \mid \overline{C}) &= P\big((A \cap B) \cup (D \cap E)\big) = \\ &= P(A \cap B) + P(D \cap E) - P(A \cap B \cap D \cap E) = \\ &= \frac{17}{81}\,. \end{aligned}$$

Sledi $P(T) = \frac{1}{3} \cdot \frac{25}{81} + \frac{2}{3} \cdot \frac{17}{81} = \frac{59}{243} \doteq 0{\cdot}243$.

18. a) Označimo z J, F in T dogodke, da je Janez, Francelj oz. Tone zadel. Naj bo še $Z = J \cup F \cup T$ dogodek, da je sploh kdo zadel. Iskana pogojna verjetnost je torej enaka:

$$P(J \mid Z) = \frac{P(J \cap Z)}{P(Z)} = \frac{P(J)}{1 - P\big(\overline{J} \cap \overline{F} \cap \overline{T}\big)} = \frac{0{\cdot}1}{1 - 0{\cdot}9 \cdot 0{\cdot}8 \cdot 0{\cdot}7} \doteq 0{\cdot}202\,.$$

b) Naj bo Z_1 dogodek, da je zajca zadel natanko en lovec. Zapišemo lahko:

$$Z_1 = (J \cap \overline{F} \cap \overline{T}) \cup (\overline{J} \cap F \cap \overline{T}) \cup (\overline{J} \cap \overline{F} \cap T)\,,$$

torej je:

$$P(Z_1) = 0{\cdot}1 \cdot 0{\cdot}8 \cdot 0{\cdot}7 + 0{\cdot}9 \cdot 0{\cdot}2 \cdot 0{\cdot}7 + 0{\cdot}9 \cdot 0{\cdot}8 \cdot 0{\cdot}3 = 0{\cdot}398\,.$$

Sledi:

$$P(J \mid Z_1) = \frac{P(J \cap Z)}{P(Z_1)} = \frac{P(J \cap \overline{F} \cap \overline{T})}{P(Z_1)} = \frac{0{\cdot}1 \cdot 0{\cdot}8 \cdot 0{\cdot}7}{0{\cdot}398} \doteq 0{\cdot}141\,.$$

19. Označimo z A, B, C in D dogodke, da zadene Andraž, Bojan, Cilka oz. Darja. Nadalje označimo z M_1 dogodek, da je v tarči natanko ena modra, z R_1 pa dogodek, da je v tarči natanko ena rdeča puščica. Iskana pogojna verjetnost je torej enaka:

$$P(B \cap D \mid M_1 \cap R_1) = \frac{P(B \cap D \cap M_1 \cap R_1)}{P(M_1 \cap R_1)} = \frac{P\big(\overline{A} \cap B \cap \overline{C} \cap D\big)}{P(M_1 \cap R_1)}\,.$$

Po izreku o neodvisnosti izpeljanih dogodkov so dogodki $\overline{A}$, B, $\overline{C}$ in D neodvisni, prav tako tudi dogodka $M_1 = (A \cap \overline{B}) \cup (\overline{A} \cap B)$ in $R_1 = (C \cap \overline{D}) \cup (\overline{C} \cap D)$. Sledi:

$$\begin{aligned}
&P(B \cap D \mid M_1 \cap R_1) = \\
&\quad = \frac{P(\overline{A})\,P(B)\,P(\overline{C})\,P(D)}{\Big(P(A)\,P(\overline{B}) + P(\overline{A})\,P(B)\Big)\Big(P(C)\,P(\overline{D}) + P(\overline{C})\,P(D)\Big)} = \\
&\quad = \frac{0{\cdot}4 \cdot 0{\cdot}7 \cdot 0{\cdot}5 \cdot 0{\cdot}9}{(0{\cdot}6 \cdot 0{\cdot}3 + 0{\cdot}4 \cdot 0{\cdot}7) \cdot (0{\cdot}5 \cdot 0{\cdot}1 + 0{\cdot}5 \cdot 0{\cdot}9)} \doteq \\
&\quad \doteq 0{\cdot}548\,.
\end{aligned}$$

20. Naj bodo H_0, H_1, H_2 in H_3 dogodki, da se je študent naučil učil 0, 1, 2 oz. 3 vprašanja, ki jih je dobil na izpitu. Velja:

$$P(H_i) = \frac{\binom{30}{i}\binom{20}{3-i}}{\binom{50}{3}}\,.$$

Nadalje naj bo A dogodek, da študent naredi izpit.

- Če se ni naučil nobenega vprašanja, mora, če želi narediti izpit, uganiti vsaj dve. Zato je $P(A \mid H_0) = 3 \cdot 0\text{'}1^2 \cdot 0\text{'}9 + 0\text{'}1^3 = 0\text{'}028$.
- Če se je študent naučil natanko eno vprašanje, mora, če želi narediti izpit, bodisi uganiti obe vprašanji, ki se ju ni naučil, bodisi ne sme pozabiti vprašanja, ki se ga je naučil, obenem pa mora uganiti še eno vprašanje, ki se ga ni učil. Zato je $P(A \mid H_1) = 0\text{'}1^2 + 2 \cdot 0\text{'}1 \cdot 0\text{'}9 \cdot 0\text{'}7 = 0\text{'}136$.
- Če se je študent naučil natanko dve vprašanji, naredi izpit natanko tedaj, ko bodisi ne pozabi nobenega vprašanja, ki se ga je učil, bodisi pozabi eno vprašanje, ki se ga je naučil, obenem pa ugane vprašanje, ki se ga ni naučil. Zato je $P(A \mid H_2) = 0\text{'}7^2 + 2 \cdot 0\text{'}7 \cdot 0\text{'}3 \cdot 0\text{'}1 = 0\text{'}532$.
- Če se je naučil vsa tri vprašanja, sme, če želi narediti izpit, pozabiti največ enega. Zato je $P(A \mid H_3) = 3 \cdot 0\text{'}7^2 \cdot 0\text{'}3 + 0\text{'}7^3 = 0\text{'}784$.

Sledi:

$$P(A) = \frac{\binom{20}{3}}{\binom{50}{3}} \cdot 0\text{'}028 + \frac{\binom{30}{1}\binom{20}{2}}{\binom{50}{3}} \cdot 0\text{'}136 + \frac{\binom{30}{2}\binom{20}{1}}{\binom{50}{3}} \cdot 0\text{'}532 + \frac{\binom{30}{3}}{\binom{50}{3}} \cdot 0\text{'}784 \doteq 0\text{'}440\,.$$

21. Označimo z J, L in S dogodke, da Janez, Lojz oz. Štefan da Mihi šmarnico. Nadalje naj bodo K_0, K_1, K_2 in K_3 dogodki, da Miha dobi 0, 1, 2 oz. 3 kozarce šmarnice, B pa dogodek, da Miho boli glava. Tedaj vemo, da je:

$$P(B \mid K_0) = 0\text{'}1\,, \quad P(B \mid K_1) = 0\text{'}4\,, \quad P(B \mid K_2) = 0\text{'}7\,, \quad P(B \mid K_3) = 1\,,$$

obenem pa je:

$$\begin{aligned}
P(K_0) &= 0\text{'}4 \cdot 0\text{'}6 \cdot 0\text{'}9 = 0\text{'}216\,,\\
P(K_1) &= 0\text{'}6 \cdot 0\text{'}6 \cdot 0\text{'}9 + 0\text{'}4 \cdot 0\text{'}4 \cdot 0\text{'}9 + 0\text{'}4 \cdot 0\text{'}6 \cdot 0\text{'}1 = 0\text{'}492\,,\\
P(K_2) &= 0\text{'}6 \cdot 0\text{'}4 \cdot 0\text{'}9 + 0\text{'}6 \cdot 0\text{'}6 \cdot 0\text{'}1 + 0\text{'}4 \cdot 0\text{'}4 \cdot 0\text{'}1 = 0\text{'}268\,,\\
P(K_3) &= 0\text{'}6 \cdot 0\text{'}4 \cdot 0\text{'}1 = 0\text{'}024\,.
\end{aligned}$$

Torej je:

$$P(B) = 0\text{'}216 \cdot 0\text{'}1 + 0\text{'}492 \cdot 0\text{'}4 + 0\text{'}268 \cdot 0\text{'}7 + 0\text{'}024 \cdot 1 = 0\text{'}43\,.$$

Sledi:

$$P(J \cap L \mid B) = \frac{P(J \cap L \cap B)}{P(B)} = \frac{P(J \cap L \cap \overline{S} \cap B) + P(J \cap L \cap S \cap B)}{P(B)}\,.$$

Seveda je $P(B \mid J \cap L \cap S) = P(B \mid K_3) = 1$. Toda velja tudi $P\big(B \mid J \cap L \cap \overline{S}\big) = 0\text{'}7$, ne le $P(B \mid K_2) = 0\text{'}7$. Sledi:

$$P(J \cap L \mid B) = \frac{0\text{'}6 \cdot 0\text{'}4 \cdot 0\text{'}9 \cdot 0\text{'}7 + 0\text{'}6 \cdot 0\text{'}4 \cdot 0\text{'}1 \cdot 1}{0\text{'}43} \doteq 0\text{'}407\,.$$

22. a) *Prvi način.* Oglejmo si nasprotni dogodek, t. j. da Albert nikoli ne bo izpuščen. Le-ta je presek padajočega zaporedja dogodkov, da bo Albert v ječi prebil več kot n noči. Verjetnosti teh dogodkov so enake:

$$\frac{1}{2} \cdot \frac{2}{3} \cdot \cdots \cdot \frac{n}{n+1} = \frac{1}{n+1},$$

torej imajo limito nič, se pravi, da je nič tudi verjetnost dogodka, da Albert nikoli ne bo izpuščen.

Drugi način. Označimo z J_n dogodek, da je Albert v ječi prespal natanko n noči. Velja:

$$P(J_n) = \frac{1}{2} \cdot \frac{2}{3} \cdot \cdots \cdot \frac{n-1}{n} \cdot \frac{1}{n+1} = \frac{1}{n(n+1)}.$$

Dogodek, da Albert nekoč pride iz ječe, je unija dogodkov $J_1, J_2, J_3, \ldots$, ki so nezdružljivi. Njegova verjetnost je enaka:

$$\sum_{n=1}^{\infty} P(J_n) = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \lim_{m\to\infty} \sum_{n=1}^{m} \left(\frac{1}{n} - \frac{1}{n+1}\right) = \lim_{m\to\infty} \left(1 - \frac{1}{m+1}\right) = 1.$$

b) Označimo z N_5 dogodek, da je Albert v dani deželi prespal natanko petkrat. Ta dogodek je možen, če je v ječi prespal trikrat, štirikrat ali petkrat. Za $n = 3, 4, 5$ velja:

$$P(N_5 \mid J_n) = \frac{1}{n+1}.$$

Po Bayesovi formuli dobimo:

$$P(J_3 \mid N_5) = \frac{\frac{1}{3 \cdot 4} \cdot \frac{1}{4}}{\frac{1}{3 \cdot 4} \cdot \frac{1}{4} + \frac{1}{4 \cdot 5} \cdot \frac{1}{5} + \frac{1}{5 \cdot 6} \cdot \frac{1}{6}} = \frac{75}{131} \doteq 0{\cdot}573.$$

23. Označimo z F_k dogodek, da Ferdinand Mirando prvič pokliče k-ti dan po zabavi. Tedaj so dogodki $F_1, F_2, \ldots$ nezdružljivi, njihova unija, ki jo označimo z F, pa je dogodek, da Ferdinand Mirando sploh pokliče. Verjetnost tega dogodka je enaka:

$$P(F) = \sum_{k=1}^{\infty} 3^{-k} = \frac{1}{2}.$$

Označimo še z M dogodek, da Miranda spozna novega fanta, preden jo Ferdinand pokliče (če je ne pokliče, je torej to dogodek, da Miranda sploh spozna novega fanta). Tedaj velja:

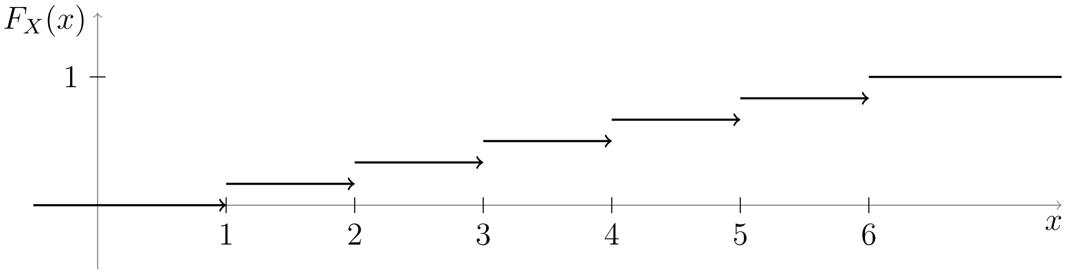
$$P(M \mid F_k) = 1 - \left(\frac{9}{10}\right)^{k-1},$$

iskana pogojna verjetnost pa je enaka:

$$P(M \mid F) = \frac{P(M \cap F)}{P(F)} = \frac{1}{P(F)} \sum_{k=1}^{\infty} P(F_k)\, P(M \mid F_k) = \frac{1}{21}.$$

4. Slučajne spremenljivke

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



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. Kumulativna porazdelitvena funkcija:

$$F_X(x) = \begin{cases} 0 & ; x \le 0 \\ \dfrac{x}{10} & ; 0 \le x \le 10 \\ 1 & ; x \ge 10 \end{cases}$$

je zvezna in odsekoma zvezno odvedljiva, torej gre res za zvezno porazdelitev. Gostota:

$$p_X(x) = \begin{cases} \dfrac{1}{10} & ; 0 < x < 2\pi \\ 0 & ; \text{sicer} \end{cases} .$$

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

5. $F_Z(z) = \begin{cases} 0 & ; z \le 0 \\ \dfrac{2z^2}{5} & ; 0 \le z \le 1 \\ \dfrac{1 + z^2}{5} & ; 1 \le z \le 2 \\ 1 & ; z \ge 2 \end{cases}$, $\quad p_Z(z) = \begin{cases} \dfrac{4z}{5} & ; 0 < z < 1 \\ \dfrac{2z}{5} & ; 1 < z < 2 \\ 0 & ; \text{sicer} \end{cases}$.

6. $P(X=k) = \binom{10}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{10-k}$. 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	$1'65 \cdot 10^{-8}$

7. $1 - \left(\frac{2}{3}\right)^6 - 6\cdot\frac{1}{3}\left(\frac{2}{3}\right)^5 - \binom{6}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^4 \doteq 0'320.$

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

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

10. Poisson (uporabimo ga na slučajni spremenljivki, ki pove, kolikokrat se voda *ne* potegne): 0'12511.
Točen rezultat: 0'12574.

11. 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 $P(175'5 < X < \infty)$: 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 $P(-\infty < X < 149'5)$: 0'19079.
Točen rezultat za $P(X < 150)$: 0'19147.

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

$$\frac{1}{2} - \Phi\left(\frac{x + \frac{1}{2} - 1600\cdot 0'1}{\sqrt{1600\cdot 0'1\cdot 0'9}}\right) \le 0'05$$

oziroma:

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

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 $y \doteq 179\text{'}24$ dobimo oceno $x = 180$.

Dejansko je verjetnost, da bo pokvarjenih izdelkov (strogo) več kot 179, enaka 0'0539, da jih bo več kot 180, pa 0'0457. Torej je ocena $x = 180$ pravilna.

12. 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.

13. Označimo z n število naročenih izdelkov, z S pa število prvovrstnih. Tedaj je $S \sim \mathrm{Bin}(n, 0\text{'}6)$. Veljati mora:

$$P(S \geq 0\text{'}59\,n) \geq 0\text{'}99\,.$$

Po Laplaceovi integralski formuli je:

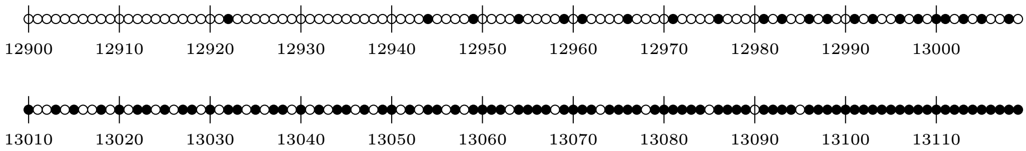
$$P(S \geq 0\text{'}59\,n) \approx \frac{1}{2} - \Phi\left(\frac{0\text{'}59\,n - 0\text{'}6\,n}{\sqrt{n \cdot 0\text{'}6 \cdot 0\text{'}4}}\right) = \frac{1}{2} + \Phi\left(\sqrt{\frac{n}{2400}}\right)\,.$$

Torej bo število naročenih izdelkov ustrezalo približno tedaj, ko bo $\sqrt{n/2400} \geq \Phi^{-1}(0\text{'}49)$ oziroma $n \geq 2400\big(\Phi^{-1}(0\text{'}49)\big)^2 \doteq 12988\text{'}55$, torej $n \geq 12989$.

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:

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

Na naslednji sliki so prikazana števila naročenih izdelkov, ki ustrezajo (polni krogci) in števila, ki ne ustrezajo (prazni krogci).



14. Označimo z n število naročenih izdelkov, z S pa število prvovrstnih. Tedaj je $S \sim \mathrm{Bin}(n, 0{\cdot}1)$. Veljati mora $P(S \geq 100) \geq 0{\cdot}95$. Tukaj je verjetnost želenega dogodka naraščajoča funkcija števila n.

Po Laplaceovi integralski formuli je:

$$P(S \geq 100) \approx \frac{1}{2} - \Phi\left(\frac{99{\cdot}5 - 0{\cdot}1\,n}{\sqrt{n \cdot 0{\cdot}1 \cdot 0{\cdot}9}}\right) = \frac{1}{2} + \Phi\left(\frac{n - 1000}{3\sqrt{n}}\right).$$

Torej bo število naročenih izdelkov ustrezalo približno tedaj, ko bo

$$\frac{n - 995}{3\sqrt{n}} \geq \Phi^{-1}(0{\cdot}45)$$

oziroma:

$$n - 3\,\Phi^{-1}(0{\cdot}45)\,\sqrt{n} - 995 \geq 0\,.$$

Če označimo $q = \Phi^{-1}(0{\cdot}45) \doteq 1{\cdot}644854$, je to natanko tedaj, ko je:

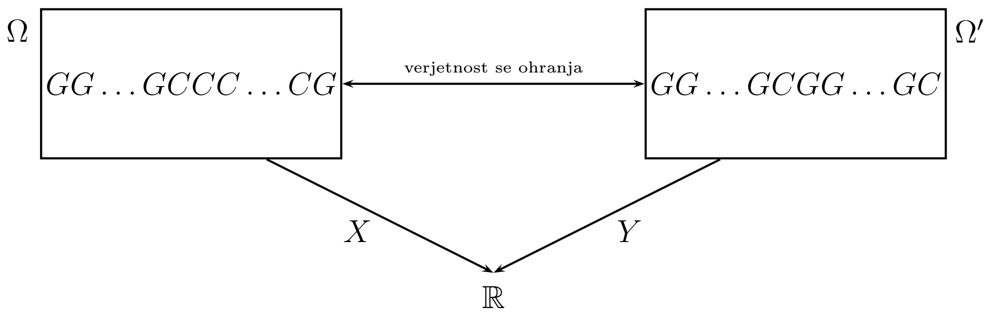
$$n \geq \left(\frac{3q + \sqrt{9q^2 + 3980}}{2}\right)^2 \doteq 1163{\cdot}32 \quad \text{oziroma} \quad n \geq 1164\,.$$

V resnici je dovolj naročiti 1161 izdelkov: verjetnost, da je prvovrstnih vsaj 100, pride pri 1160 izdelkih 0˙9493, pri 1161 izdelkih pa 0˙9502.

15. $P(X = k) = \dfrac{1}{6}\left(\dfrac{5}{6}\right)^{k-1}$; $k \in \mathbb{N}$.

16. $P(X = k) = \dbinom{k-1}{9}\left(\dfrac{1}{6}\right)^{10}\left(\dfrac{5}{6}\right)^{k-10}$; $k = 10, 11, 12, \ldots$.

17. $P(X = k) = (k-1)\,2^{-k}$ za $k = 2, 3, \ldots$. Porazdelitev je Pascalova $\mathrm{NegBin}(2, 1/2)$. 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. Natančneje, če Y označuje število metov, dokler ne padeta dve cifri, sta X in Y enako porazdeljeni – pišemo $X \stackrel{\mathrm{d}}{=} Y$. Slika:



Če kovanec ni pošten, zgornja verjetnostna razlaga ne zdrži in tudi formula za porazdelitev je znatno bolj zapletena: če je p verjetnost, da pri posameznem metu pade grb, velja:

$$P(X=k)=\sum_{i=0}^{k-2} p^{i+1}(1-p)^{k-i-1}=p(1-p)\,\frac{(1-p)^{k-1}-p^{k-1}}{1-2p}\,.$$

18. $P(X=k)=2^{-(k-1)}$ za $k=2,3,\ldots$, $X-1\sim\mathrm{Geo}(1/2)$.
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).

Če kovanec ni pošten, zgornja razlaga spet ne zdrži; če je p verjetnost, da pri posameznem metu pade grb, velja:

$$P(X=k)=p(1-p)\big[p^{k-2}+(1-p)^{k-2}\big]\,.$$

19. Velja:

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

oziroma:

$$X\sim\begin{pmatrix} 0 & 1 & 2 & 3 & 4\\ \frac{18}{260} & \frac{84}{260} & \frac{108}{260} & \frac{45}{260} & \frac{5}{260}\end{pmatrix}\doteq\begin{pmatrix} 0 & 1 & 2 & 3 & 4\\ 0{\cdot}0692 & 0{\cdot}3231 & 0{\cdot}4154 & 0{\cdot}1731 & 0{\cdot}0192\end{pmatrix}.$$

20. $X\sim\mathrm{Hip}(4,3,12)=\mathrm{Hip}(3,4,12)$.

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

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

Število škatel, v katerih ni nobene kroglice, je porazdeljeno hipergeometrijsko $\mathrm{Hip}(8,9,12)=\mathrm{Hip}(9,8,12)$.

21. Označimo s H_j dogodek, da se študent ni učil j vprašanj, ki jih je dobil ($j=0,1,2$). Tedaj velja:

$$P(H_j)=\frac{\binom{5}{j}\binom{5}{2-j}}{\binom{10}{2}}\,,\qquad P(U=i\mid H_j)=\binom{j}{i}\cdot 0{\cdot}2^i\cdot 0{\cdot}8^{j-i}$$

oziroma:

$$
\begin{array}{llll}
P(H_0) = \dfrac{10}{45}\,, & P(U = 0 \mid H_0) = 1\,, & P(U = 1 \mid H_0) = 0\,, & P(U = 2 \mid H_0) = 0\,, \\
P(H_1) = \dfrac{25}{45}\,, & P(U = 0 \mid H_1) = 0'8\,, & P(U = 1 \mid H_1) = 0'2\,, & P(U = 2 \mid H_1) = 0\,, \\
P(H_2) = \dfrac{10}{45}\,, & P(U = 0 \mid H_2) = 0'64\,, & P(U = 1 \mid H_2) = 0'32\,, & P(U = 2 \mid H_2) = 0'04\,.
\end{array}
$$

Torej je:

$$
\begin{aligned}
P(U = 0) &= \frac{10}{45} \cdot 1 + \frac{25}{45} \cdot 0'8 + \frac{10}{45} \cdot 0'64 \doteq 0'8089\,, \\
P(U = 1) &= \frac{10}{45} \cdot 0 + \frac{25}{45} \cdot 0'2 + \frac{10}{45} \cdot 0'32 \doteq 0'1822\,, \\
P(U = 2) &= \frac{10}{45} \cdot 0 + \frac{25}{45} \cdot 0 + \frac{10}{45} \cdot 0'04 \doteq 0'0089\,,
\end{aligned}
$$

oziroma približno:

$$
U \sim \begin{pmatrix} 0 & 1 & 2 \\ 0'8089 & 0'1822 & 0'0089 \end{pmatrix}.
$$

Opomba: seveda bi bilo dovolj izračunati $P(U = 1)$ in $P(U = 2)$ in nato upoštevati, da je vsota verjetnosti enaka 1.

22. Diskretnost skupaj s premo sorazmernostjo pomeni, da je:

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

za neko konstanto c, poleg tega pa še $\sum_{k=1}^{10} P(X = k) = 1$. Od tod dobimo $c = 1/55$ in $P(X > 3) = 49/55$.

23. Velja $c = \lambda$ in:

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

24. $Y \sim \begin{pmatrix} 0 & 1 & 4 \\ 0'1 & 0'7 & 0'2 \end{pmatrix}$

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

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

$$
\begin{aligned}
P(D = d) = P(d \le b^U < d + 1) &= P(\log_b d \le U \le \log_b (d + 1)) = \\
&= \log_b (d + 1) - \log_b d = \log_b \left(1 + \frac{1}{d}\right)
\end{aligned}
$$

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 $\mathrm{E_z}(0,1)$. Porazdelitvena shema za Benfordovo porazdelitev pri $b = 10$:

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

27. Nalogo rešimo prek kumulativne porazdelitvene funkcije. Očitno je $Y \ge 0$. Za $y \ge 0$ velja:

$$F_Y(y) = P(Y \le y) = P\big(-\sqrt{y} - a \le X \le \sqrt{y} - a\big).$$

To moramo izračunati na podlagi dejstva, da je X porazdeljena eksponentno. Ločimo več primerov. Za $a \ge 0$ velja:

$$F_Y(y) = \begin{cases} 0 & ; \ y \le a^2 \\ 1 - e^{\lambda(a-\sqrt{y})} & ; \ y \ge a^2 \end{cases}.$$

Ker je ta funkcija zvezna in odsekoma zvezno odvedljiva, je Y porazdeljena zvezno z gostoto:

$$p_Y(y) = \begin{cases} 0 & ; \ y \le a^2 \\ \dfrac{\lambda}{2\sqrt{y}}\, e^{\lambda(a-\sqrt{y})} & ; \ y \ge a^2 \\ 0 & ; \ \text{sicer} \end{cases}.$$

Naj bo zdaj $a < 0$. V tem primeru pa dobimo:

$$F_Y(y) = \begin{cases} 0 & ; \ y \le 0 \\ e^{\lambda(a+\sqrt{y})} - e^{\lambda(a-\sqrt{y})} & ; \ 0 \le y \le a^2 \\ 1 - e^{\lambda(a-\sqrt{y})} & ; \ y \ge a^2 \end{cases},$$

kar je spet zvezna in odsekoma zvezno odvedljiva funkcija, ki nam da gostoto:

$$p_Y(y) = \begin{cases} 0 & ; \ y < 0 \\ \dfrac{\lambda}{2\sqrt{y}}\left(e^{\lambda(a+\sqrt{y})} + e^{\lambda(a-\sqrt{y})}\right) & ; \ 0 < y < a^2 \\ \dfrac{\lambda}{2\sqrt{y}}\, e^{\lambda(a-\sqrt{y})} & ; \ y > a^2 \end{cases},$$

28. Naj bo $U \sim \mathrm{E_z}(0,1)$. Če postavimo:

$$X := \begin{cases} -1 & ; \ U < 0\text{'}4 \\ 0 & ; \ 0\text{'}4 \le U < 0\text{'}5 \\ 1 & ; \ 0\text{'}5 \le U < 0\text{'}8 \\ 2 & ; \ U \ge 0\text{'}8 \end{cases},$$

je $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 0{\cdot}4 & 0{\cdot}1 & 0{\cdot}3 & 0{\cdot}2 \end{pmatrix}$. Za generiranje eksponentne porazdelitve pa si pomagamo s kumulativno porazdelitveno funkcijo: slučajna spremenljivka Y je porazdeljena eksponentno $\mathrm{Exp}(\lambda)$, če za vse $y \ge 0$ velja:

$$P(Y \le y) = 1 - e^{-\lambda y}\,.$$

Zdaj pa porazdelitev slučajne spremenljivke U zapišimo v obliki:

$$P\big(U \le 1 - e^{-\lambda y}\big) = 1 - e^{-\lambda y}$$

oziroma:

$$P\left(-\frac{\ln(1-U)}{\lambda} \le y\right) = 1 - e^{-\lambda y}\,.$$

Če torej postavimo $Y := -\dfrac{\ln(1-U)}{\lambda}$, bo gotovo $Y \sim \mathrm{Exp}(\lambda)$.

Opomba. Definiciji slučajnih spremenljivk X in Y nista edini izbiri: lahko bi npr. definirali tudi $Y = -(\ln U)/\lambda$.

Opomba. Če je X porazdeljena zvezno, je $F_X(X) \sim \mathrm{E_z}(0,1)$. Račun $F_X(X) = P(X \le X) = 1$ ni pravilen, ker gre tu za dva različna X-a. To je podobno, kot če naprogramiramo:

```
function f(k)
{
  x = ... ;
  ...
}

...

x = ... ;
f(x);
```

in je spremenljivka `x` v klicu `f(x)` druga kot tista znotraj definicije funkcije `f`. Če se vrnemo na slučajne spremenljivke in je denimo X porazdeljena eksponentno $\mathrm{Exp}(\lambda)$, velja $F_X(X) = 1 - e^{-\lambda X}$.

29. Kvantil za verjetnost $0{\cdot}3$ je natančno določen: $q_{0.3} = 3$.
Kvantil za verjetnost $0{\cdot}5$ je lahko kar koli iz intervala $[4,5]$.

30. Kvantil za verjetnost p je enak $\dfrac{p}{1-p}$. Torej je $m = 1$ in $s = 4/3$.

31. a) Dovolj je obravnavati le implikaciji na levi, saj sta jima implikaciji na desni ekvivalentni. Brž ko je $x < q$, je $F(x) = P(X \le x) \le P(X < q) \le p$, torej implikacija velja. Toda okrepiti je ne moremo na nobeni strani: če je $P(X = 0) = P(X = 1) = 1/2$, je $q = 1$ kvantil slučajne spremenljivke X za verjetnost $p = 3/4$

(tretji kvartil). Toda če je $x = q = 1$, je $F(x) = 1 > p$, zato implikaciji na levi strani ne moremo dodati enačaja. Če pa postavimo $p = 1/2$, je $q = 2/3$ ena od median, toda za $x = 1/3 < q$ velja $F(x) = 1/2 = p$, zato implikaciji na desni strani enačaja ne moremo odvzeti.

Oglejmo si še spodnjo implikacijo. Brž ko je $x \le q$, je $F(x) = P(X \le x) \ge P(X \le q) \ge p$, torej implikacija velja, in sicer tudi, če ji na levi strani dodamo enačaj. Spet pa ji enačaja na desni strani ne moremo odvzeti: $P(X = 0) = P(X = 1) = 1/2$ in $p = 1/2$, je $q = 1/3$ ena od median, a za $x = 2/3 > q$ velja $F(x) = 1/2 = p$.

b) Naj bo $u < v$ ter $x = Q(u)$ in $y = Q(v)$. Tedaj je:

$$P(X < x) \le u < v \le P(X \le y)\,,$$

od koder sledi $x \le y$.

c) Naj bo $x \in \mathbb{R}$ in $\varepsilon > 0$. Tedaj velja:

$$\begin{aligned}
&P(Q(U) \le x - \varepsilon) \le P(Q(U) < x) \le P(U \le F(x)) = F(x)\,,\\
&P(Q(U) \ge x + \varepsilon) \le P(Q(U) > x) \le P(U \ge F(x)) = 1 - F(x)\,,\\
&P(Q(U) < x + \varepsilon) \ge F(x)\,.
\end{aligned}$$

Ko pošljemo ε proti nič, dobimo $P(Q(U) < x) \le F(x) \le P(Q(U) \le x)$. Zdaj pa spet vzemimo $\varepsilon > 0$ in pišimo:

$$F(x) \le P(Q(U) \le x) \le P(Q(U) < x + \varepsilon) \le F(x + \varepsilon)\,.$$

Ko pošljemo ε proti nič in upoštevamo desno zveznost kumulativne porazdelitvene funkcije F, dobimo, da mora biti $P(Q(U) \le x) = F(x)$, torej je $Q(U)$ res porazdeljena enako kot X.

32. 0˙93319.

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

34. $1/2 - \Phi(9/5) \doteq 0{,}03593$.

35. $Y \sim \mathrm{Gama}(a, \lambda/k)$.

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

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

$$p_Z(z) = \begin{cases} \dfrac{1}{2\sqrt{2\pi z}}\Big(e^{-(1+\sqrt{z})^2/2} + e^{-(1-\sqrt{z})^2/2}\Big) & ;\ z > 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

Opomba: slučajna spremenljivka Y ima porazdelitev *hi hvadrat z eno prostostno stopnjo*, ki jo označimo s $\chi^2(1)$ (glej 5. razdelek).

5. Slučajni vektorji

1. Navzkrižno in robni porazdelitvi lahko predstavimo s tabelo:

	$M = 0$	$M = 1$	$M = 2$	
$R = 0$	$\frac{10}{120}$	$\frac{20}{120}$	$\frac{5}{120}$	$\frac{35}{120}$
$R = 1$	$\frac{30}{120}$	$\frac{30}{120}$	$\frac{3}{120}$	$\frac{63}{120}$
$R = 2$	$\frac{15}{120}$	$\frac{6}{120}$	0	$\frac{21}{120}$
$R = 3$	$\frac{1}{120}$	0	0	$\frac{1}{120}$
	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$	1

Slučajni spremenljivki R in M sta odvisni. Velja:

$$R + M \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{10}{120} & \frac{50}{120} & \frac{50}{120} & \frac{10}{120} \end{pmatrix}.$$

Navzkrižno porazdelitev lahko dobimo iz formule:

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

vse ostale pa iz dejstva, da gre za hipergeometrijsko porazdelitev: $R \sim \mathrm{Hip}(3, 3, 10)$, $M \sim \mathrm{Hip}(3, 2, 10)$, $R + M \sim \mathrm{Hip}(3, 5, 10)$.

2. $R + M \sim \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ \frac{245}{1800} & \frac{686}{1800} & \frac{623}{1800} & \frac{217}{1800} & \frac{28}{1800} & \frac{1}{1800} \end{pmatrix}$.

Nauk: za porazdelitev funkcije dveh slučajnih spremenljivk ni dovolj poznati le porazdelitve posameznih slučajnih spremenljivk: v splošnem je potrebno poznati navzkrižno porazdelitev.

3. a) Najprej preverimo, da je vsota deklariranih verjetnosti res ena. Ker je to res, je dovolj preveriti, da so vse deklarirane verjetnosti nenegativne. Po ureditvi dobimo, da to velja, če je $a \ge 0$ in $2a^2 + a - 5 \le 0$, torej če je $0 \le a \le \frac{-1+\sqrt{41}}{4} \doteq 1{\cdot}351$.

b) Velja:

$$X \sim \begin{pmatrix} 1 & 2 \\ \frac{5}{6} - \frac{1}{6}a & \frac{1}{6} + \frac{1}{6}a \end{pmatrix}, \qquad Y \sim \begin{pmatrix} 1 & 2 \\ \frac{5}{6} - \frac{1}{3}a^2 & \frac{1}{6} + \frac{1}{3}a^2 \end{pmatrix}$$

Slučajni spremenljivki sta torej enako porazdeljeni, če je $\frac{1}{6}a = \frac{1}{3}a^2$, torej če je $a = 0$ ali $a = \frac{1}{2}$.

c) Slučajni spremenljivki X in Y sta skoraj gotovo enaki, če sta verjetnosti v poljih, kjer nista enaki, enaki nič. To je res za $a = 0$.

d) Slučajni spremenljivki sta neodvisni, če je izpolnjen naslednji sistem enačb:

$$\begin{aligned}
\frac{5}{6} - \frac{1}{6}\,a - \frac{1}{3}\,a^2 &= \left(\frac{5}{6} - \frac{1}{6}\,a\right)\left(\frac{5}{6} - \frac{1}{3}\,a^2\right) \\
\frac{1}{3}\,a^2 &= \left(\frac{5}{6} - \frac{1}{6}\,a\right)\left(\frac{1}{6} + \frac{1}{3}\,a^2\right) \\
\frac{1}{6}\,a &= \left(\frac{5}{6} - \frac{1}{3}\,a^2\right)\left(\frac{1}{6} + \frac{1}{6}\,a\right) \\
\frac{1}{6} &= \left(\frac{1}{6} + \frac{1}{6}a\right)\left(\frac{1}{6} + \frac{1}{3}a^2\right).
\end{aligned}$$

Po ureditvi dobimo, da je to natanko tedaj, ko je $2a^3 + 2a^2 + a - 5 = 0$, kar je ekvivalentno $(a-1)(2a^2 + 4a + 5) = 0$, to pa je res natanko za $a = 1$.

4. 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 $Y - X \sim \begin{pmatrix} -1 & 0 & 1 & 2 & 3 \\ 0'05 & 0'2 & 0'35 & 0'3 & 0'1 \end{pmatrix}$

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

Če so $X_1, \ldots X_n$ neodvisne in imajo vse Bernoullijevo porazdelitev $\mathrm{Ber}(p)$, je njihova vsota porazdeljena binomsko $\mathrm{Bin}(n, p)$. Vzemimo namreč zaporedje n neodvisnih slučajnih poskusov, pri katerih vsak uspe z verjetnostjo p. Naj bo:

$$X_i = \begin{cases} 1 & ; \text{če } i\text{-ti poskus uspe} \\ 0 & ; \text{sicer} \end{cases}$$

Tedaj je $S = X_1 + X_2 + \cdots + X_n$ ravno število uspelih poskusov, za to pa vemo, da je porazdeljeno binomsko $\mathrm{Bin}(n, p)$.

6. *Prvi način.* Za $n \in \mathbb{N}_0$ velja:

$$\begin{aligned} P(U = n) &= \sum_{k=0}^{n} P(S = k, T = n - k) = \\ &= \sum_{k=0}^{n} P(S = k)\, P(T = n - k) = \\ &= \sum_{k=0}^{n} \frac{\lambda^k \mu^{n-k} e^{-\lambda-\mu}}{k!\,(n-k)!} = \\ &= \frac{e^{-\lambda-\mu}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \mu^{n-k} = \\ &= \frac{(\lambda+\mu)^n\, e^{-\lambda-\mu}}{n!}\,, \end{aligned}$$

torej je $U \sim \mathrm{Poi}(\lambda + \mu)$.

Drugi način. Naj bosta $S_1, S_2, S_3, \ldots$ in $T_1, T_2, T_3, \ldots$ zaporedji slučajnih spremenljivk, pri čemer naj bosta S_k in T_k neodvisni za vsak k ter še $S_k \sim \mathrm{Bin}(m_k, p_k)$ in $T_k \sim \mathrm{Bin}(n_k, p_k)$. Tedaj je $S_k + T_k \sim \mathrm{Bin}(m_k + n_k, p_k)$.

Če gre m_k in n_k proti neskončno ter $m_k p_k \to \lambda$ in $n_k p_k \to \mu$ (taka zaporedja m_k, n_k in p_k se dajo konstruirati za poljubna λ in μ), se po Poissonovem obrazcu porazdelitev slučajnih spremenljivk S_k bliža Poissonovi porazdelitvi $\mathrm{Poi}(\lambda)$, porazdelitev slučajnih spremenljivk T_k pa Poissonovi porazdelitvi $\mathrm{Poi}(\mu)$. Še več: ko gre k proti neskončno, se porazdelitev slučajnega vektorja (S_k, T_k) bliža porazdelitvi slučajnega vektorja (S, T). Zato se morajo tudi porazdelitve vsot $S_k + T_k$ bližati porazdelitvi vsote $S + T$. Ker pa se porazdelitve teh vsot, ki so binomske, spet po Poissonovem obrazcu bližajo tudi Poissonovi porazdelitvi $\mathrm{Poi}(\lambda+\mu)$, od tod zaključimo, da mora biti $S + T \sim \mathrm{Poi}(\lambda + \mu)$.

7. Zamislimo si, da izvajamo zaporedje neodvisnih slučajnih poskusov, pri katerih vsak uspe z verjetnostjo p, dokler ne uspe n poskusov. Naj bo X_1 število poskusov do vključno prvega uspelega. Očitno je $X_1 \sim \mathrm{Geo}(p)$. Nadalje naj bo za $i = 2, 3, \ldots n$ slučajna spremenljivka X_i definirana kot število poskusov od nevključno $(i-1)$-tega do vključno i-tega uspelega. Tedaj vsota $S := X_1 + X_2 + \cdots + X_n$ predstavlja število vseh izvedb poskusa, za le-to pa vemo, da ima Pascalovo porazdelitev $\mathrm{NegBin}(n, p)$.

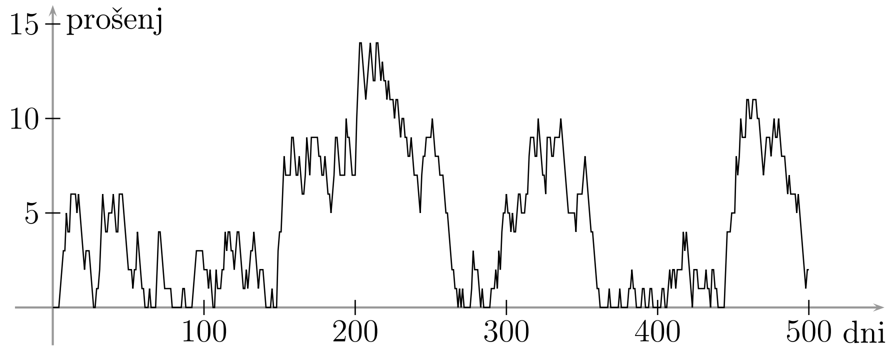
Znano je, da se zaporedje poskusov, ki sledijo i-temu uspelemu poskusu, pogojno glede na zgodovino (dogajanje do vključno i-tega uspelega poskusa) obnaša enako kot celotno zaporedje poskusov od začetka (za tem tiči tako imenovana *krepka časovno homogena lastnost Markova*). Zato so slučajne spremenljivke $X_1, X_2, \ldots X_n$ neodvisne in vse porazdeljene geometrijsko $\mathrm{Geo}(p)$. S tem je natanko določena porazdelitev slučajnega vektorja $(X_1, X_2, \ldots X_n)$.

Ker je porazdelitev vsote $S = X_1 + X_2 + \cdots + X_n$ odvisna zgolj od porazdelitve slučajnega vektorja $(X_1, X_2, \ldots X_n)$, lahko zaključimo, da, brž ko so slučajne spremenljivke $X_1, X_2, \ldots X_n$ neodvisne in porazdeljene geometrijsko $\mathrm{Geo}(p)$, ima njihova vsota Pascalovo porazdelitev $\mathrm{NegBin}(n, p)$.

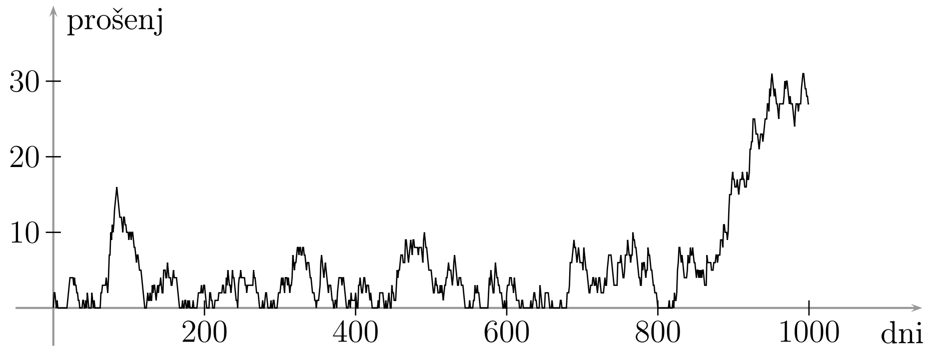
8. $1 - e^{-2\lambda}\left(1 + 2\lambda + 2\lambda^2 + \dfrac{7\lambda^3}{6}\right)$.

Simulacije nadaljnjega obnašanja števila prošenj, ki ostanejo na kupu, za različne λ:

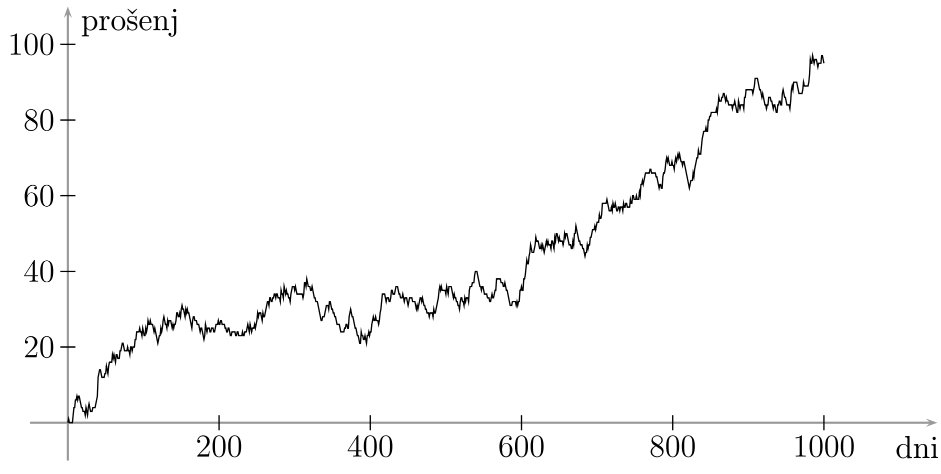
$\lambda = 0{,}9$:



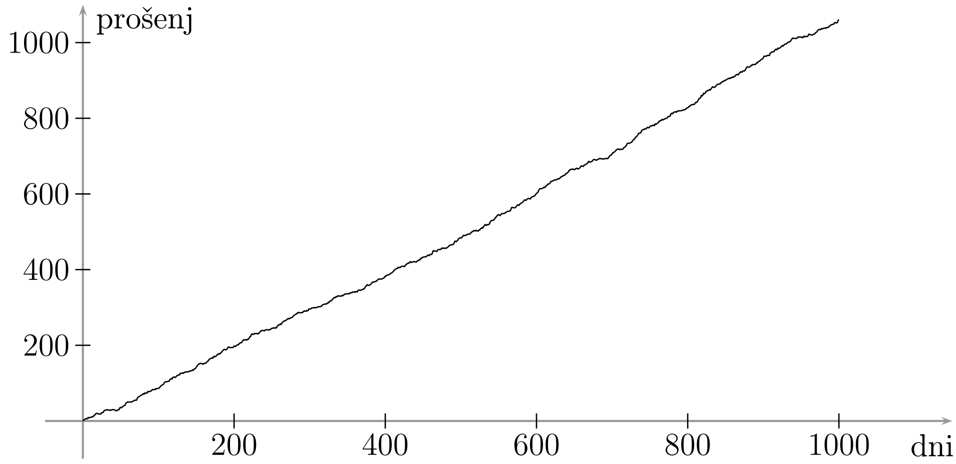
$\lambda = 1$:



$\lambda = 1{,}1$:



$\lambda = 2$:



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

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

11. Velja:
$$\boldsymbol{X}_1 + \boldsymbol{X}_2 = \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix},$$
kar je porazdeljeno n-razsežno normalno s pričakovano vrednostjo $\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$ in kovariančno matriko $\begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{I}_n \end{bmatrix} = \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2$.

12. $1/2 - \Phi(1/5) \doteq 0{\cdot}42074$.

13. Verjetnost bomo izrazili s standardnim dvorazsežnim normalnim vektorjem: poiskali bomo matriko $\mathbf{A}$, za katero velja $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \begin{bmatrix} 8 & 2 \\ 2 & 5 \end{bmatrix} =: \boldsymbol{\Sigma}$. Tedaj bo $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \stackrel{\mathrm{d}}{=} \mathbf{A} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$, kjer sta Z_1 in Z_2 neodvisni standardni enorazsežni normalni spremenljivki. To gre lahko na več načinov.

Prvi način: postavimo $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$, kar izračunamo s pomočjo lastnih vrednosti. Matrika $\boldsymbol{\Sigma}$ ima normiran lastni vektor $\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ za lastno vrednost 4 in normiran lastni vektor $\frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ za lastno vrednost 9. Sledi:

$$\boldsymbol{\Sigma}^{1/2} = \frac{1}{5} \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 14 & 2 \\ 2 & 11 \end{bmatrix}.$$

Torej je:

$$\begin{aligned} P(X_1 > 0, X_2 > 0) &= P(14Z_1 + 2Z_2 > 0,\ 2Z_1 + 11Z_2 > 0) = \\ &= \frac{1}{2\pi} \iint\limits_{\substack{14z_1+2z_2>0 \\ 2z_1+11z_2>0}} e^{-(z_1^2+z_2^2)/2} \, \mathrm{d}z_1 \, \mathrm{d}z_2 = \\ &= \frac{1}{2\pi} \iint\limits_{\substack{r>0 \\ -\operatorname{arctg}(2/11)<\varphi<\pi-\operatorname{arctg} 7}} r\, e^{-r^2/2} \, \mathrm{d}r \, \mathrm{d}\varphi = \\ &= \frac{1}{2} - \frac{\operatorname{arctg} 7 - \operatorname{arctg} \frac{2}{11}}{2\pi} \doteq \\ &\doteq 0{\cdot}301\,. \end{aligned}$$

Drugi način: z razcepom Choleskega dobimo $\mathbf{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} 4 & 0 \\ 1 & 3 \end{bmatrix}$. Dobimo:

$$\begin{aligned} P(X_1 > 0, X_2 > 0) &= P(Z_1 > 0,\ Z_1 + 3Z_2 > 0) = \\ &= \frac{1}{2\pi} \iint\limits_{\substack{z_1>0 \\ z_1+3z_2>0}} e^{-(z_1^2+z_2^2)/2} \, \mathrm{d}z_1 \, \mathrm{d}z_2 = \\ &= \frac{1}{2\pi} \iint\limits_{\substack{r>0 \\ -\operatorname{arctg}(1/3)<\varphi<\pi/2}} r\, e^{-r^2/2} \, \mathrm{d}r \, \mathrm{d}\varphi = \\ &= \frac{1}{4} + \frac{\operatorname{arctg} \frac{1}{3}}{2\pi} \doteq \\ &\doteq 0{\cdot}301\,. \end{aligned}$$

Opomba: seveda se rezultata ujemata. Kot stranski produkt smo dobili trigonometrijsko zvezo:

$$\frac{1}{2} - \frac{\operatorname{arctg} 7 - \operatorname{arctg} \frac{2}{11}}{2\pi} = \frac{1}{4} + \frac{\operatorname{arctg} \frac{1}{3}}{2\pi}$$

ali ekvivalentno:

$$\frac{\pi}{2} - \operatorname{arctg}\frac{1}{3} - \operatorname{arctg} 7 + \operatorname{arctg}\frac{2}{11} = 0$$

ali ekvivalentno:

$$\operatorname{arctg} 3 + \operatorname{arctg}\frac{2}{11} = \operatorname{arctg} 7\,.$$

Ta enakost pa sledi iz znane zveze $\operatorname{arctg} x + \operatorname{arctg} y = \operatorname{arctg}\dfrac{x+y}{1-xy}$, ki velja, brž ko je $xy < 1$.

14. Slučajni vektor:

$$\begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 + \mathbf{C}\boldsymbol{X}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$$

je porazdeljen $(m+n)$-razsežno normalno s pričakovano vrednostjo:

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 + \mathbf{C}\boldsymbol{\mu}_1 \end{bmatrix}$$

in kovariančno matriko:

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{C}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} =$$
$$= \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{11}\mathbf{C}^{\mathrm{T}} \\ \mathbf{\Sigma}_{21} + \mathbf{C}\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{22} + \mathbf{C}\mathbf{\Sigma}_{12} + \mathbf{\Sigma}_{21}\mathbf{C}^{\mathrm{T}} + \mathbf{C}\mathbf{\Sigma}_{11}\mathbf{C}^{\mathrm{T}} \end{bmatrix}.$$

Torej sta slučajna vektorja $\boldsymbol{X}_1$ in $\boldsymbol{X}_2 + \mathbf{C}\boldsymbol{X}_1$ neodvisna natanko tedaj, ko je $\mathbf{C} = -\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$.

15. Slučajni vektor (X, Y) ima dvorazsežno gostoto:

$$p_{X,Y}(x,y) = \frac{1}{2\pi}\, e^{-(x^2+y^2)/2}\,.$$

Postavimo lahko $h(x,y) = \bigl(\sqrt{x^2+y^2}, \arg(x,y)\bigr)$, kar je bijektivna preslikava iz $\mathbb{R}^2 \setminus \bigl((-\infty,0] \times \{0\}\bigr)$ v $(0,\infty) \times (-\pi,\pi)$ (poltrak, ki smo ga izločili, ima volumen nič, zato se porazdelitev ohrani). Velja:

$$h^{-1}(r,\theta) = \bigl(r\cos\theta, r\sin\theta\bigr)\,, \qquad J(h^{-1})(r,\theta) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r$$

Sledi:

$$p_{R,\Theta}(r,\theta) = \begin{cases} \dfrac{r}{2\pi}\, e^{-r^2/2} & ;\ r > 0, -\pi < \theta < \pi \\ 0 & ;\ \text{sicer} \end{cases}.$$

Opazimo, da sta slučajni spremenljivki R in Θ neodvisni. Slučajna spremenljivka Θ je porazdeljena enakomerno na intervalu $(-\pi,\pi)$, slučajna spremenljivka R pa je porazdeljena zvezno z gostoto:

$$p_R(r) = \begin{cases} r\, e^{-r^2/2} & ;\ r > 0 \\ 0 & ;\ \text{sicer} \end{cases}.$$

Tako je možno generirati psevdonaključna števila z normalno porazdelitvijo. Če sta namreč U in V neodvisni slučajni spremenljivki, porazdeljeni enakomerno na intervalu $(0, 1)$, izračunamo:

$$R = \sqrt{2 \ln U}\,, \qquad \Theta = \pi(2V - 1)\,,$$
$$X = R \cos \Theta\,, \qquad Y = R \sin \Theta\,.$$

Da se preveriti, da imata v tem primeru R in Θ ustrezni porazdelitvi. Tako sta potem tudi X in Y neodvisni in porazdeljeni standardno normalno.

16. Slučajni vektor (X, Y) ima dvorazsežno gostoto:

$$p_{X,Y}(x, y) = \frac{1}{2\pi}\, e^{-(x^2+(y-1)^2)/2}\,.$$

Tega preslikamo s preslikavo $h(x, y) = \left(xy, \frac{y}{x}\right)$ (seveda je slučajni vektor (X, Y) skoncentriran na množici $(\mathbb{R} \setminus \{0\}) \times \mathbb{R}$, na kateri je preslikava A definirana). Velja:

$$Jh = \begin{vmatrix} y & x \\ -\frac{y}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{2y}{x}\,.$$

Pišimo $U = XY$ in $V = Y/X$. Če je $u > 0$ in $v > 0$, ima sistem enačb $h(x, y) = (u, v)$ dve rešitvi:

$$x = \sqrt{\frac{u}{v}}\,, \quad y = \sqrt{uv}$$

in

$$x = -\sqrt{\frac{u}{v}}\,, \quad y = -\sqrt{uv}$$

Če je $u < 0$ in $v < 0$, ima sistem prav tako dve rešitvi:

$$x = \sqrt{\frac{u}{v}}\,, \quad y = -\sqrt{uv}$$

in

$$x = -\sqrt{\frac{u}{v}}\,, \quad y = \sqrt{uv}$$

Primer, ko je $u = 0$ ali $v = 0$, lahko zanemarimo, ker ima Lebesgueovo mero nič. Sicer pa sistem nima rešitev. Opazimo še, da vselej velja $Jh(x, y) = 2v$. Ko vse skupaj sestavimo, dobimo:

$$p_{U,V}(u, v) = \begin{cases} \dfrac{1}{4\pi|v|}\left(e^{-(u/v+(\sqrt{uv}-1)^2)/2} + e^{-(u/v+(\sqrt{uv}+1)^2)/2}\right) & ;\ u, v > 0 \text{ ali } u, v < 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

17. $p_Z(z) = \int_{-\infty}^{\infty} p_X(x)\, p_Y(x-z)\, \mathrm{d}x = \dfrac{\lambda}{2} e^{-\lambda|z|}$.
Tej porazdelitvi pravimo *Laplaceova porazdelitev.*

18. $c = 2, \quad p_Z(z) = \int_{-\infty}^{\infty} p_{X,Y}(x, xz)\, |x|\, \mathrm{d}x = \begin{cases} (1+z)^{-2} & ; \ z > 0 \\ 0 & ; \ \text{sicer} \end{cases},$
$P(X < 2Y) = P(Z > 1/2) = 2/3$.

19. $F_Z(z) = \begin{cases} 0 & ; \ z \le 0 \\ z(1 - \ln z) & ; \ 0 \le z \le 1 \\ 1 & ; \ z \ge 1 \end{cases}, \qquad p_Z(z) = \begin{cases} -\ln z & ; \ 0 < z < 1 \\ 0 & ; \ \text{sicer} \end{cases}$

20. *Prvi način*: s pomočjo transformacije gostote. Velja:

$$p_Z(z) = \int_{-\infty}^{\infty} \sum_{y;\, x^2+y^2=z} \frac{p_{X,Y}(x,y)}{|2y|}\, \mathrm{d}x\,.$$

Enačba $x^2 + y^2 = z$ je rešljiva na y natanko tedaj, ko je $z - x^2 \ge 0$, to pa je natanko tedaj, ko je $z \ge 0$ in $-\sqrt{z} \le x \le \sqrt{z}$. Rešitvi prej omenjene enačbe sta $y = \sqrt{z - x^2}$ in $x = -\sqrt{z - x^2}$. Torej za $z > 0$ velja:

$$\begin{aligned} p_Z(z) &= \int_{-\sqrt{z}}^{\sqrt{z}} \left(\frac{p_{X,Y}\big(x, \sqrt{z - x^2}\big)}{\big|2\sqrt{z - x^2}\big|} + \frac{p_{X,Y}\big(x, -\sqrt{z - x^2}\big)}{\big|-2\sqrt{z - x^2}\big|} \right) \mathrm{d}x = \\ &= \frac{C}{z^{5/2}(1+z)} \int_{-\sqrt{z}}^{\sqrt{z}} x^2 \sqrt{z - x^2}\, \mathrm{d}x\,. \end{aligned}$$

Z upoštevanjem sodosti in substitucijo $t = x^2/z$ dobimo:

$$p_Z(z) = \frac{C_1}{(1+z)\sqrt{z}}\,,$$

kjer je:

$$C_1 = C \int_0^1 \sqrt{t(1-t)}\, \mathrm{d}t\,.$$

S substitucijo $t = 1/(1+z)$ izračunamo:

$$\int_0^{\infty} \frac{\mathrm{d}z}{(1+z)\sqrt{z}} = \int_0^1 \frac{\mathrm{d}t}{\sqrt{t(1-t)}} = \mathrm{B}\big(\tfrac{1}{2}, \tfrac{1}{2}\big) = \frac{\left[\Gamma\big(\tfrac{1}{2}\big)\right]^2}{\Gamma(1)} = \pi\,,$$

od koder sledi $C_1 = 1/\pi$, torej:

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

Sedaj lahko izračunamo tudi konstanto C. Iz:

$$\int_0^1 \sqrt{t(1-t)}\,\mathrm{d}t = \mathrm{B}\big(\tfrac{3}{2},\tfrac{3}{2}\big) = \frac{\Big[\Gamma\big(\frac{3}{2}\big)\Big]^2}{\Gamma(3)} = \frac{\pi}{8}$$

sledi $C = 8C_1/\pi = 8/\pi^2$.
Drugi način: z uporabo kumulativne porazdelitvene funkcije in polarnih koordinat. Za $z \ge 0$ velja:

$$F_Z(z) = \iint_{x^2+y^2<z} \frac{Cx^2y^2}{(1+x^2+y^2)(x^2+y^2)^{5/2}}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Z uporabo polarnih koordinat dobimo:

$$F_Z(z) = C_1 \int_0^{\sqrt{z}} \frac{\mathrm{d}r}{1+r^2} = C_2 \operatorname{arctg}\sqrt{r}\,,$$

kjer je:

$$C_2 = C\int_0^{2\pi} \cos^2\varphi\sin^2\varphi\,\mathrm{d}\varphi\,.$$

Ker mora biti $\lim_{z\to\infty} F_Z(z) = 1$, velja $C_2 = 2/\pi$. Za $z > 0$ torej velja:

$$p_Z(z) = F_Z'(z) = \frac{1}{\pi(1+z)\sqrt{z}}\,,$$

kar je isto kot prej. Sedaj lahko ponovno izračunamo tudi konstanto C. Iz:

$$\int_0^{2\pi} \cos^2\varphi\sin^2\varphi\,\mathrm{d}\varphi = \frac{1}{4}\int_0^{2\pi}\sin^2(2\varphi)\,\mathrm{d}\varphi = \frac{\pi}{4}$$

sledi $C = 4C_2/\pi = 8/\pi^2$, spet isto kot prej.

21. *Prvi način.* Najprej bomo izračunali kumulativno porazdelitveno funkcijo slučajne spremenljivke T_n. Dogodek $T_n \le t$ se zgodi, če se je do časa T_n zgodilo že vsaj n danih pojavov, torej za $t \ge 0$ velja:

$$F_{T_n}(t) = P(T_n \le t) = 1 - e^{-\lambda t}\left(1 + \lambda t + \frac{\lambda^2 t^2}{2!} + \cdots + \frac{\lambda^{n-1}t^{n-1}}{(n-1)!}\right).$$

Po odvajanju dobimo:

$$p_{T_n}(t) = \begin{cases} \dfrac{\lambda^n}{(n-1)!}\,t^{n-1}\,e^{-\lambda t} & ;\ t \ge 0 \\ 0 & ;\ \text{sicer} \end{cases},$$

torej je $T_n \sim \mathrm{Gama}(n,\lambda)$. Za $n = 1$ velja $\mathrm{Gama}(1,\lambda) = \mathrm{Exp}(\lambda)$.

Drugi način. Označimo z N_t število pojavov, ki se je zgodilo do vključno časa t. Za $0 \le t < t + h$ velja:

$$F_{T_n}(t+h) - F_{T_n}(t) = P(N_{t+h} \ge n) - P(N_t \ge n)\,.$$

Situacija, ko je $N_t \sim \operatorname{Poi}(\lambda t)$ in $N_{t+h} \sim \operatorname{Poi}(\lambda(t+h))$, nastopi tudi v primeru, ko je razlika $N_{t+h} - N_t$ neodvisna od N_t in porazdeljena po Poissonu $\operatorname{Poi}(\lambda h)$; to sicer velja tudi pri Poissonovem procesu, a pri izračunu zgornje verjetnosti lahko to privzamemo brez škode za splošnost. Tedaj velja:

$$\begin{aligned}
F_{T_n}(t+h) - F_{T_n}(t) &= P(N_t < n, N_{t+h} \ge n) = \\
&= \sum_{i=0}^{n-1} \sum_{j=n}^{\infty} P(N_t = i, N_{t+h} = j) = \\
&= \sum_{i=0}^{n-1} \sum_{j=n}^{\infty} P(N_t = i, N_{t+h} - N_t = j - i) = \\
&= \sum_{i=0}^{n-1} \sum_{j=n}^{\infty} \frac{\lambda^j t^i h^{j-i} e^{-\lambda(t+h)}}{i!\,(j-i)!}\,.
\end{aligned}$$

Opazimo, da je potenca pri h vselej najmanj 1. Če delimo s h in naredimo limito, ko gre h proti nič, na levi strani dobimo natanko odvod, to je gostoto porazdelitve. Na desni strani pa ostanejo le členi s h^1 (utemeljitev konvergence bomo opustili), to pa je le člen z $i = n - 1$ in $n = k$. Torej velja:

$$p_{T_n}(t) = \frac{\lambda^n t^{n-1}\, e^{-\lambda t}}{(n-1)!}\,,$$

kar je isto kot prej.

22. Za $a, b \in \mathbb{N}$ se da porazdelitev izpeljati z verjetnostnim premislekom. Vzemimo Poissonov tok pojavov z intenziteto λ ter naj bo S čas, ob katerem se zgodi a-ti, T pa čas, ki mine od a-tega do $(a+b)$-tega pojava. Tedaj iz prejšnje naloge dobimo, da je $S \sim \operatorname{Gama}(a, \lambda)$. Nadalje je $S + T$ čas, ob katerem se zgodi $(a+b)$-ti pojav, zato je $S + T \sim \operatorname{Gama}(a+b, \lambda)$.

Tudi za Poissonov tok velja *krepka časovno homogena lastnost Markova*: pogojno na zgodovino, t. j. dogajanje do časa S, se nadaljevanje Poissonovega toka obnaša enako kot Poissonov tok od začetka. Zato ima ob zgornji definiciji slučajna spremenljivka T porazdelitev $\operatorname{Gama}(b, \lambda)$ in je neodvisna od S. S tem je navzkrižna porazdelitev slučajnih spremenljivk S in T natančno določena.

Ker je porazdelitev vsote natančno določena z navzkrižno porazdelitvijo, lahko sklepamo, da, brž ko sta slučajni spremenljivki $S \sim \operatorname{Gama}(a, \lambda)$ in $T \sim \operatorname{Gama}(b, \lambda)$ neodvisni in je $a, b \in \mathbb{N}$, velja $U = S + T \sim \operatorname{Gama}(a+b, \lambda)$.

Za splošni primer pa bomo porazdelitev izpeljali računsko. Za $u > 0$ velja:

$$p_U(u) = \int_{-\infty}^{\infty} p_S(s)\, p_T(u-s)\, \mathrm{d}s = \frac{\lambda^{a+b}\, e^{-\lambda u}}{\Gamma(a)\,\Gamma(b)} \int_0^u s^{a-1} (u-s)^{b-1}\, \mathrm{d}s\,.$$

S substitucijo $t = s/u$ dobimo:

$$p_U(u) = C\, u^{a+b-1}\, e^{-\lambda u}\,,$$

kjer je:

$$C = \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)} \int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t = \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)}\,\mathrm{B}(a,b)\,.$$

Od tod dobimo, da je res $U \sim \mathrm{Gama}(a+b,\lambda)$. Ker od tod sledi, da mora biti $C = \lambda^{a+b}/\,\Gamma(a+b)$, sledi še znana zveza:

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}\,,$$

23. Za $0 \le t \le 1$ označimo z N_t število gostov, ki so prišli na zabavo do časa t. Tedaj je $\{T_k \le t\} = \{N_t \ge k\}$.

Prvi način. Ker je $N_t \sim \mathrm{Bin}(n,t)$, ima T_k kumulativno porazdelitveno funkcijo:

$$F_{T_k}(t) = \sum_{l=k}^{n} \binom{n}{l} t^l (1-t)^{n-l}\,; \quad 0 \le t \le 1\,.$$

Z odvajanjem dobimo še porazdelitveno gostoto (pazimo na zadnji člen):

$$\begin{aligned}
p_{T_k}(t) &= \sum_{l=k}^{n} \binom{n}{l} \Big[l\, t^{l-1}(1-t)^{n-l} - (n-l)t^l(1-t)^{n-l-1} \Big] = \\
&= \sum_{l=k}^{n} \binom{n}{l} l\, t^{l-1}(1-t)^{n-l} - \sum_{l=k}^{n-1} \binom{n}{l} (n-l)t^l(1-t)^{n-l-1} = \\
&= \sum_{l=k}^{n} \frac{n!}{(l-1)!\,(n-l)!}\, t^{l-1}(1-t)^{n-l} - \sum_{l=k}^{n-1} \frac{n!}{l!\,(n-l-1)!}\, t^l(1-t)^{n-l-1} = \\
&= \sum_{l=k}^{n} \frac{n!}{(l-1)!\,(n-l)!}\, t^{l-1}(1-t)^{n-l} - \sum_{j=k+1}^{n} \frac{n!}{(j-1)!\,(n-j)!}\, t^{j-1}(1-t)^{n-j} = \\
&= \frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k}\,.
\end{aligned}$$

Dobili smo *porazdelitev beta*: $T_k \sim \mathrm{Beta}(k, n-k+1)$ (glej tudi 25. nalogo).

Drugi način. Gostoto zlahka dobimo z naslednjim fizikalnim, a matematično ne ravno korektnim premislekom: dogodek $\{T_k \in [t, t+\mathrm{d}t]\}$ pomeni, da je do časa t prišlo $k-1$ gostov, en gost je prišel v infinitezimalnem časovnem intervalu od t do $t+\mathrm{d}t$, preostalih $n-k$ gostov pa je prišlo od časa $t+\mathrm{d}t$ naprej. Torej je:

$$P(T_k \in [t, t+\mathrm{d}t]) = \frac{n!}{(k-1)!\,1!\,(n-k)!} t^{k-1}\,\mathrm{d}t\,(1-t)^{n-k}\,,$$

od koder sledi:

$$p_{T_k}(t) = \frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k}\,,$$

kar je isto kot prej. Korektna izpeljava pa je naslednja: za $0 \le t < t + h \le 1$ velja:

$$F_{T_k}(t+h) - F_{T_k}(t) = P(N_{t+h} \ge k) - P(N_t \ge k)\,.$$

Ker je $\{N_t \ge k\} \subseteq \{N_{t+h} \ge k\}$, je tudi:

$$\begin{aligned} F_{T_k}(t+h) - F_{T_k}(t) &= P(N_t < k, N_{t+h} \ge k) = \\ &= P(N_t < k, N_1 - N_{t+h} \le n - k) = \\ &= \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} P(N_t = i, N_{t+h} - N_t = n - i - j, N_1 - N_{t+h} = j) = \\ &= \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \frac{n!}{i!\,(n-i-j)!\,j!}\, t^i h^{n-i-j} (1-t-h)^j\,. \end{aligned}$$

Opazimo, da je eksponent pri h vselej najmanj 1. Če delimo s h in naredimo limito, ko gre h proti nič, na levi strani dobimo natanko *desni* odvod funkcije F_{T_k} v točki t. Na desni strani pa ostanejo le členi s h^1, to pa je le člen z $i = k - 1$ in $j = n - k$. Torej je prej omenjeni desni odvod enak:

$$\frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k}\,,$$

kar je spet isto kot prej. Dokazati moramo le še, da enako velja tudi za *levi* odvod. Za $0 \le t - h < t \le 1$ podobno kot prej izračunamo:

$$F_{T_k}(t) - F_{T_k}(t-h) = \sum_{i=0}^{k-1} \sum_{j=0}^{n-k} \frac{n!}{i!\,(n-i-j)!\,j!}\, (t-h)^i h^{n-i-j} (1-t)^j\,.$$

in limita, ko gre h proti nič, je ista kot prej. To pomeni, da je F_{T_k} na celotnem intervalu $(0, 1)$ odvedljiva, in ni se težko prepričati, da je odvod zvezen. Poleg tega je F_{T_k} povsod zvezna. Sledi:

$$p_{T_k}(t) = \frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k}\,.$$

24. a) Če z p_X in p_Y označimo ustrezni gostoti:

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

velja:

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

b) Spomnimo se, da je:

$$\frac{X_1^2+\cdots+X_n^2}{\sigma^2} \sim \chi^2(n) = \mathrm{Gama}\left(\frac{n}{2},\frac{1}{2}\right) \,,$$

torej je (glej 35. nalogo iz 4. razdelka):

$$\frac{X_1^2+\cdots+X_n^2}{n} \sim \mathrm{Gama}\left(\frac{n}{2},\frac{n}{2\sigma^2}\right) \,.$$

Tako lahko primer prevedemo na prejšnjo točko in dobimo, da ima Studentova porazdelitev z n prostostnimi stopnjami gostoto:

$$p_T(t) = \frac{1}{\mathrm{B}\big(\frac{n}{2},\frac{1}{2}\big)\sqrt{n}\left(1+\dfrac{t^2}{n}\right)^{(n+1)/2}} \,,$$

kar je neodvisno od σ. Ko gre n proti neskončno, to konvergira proti gostoti standardne normalne porazdelitve.

25. Najprej opazimo, da Z skoraj gotovo zavzame vrednosti na intervalu $(0,1)$. Za $0<z<1$ ima enačba $x/(x+y)=z$ enolično rešitev $y=x(1-z)/z$. Za ta y je:

$$\frac{\partial}{\partial y}\frac{x}{x+y} = -\frac{x}{(x+y)^2} = -\frac{x}{z^2} \,.$$

Po krajšem računu dobimo, da za $0<z<1$ velja:

$$\begin{aligned} p_Z(z) &= \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)} \frac{(1-z)^{b-1}}{z^{b+1}} \int_0^\infty x^{a+b-1} e^{-\lambda x/z}\,\mathrm{d}x = \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} z^{a-1}(1-z)^{b-1} \,. \end{aligned}$$

Gostota je ravno integrand v integralski definiciji funkcije beta in tudi dobljeno porazdelitev imenujemo *porazdelitev beta* in pišemo $Z \sim \mathrm{Beta}(a,b)$. Ker je

$\int_{-\infty}^{\infty} p_Z(z)\,\mathrm{d}z = \int_0^1 p_Z(z)\,\mathrm{d}z = 1$, se predfaktor izraža tudi s funkcijo beta, tako da velja:

$$p_Z(z) = \begin{cases} \dfrac{1}{\mathrm{B}(a,b)}\, z^{a-1}(1-z)^{b-1} & ;\ 0 < z < 1 \\ 0 & ;\ \text{sicer} \end{cases}.$$

in spet sledi dobro znana zveza $\mathrm{B}(a,b) = \dfrac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}$.

26. *Prvi način.* Če ustrezni gostoti označimo z p_X in p_Y, za $q > 0$ velja:

$$\begin{aligned} p_Q(q) &= \int_0^\infty p_X(qy)\,p_Y(y)\,y\,\mathrm{d}y = \\ &= \frac{\lambda^a \mu^b}{\Gamma(a)\,\Gamma(b)}\, q^{a-1} \int_0^\infty y^{a+b-1} e^{-(\lambda q+\mu)y}\,\mathrm{d}y = \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\, \frac{\lambda^a\mu^b}{(\lambda q+\mu)^{a+b}}\, q^{a-1} = \\ &= \frac{1}{\mathrm{B}(a,b)\left(1+\dfrac{\mu}{\lambda q}\right)^a \left(1+\dfrac{\lambda q}{\mu}\right)^b q}\,. \end{aligned}$$

Torej je:

$$p_Q(q) = \begin{cases} \left[\mathrm{B}(a,b)\left(1+\frac{\mu}{\lambda q}\right)^a\left(1+\frac{\lambda q}{\mu}\right)^b q\right]^{-1} & ;\ q > 0 \\ 0 & ;\ \text{sicer} \end{cases}.$$

Drugi način: prevedemo na prejšnjo nalogo, najprej za poseben primer, ko je $\lambda = \mu$. Če definiramo $Z := X/(X+Y) \sim \mathrm{Beta}(a,b)$, velja $Q = h(Z)$, kjer je:

$$h(z) = \frac{z}{z-1}\,, \quad h^{-1}(q) = \frac{q}{q+1}\,, \quad (h^{-1})'(q) = -\frac{1}{(q+1)^2}\,,$$

torej za $q > 0$ velja:

$$p_Q(q) = p_Z\left(\frac{q}{q+1}\right)\frac{1}{(q+1)^2} = \frac{q^{a-1}}{\mathrm{B}(a,b)(q+1)^{a+b}}$$

ali tudi:

$$p_Q(q) = \frac{1}{\mathrm{B}(a,b)\left(1+\frac{1}{q}\right)^a(1+q)^b\,q}\,.$$

Za splošni primer definiramo:

$$X' := \lambda X \sim \mathrm{Gama}(a,1)\,, \quad Y' := \mu Y \sim \mathrm{Gama}(b,1) \quad \text{in} \quad Q' := \frac{X'}{Y'} = \frac{\lambda}{\mu}\,Q\,.$$

Tedaj velja $p_Q(q) = \dfrac{\lambda}{\mu}\, p_{Q'}\left(\dfrac{\lambda}{\mu}\,q\right)$, ker je isto kot prej.

27. Naj bosta U in V neodvisni ter $U \sim \chi^2(m)$ in $V \sim \chi^2(n)$. Tedaj je

$$F := \frac{n}{m}\frac{U}{V} \sim \mathrm{F}(m, n) \quad \text{in} \quad B := \frac{U}{U+V} \sim \mathrm{Beta}\left(\frac{m}{2}, \frac{n}{2}\right).$$

Med tema slučajni spremenljivki velja funkcijska zveza:

$$B = \frac{mF}{mF+n} \quad \text{oziroma} \quad F = \frac{n}{m}\frac{B}{1-B}.$$

6. Matematično upanje in sorodne karakteristike

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

2. $E(X) = 511, \quad D(X) = 129, \quad \sigma(X) \doteq 11{\cdot}36.$

3. Matematično upanje:

$$E(X) = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

ne obstaja. Pač pa obstaja matematično upanje:

$$E\big(\sqrt{X}\big) = \frac{1}{2} + \frac{1}{2\sqrt{2}} + \frac{1}{4} + \frac{1}{4\sqrt{2}} + \cdots = \frac{1}{2 - \sqrt{2}} \doteq 1{\cdot}707\,.$$

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

5. Matematično upanje bi bilo enako:

$$\int_0^\infty \frac{x}{(1+x)^2}\,\mathrm{d}x = \lim_{b\to\infty} \left(\ln(1+b) - \frac{b}{1+b}\right)\Big|_0^\infty\,,$$

kar ne obstaja.

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

7. $E(1 + 4X^2) = 2, \quad D(X) = 1/4.$

8. $3{\cdot}55.$

9. $1/2.$

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

11. Zaradi enakomerne porazdelitve smeri lahko vesoljsko smer, iz katere pade meteorit, fiksiramo. Ravne trajektorije, ki sekajo planet, ustrezajo točkam na krogu. Za enoto lahko brez škode za splošnost vzamemo kar polmer planeta. Če trajektorija ustreza točki (X, Y) iz enotskega kroga, meteorit pade pod kotom $\arccos\sqrt{X^2 + Y^2}$. Tako je pričakovani kot enak:

$$\bar{\alpha} := E\left[\arccos\sqrt{X^2 + Y^2}\right],$$

kjer je točka (X, Y) porazdeljena enakomerno po enotskem krogu. Sledi:

$$\bar{\alpha} = \frac{1}{\pi}\int_{x^2+y^2\le 1} \arccos\sqrt{x^2 + y^2}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Prevedba na polarne koordinate nam da:

$$\bar{\alpha} = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 r \arccos r \, \mathrm{d}r \, \mathrm{d}\varphi = 2 \int_0^1 r \arccos r \, \mathrm{d}r \, \mathrm{d}\varphi \, .$$

S substitucijo $r = \cos t$ dobimo:

$$\bar{\alpha} = 2 \int_0^{\pi/2} t \cos t \sin t \, \mathrm{d}t = \int_0^{\pi/2} t \sin(2t) \, \mathrm{d}t = - \left. \frac{t \cos(2t)}{2} \right|_0^{\pi/2} + \frac{1}{2} \int_0^{\pi/2} \cos(2t) \, \mathrm{d}t = \\ = \frac{\pi}{4} \, .$$

Opomba. Porazdelitev smeri, iz katere pade meteorit, je enakomerna *le za cel planet*, ni pa enakomerna za posamezno točko, na katero pade meteorit. Če si namreč zamislimo neko majhno površino okrog izbrane točke, vsaki smeri ustreza prizma žarkov, ki ponazarjajo trajektorije meteoritov, ki padejo iz izbrane smeri na to površino. Toda smeri, ki je pravokotna na površino, ustreza prizma z večjim presekom kot pa smeri, ki je na površino poševna. Zato bi, če bi gledali porazdelitev smeri za posamezno točko, bolj poševne smeri (t. j. tiste, ki ustrezajo manjšim kotom) imele manjšo gostoto verjetnosti.

Opomba. Pričakovana vrednost, ki smo jo dobili, se ujema s pričakovano vrednostjo kota, ki bi bil porazdeljen enakomerno na intervalu od 0 do $\pi/2$. Vendar pa kot, pod katerim pade meteorit, ni porazdeljen enakomerno, saj za $0 \le t \le \pi/2$ velja:

$$F_\alpha(t) = P(\alpha \le t) = P\big(\arccos \sqrt{X^2 + Y^2} \le t\big) = P\big(\sqrt{X^2 + Y^2} \ge \cos t\big) = 1 - \cos t \, ,$$

od koder sledi $p_\alpha(t) = \sin t \cos t$.

12. Velja $X = \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \mathbf{1}_{A_3} + 2 \cdot \mathbf{1}_{A_4} + 3 \cdot \mathbf{1}_{A_5}$, kjer je:

$$\begin{aligned} A_1 &= \{\text{prva ploščica z oznako } \boxed{1} \text{ je bila odkrita}\} \, , \\ A_2 &= \{\text{druga ploščica z oznako } \boxed{1} \text{ je bila odkrita}\} \, , \\ A_3 &= \{\text{tretja ploščica z oznako } \boxed{1} \text{ je bila odkrita}\} \, , \\ A_4 &= \{\text{ploščica z oznako } \boxed{2} \text{ je bila odkrita}\} \, , \\ A_5 &= \{\text{ploščica } \boxed{3} \text{ je bila odkrita}\} \, . \end{aligned}$$

Zaradi simetrije so verjetnosti vseh dogodkov A_i enake $1/2$. Sledi $E(S) = \frac{1}{2}(1 + 1 + 1 + 2 + 3) = 4$.

13. Označimo cene v posameznih trgovinah z X_1, X_2 in X_3, poleg tega pa naj bo še Z_i indikator dogodka, da iskani čevlji v i-ti trgovini stanejo največ 50 evrov. Tedaj je:

$$\begin{aligned} C &= C \, \mathbf{1}(X_1 \le 50) + C \, \mathbf{1}(X_1 > 50, X_2 < 50) + C \, \mathbf{1}(X_1 > 50, X_2 > 50) = \\ &= X_1 \, \mathbf{1}(X_1 \le 50) + X_2 \, \mathbf{1}(X_1 > 50) \, \mathbf{1}(X_2 \le 50) + X_3 \, \mathbf{1}(X_1 > 50) \, \mathbf{1}(X_2 > 50) \, . \end{aligned}$$

Sledi:

$$\begin{aligned} E(C) &= E\big[X_1\,\mathbf{1}(X_1 \le 50)\big] + E\big[\mathbf{1}(X_1 > 50)\big]\,E\big[X_2\,\mathbf{1}(X_2 \le 50)\big] + \\ &\qquad + E\big[\mathbf{1}(X_1 > 50)\big]\,E\big[\mathbf{1}(X_2 > 50)\big]\,E(X_3) = \\ &= E\big[X_1\,\mathbf{1}(X_1 \le 50)\big] + P(X_1 > 50)\,E\big[X_2\,\mathbf{1}(X_2 \le 50)\big] + \\ &\qquad + P(X_1 > 50)\,P(X_2 > 50)\,E(X_3) = \\ &= \frac{36+45}{3} + \frac{1}{3}\int_{40}^{50} x\,\frac{60-x}{200}\,\mathrm{d}x + \frac{1}{3}\int_{50}^{60}\frac{60-x}{200}\,\mathrm{d}x \cdot 54 = \\ &= 42\tfrac{11}{18} \doteq 42\text{'}61\,. \end{aligned}$$

14. Spomnimo se, da ima hipergeometrijsko porazdelitev $\mathrm{Hip}(s; r, n)$ slučajna spremenljivka N, ki šteje število rdečih kroglic med s kroglicami, ki jih na slepo in brez vračanja izvlečemo iz posode, v kateri je n kroglic, od tega r rdečih. Pišemo lahko $N = X_1 + X_2 + \cdots + X_s$, kjer je X_i indikator dogodka, da je i-ta izvlečena kroglica rdeča. Očitno je $E(X_i) = r/n$, torej je $E(N) = rs/n$.

15. Velja $\mathbf{1}_{\overline{A}} = 1 - \mathbf{1}_A$ in $\mathbf{1}_{A\cap B} = \mathbf{1}_A\,\mathbf{1}_B$. Načelo vključitev in izključitev predstavlja formulo za verjetnost unije dogodkov, toda glede na pravkar izpeljani zvezi nam pri uniji bolj prav pride zapis:

$$A_1 \cup A_2 \cup \cdots \cup A_n = \overline{\overline{A}_1 \cap \overline{A}_2 \cap \cdots \cap \overline{A}_n}\,.$$

Sledi:

$$\begin{aligned} \mathbf{1}_{A_1\cup A_2\cup\cdots\cup A_n} &= 1 - (1-\mathbf{1}_{A_1})(1-\mathbf{1}_{A_2})\cdots(1-\mathbf{1}_{A_n}) = \\ &= 1 - 1 + \mathbf{1}_{A_1} + \mathbf{1}_{A_2} + \cdots + \mathbf{1}_{A_n} - \\ &\qquad - \mathbf{1}_{A_1\cap A_2} - \mathbf{1}_{A_1\cap A_3} - \cdots - \mathbf{1}_{A_{n-1}\cap A_n} + \\ &\qquad + \mathbf{1}_{A_1\cap A_2\cap A_3} + \mathbf{1}_{A_1\cap A_2\cap A_4} + \cdots + \mathbf{1}_{A_{n-2}\cap A_{n-1}\cap A_n} - \\ &\qquad\qquad \cdots \\ &\qquad + (-1)^{n+1}\,\mathbf{1}_{A_1\cap A_2\cap\cdots\cap A_n} \end{aligned}$$

Načelo vključitev in izključitev od tod takoj dobimo z uporabo matematičnega upanja in upoštevanjem njegove linearnosti.

16. Zahtevano sledi iz formule:

$$N = \sum_{n=1}^{\infty} \mathbf{1}(N \ge n) = \sum_{n=0}^{\infty} \mathbf{1}(N > n)$$

17. Označimo število vseh metov z N. Iz prejšnje naloge sledi:

$$a := E(N) = \sum_{n=0}^{\infty} p_n\,,$$

kjer je $p_n = P(N > n)$. V 8. nalogi v 3. razdelku smo izračunali, da repne verjetnosti p_n zadoščajo rekurzivni zvezi:

$$p_n = \frac{p_{n-1}}{2} + \frac{p_{n-2}}{4}$$

z začetnima pogojema $p_0 = p_1 = 1$. Iz te zveze jih lahko celo eksplicitno izračunamo (izražajo se s Fibonaccijevim zaporedjem) in z njihovim seštetjem potem dobimo matematično upanje. Vendar pa se da to narediti še hitreje, saj lahko iz rekurzivne zveze izpeljemo $a - 2 = \frac{1}{2}(a-1) + \frac{1}{4}a$, kar nam da $a = 6$.

18. Za $i = 1, \ldots 8$ naj bi X_i indikator dogodka, da i-ti igralec stavi svojo ženo. Tedaj je očitno:

$$E(X_i) = p_0 := P(i\text{-ti igralec stavi svojo ženo}) = \frac{4}{52} \cdot \frac{48}{51} \cdot \frac{47}{50} \doteq 0\dot{}06805\,.$$

Ker je $S = X_1 + \cdots + X_8$, je tudi $E(S) = 8p_0 \doteq 0\dot{}5444$. 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:

$$\begin{aligned} p_1 &= 0 \\ p_2 &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{48}{50} \cdot \frac{47}{49} \cdot \frac{46}{48} \doteq 0\dot{}00399\,, \\ p_3 = p_4 &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{48}{50} \cdot \frac{47}{49} \cdot \frac{46}{48} \cdot \frac{45}{47} \doteq 0\dot{}00382\,. \end{aligned}$$

Torej je:

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

19. Velja $E(X) = 0\dot{}4$, $E(Y) = 2$ in $E(XY) = 0\dot{}8$, torej sta X in Y res nekorelirani. Sta pa odvisni, kar lahko razberemo iz poljubne navzkrižne verjetnosti. To sledi tudi iz dejstva, da je pri $X = 1$ skoraj gotovo $Y = 2$, medtem ko pri $X = 0$ skoraj gotovo velja $Y \neq 2$.

20. a) $0 \le t \le 3/2$.
b) $t = 0$, $t = 1$.
c) $t = 1$.

21. $K(X,Y) = -93$, $\quad r(X,Y) \doteq -0{\cdot}3952$.

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

23. a) Če indeksa i in j oba označujeta žensko, imamo tri možnosti.

- Za $i = j$ je $P(A_i \cap A_j) = P(A_i) = \dfrac{5^2}{6^3} = \dfrac{25}{216}$.
- Če je med ženskama natanko en moški, je $P(A_i \cap A_j) = \dfrac{5^3}{6^5} = \dfrac{125}{7776}$.
- Če ženski sedita nasproti, je $P(A_i \cap A_j) = \left(\dfrac{5^2}{6^3}\right)^2 = \dfrac{625}{46656}$.

Podobno velja za dva moška: za $i = j$ je $P(B_i) = \dfrac{25}{216}$, če je med moškima natanko ena ženska, je $P(B_i \cap B_j) = \dfrac{125}{7776}$, za moška, ki sedita nasproti, pa je $P(B_i \cap B_j) = \dfrac{625}{46656}$.

Če pa i označuje žensko in j moškega, imamo dve možnosti:

- Če sta si sosedna, je $P(A_i \cap B_j) = \dfrac{5^2}{6^4} = \dfrac{25}{1296}$.
- Če med njima sedi en moški in ena ženska, je $P(A_i \cap B_j) = \left(\dfrac{5^2}{6^3}\right)^2 = \dfrac{625}{46656}$.

b) Naj bo X_i indikator dogodka A_i, Y_i pa indikator dogodka B_i. Tedaj velja:

$$X = \sum_{i \in Z} X_i \,, \qquad Y = \sum_{i \in M} Y_i \,,$$

kjer je Z množica žensk, M pa množica moških. Velja:

$$D(X) = \sum_{i \in Z} \sum_{i \in Z} K(X_i, X_j) \,.$$

Obstajajo 4 urejeni pari (i,j) dveh žensk, za katere je $i = j$. Za te pare je:

$$K(X_i, X_j) = P(A_i) - \big(P(A_i)\big)^2 = \frac{25}{216} - \left(\frac{25}{216}\right)^2 = \frac{4775}{46656} \,.$$

Nadalje obstaja 8 urejenih parov (i,j) dveh žensk, med katerima je natanko en moški. Za te pare je:

$$K(X_i, X_j) = P(A_i \cap A_j) - P(A_i)\; P(A_j) = \frac{125}{7776} - \left(\frac{25}{216}\right)^2 = \frac{125}{46656} \,.$$

Za preostale pare dveh žensk pa sta dogodka A_i in A_j neodvisna, z njima pa tudi slučajni spremenljivki X_i in X_j, zato je $K(X_i, X_j) = 0$. Sledi:

$$D(X) = 4 \cdot \frac{4775}{46656} + 8 \cdot \frac{125}{46656} = \frac{5025}{11664} \,.$$

Podobno je tudi $D(X) = \frac{5025}{11664}$. Naslednji korak je izračun kovariance:

$$K(X,Y) = \sum_{i \in Z} \sum_{j \in M} K(X_i, Y_j)\,.$$

Obstaja 8 urejenih parov (i, j), kjer je i označuje žensko, j pa moškega, ki sedi poleg nje. Za te pare je:

$$K(X_i, Y_j) = P(A_i \cap B_j) - P(A_i)\ P(B_j) = \frac{25}{1296} - \left(\frac{25}{216}\right)^2 = \frac{275}{46656}\,.$$

Za preostale pare (i, j) pa sta slučajni spremenljivki X_i in X_j neodvisni. Torej je:

$$K(X,Y) = 8 \cdot \frac{275}{46656} = \frac{275}{5832}$$

Končno je:

$$r(X,Y) = \frac{\frac{275}{5832}}{\frac{5025}{11664}} = \frac{22}{201} \doteq 0{\cdot}104\,.$$

Opomba. Izračun verjetnosti $P(X = r)$, $P(Y = s)$ in $P(X = r, Y = s)$, s pomočjo katerih bi lahko bolj neposredno izračunali korelacijski koeficient, še malo ni preprost. Dogodek $\{X = 1\}$ se npr. lahko zgodi, če je prva ženska vrgla šestico, te pa ni vrgel nihče drug (ne ženska ne moški), lahko pa se zgodi tudi, če vse ženske vržejo šestico in jo vržeta tudi dva moška, med katerima je ena ženska, preostala dva moška pa je ne vržeta. Vmes je še veliko drugih možnosti.

24. Če je $S \sim \text{Bin}(n, p)$, je $E(S) = np$ in $D(S) = np(1 - p)$.

25. Če je $S \sim \text{Poi}(\lambda)$, je $E(S) = \lambda$ in $D(S) = \lambda$.

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

27. $D(3X - Y) = 64, \quad E\big((X - 2Y)^2\big) = 95.$

28. Če je $X \sim \text{Geo}(p)$, je $E(X) = 1/p$ in $D(X) = q/p^2$.
Če je $S \sim \text{NegBin}(n, p)$, je $E(X) = n/p$ in $D(X) = nq/p^2$.

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

30. Naj bo slučajni vektor $\boldsymbol{X}$ porazdeljen večrazsežno normalno s pričakovano vrednostjo $\boldsymbol{\mu}$ in kovariančno matriko $\boldsymbol{\Sigma}$ (obojim iz definicije večrazsežne normalne porazdelitve). Dokazati moramo, da je $\boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{\mu}$ in $\mathbf{K}(\boldsymbol{X}) = \boldsymbol{\Sigma}$. Najprej to dokažimo za primer, ko je $\boldsymbol{X}$ porazdeljen *standardno* normalno. Če je torej $\boldsymbol{X} = (X_1, \ldots X_n)$ porazdeljen standardno n-razsežno normalno, so njegove komponente neodvisne in porazdeljene standardno enorazsežno normalno, torej je $E(X_i) = 0$, $D(X_i) = 1$ in $K(X_i, X_j) = 0$ za poljubna $i \neq j$. To pomeni, da je $\boldsymbol{E}(\boldsymbol{X}) = \mathbf{0}$ in $\mathbf{K}(\boldsymbol{X}) = \mathbf{I}_n$, tako kot mora biti.

V splošnem pa je X enako porazdeljen kot $\mathbf{A}\boldsymbol{Z}_n + \boldsymbol{\mu}$, kjer je $\mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}$ (npr. $\mathbf{A} = \boldsymbol{\Sigma}^{1/2}$). Sledi:

$$\boldsymbol{E}(\boldsymbol{X}) = \mathbf{A}\,\boldsymbol{E}(\boldsymbol{Z}_n) + \boldsymbol{\mu} = \boldsymbol{\mu}\,, \qquad \mathbf{K}(\boldsymbol{X}) = \mathbf{A}\,\mathbf{K}(\boldsymbol{X})\mathbf{A}^{\mathrm{T}} = \mathbf{A}\mathbf{A}^{\mathrm{T}} = \boldsymbol{\Sigma}\,,$$

kot mora biti.

31. V matrični obliki lahko pišemo $\langle \boldsymbol{X}, \boldsymbol{u} \rangle = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{X}$, torej je:

$$D\big(\langle \boldsymbol{X}, \boldsymbol{u} \rangle\big) = \boldsymbol{u}^{\mathrm{T}}\,\mathbf{K}(\boldsymbol{X})\boldsymbol{u} = \langle \mathbf{K}\boldsymbol{u}, \boldsymbol{u} \rangle\,.$$

Ta izraz je minimalen oz. maksimalen, če je $\boldsymbol{u}$ lastni vektor matrike $\mathbf{K}$ za minimalno oz. maksimalno lastno vrednost. Iz karakterističnega polinoma:

$$\det(\mathbf{K} - \lambda\mathbf{I}) = -\lambda^3 + 12\lambda^2 - 21\lambda + 10 = -(\lambda - 1)^2(\lambda - 10)$$

dobimo dvakratno lastno vrednost $\lambda_{1,2} = 1$ in še lastno vrednost $\lambda_3 = 10$. Najmanjša možna disperzija je 1, dosežena pa je, če je $\boldsymbol{u}$ lastni vektor matrike $\mathbf{K}$ za to lastno vrednost. To pa so vsi (enotski) vektorji, pravokotni na vektor $(2, 1, 2)$. Največja možna disperzija pa je 10 in dosežena je za $\boldsymbol{u} = \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)$ in za $\boldsymbol{u} = \left(-\frac{2}{3}, -\frac{1}{3}, -\frac{2}{3}\right)$.

7. Pogojne porazdelitve in pogojevanje na slučajne spremenljivke

1. Naj bo B dogodek, da je po Pepčkovi zamenjavi na obeh prvih kovancih grb. Če se zgodi B, je očitno $X > 0$. Iz:

$$P\big(\{X = 1\} \cap B\big) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{12}, \quad P\big(\{X = 2\} \cap B\big) = \frac{1}{4}\left(\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2}\right) = \frac{1}{6}$$

najprej dobimo $P(B) = 1/4$, nato pa še pogojno porazdelitev, ki je določena s shemo $\begin{pmatrix} 1 & 2 \\ 1/3 & 2/3 \end{pmatrix}$.

2. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{27}{65} & \frac{18}{65} & \frac{12}{65} & \frac{8}{65} \end{pmatrix}$.

3. Pogojno na $Y = 2$ je $X \sim \begin{pmatrix} 1 & 2 \\ 1/4 & 3/4 \end{pmatrix}$.

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

Pogojno na $Y \ge X$ je $Y \sim \begin{pmatrix} 0 & 1 & 2 \\ 1/7 & 2/7 & 4/7 \end{pmatrix}$.

4. To je zvezna porazdelitev z gostoto:

$$p_{X \mid X \ge a}(x) = \begin{cases} \lambda\, e^{-\lambda(x-a)} & ;\ x > a \\ 0 & ;\ \text{sicer} \end{cases}.$$

To pomeni, da je slučajna spremenljivka $X - a$ pogojno glede na dogodek $\{X \ge a\}$ spet porazdeljena eksponentno $\mathrm{Exp}(\lambda)$. Pravimo, da je eksponentna porazdelitev *pozabljiva* (beseda dobi smisel, če si zalogo vrednosti slučajne spremenljivke predstavljamo kot čas).

5. To je porazdelitev z gostoto:

$$p_{X \mid X < Y}(x) = \begin{cases} 2(1 - x) & ;\ 0 < x < 1 \\ 0 & ;\ \text{sicer} \end{cases}.$$

6. Najprej bomo izračunali pogojno kumulativno porazdelitveno funkcijo. Označimo s T_1 dogodek, da je Tone vprašal, ali je $X < 2/3$, s T_2 dogodek, da je Tone vprašal, ali je $X > 1/3$, z D pa dogodek, da je bil odgovor na vprašanje pritrdilen. Pišimo:

$$F_{X \mid D}(x) = \frac{P(\{X \le x\} \cap D)}{P(D)},$$

$$P(\{X \le x\} \cap D) = P(T_1)\, P\big(\{X \le x\} \cap D \mid T_1\big) + \\ + P(T_2)\, P\big(\{X \le x\} \cap D \mid T_2\big).$$

Sedaj ločimo tri primere. Za $0 \le x \le 1/3$ velja:

$$\begin{aligned} P\big(\{X \le x\} \cap D \bigm| T_1\big) &= P(X \le x) = x\,, \\ P\big(\{X \le x\} \cap D \bigm| T_2\big) &= 0\,, \end{aligned}$$

za $1/3 \le x \le 2/3$ velja:

$$\begin{aligned} P\big(\{X \le x\} \cap D \bigm| T_1\big) &= P(X \le x) = x\,, \\ P\big(\{X \le x\} \cap D \bigm| T_2\big) &= P\big(\tfrac{1}{3} < X \le x\big) = x - \tfrac{1}{3}\,, \end{aligned}$$

za $2/3 < x \le 1$ pa velja:

$$\begin{aligned} P\big(\{X \le x\} \cap D \bigm| T_1\big) &= \tfrac{2}{3}\,, \\ P\big(\{X \le x\} \cap D \bigm| T_2\big) &= P\big(\tfrac{1}{3} < X \le x\big) = x - \tfrac{1}{3}\,. \end{aligned}$$

Če v vse skupaj vstavimo $x = 1$ in seštejemo, najprej dobimo $P(D) = 2/3$. Od tod dobimo pogojno kumulativno porazdelitveno funkcijo:

$$F_{X|D}(x) = \begin{cases} 0 & ;\ x \le 0 \\ \frac{3}{4}\,x & ;\ 0 \le x \le \frac{1}{3} \\ \frac{3}{2}\,x - \frac{1}{4} & ;\ \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{3}{4}\,x + \frac{1}{4} & ;\ \frac{2}{3} \le x \le 1 \\ 1 & ;\ x \ge 1 \end{cases}$$

in nazadnje še pogojno gostoto:

$$p_{X|D}(x) = \begin{cases} 0 & ;\ x < 0 \\ \frac{3}{4} & ;\ 0 < x < \frac{1}{3} \\ \frac{3}{2} & ;\ \frac{1}{3} < x < \frac{2}{3} \\ \frac{3}{4} & ;\ \frac{2}{3} < x < 1 \\ 0 & ;\ x > 1 \end{cases}\,.$$

7. Za X in B iz 1. naloge velja $E(X \mid D) = 5/3$ in $D(X \mid D) = 2/9$. Za X in D iz 6. naloge pa velja $E(X \mid D) = 1/2$ in $D(X \mid D) = 7/108$.

8. Če je Z indikator dogodka, da je $2X > Y$, velja:

$$E(XY \mid 2X > Y) = \frac{E(XYZ)}{P(2X > Y)}\,.$$

Po izreku o polni verjetnosti je:

$$\begin{aligned} P(2X > Y) &= P(Y = 0)\,P(2X > Y \mid Y = 0) + P(Y = 0)\,P(2X > Y \mid Y = 0) + \\ &\qquad + P(Y = 2)\,P(2X > Y \mid Y = 2) = \\ &= P(Y = 0)\,P(2X > 0) + P(Y = 1)\,P(2X > 1) + \\ &\qquad + P(Y = 2)\,P(2X > 2) = \\ &= \frac{1}{2}\,. \end{aligned}$$

Nadalje je $XYZ = f(X)\,g(Y)$, kjer je:

$$f(x) = \begin{cases} x & ; x > 1/2 \\ 0 & ; \text{sicer} \end{cases}, \qquad g(y) = \begin{cases} 1 & ; y = 1 \\ 0 & ; \text{sicer} \end{cases},$$

Sledi:

$$E(XY \mid 2X > Y) = 2\,E\big[f(X)\big]\,E\big[g(Y)\big] = \frac{2}{3}\int_{1/2}^{1} x\,\mathrm{d}x = \frac{1}{4}\,.$$

9. Naj bo B dogodek, da je bila premeščena kroglica bela, R pa dogodek, da je bila rdeča. Nadalje naj bo X število belih izvlečenih kroglic. Tedaj velja:

$$E(X \mid B) = \frac{12}{10}\,, \quad E(X \mid R) = \frac{9}{10}\,.$$

Sledi:

$$E(X) = P(B)\,E(X \mid B) + P(R)\,E(X \mid R) = \frac{6}{10}\cdot\frac{12}{10} + \frac{4}{10}\cdot\frac{9}{10} = 1{\cdot}08\,.$$

10. Ta naloga se je sicer že pojavila kot 17. naloga v 6. razdelku, vendar pa jo lahko zdaj rešimo hitreje, bolj neposredno. Označimo spet z N število vseh metov in definirajmo naslednje tri hipoteze:

$$\begin{aligned} H_1 &= \{\text{v prvem metu pade grb}\}\,. \\ H_2 &= \{\text{v prvem metu pade cifra, v drugem pa grb}\}\,. \\ H_3 &= \{\text{v prvih dveh metih pade cifra}\}\,. \end{aligned}$$

Če se zgodi H_3, je očitno $N = 2$ in torej tudi $E(N \mid H_3) = 2$. Če pa se zgodita H_1 ali H_2, je nadaljnje dogajanje spet zaporedje neodvisnih metov kovanca, zato je $E(N \mid H_1) = 1 + E(N)$ in $E(N \mid H_2) = 2 + E(N)$. Sledi:

$$E(N) = \frac{1}{2}\big(1 + E(N)\big) + \frac{1}{4}\big(2 + E(N)\big) + \frac{1}{4}\cdot 2\,,$$

od koder sledi $E(N) = 6$.

11. Če z Y označimo število vseh metov, z A pa dogodek, da ne pade nobena enojka, je $P(A \mid Y) = \left(\frac{4}{5}\right)^{Y-1}$.

12. a) $$P(A \mid Y) = \frac{\frac{1}{4}\cdot\frac{\binom{3}{Y}\binom{12}{4-Y}}{\binom{15}{4}}}{\frac{\binom{4}{Y}\binom{12}{4-Y}}{\binom{16}{4}}} = \frac{\frac{1}{4}\cdot\frac{\binom{4}{Y}\binom{11}{3-Y}}{\binom{15}{3}}}{\frac{\binom{4}{Y}\binom{12}{4-Y}}{\binom{16}{4}}} = \frac{4-Y}{12}.$$

b) Ker je $Y \sim \mathrm{Hip}(4, 4, 16)$, je $E\big(P(A \mid Y)\big) = \big(4 - E(Y)\big)/12 = 1/4$. Opazimo, da je to enako ravno $P(A)$.

c) Če Y zavzame vrednosti $b_1, b_2, b_3, \ldots$, iz formule za matematično upanje funkcije slučajne spremenljivke in izreka o polni verjetnosti dobimo:

$$\begin{aligned} E\big(P(A \mid Y)\big) &= P(Y = b_1)\,P(A \mid Y = b_1) + P(Y = b_2)\,P(A \mid Y = b_2) + \cdots = \\ &= P(A)\,. \end{aligned}$$

13. Tako kot v 11. nalogi označimo z Y število vseh metov, z A pa dogodek, da ne pade nobena enojka. Izračunali smo že, da je $P(A \mid Y) = \left(\frac{4}{5}\right)^{Y-1}$. Torej je:

$$P(A) = E\left[\left(\frac{4}{5}\right)^{Y-1}\right] = \sum_{k=1}^{\infty}\left(\frac{4}{5}\right)^{Y-1} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{k-1} = \frac{1}{2}\,.$$

14. a) Porazdelitev je hipergeometrijska $\operatorname{Hip}(2, Y-1, 9)$ oziroma:

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{72}(10-Y)(9-Y) & \frac{1}{36}(Y-1)(10-Y) & \frac{1}{72}(Y-1)(Y-2) \end{pmatrix}.$$

b) $E(X \mid Y) = \frac{2}{9}(Y-1)$, $\quad E\big(E(X \mid Y)\big) = \frac{2}{9}\big(E(Y)-1\big) = 1$. Iz simetrije lahko sklepamo, da je tudi $E(X) = 1$: če z Z označimo število zelenih kroglic *za* rdečo, je zaradi simetrije $E(Z) = E(X)$, velja pa tudi $X + Z = 2$.
c) Pišimo $E(X \mid Y) = g(Y)$, kjer je $g(y) = E(X \mid Y = y)$. Če Y zavzame vrednosti $b_1, b_2, b_3, \ldots$, iz izreka o polnem matematičnem upanju dobimo:

$$\begin{aligned} E\big(E(X \mid Y)\big) = E\big[g(Y)\big] = \sum_k P(Y = b_k)\, g(b_k) = \sum_k P(Y = b_k)\, E(X \mid Y = b_k) = \\ = E(X)\,. \end{aligned}$$

15. Računamo:

$$\begin{aligned} P(X = k \mid Z = n) &= \frac{P(X = k, Z = n)}{P(Z = n)} = \\ &= \frac{P(X = k, Y = n-k)}{P(Z = n)} = \\ &= \frac{P(X = k)\, P(Y = n-k)}{P(Z = n)} = \\ &= \frac{n!}{k!(n-k)!}\,\frac{\lambda^k \mu^{n-k}}{(\lambda+\mu)^n} = \\ &= \binom{n}{k}\left(\frac{\lambda}{\lambda+\mu}\right)^k\left(\frac{\mu}{\lambda+\mu}\right)^{n-k}, \end{aligned}$$

torej je iskana pogojna porazdelitev binomska $\operatorname{Bin}\left(Z, \dfrac{\lambda}{\lambda+\mu}\right)$.

16. Velja:

$$\begin{aligned} E(X) &= E\big[E(X \mid Y)\big] = E(Y) = 2 \cdot \frac{1}{2} = 1\,, \\ E(X^2 \mid Y) &= D(X \mid Y) + \big(E(X \mid Y)\big)^2 = Y^2 + Y + 1\,, \\ E(X^2) &= E\big[E(X^2 \mid Y)\big] = E(Y^2) + E(Y) + 1 = D(Y) + \big(E(Y)\big)^2 + E(Y) + 1 = \\ &= 2 \cdot \frac{1}{2} \cdot \frac{1}{2} + 3 = \frac{7}{2}\,, \\ D(X) &= E(X^2) - \big(E(X)\big)^2 = \frac{5}{2}\,. \end{aligned}$$

17. Tako kot v 11. nalogi označimo z Y število vseh metov, z X pa označimo število vseh enojk. Pogojno na Y je $X \sim \mathrm{Bin}(Y-1, 1/5)$. Ob dogovoru, da je $\binom{m}{k} = 0$, brž ko je $m \notin \{0, 1, \ldots m\}$, torej velja:

$$P(X = n \mid Y) = \binom{Y-1}{n} \left(\frac{1}{5}\right)^n \left(\frac{4}{5}\right)^{Y-1-n}.$$

Za $n = 0, 1, 2, \ldots$ torej velja:

$$\begin{aligned} P(X = n) &= E\left[\binom{Y-1}{n} \left(\frac{1}{5}\right)^n \left(\frac{4}{5}\right)^{Y-1-n}\right] = \\ &= \frac{1}{6} \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{k-1} \binom{k-1}{n} \left(\frac{1}{5}\right)^n \left(\frac{4}{5}\right)^{k-1-n} = \\ &= \frac{1}{6} \left(\frac{1}{4}\right)^n \sum_{k=n+1}^{\infty} \binom{k-1}{n} \left(\frac{2}{3}\right)^{k-1}. \end{aligned}$$

Zdaj pa se spomnimo gostote Pascalove porazdelitve: če je $Z \sim \mathrm{NegBin}(n+1, p)$, velja:

$$P(Z = k) = \binom{k-1}{n} p^{n+1} (1-p)^{k-n-1}\,; \quad k = n+1, n+2, \ldots$$

Sledi:

$$\sum_{k=n+1}^{\infty} \binom{k-1}{n} (1-p)^{k-1} = \frac{(1-p)^n}{p^{n+1}}.$$

Torej je:

$$\sum_{k=n+1}^{\infty} \binom{k-1}{n} \left(\frac{2}{3}\right)^{k-1} = 3 \cdot 2^n$$

in končno:

$$P(X = n) = \frac{1}{2^{n+1}}.$$

Dobili smo geometrijsko porazdelitev $\mathrm{Geo}(1/2)$, pomaknjeno za ena v levo.

18. Pogojno na N je $S \sim \mathrm{Bin}(N, p)$, t. j.:

$$P(S = k \mid N) = \binom{N}{k} p^k (1-p)^{N-k}.$$

Sledi:

$$\begin{aligned}
P(S = k) &= E\left[\binom{N}{k} p^k (1-p)^{N-k}\right] = \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n\, e^{-\lambda}}{n!} \binom{n}{k} p^k (1-p)^{n-k} = \\
&= \frac{\lambda^n p^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} = \\
&= \frac{(p\lambda)^k\, e^{-p\lambda}}{k!}\,.
\end{aligned}$$

Torej je $S \sim \mathrm{Poi}(p\lambda)$. Podobno je tudi $T \sim \mathrm{Poi}((1-p)\lambda)$. Nadalje velja še:

$$\begin{aligned}
P(S = k, T = l) = P(S = k, N = k + l) &= P(N = k + l)\, P(S = k \mid N = k + l) = \\
&= \frac{\lambda^{k+l}\, e^{-\lambda}}{(k+l)!} \binom{k+l}{k} p^k (1-p)^l = \frac{\lambda^{k+l}\, p^k (1-p)^l\, e^{-\lambda}}{k!\, l!} = \\
&= P(S = k)\, P(T = l)\,,
\end{aligned}$$

torej sta S in T res neodvisni.

Neodvisnost slučajnih spremenljivk S in T pa lahko izpeljemo tudi s sklicevanjem na 15. nalogo: iz le-te namreč sledi, da obstajajo slučajne spremenljivke S, T in N, za katere velja, da je $S \sim \mathrm{Poi}(p\lambda)$ in $N \sim \mathrm{Poi}(\lambda)$, da sta S in T neodvisni, da je $N = S + T$ in da pogojno na N velja $S \sim \mathrm{Bin}(N, p)$. Ker je porazdelitev slučajnega vektorja (S, T, N) natančno določena s porazdelitvijo slučajne spremenljivke N, pogojno porazdelitvijo slučajne spremenljivke S glede a N in zvezo $N = S + T$, od tod sledi, da S in T morata biti neodvisni, brž ko izpolnjujeta pogoje naloge.

Če je $N = n$ kar konstanta, sta S in T odvisni, brž ko je $n \geq 1$ in $0 < p < 1$: v tem primeru namreč S in T zavezameta vsaj dve vrednosti, a če $S = k$, je nujno $T = n - k$.

19. Pogojno glede na N ima slučajna spremenljivka S porazdelitev $\mathrm{Gama}(N, \lambda)$, torej za $s > 0$ velja:

$$p_{S|N}(s) = \frac{\lambda^N s^{N-1}}{(N-1)!}\, e^{-\lambda s}\,.$$

Sledi:

$$p_S(s) = E\left[\frac{\lambda^N s^{N-1}}{(N-1)!}\, e^{-\lambda s}\right] = p\, e^{-\lambda s} \sum_{n=1}^{\infty} \frac{\lambda^n s^{n-1} (1-p)^{n-1}}{(n-1)!} = \lambda p\, e^{-p\lambda s}\,.$$

Torej je $S \sim \mathrm{Exp}(p\lambda)$.

20. $E(X \mid Y = 2) = 7/4, \quad E(X^2 \mid Y = 2) = 13/4,$
$E(X^2 Y^2 \mid Y = 2) = 4\, E(X^2 \mid Y = 2) = 13.$

21. Tako kot v 17. nalogi označimo z Y število vseh metov, z X pa označimo število vseh enojk. Pogojno na Y je $X \sim \mathrm{Bin}(X-1, 1/5)$, torej je tudi $E(X \mid Y) = \frac{Y-1}{5}$. Pričakovani delež enojk med vsemi meti je:

$$E\left(\frac{X}{Y}\right) = E\left(\frac{Y-1}{5Y}\right) = \frac{1}{5}\left[1 - E\left(\frac{1}{Y}\right)\right] = \frac{1}{5} - \frac{1}{30}\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{5}{6}\right)^{k-1} = \frac{1}{5} - \frac{\ln 6}{25} \doteq$$
$$\doteq 0{\cdot}128\,.$$

22. Ker slučajna spremenljivka N zavzame le vrednosti v naravnih številih, je dovolj za vsako naravno število n poiskati pogojno verjetnost $h_n(S) = P(N = n \mid S)$. To bo natanko tedaj, ko bo za vsako primerno merljivo funkcijo g veljala zveza:

$$E\big[h_n(S)\, g(S)\big] = E\big[Z_n\, g(S)\big]\,,$$

kjer je Z_n indikator dogodka, da je $N = n$. Z drugimi besedami, veljati mora:

$$E\big[h_n(S)\, g(S)\big] = P(N = n)\, E\big[g(S) \mid N = n\big]\,.$$

Ko vstavimo brezpogojno in pogojno gostoto, dobimo:

$$p\lambda \int_0^\infty h_n(s)\, g(s)\, e^{-p\lambda s}\, \mathrm{d}s = p(1-p)^{n-1}\,\frac{\lambda^n}{(n-1)!}\int_0^\infty g(s)\, s^{n-1}\, e^{-\lambda s}\, \mathrm{d}s\,.$$

Ta pogoj bo izpolnjen za regresijsko funkcijo:

$$h_n(s) = \frac{\lambda^{n-1}(1-p)^{n-1} s^{n-1}}{(n-1)!}\, e^{-\lambda(1-p)s}\,.$$

Z drugimi besedami, velja:

$$P(N = n \mid S) = \frac{\lambda^{n-1}(1-p)^{n-1} S^{n-1}}{(n-1)!}\, e^{-\lambda(1-p)S}\,.$$

Pogojno na S ima torej slučajna spremenljivka $N-1$ Poissonovo porazdelitev $\mathrm{Poi}\big(\lambda(1-p)S\big)$.

23. Dejstvo, da je X pogojno na U porazdeljena binomsko $\mathrm{Bin}(n, U)$, lahko zapišemo takole:

$$E\big[\mathbf{1}(X = k) \bigm| U\big] = P(X = k \mid U) = \binom{n}{k} U^k (1-U)^{n-k}\,; \qquad k = 0, 1, 2, \ldots n\,.$$

Sledi:

$$E\big[h(U)\,\mathbf{1}(X = k)\big] = \binom{n}{k}\int_0^1 h(u)\, u^k (1-u)^{n-k}\, \mathrm{d}u\,.$$

Za $h \equiv 1$ dobimo:

$$P(X = k) = \binom{n}{k}\,\mathrm{B}(k+1, n-k+1) = \frac{1}{n+1}\,.$$

Slučajna spremenljivka X je torej porazdeljena diskretno enakomerno na $\{0, 1, \dots n\}$. Nadalje velja:

$$\begin{aligned} E\big[h(U) \bigm| X = k\big] &= \frac{E\big[h(U)\,\mathbf{1}(X = k)\big]}{P(X = k)} = \\ &= \frac{1}{\mathrm{B}(k+1, n-k+1)} \int_0^1 h(u)\,u^k (1-u)^{n-k}\,\mathrm{d}u\,. \end{aligned}$$

Sledi, da ima slučajna spremenljivka U pogojno na X porazdelitev $\mathrm{Beta}(X+1, n-X+1)$.

24. Iz $E(X \mid Y) \sim \begin{pmatrix} 3/4 & 1 & 7/4 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$ dobimo $E\big(E(X \mid Y)\big) = 7/6$ in $D\big(E(X \mid Y)\big) = 13/72$.
Nadalje iz $D(X \mid Y) \sim \begin{pmatrix} 3/16 & 1 \\ 2/3 & 1/3 \end{pmatrix}$ dobimo $E\big(D(X \mid Y)\big) = 11/24$.
Iz $X \sim \begin{pmatrix} 0 & 1 & 2 \\ 1/4 & 1/3 & 5/12 \end{pmatrix}$ pa dobimo $E(X) = 7/6$ in $D(X) = 23/36$.
Seveda je $E\big(E(X \mid Y)\big) = E(X)$, opazimo pa še, da je $D(X) = D\big(E(X \mid Y)\big) + E\big(D(X \mid Y)\big)$.

25. Velja:

$$\begin{aligned} D\big(E(X \mid Y)\big) &= E\Big[\big(E(X \mid Y)\big)^2\Big] - \big(E(X)\big)^2 \\ D(X \mid Y) &= E(X^2 \mid Y) - \big(E(X \mid Y)\big)^2\,, \\ E\big(D(X \mid Y)\big) &= E(X^2) - E\Big[\big(E(X \mid Y)\big)^2\Big]\,. \end{aligned}$$

Ko seštejemo, dobimo:

$$D\big(E(X \mid Y)\big) + E\big(D(X \mid Y)\big) = E(X^2) - \big(E(X)\big)^2 = D(X)\,.$$

26. Iz Poissonove porazdelitve (glej 25. nalogo iz 6. razdelka) dobimo, da je $E(X^2 \mid Y) = Y^2 + Y$. Iz eksponentne porazdelitve (glej 4. nalogo iz 6. razdelka) zdaj dobimo:

$$E(X^2) = E\big[E(X^2 \mid Y)\big] = E(Y^2 + Y) = \frac{2}{\lambda^2} + \frac{1}{\lambda}\,.$$

27. Najprej izračunamo:

$$E\big[a^X \mid Y\big] = \sum_{k=0}^{\infty} a^k\,\frac{Y^k e^{-Y}}{k!} = e^{(a-1)Y}\,.$$

To obstaja za vse $a \in \mathbb{R}$. Formalno gledano je potem:

$$E(a^X) = E\big[e^{(a-1)Y}\big] = \lambda \int_0^{\infty} e^{(a-1)y}\,e^{-\lambda y}\,\mathrm{d}y = \frac{\lambda}{\lambda - a + 1}\,,$$

in sicer za $a < \lambda + 1$. Toda v resnici $E(a^X)$ za $a \le -\lambda - 1$ ne obstaja, ker ne obstaja $E(|a^X|)$. Iskano matematično upanje torej obstaja za $|a| < \lambda + 1$.

28. Za $t > 0$ ima slučajna spremenljivka T pogojno gostoto:

$$p_{T|N=n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}\, e^{-\lambda t} .$$

Brezpogojno gostoto dobimo s seštetjem:

$$p_T(t) = \sum_{n=1}^{\infty} P(N=n)\, p_{T|N=n}(t) = \sum_{n=1}^{\infty} p(1-p)^{n-1} \frac{\lambda^n t^{n-1}}{(n-1)!}\, e^{-\lambda t} = p\lambda\, e^{-p\lambda t} .$$

Torej je $T \sim \mathrm{Exp}(p\lambda)$.

29. 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 X ima torej Y enakomerno porazdelitev $\mathrm{E_z}(0, X)$. Torej ima Y/X pogojno na X enakomerno porazdelitev $\mathrm{E_z}(0, 1)$, ne glede na X. To pa pomeni, da sta Y/X in X neodvisni.

Pogojno glede na Y pa ima X eksponentno porazdelitev $\mathrm{Exp}(1)$, pomaknjeno za Y v desno. Torej ima $X - Y$ pogojno glede na Y eksponentno porazdelitev $\mathrm{Exp}(1)$, ne glede na Y. To pa pomeni, da sta $X - Y$ in Y neodvisni.

30. *Prvi način.* Iz pogojne gostote:

$$p_{Y|X}(y \mid x) = \frac{|x|}{\sqrt{2\pi}}\, e^{-x^2 y^2}$$

izračunamo dvorazsežno gostoto:

$$p_{X,Y}(x, y) = p_X(x)\, p_{Y|X}(y \mid x) = \frac{|x|}{2\pi}\, e^{-x^2(1+y^2)/2} ,$$

z integriranjem katere določimo porazdelitev gostoto spremenljivke Y:

$$p_Y(y) = \int_{-\infty}^{\infty} p_{X,Y}(x, y)\, \mathrm{d}x = \frac{1}{\pi(1+y^2)} .$$

Slučajna spremenljivka Y ima torej Cauchyjevo porazdelitev.

Drugi način (v resnici le drugače pisan prvi način). Uporabimo formulo:

$$p_Y(y) = E\big[p_{Y|X}(y)\big] = E\, \frac{|X|}{\sqrt{2\pi}}\, e^{-X^2 y^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |x|\, e^{-x^2(1+y^2)/2}\, \mathrm{d}x ,$$

ki nam seveda da isto kot prej.

Tretji način. Če postavimo $U := XY$, iz formule za transformacijo enorazsežne normalne porazdelitve, uporabljene pogojno na X, dobimo, da je pogojna porazdelitev slučajne spremenljivke U glede na $X = x$ standardna normalna, ne glede na x. To pa pomeni, da je U neodvisna od X. Po formuli za gostoto funkcije dveh slučajnih spremenljivk dobimo:

$$p_Y(y) = \int_{-\infty}^{\infty} p_X(x)\, p_U(xy)\, |x| \, \mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| \, e^{-x^2(1+y^2)/2} \, \mathrm{d}x \,,$$

kar je spet isto kot prej.

31. Za $0 \le x, y \le 1$ velja $p_{X|Y}(x \mid y) = \dfrac{6(x^2 + xy)}{2 + 3y}$, od koder dobimo

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

32. a) Z deljenjem gostot dobimo:

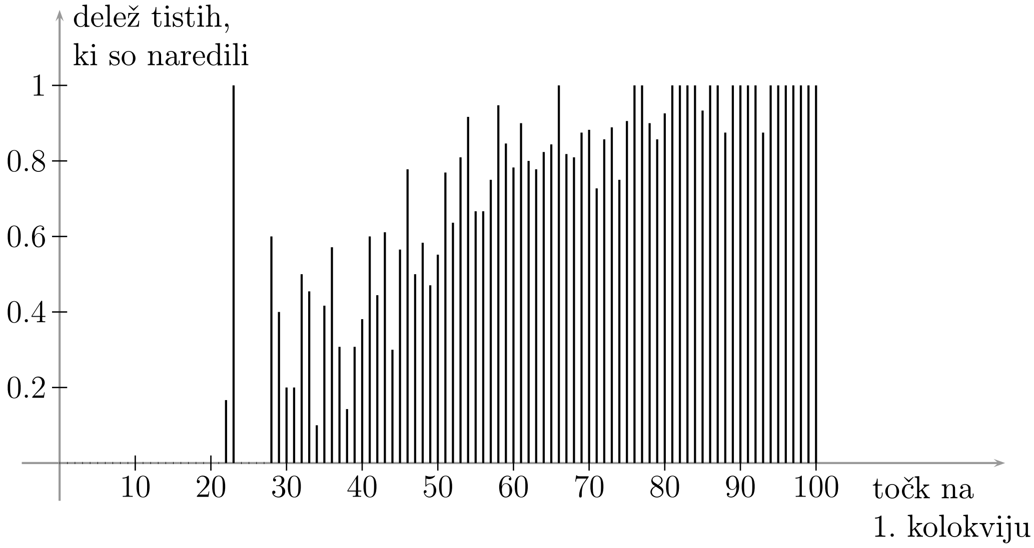
$$\begin{aligned}
p_{X_2|X_1}(x_2 \mid x_1) &= \frac{1}{2\pi\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{(x_1-\mu_1)^2}{2(1-\rho^2)\sigma_1^2} + \frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{(1-\rho^2)\sigma_1\sigma_2} - \frac{(x_2-\mu_2)^2}{2(1-\rho^2)\sigma_2^2} + \frac{(x_1-\mu_1)^2}{2\sigma_1^2}} = \\
&= \frac{1}{2\pi\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{\sigma_2^2(x_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2 - \sigma_2^2(1-\rho^2)(x_1-\mu_1)^2}{2(1-\rho^2)\sigma_1^2\sigma_2^2}} = \\
&= \frac{1}{2\pi\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{\rho^2\sigma_2^2(x_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2}{2(1-\rho^2)\sigma_1^2\sigma_2^2}} = \\
&= \frac{1}{2\pi\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{(\sigma_1(x_2-\mu_2) - \rho\sigma_2(x_1-\mu_1))^2}{2(1-\rho^2)\sigma_1^2\sigma_2^2}} = \\
&= \frac{1}{2\pi\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{((x_2-\mu_2) - \rho\sigma_2(x_1-\mu_1)/\sigma_1)^2}{2(1-\rho^2)\sigma_2^2}} .
\end{aligned}$$

Iskana pogojna porazdelitev je torej normalna $\mathrm{N}\left(\mu_2 + \dfrac{\rho\sigma_2}{\sigma_1}(X_1 - \mu_1),\, \sqrt{1-\rho^2}\,\sigma_2\right)$.

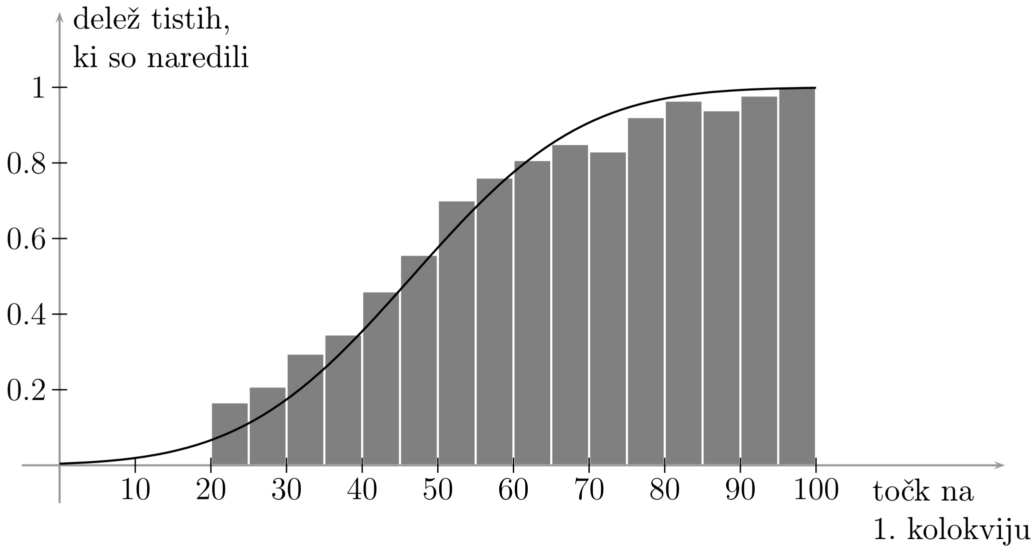
b) $\dfrac{1}{2} - \Phi\left(\dfrac{75 - 58{}^{\cdot}0 - 0{}^{\cdot}291 \cdot 25{}^{\cdot}3 \cdot (25 - 59{}^{\cdot}2)/20{}^{\cdot}1}{25{}^{\cdot}3 \cdot \sqrt{1 - 0{}^{\cdot}291^2}}\right) \doteq 0{}^{\cdot}111.$

V resnici je 7 kandidatov na prvem kolokviju zbralo 25 točk in nihče izmed njih ni naredil. Prav tako ni naredil nihče, ki je zbral 26 ali 27 točk, pač pa je naredil en kandidat od 6, ki so zbrali 22 točk, in 3 kandidati od 5, ki so zbrali 28 točk.

Diagram pogojnih deležev:

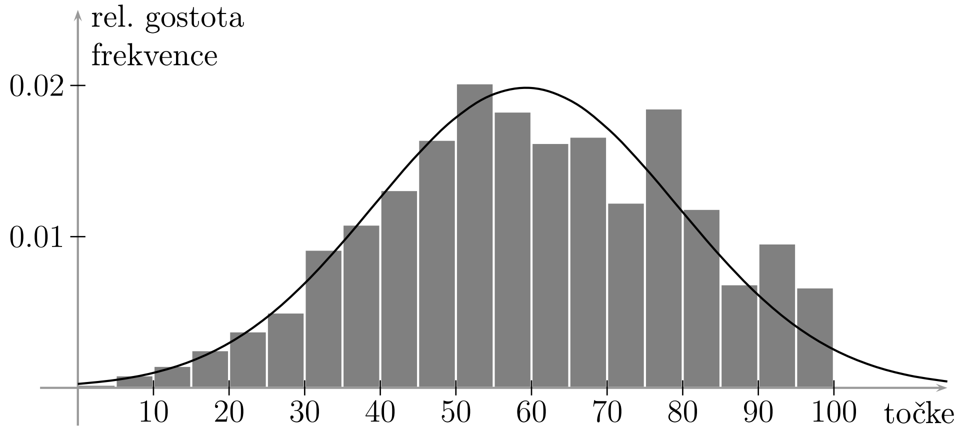


Zaradi velike zaloge vrednosti rezultatov dobimo veliko boljši pregled, če gledamo pogojne verjetnosti glede na 5 točk široke razpone. Krivulja ponazarja ocene za deleže (verjetnosti) na podlagi Gaussovega modela, t. j. modela z dvorazsežno normalno porazdelitvijo:

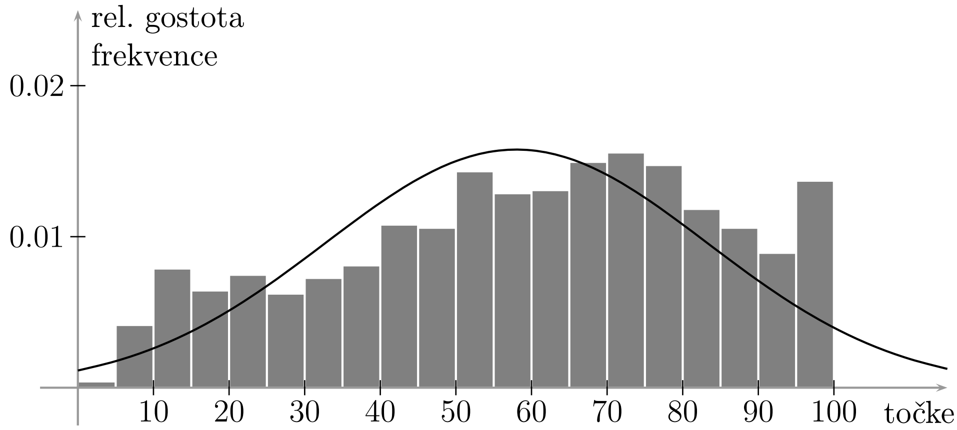


Histogram se kar dobro prilega krivulji glede na to, da se histogram frekvenčnih porazdelitev vsaj za drugi kolokvij ne prilega tako dobro normalni porazdelitvi. Prikazane so *relativne gostote frekvenc*, t. j. deleži ustrezajo ploščinam pravokotnikov v histogramu; višine pravokotnikov so tako primerljive z gostoto zvezne porazdelitve, s katero aproksimiramo.

Histogram za prvi kolokvij:



Histogram za drugi kolokvij:



8. Rodovne, momentno-rodovne in karakteristične funkcije

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

2. $G(s) = \dfrac{ps}{1-(1-p)s}$, $\quad E(X) = \dfrac{1}{p}$, $\quad D(X) = \dfrac{(1-p)}{p^2}$.

3. Iz $G_X(s) = 1$ dobimo $a = e^{-2}$. Iz razvoja:

$$G_X(s) = e^{-2}\left[1 + (s+s^2) + \frac{(s+s^2)^2}{2} + \cdots\right] = e^{-2}\left[1 + s + \frac{3s^2}{2} + \cdots\right]$$

dobimo $P(X=2) = \dfrac{3\,e^{-2}}{2} \doteq 0{\cdot}203$.

4. Če z Z označimo vsoto, je rodovna funkcija te slučajne spremenljivke enaka:

$$\begin{aligned} G_Z(z) &= \frac{s}{2-s}\cdot\frac{s}{3-2s} = \frac{2s^2}{3-2s} - \frac{s^2}{2-s} = \\ &= \frac{2s^2}{3}\left[1 + \frac{2s}{3} + \left(\frac{2s}{3}\right)^2 + \cdots\right] - \\ &\qquad - \frac{s^2}{2}\left[1 + \frac{s}{2} + \left(\frac{s}{2}\right)^2 + \cdots\right], \end{aligned}$$

torej je:

$$P(Z=k) = \left(\frac{2}{3}\right)^{k-1} - \frac{1}{2^{k-1}}\,; \quad k = 2, 3, 4, \ldots$$

Nalogo bi seveda lahko rešili tudi neposredno, brez uporabe rodovnih funkcij.

5. 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:

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

6. 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:

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

7. 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:

$$G(s) = G_1\bigl(G_2(s)\bigr) \,,$$

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

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

8. *Prvi način.* Če vsoto označimo z Z, velja:

$$G_Z(s) = \frac{a\,\frac{bs}{1-(1-b)s}}{1-(1-a)\,\frac{bs}{1-(1-b)s}} = \frac{abs}{1-(1-ab)s} \,,$$

torej je $Z \sim \mathrm{Geo}(ab)$.

Drugi način. Oglejmo si dve neodvisni zaporedji neodvisnih poskusov, recimo jima serija A in serija B. Vsak poskus iz serije A uspe z verjetnostjo a, vsak poskus iz serije B pa z verjetnostjo b. Seriji naj se izvajata vzporedno, torej se hkrati izvedeta n-ti poskus iz serije A in n-ti poskus iz serije B.

Naj bo X_i število poskusov od nevključno $(i-1)$-tega do vključno i-tega poskusa iz serije A (X_1 pa naj bo število poskusov iz serije A do vključno prvega uspešnega). Nadalje z N označimo število poskusov iz serije A do vključno prvega, pri katerem je uspel tudi istoležni poskus iz serije B. Tedaj so $X_1, X_2, \ldots$ neodvisne in porazdeljene geometrijsko $\mathrm{Geo}(a)$, N pa je od njih neodvisna in porazdeljena geometrijsko $\mathrm{Geo}(b)$. Vsota $X_1 + \cdots + X_N$ pa je zaporedna številka prvega poskusa, pri katerem je uspela tako serija A kot tudi serija B in je zato porazdeljena geometrijsko $\mathrm{Geo}(ab)$.

9. Naj bo G_1 rodovna funkcija slučajnih spremenljivk X_i. Ker so slučajne spremenljivke X_i porazdeljene geometrijsko, je:

$$G_1(s) = \frac{\frac{2}{3}s}{1-\frac{1}{3}s} = \frac{2s}{3-s} \,.$$

Rodovna funkcija slučajne spremenljivke N je funkcija:

$$s \mapsto \frac{\frac{3}{4}s}{1-\frac{1}{4}s} = \frac{3s}{4-s} \,.$$

Če z G_2 označimo rodovno funkcijo slučajne spremenljivke $2N$, je torej $G_2(s) = \dfrac{3s^2}{4-s^2}$. Rodovna funkcija slučajne vsote S je tako funkcija:

$$G_2(G_1(s)) = \frac{3\left(\frac{2s}{3-s}\right)^2}{4-\left(\frac{2s}{3-s}\right)^2} = \frac{s^2}{3-2s} = \frac{\frac{1}{3}s^2}{1-\frac{2}{3}s}\,,$$

torej ima S geometrijsko porazdelitev Geo(1/3), pomaknjeno za 1 v desno. Z drugimi besedami, $S-1 \sim \mathrm{Geo}(1/3)$.

10. 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 $p = 1/3$.

11. $M_X(t) = \begin{cases} \dfrac{e^{bt}-e^{at}}{t(b-a)} & ; t \neq 0 \\ 1 & ; t = 0 \end{cases}$.

Definirana je za vse t.

12. $M_X(t) = \exp\left(e^{\lambda e^t} - 1\right)$, definirana je za vse t.

13. $z_r = r!/\lambda^r$,

$M_X(t) = \dfrac{\lambda}{\lambda - t}$, definirana je za $t < \lambda$.

$m_3 = 2/\lambda^3$.

14. Za $Z \sim \mathrm{N}(0,1)$ velja:

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - t^2/2}\,\mathrm{d}t\,.$$

S substitucijo $u = t - x$ dobimo:

$$M_Z(t) = \frac{e^{u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\,\mathrm{d}u = e^{t^2/2} = 1 + \frac{t^2}{2} + \frac{t^4}{2^2\cdot 2!} + \frac{t^6}{2^3\cdot 3!} + \cdots\,.$$

Vsi momenti lihih redov so torej enaki 0. Če je n sod, pa je:

$$z_n = \frac{n!}{2^{n/2}\cdot (n/2)!} = 1\cdot 3\cdot\cdots\cdot(n-1) = (n-1)!!\,.$$

15. Za $X \sim \mathrm{N}(\mu,\sigma)$ lahko zapišemo $X = \sigma Z + \mu$, kjer je $Z \sim \mathrm{N}(0,1)$. Sledi:

$$M_X(t) = E\left[e^{t(\sigma X+\mu)}\right] = e^{\mu t}\, M_Z(\sigma t) = e^{\sigma^2 t^2/2 + \mu t}\,.$$

Od tod in iz izreka o enoličnosti sledi, da, če sta $X_1 \sim \mathrm{N}(\mu_1,\sigma_1)$ in $X_2 \sim \mathrm{N}(\mu_2,\sigma_2)$ neodvisni slučajni spremenljivki, velja $X_1 + X_2 \sim \mathrm{N}\left(\mu_1+\mu_2, \sqrt{\sigma_1^2+\sigma_2^2}\right)$.

16. Če je $X \sim \text{Gama}(a, \lambda)$, velja $M_X(t) = \left(\dfrac{\lambda}{\lambda - t}\right)^a$; momentno-rodovna funkcija je definirana za $t < \lambda$ (v kompleksnem pa za $\operatorname{Re} t < \lambda$).

Od tod in iz izreka o enoličnosti sledi, da, če sta $X_1 \sim \text{Gama}(a_1, \lambda)$ in $X_2 \sim \text{Gama}(a_2, \lambda)$ neodvisni slučajni spremenljivki, velja $X_1 + X_2 \sim \text{Gama}(a_1 + a_2, \lambda)$.

17. $\kappa_1 = \mu, \quad \kappa_2 = \sigma^2, \quad \kappa_r = 0$ za vse $r \ge 3$.

19. $\kappa_2 = m_2 = \sigma^2, \quad \kappa_3 = m_3, \quad \kappa_4 = m_4 - 3m_2^2, \quad A = m_3/\sigma^3, \quad K = m_4/\sigma^4 - 3$.

20. $\kappa_1 = \dfrac{a+b}{2}, \quad \kappa_2 = \dfrac{(b-a)^2}{12}, \quad \kappa_3 = 0, \quad \kappa_4 = -\dfrac{(b-a)^4}{120}, \quad A = 0, \quad K = -\dfrac{6}{5}$.

21. $m_2 = m_3 = \lambda, \quad m_4 = \lambda + 3\lambda^2$.

22. Spomnimo se, da je $E(S) = 1600 \cdot 0'1 = 160$. Neposredno po neenačbi Markova lahko ocenimo $P(S > 250) = P(S \ge 251) \le \dfrac{160}{251} \doteq 0'638$ (zaokroženo navzgor).

Za oceno po neenačbi Čebiševa potrebujemo $D(S) = 1600 \cdot 0'1 \cdot 0'9 = 144$. Ocenimo:

$$P(S > 250) = P(S - 160 \ge 91) \le P(|S - 160| \ge 91) \le \frac{144}{91^2} \doteq 0'0174$$

(spet zaokroženo navzgor).

Končno ocenimo še z uporabo momentno-rodovne funkcije. Za $S \sim \text{Bin}(n, p)$ iz pravila za vsote neodvisnih slučajnih spremenljivk dobimo
$M_S(t) = \left(1 - p + p\,e^t\right)^n$. Če je $x < n$, je izraz $e^{-tx}\,M(t)$ najmanjši pri:

$$t = \ln \frac{(1-p)x}{p(n-x)}\,.$$

Za $n = 1600$ in $p = 0'1$ dobimo $t \doteq 0'5156$ in $P(S > 250) = P(S \ge 251) \le 1'45 \cdot 10^{-11}$.

Če bi šli po Laplaceovi integralski formuli in upoštevali $\frac{1}{2} - \Phi(x) \approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$, bi dobili naslednjo aproksimacijo, ki ni ocena navzgor:

$$P(S > 250) \approx \frac{1}{2} - \Phi\left(\frac{90'5}{12}\right) \approx \frac{12}{90'5\sqrt{2\pi}}\, e^{-90.5^2/288} \doteq 2'36 \cdot 10^{-14}\,.$$

Točen rezultat: $9'705 \cdot 10^{-13}$.

23. Najprej izpeljimo oceni za splošni primer. Naj bo $x > 0$. Iz simetrije in neenačbe Čebiševa dobimo:

$$P(S_n > x) = \frac{1}{2}\, P(|S_n| > x) \le \frac{D(S_n)}{2x^2} = \frac{n}{x^2}\,,$$

Momentno-rodovna funkcija pa je enaka:

$$M_{S_n}(t) = \frac{1}{1 - t^2}$$

in je definirana za $-1 < t < 1$, izraz $e^{-tx} M_{S_n}(t)$ pa je minimalen pri

$$t = \frac{\sqrt{n^2 + x^2} - n}{x} .$$

Za $P(S_5 > 5)$ iz neenačbe Čebiševa dobimo zgornjo mejo 0'2, iz momentno-rodovne funkcije pa (za $t = \sqrt{2} - 1$) 0'323. Ocena iz neenačbe Čebiševa je torej tu ostrejša od ocene iz momentno-rodovnih funkcij. Točen rezultat pa je 0'0555.

Za $P(S_{20} > 20)$ iz neenačbe Čebiševa dobimo zgornjo mejo 0'05, iz momentno-rodovne funkcije pa (prav tako za $t = \sqrt{2} - 1$) 0'0109. Zdaj je torej ocena iz momentno-rodovnih funkcij ostrejša od ocene iz neenačbe Čebiševa. Točen rezultat pa je 0'00114.

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

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

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

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

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

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

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

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

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

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

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

9. Limitni izreki

1. Označimo z ℓ_n število vseh zaporednih cifer od n-tega meta nazaj (če torej v n-tem metu pade grb, je $\ell_n = 0$). Tedaj za vsak $\varepsilon > 0$ velja:

$$P(\ell_n \geq (1+\varepsilon)\log_2 n) \leq n^{-(1+\varepsilon)}\,.$$

Ker vrsta iz izrazov na dseni konvergira, se po prvi Borel–Cantellijevi lemi skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\ell_n \geq (1+\varepsilon)\log_2 n$. A če je to res velja tudi $\limsup_{n\to\infty} L_n/\log_2 n \leq 1+\varepsilon$. Ker to za vsak ε skoraj gotovo velja, skoraj gotovo velja tudi:

$$\limsup_{n\to\infty} \frac{L_n}{\log_2 n} \leq 1\,.$$

Za oceno limite v nasprotno smer razdelimo mete na bloke dolžine $\lfloor(1-\varepsilon)\log_2 n\rfloor + 1$. Na vsakem od teh blokov je verjetnost, da bodo padle same cifre, enaka $2^{-(\lfloor(1-\varepsilon)\log_2 n\rfloor+1)} \geq n^{-(1-\varepsilon)}/2$. Če je $L_n \leq (1-\varepsilon)\log_2 n$, na vseh blokih velja, da ne padejo same cifre, torej je:

$$\begin{aligned} P(L_n \leq (1-\varepsilon)\log_2 n) &\leq \left(1 - \frac{n^{-(1-\varepsilon)}}{2}\right)^{\lfloor n/(\lfloor(1-\varepsilon)\log_2 n\rfloor+1)\rfloor} \leq \\ &\leq \exp\left\{-\frac{n^{-(1-\varepsilon)}}{2}\left(\frac{n}{\lfloor(1-\varepsilon)\log_2 n\rfloor + 1} - 1\right)\right\}\,. \end{aligned}$$

Za dovolj velike n velja tudi:

$$P(L_n \leq (1-\varepsilon)\log_2 n) \leq \exp\left(-\frac{n^{\varepsilon}}{2\log_2 n}\right)\,.$$

Ker vrsta iz izrazov na desni spet konvergira, se skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\{L_n \leq (1-\varepsilon)\log_2 n\}$, a če je to res, je tudi $\liminf_{n\to\infty} L_n/\log_2 n \geq 1-\varepsilon$. Ker to za vsak $\varepsilon > 0$ skoraj gotovo velja, skoraj gotovo velja tudi:

$$\liminf_{n\to\infty} \frac{L_n}{\log_2 n} \geq 1\,.$$

Iz obeh ocen pa že sledi želeni rezultat.

2. Zaporedje konvergira skoraj gotovo in posledično tudi v verjetnosti in šibko.

Konvergenca v verjetnosti sledi iz neenačbe Čebiševa: iz $E(S_n) = 0$ in $D(S_n/n) = 1/n$ sledi:

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq \frac{1}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0\,.$$

Za dokaz skoraj gotove konvergence pa se moramo malo bolj potruditi. Spomnimo se na momentno-rodovne funkcije in uporabo neenačbo Markova za eksponentno funkcijo. Iz:

$$E\left[e^{tS_n}\right] = \left(\frac{e^t + e^{-t}}{2}\right)^n$$

izpeljemo:

$$P\left(\frac{S_n}{n} \geq \varepsilon\right) = P(S_n \geq n\varepsilon) \leq e^{-nt\varepsilon}\left(\frac{e^t + e^{-t}}{2}\right)^n = \left(e^{-t\varepsilon}\,\frac{e^t + e^{-t}}{2}\right)^n .$$

Za vsak $\varepsilon > 0$ obstaja tak t, da je $e^{-t\varepsilon}\,\frac{e^t+e^{-t}}{2} < 1$. V tem primeru vrsta iz izrazov na desni konvergira, torej se po prvi Borel–Cantellijevi lemi skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\{S_n/n \geq \varepsilon\}$. Zaradi simetrije to velja tudi za dogodke $\{S_n/n \leq -\varepsilon\}$, torej velja tudi za dogodke $\{|S_n/n| \geq \varepsilon\}$. Torej skoraj gotovo velja, da se za vsak $m \in \mathbb{N}$ zgodi kvečjemu končno mnogo dogodkov $\{|S_n/n| \geq 1/m\}$. A če je to res, zaporedje S_n/n konvergira proti nič. Sklep: zaporedje slučajnih spremenljivk S_n/n skoraj gotovo konvergira proti nič.

3. Velja $E(X_n) = 0$ in $D(X_n) = n/\ln n$, od koder sledi:

$$E(S_n = 0)\,, \qquad D(S_n) = \sum_{k=2}^{n} \frac{k}{\ln k} \leq \frac{n^2}{\ln n}$$

(slednja ocena velja za $n \geq 3$). Po neenačbi Čebiševa dobimo:

$$P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) = P(|S_n| \geq n\varepsilon) \leq \frac{1}{\varepsilon^2 \ln n}\,,$$

kar gre proti nič, ko gre n proti neskončno, zato zaporedje šibko konvergira proti nič.

V nadaljevanju bomo pokazali, da zaporedje S_n/n skoraj gotovo ni Cauchyjevo, ker skoraj gotovo za neskončno mnogo indeksov n velja $\left|\frac{S_{n+1}}{n+1} - \frac{S_n}{n}\right| \geq \frac{1}{2}$. Iz zapisa:

$$\left|\frac{S_{n+1}}{n+1} - \frac{S_n}{n}\right| = \frac{X_{n+1}}{n+1} - \frac{S_n}{n(n+1)}$$

namreč vidimo, da se to zagotovo zgodi, če je $|X_{n+1}| = n+1$ in $|S_n| \leq n(n+1)/2$. Po neenačbi Čebiševa ocenimo $P\big(|S_n| \leq \frac{n(n+1)}{2}\big) \leq \frac{4}{(n+1)^2 \ln n}$ in iz prve Borel–Cantellijeve leme sledi, da se skoraj gotovo zgodi kvečjemu končno mnogo dogodkov $\{|S_n| \leq n(n+1)/2\}$. Nadalje je $P(|X_{n+1}| = n+1) = 1/((n+1)\ln(n+1))$ in ker vrsta iz desnih strani divergira, se po drugi Borel–Cantellijevi lemi skoraj gotovo zgodi neskončno mnogo dogodkov $\{|X_{n+1}| = n+1\}$. Če oboje združimo, dobimo, da se skoraj gotovo zgodi neskončno mnogp dogodkov $\left|\frac{S_{n+1}}{n+1} - \frac{S_n}{n}\right| \geq \frac{1}{2}$, torej je zporedje S_n/n skoraj gotovo divergentno.

4. Za $k \in \mathbb{Z}$ označimo s π_k verjetnost, da slučajni sprehod pride v točko k; s π_0 označimo verjetnost, da se sprehod še kdaj vrne v izhodišče, in ta verjetnost nas zanima. Iz simetrije in dinamike slučajnega sprehoda (dogajanje razdelimo glede na *prvi* korak) razberemo:

$$\pi_k = \pi_{-k}\,, \tag{1}$$
$$\pi_k = \frac{\pi_{k-1} + \pi_{k+1}}{2}\,; \quad k = 2, 3, 4, \ldots \tag{2}$$
$$\pi_1 = \frac{1 + \pi_2}{2}\,, \tag{3}$$
$$\pi_0 = \frac{\pi_{-1} + \pi_1}{2} = \pi_1\,. \tag{4}$$

Enačba (2) je diferenčna enačba, katere rešitve so linearne kombinacije členov $k^r \lambda^k$, kjer je λ rešitev karakteristične enačbe $\lambda^2 - 2\lambda + 1 = 0$, ki ima dvojno rešitev $\lambda_{1,2} = 1$, torej so za $k \in \mathbb{N}$ rešitve oblike $\pi_k = C_1 + C_2 k$. A le za $C_2 = 0$ je možno, da so vse verjetnosti π_k v intervalu $[0, 1]$, torej so te verjetnosti vse enake. Iz zveze (3) dobimo, da je $\pi_k = 1$ za vse $k \in \mathbb{N}$, iz zvez (1) in (4) pa potem še, da je tudi $\pi_0 = 1$. Sklep: standardni slučajni sprehod se skoraj gotovo vrne v izhodišče (in tudi obišče vsa stanja).

5. 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:

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

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:

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

6. 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 4. 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.

7. Vzemimo slučajne spremenljivke:

$$X_1 = X_2 = \cdots \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} .$$

in postavimo še $X := -X_1$. Tedaj so vse slučajne spremenljivke enako porazdeljene, torej zaporedje X_n zagotovo konvergira proti X v porazdelitvi. Po drugi strani pa je:

$$P(|X_n - X| < 2) = P(X_n = X) = 0$$

za vse n, zato zaporedje ne more konvergirati v verjetnosti.

Če pa privzamemo, da je $X = c$ konstanta, konvergenca v porazdelitvi pomeni, da je:

$$\lim_{n\to\infty} P(X_n \le x) = 0 \quad \text{za } x < c$$
$$\lim_{n\to\infty} P(X_n \le x) = 1 \quad \text{za } x > c\,,$$

kar je ekvivalentno trditvi, da za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} P(X_n \le c - \varepsilon) = \lim_{n\to\infty} P(X_n > c + \varepsilon) = 0\,,$$

kar je ekvivalentno tudi trditvi, da za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} P(X_n \le c - \varepsilon) = \lim_{n\to\infty} P(X_n \ge c + \varepsilon) = 0\,,$$

to pa je ekvivalentno trditvi, da za vsak $\varepsilon > 0$ velja:

$$\lim_{n\to\infty} P\big(|X_n - c| \ge \varepsilon\big) = 0\,,$$

le-ta pa pomeni ravno konvergenco v verjetnosti.

8. Porazdelitve opišimo s kumulativnimi porazdelitvenimi funkcijami F_n, limitna porazdelitev pa naj ima kumulativno porazdelitveno funkcijo F. Slučajne spremenljivke konstruiramo tako kot v 31. nalogi iz 4. razdelka: naj bodo Q_n oz. Q merljive izbire kvantilnih funkcij, ki pripadajo danim kumulativnim porazdelitvenim funkcijam (ena od takih izbir je, če vzamemo spodnje kvantile, t. j. $Q(\omega) = \inf\{x\,;\, F(x) \ge \omega\}$ in analogno tudi $Q_n(\omega)$). Nadalje naj bo U porazdeljena enakomerno na intervalu $(0, 1)$. Tedaj imajo slučajne spremenljivke $Q_n(U)$ in $Q(U)$ kumulativne porazdelitvene funkcije F_n in F.

Funkcija Q je naraščajoča, zato ima kvečjemu števno mnogo točk nezveznosti. Dovolj bo torej pokazati, da je $\lim_{n\to\infty} Q_n(u) = Q(u)$, brž ko je Q zvezna v u.

Naj bo u taka točka in $x = Q(u)$. Nadalje naj bo $\varepsilon > 0$. Torej obstajata tak $u^+ > u$, da je $Q(u^+) < x + \varepsilon$, in tak $u^- < u$, da je $Q(u^-) > x - \varepsilon$. Tudi funkcija F ima kvečjemu števno mnogo točk nezveznosti, zato nadalje obstajata taka $x < x^+ < x + \varepsilon$ in $x - \varepsilon < x^- < x$, da je F zvezna v x^+ in x^- ter da je $Q(u^+) < x^+$ in $Q(u^-) > x^-$. Iz implikacij iz točke a) 31. naloge iz 4. razdelka dobimo, da potem velja $F(x^+) \ge u^+ > u$ in $F(x^-) \le u^- < u$. Zaradi zveznosti in šibke konvergence za dovolj velike n velja tudi $F_n(x^+) > u$ in $F_n(x^-) < u$. Spet iz implikacij dobimo $Q_n(u) \le x^+ < x + \varepsilon$ in $Q_n(u) \ge x^- > x - \varepsilon$. Ker to velja za vsak $\varepsilon > 0$, zaporedje $Q_n(u)$ konvergira proti x.

9. Za skoraj gotovo konvergenco je trditev pravilna, saj je:

$$\left\{\lim_{n\to\infty} X_n = X\right\} \cap \left\{\lim_{n\to\infty} Y_n = Y\right\} \subseteq \left\{\lim_{n\to\infty} (X_n + Y_n) = X + Y\right\}.$$

Prav tako je trditev pravilna za konvergenco v verjetnosti, saj je:

$$\left\{\lim_{n\to\infty} |X_n - X| < \frac{\varepsilon}{2}\right\} \cap \left\{\lim_{n\to\infty} |Y_n - Y| < \frac{\varepsilon}{2}\right\} \subseteq \left\{\lim_{n\to\infty} \bigl|(X_n + Y_n) - (X + Y)\bigr| < \varepsilon\right\}.$$

Pač pa trditev ni pravilna za konvergenco v porazdelitvi: vzemimo slučajne spremenljivke:

$$X_1 = X_2 = \cdots = Y_1 = Y_2 = \cdots \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

ter še slučajni spremenljivki:

$$X \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad Y = -X$$

Ker so vse omenjene slučajne spremenljivke enako porazdeljene, zaporedje X_n v porazdelitvi zagotovo konvergira proti X, Y_n pa proti Y. Pač pa je:

$$X_n + Y_n \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \text{medtem ko je} \quad X + Y = 0\,,$$

zato zaporedje $X_n + Y_n$ zagotovo ne konvergira v porazdelitvi proti $X + Y$.

10. Dokazati moramo, da za vsak z, kjer je F_{X+c} zvezna, velja:

$$\lim_{n\to\infty} P(X_n + Y_n \le z) = P(X + x \le z)\,,$$

kar je ekvivalentno trditvi, da za vsak x, kjer je F_X zvezna, velja:

$$\lim_{n\to\infty} P(X_n + Y_n - c \le x) = P(X \le x)\,. \tag{$*$}$$

Naj bo $\varepsilon > 0$ in naj bo F_X zvezna v $x + \varepsilon$ in $x - \varepsilon$. Iz zveze:

$$\begin{aligned} P(X_n + Y_n - c \le x) &= P(X_n + Y_n - c \le x, |Y_n - c| < \varepsilon) + \\ &\quad + P(X_n + Y_n - c \le x, |Y_n - c| \ge \varepsilon) \le \\ &\le P(X_n \le x + \varepsilon) + P(|Y_n - c| \ge \varepsilon) \end{aligned}$$

in konvergence zaporedij sledi:

$$\limsup_{n\to\infty} P(X_n + Y_n - c \le x) \le P(X \le x + \varepsilon)\,.$$

V limiti, ko gre ε proti nič, ob upoštevanju dejstva, da je F_X nezvezna v kvečjemu števno mnogo točkah, dobimo:

$$\limsup_{n\to\infty} P(X_n + Y_n - c \le x) \le P(X \le x)\,. \tag{$**$}$$

Naj bo zdaj spet $\varepsilon > 0$ in F_X zvezna v $x - \varepsilon$. Podobno kot prej iz zveze:

$$\begin{aligned} P(X_n \le x - \varepsilon) = P(X_n \le x - \varepsilon, |Y_n - c| < \varepsilon) + \\ + P(X_n \le x - \varepsilon, |Y_n - c| \ge \varepsilon) \le \\ \le P(X_n + Y_n - c \le x) + P(|Y_n - c| \ge \varepsilon) \end{aligned}$$

in konvergence zaporedij sledi:

$$\liminf_{n\to\infty} P(X_n + Y_n - c \le x) \ge P(X \le x - \varepsilon)\,.$$

V limiti, ko gre ε proti nič, ob upoštevanju zveznosti funkcije F_X v x in spet dejstva, da je F_X nezvezna v kvečjemu števno mnogo točkah, dobimo:

$$\liminf_{n\to\infty} P(X_n + Y_n - c \le x) \ge P(X \le x)\,. \qquad (***)$$

Iz $(**)$ in $(***)$ pa že sledi $(*)$.

11. Pišemo lahko $S = X_1 + X_2 + \cdots + X_{105}$, kjer je X_i število pik pri i-tem metu. Slučajne spremenljivke X_i so porazdeljene enakomerno na množici $\{1, 2, 3, 4, 5, 6\}$, od koder sledi $E(X_i) = 7/2$ in $D(X_i) = 35/12$. Torej je $E(S) = 735/2 = 367{\cdot}5$ in $D(S) = 1225/4$, torej $\sigma(S) = 17{\cdot}5$. Na prvi pogled se zdi smiselna aproksimacija:

$$P(S < 350) \approx \Phi\left(\frac{350 - 367{\cdot}5}{17{\cdot}5}\right) + \frac{1}{2} \doteq 0{\cdot}15866\,,$$

vendar je prav tako smiselno aproksimirati tudi:

$$P(S < 350) = P(S \le 349) \approx \Phi\left(\frac{349 - 367{\cdot}5}{17{\cdot}5}\right) + \frac{1}{2} \doteq 0{\cdot}14522\,.$$

Kadar slučajna spremenljivka zavzame vrednosti na mreži oblike $a\mathbb{Z} + b$, se v povprečju najbolj splača aproksimirati na sredini, torej:

$$P(S < 350) = P(S < 349{\cdot}5) \approx \Phi\left(\frac{349{\cdot}5 - 367{\cdot}5}{17{\cdot}5}\right) + \frac{1}{2} \doteq 0{\cdot}15184\,.$$

Točen rezultat: $0{\cdot}1520449$.

12. Rezultat, dobljen s pomočjo centralnega limitnega izreka:
$\Phi\left(\frac{9{\cdot}5}{\sqrt{120}}\right) + \Phi\left(\frac{29{\cdot}5}{\sqrt{120}}\right) \doteq 0{\cdot}80355$.
Točen rezultat: $0{\cdot}80439$.

13. Naša bilanca se izraža kot vsota 50 neodvisnih slučajnih spremenljivk s porazdelitvijo:

$$\begin{pmatrix} -1 & 0 & 9 \\ 0{\cdot}5 & 0{\cdot}45 & 0{\cdot}05 \end{pmatrix}.$$

Posamezen sumand ima matematično upanje $-0^{\cdot}05$ in disperzijo $4^{\cdot}5475$; cela bilanca pa ima matematično upanje $-2^{\cdot}5$, disperzijo $227^{\cdot}375$ in standardni odklon $15^{\cdot}08$. Verjetnost dobička aproksimiramo kot:

$$\frac{1}{2} - \Phi\left(\frac{0^{\cdot}5 + 2^{\cdot}5}{15^{\cdot}08}\right) \doteq 0^{\cdot}4211\,.$$

Točen rezultat pa je $0^{\cdot}3957429$. Napaka je tokrat večja, kot bi pričakovali iz prejšnjih primerov, to pa zato, ker ima porazdelitev vrednost, ki zelo odstopa.

14. 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 normalno $\mathrm{N}\big(100, \sqrt{200}\big)$.
$P(X > 110)$: CLI: $0^{\cdot}23975$, točen rezultat: $0^{\cdot}23220$.
$P(90 < X < 110)$: CLI: $0^{\cdot}52050$, točen rezultat: $0^{\cdot}52099$.

Opomba. Porazdelitev hi kvadrat je poseben primer porazdelitve gama, ta pa je *neskončno deljiva*, kar pomeni, da se da zapisati kot porazdelitev vsote *poljubnega* števila neodvisnih in enako porazdeljenih slučajnih spremenljivk. Natančneje, če so $X_1, X_2, \ldots X_n$ neodvisne s porazdelitvijo $\mathrm{Gama}(50/n, 1/2)$, ima njihova vsota porazdelitev hi kvadrat s 100 prostostnimi stopnjami. A to ne pomeni, da je porazdelitev hi kvadrat poljubno blizu normalni (oziroma kar normalna), ker se z rastočim n veča tudi razmerje γ_1/σ_1.

15. Iz centralnega limitnega izreka dobimo:

$$P(S_n > x) \approx \frac{1}{2} - \Phi\left(\frac{x}{\sqrt{2n}}\right)\,.$$

$P(S_5 > 5)$: CLI: $0^{\cdot}0569$, točen rezultat: $0^{\cdot}0555$.
$P(S_{20} > 20)$: CLI: $0^{\cdot}00078$, točen rezultat: $0^{\cdot}00114$.

Približek za $P(S_{20} > 20)$ ima veliko relativno napako, ker se že nahajamo v območju *velikih odklonov.*

16. Iz $E(U_i) = 1/2$, $E(U_i^2) = 1/3$, $E(V_i) = 4/3$, $E(V_i^2) = 2$ in neodvisnosti dobimo $E(X_i) = E(U_i)\,E(V_i) = 2/3$, $E(X_i^2) = E(U_i^2)\,E(V_i^2) = 2/3$ in $D(X_i) = 2/3 - (2/3)^2 = 2/9$. Sledi $E(S) = 200/3$ in zaradi neodvisnosti še $D(S) = 200/9$. Ker je S vsota velikega števila neodvisnih slučajnih spremenljivk, je po centralnem limitnem izreku:

$$P(S < 60) \approx \Phi\left(\frac{60 - \frac{200}{3}}{\sqrt{\frac{200}{9}}}\right) + \frac{1}{2} = \frac{1}{2} - \Phi\big(\sqrt{2}\big) \doteq 0^{\cdot}0787\,.$$

Točen rezultat: $0^{\cdot}07642$.

17. Iz:

$$E(X_i) = 1\,, \quad D(X_i) = 2p\,, \quad E(S) = 100\,, \quad D(S) = 200p$$

in centralnega limitnega izreka dobimo:

$$P(S < 90) \approx \Phi\left(\frac{89'5 - 100}{\sqrt{200p}}\right) + \frac{1}{2}$$

Torej bo $\Phi\left(\frac{10'5}{\sqrt{200p}}\right) \approx 0'45$ oziroma $\frac{10'5}{\sqrt{200p}} \approx 1'645$ oziroma $p \approx 0'204$.
Točen rezultat: 0'20414.

18. 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 $\mathrm{Bin}(450, 0'2)$ aproksimiramo z normalno $\mathrm{N}\big(90, \sqrt{72}\big)$, dobimo približen rezultat 0'10796.

Drugi način. Število potopov je porazdeljeno po Pascalovi porazdelitvi $\mathrm{NegBin}(80, 0'2)$, 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.

19. a) Velja $E(X_k) = k$, $D(X_k) = k^2$ in $E\big(|X_k - E(X_k)|^3\big) = k^3\gamma_1^3$, kjer je:

$$\gamma_1^3 = E\big(|X_1 - E(X_1)|^3\big) = \int_0^\infty |x-1|^3\, e^{-x}\, \mathrm{d}x = \frac{12}{e} - 2 \doteq 2'415\,.$$

Torej je:

$$E(S_n) = \frac{n(n+1)}{2}\,, \qquad D(S_n) = \frac{n(n+1)(2n+1)}{6}\,,$$
$$\sum_{k=1}^{n} E(|X_k - E(X_k)|^3) = \frac{n^2(n+1)^2}{4}\,\gamma_1^3\,.$$

Ker razmerje Ljapunova $\gamma_1^3\sqrt{\frac{27n(n+1)}{2(2n+1)^3}}$ konvergira proti nič, ko gre n proti neskončno, intervalske verjetnosti enakomerno konvergirajo proti ustreznim verjetnostim za normalno porazdelitev, od tod pa sledi, da standardizirane vsote šibko konvergirajo proti standardni normalni porazdelitvi.

b) Približno velja:

$$P(S_{100} < 6000) \approx \frac{1}{2} + \Phi\left(\frac{6000 - 5050}{\sqrt{338350}}\right) \doteq 0'948787\,.$$

Iz Tjurinove ocene po krajšem računu dobimo, da je napaka, ki smo jo naredili, navzgor omejena z 0'175. Dejanska napaka, ki smo jo naredili, pa je dosti manjša: s simulacijo Monte Carlo ($2 \cdot 10^8$ ponovitev) dobimo približek 0'9422 (95% interval zaupanja: $(0'9421832, 0'9422002)$).

20. Ker sumand X_1 izstopa (**prispeva celo več kot polovico disperzije**), ga moramo obravnavati posebej. Če z S' označimo vsoto preostalih, t. j. $S' = X_2 + X_3 + \cdots + X_{100}$, velja:

$$
\begin{aligned}
P(S < 41) &= P(X_1 = -1)\, P(S < 41 \mid X_1 = -1) + P(X_1 = 1)\, P(S < 41 \mid X_1 = 1) = \\
&= \frac{1}{3}\, P(S' < 51) + \frac{2}{3}\, P(S' < 31)\,.
\end{aligned}
$$

Ker slučajna spremenljivka S' zavzame vrednosti na lihih številih, je meji za S' v zgornji formuli smiselno znižati za 1. Iz:

$$
E(X_i) = \frac{1}{3}\,, \quad D(X_i) = \frac{8}{9}\,; \qquad E(S') = 33\,, \quad D(S') = 88
$$

dobimo:

$$
P(S < 41) \approx \frac{1}{3}\left[\Phi\left(\frac{50-33}{\sqrt{88}}\right) + \frac{1}{2}\right] + \frac{2}{3}\left[\Phi\left(\frac{30-33}{\sqrt{88}}\right) + \frac{1}{2}\right] \doteq 0^{\cdot}57138\,.
$$

Točen rezultat: $0^{\cdot}5695804$.

Opomba. Če bi centralni limitni izrek uporabili neposredno na S, bi dobili približek $0^{\cdot}692$, kar precej odstopa od prave vrednosti.

21. 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 $a \le b$ velja:

$$
\lim_{n\to\infty} P\left(a \le \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \le b\right) = \Phi(b) - \Phi(a)\,,
$$

brž ko je $a \le b$. Torej bi moralo veljati tudi:

$$
\lim_{n\to\infty} P\left(-\frac{1}{2} \le \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \le \frac{1}{2}\right) = 2\,\Phi\left(\frac{1}{2}\right) \doteq 0^{\cdot}3829 < 0^{\cdot}4\,. \qquad (*)
$$

Toda že za $n \ge 4$ velja:

$$
\begin{aligned}
P\left(-\frac{1}{2} \le \frac{X_1 + X_2 + \ldots + X_n}{\sqrt{n}} \le \frac{1}{2}\right) &\ge P(X_2 = X_3 = \cdots = X_n = 0) = \\
&= \left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right)\cdots\left(1 - \frac{1}{n^2}\right) = \\
&= \frac{n+1}{2n} \ge \frac{1}{2}\,.
\end{aligned}
$$

torej $(*)$ ne more veljati.

22. a) Lahko si predstavljamo, da iz posode, v kateri je n kroglic, od tega r rdečih, na slepo in brez vračanja izvlečemo s kroglic, ter da je N število rdečih med izvlečenimi. Torej je $N = X_1 + X_2 + \cdots + X_s$, kjer je X_i indikator dogodka, da je bila i-ta izvlečena kroglica rdeča. Iz $E(X_i) = \frac{r}{n}$ sledi $E(N) = \frac{rs}{n}$. Nadalje za različna i in j velja $E(X_j) = \frac{r(r-1)}{n(n-1)}$. Sledi:

$$D(N) = \frac{rs}{n} + \frac{r(r-1)s(s-1)}{n(n-1)} - \frac{r^2s^2}{n^2} = \frac{rs(n-r)(n-s)}{n^2(n-1)}\,.$$

b) Slučajno spremenljivko N lahko zapišemo kot vsoto indikatorjev na vsaj dva načina: kot vsoto indikatorjev 100 dogodkov, da je posamezna izvlečena kroglica rdeča, ali kot vsoto indikatorjev 200 dogodkov, da je posamezna rdeča kroglica izvlečena. Čeprav ti indikatorji niso neodvisni, normalna aproksimacija vseeno daje dobre rezultate.

Iz $E(N) = 40$ in $\sigma(N) \doteq 4\text{'}386169$ sklepamo:

$$P(N > 45) \approx \frac{1}{2} - \Phi\left(\frac{45\text{'}5 - 40}{4\text{'}386169}\right) \doteq 0\text{'}10493\,.$$

Točen rezultat: 0'1050956.

23. Centralnega limitnega izreka ne moremo uporabiti neposredno na S, ker produkt slučajne spremenljivke X in normalno porazdeljene slučajne spremenljivke, četudi neodvisne od X, ni nujno normalno porazdeljen. Prav tako centralni limitni izrek ne velja za vsoto $XY_1 + XY_2 + \cdots + XY_{100}$, saj so seštevanci odvisni (čeprav se da centralni limitni izrek posplošiti tudi na vsote slučajnih spremenljivk z določeno vrsto odvisnosti, je odvisnost prej omenjenih seštevancev premočna). Pravilno pa bo iskano verjetnost računati s pomočjo pogojnih verjetnosti glede na X. Če pišemo $T := Y_1 + Y_2 + \cdots + Y_{100}$, po izreku o polni verjetnosti velja:

$$\begin{aligned} P(400 < S < 500) &= P(X = 1)\, P(400 < S < 500 \mid X = 1) + \\ &\qquad + P(X = 2)\, P(400 < S < 500 \mid X = 2) = \\ &= \frac{2}{3}\, P(400 < T < 500) + \frac{1}{3}\, P(200 < T < 250)\,. \end{aligned}$$

Za slučajno spremenljivko T pa centralni limitni izrek velja: iz $E(T) = 300$ in $D(T) = 10000$ dobimo:

$$P(a < T < b) \approx \Phi\left(\frac{b-300}{100}\right) - \Phi\left(\frac{a-300}{100}\right),$$

torej je:

$$P(400 < S < 500) \approx \frac{2}{3}\big[\Phi(2) - \Phi(1)\big] + \frac{1}{3}\big[\Phi(1) - \Phi(0\text{'}5)\big] \doteq 0\text{'}141\,.$$

Oglejmo si še, koliko bi znašala iskana verjetnost za normalno slučajno spremenljivko z enakim matematičnim upanjem in disperzijo kot S. Za ta namen izračunamo:

$$E(S) = E(X)\, E(T) = 400\,.$$

Nadalje je $E(T^2) = D(T) + \big(E(T)\big)^2 = 100000$ in zato:

$$\begin{aligned} E(S^2) &= E(X^2)\,E(T^2) = 200000\,, \\ D(S) &= E(S^2) - \big(E(S)\big)^2 = 40000\,. \end{aligned}$$

Za normalno slučajno spremenljivko bi bila torej iskana verjetnost enaka:

$$\Phi(0\text{˙}5) \doteq 0\text{˙}191\,,$$

kar se občutno razlikuje od pravega rezultata.

10. Zadostne in postranske statistike

1. Uspešnost poskusov ponazorimo s slučajnim vektorjem $(X_1, \ldots X_n)$, kjer je $X_i = 1$, če i-ti poskus uspe, in 0, če ne uspe. Tedaj je S število poskusov. Pogojno na S je slučajni vektor $(X_1, \ldots X_n)$ porazdeljen enakomerno na vseh n-tericah, ki imajo natanko S enic in $n - S$ ničel. To velja ne glede na θ, zato je S zadostna statistika.

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 polni 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.

Primer: obstajajo natanko štiri zaporedja dolžine 8, za katera je $x_1 = 0$, $n_{00} = 3$, $n_{01} = n_{10} = 1$ in $n_{11} = 2$, in sicer:

01110000
00111000
00011100
00001110

Brezpogojna verjetnost vsakega od teh zaporedij je $(1-\theta)\bigl(1-\frac{\theta}{2}\bigr)^3\frac{\theta(1-\theta)}{4}\bigl(\frac{1+\theta}{2}\bigr)^2$, pogojna verjetnost glede na dogodek $\{X_0 = 0, N_{00} = 3, N_{01} = N_{10} = 1, N_{11} = 2\}$ pa je enaka $1/4$.

3. (1) $\Rightarrow$ (2). V diskretnem primeru zadostnost pomeni, da so pogojne verjetnosti $P_\theta(X = x \mid T = \tau)$ neodvisne od θ. Seveda ima to smisel gledati le za $\tau = t(x)$. To torej pomeni, da obstaja taka funkcija ρ, da je:

$$P_\theta(X = x) = P_\theta(T = t(x))\,\rho(x) \qquad (**)$$

(brž ko je $P_\theta(T = \tau) > 0$, je torej $\rho(x) = (P_\theta(X = x \mid T = \tau))$. Če postavimo še $g(\tau, \theta) := P_\theta(T = \tau)$, imamo zahtevano izražavo $(*)$.

(2) $\Rightarrow$ (3): očitno.

(3) $\Rightarrow$ (1): Dokazati moramo, da za vsak τ, za katerega obstaja tak $\theta \in \Theta$, da je $P_\theta(T = \tau) > 0$, in za katerega je $t^{-1}(\{\tau\}) = \{x_0, x_1, x_2, \ldots\}$, obstajajo taki $p_0, p_1, p_2, \ldots$ (neodvisni od θ), da je $P_\theta(X = x_i \mid T = \tau) = p_i$ za vsak i in vsak $\theta \in \Theta$, za katerega je $P_\theta(T = \tau) > 0$. Ekvivalentno, za vsak $\theta \in \Theta$ mora veljati $P_\theta(X = x_i) = p_i\,P_\theta(T = \tau)$. Če je $P_\theta(T = \tau) > 0$, obstaja tudi tak i, da je $P_\theta(X = x_i) > 0$, in brez škode za splošnost lahko privzamemo, da je $i = 0$. Ker sta za vsak $j = 1, 2, \ldots$ verjetnosti $P_\theta(X = x_0)$ in $P_\theta(X = x_j)$ sorazmerni, obstaja tak k_j, da je $P_\theta(X = x_j) = k_j\,P_\theta(X = x_0)$. za vse $\theta \in \Theta$. Če označimo še $k_0 = 1$ in $s := k_0 + k_1 + k_2 + \cdots$, potem velja $P_\theta(T = \tau) = P_\theta(X = x_0)$ in lahko postavimo $p_i := k_i/s$.

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. a) $b = 1 - 5a - 6a^2 = (1 + a)(1 - 6a)$.
b) $0 \le a \le 1/6$.
c) Če sorazmernost vzamemo tako kot v 3. nalogi, dobimo, da sta $P(X = 1)$ in $P(X = 4)$ sorazmerni, nadalje so sorazmerne verjetnosti $P(X = 2)$, $P(X = 3)$ in $P(X = 5)$, drugje pa ni sorazmernosti. Statistika je torej zadostna natanko tedaj, ko poljubne tri točke iz množic $\{1, 4\}$, $\{2, 3, 5\}$ in $\{6\}$, kjer iz vsake množice vzamemo po eno točko, preslika v tri različne točke. Ekvivalentno, statistika je zadostna

natanko tedaj, ko še vedno enolično določa, kateri od podmnožic $\{1,4\}$, $\{2,3,5\}$ in $\{6\}$ pripada opažanje. Minimalna zadostna statistika bo torej npr.:

$$t(1) = t(4) = \tau_1\,, \quad t(2) = t(3) = t(5) = \tau_2\,, \quad t(6) = \tau_3\,,$$

kjer so τ_1, τ_2 in τ_3 same različne vrednosti.

d) Zdaj pa so sorazmerne vse verjetnosti $P(X = i)$, kjer je $i = 1,2,3,4,5$, drugih sorazmernosti pa ni. Minimalna zadostna statistika bo torej npr.:

$$t(1) = t(2) = t(3) = t(4) = t(5) = \tau_1\,, \quad t(6) = \tau_2\,.$$

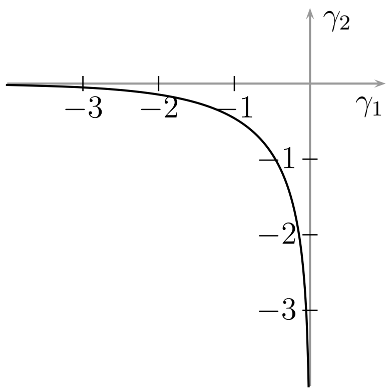
6. Najprej za $X \sim \begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}$ in $x \in \{0,1\}$ pišimo:

$$P(X = x) = (1-\theta)^{1-x}\theta^x = e^{(1-x)\ln(1-\theta)+x\ln\theta}\,,$$

kar nam da dvoparametrično eksponentno družino:

$$\gamma_1 = \ln(1-\theta)\,,\ \gamma_2 = \ln\theta\,, \quad h_1(x) = 1-x\,,\ h_2(x) = x\,, \quad \rho(x) = 1\,, \quad g(\gamma_1,\gamma_2) = 1$$

s parametričnim prostorom $(\gamma_1,\gamma_2) \in \mathbb{R}^2\,;\ e^{\gamma_1} + e^{\gamma_2} = 1$:



Družino pa lahko zapišemo tudi enoparametrično. Velja namreč tudi:

$$P(X = x) = (1-\theta)\exp\left(x\ln\frac{\theta}{1-\theta}\right)\,,$$

kar nam da:

$$\gamma = \ln\frac{\theta}{1-\theta}\,, \quad h(x) = x\,, \quad \rho(x) = 1\,, \quad g(\gamma) = \frac{1}{1+e^\gamma}$$

in parametrični prostor je kar cela realna os.

Podobno, če je $X \sim \mathrm{Bin}(n,p)$, lahko za $x \in \{0,1,\ldots n\}$ pišemo:

$$P(X = x) = \binom{n}{x}(1-\theta)^n \exp\left(x\ln\frac{\theta}{1-\theta}\right)\,,$$

kar je spet zapis enoparametrične eksponentne družine z:

$$\gamma = \ln\frac{\theta}{1-\theta}\,, \quad h(x) = x\,, \quad \rho(x) = \binom{n}{x}, \quad g(\gamma) = \frac{1}{\left(1+e^{\gamma}\right)^n}$$

in parametrični prostor je spet cela realna os.

7. Gre za eksponentno družino, saj lahko verjetnostno funkcijo zapišemo v obliki:

$$P_\lambda(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = \frac{e^{-n\lambda}}{x_1!\,x_2!\cdots x_n!}\,e^{(x_1+x_2+\cdots+x_n)\ln\lambda}\,.$$

Ker je parametrični prostor cela realna os, je vzorčna vsota res minimalna zadostna statistika.

8. Najprej verjetnostno gostoto za eno samo opažanje pri splošni normalni porazdelitvi $\mathrm{N}(\mu, \sigma)$ zapišemo v obliki:

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

iz katere razberemo, da gre za eksponentno družino z naravnima parametroma $-1/(2\sigma^2)$ in μ/σ^2 ter pripadajočo naravno zadostno statistiko (X^2, X).

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) Minimalna zadostna statistika je $\sum_{i=1}^n X_i^2$.

e) 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.

9. Pri porazdelitvi $\mathrm{Gama}(\lambda, a)$ je porazdelitvena gostota za eno samo opažanje enaka:

$$p(x \mid \lambda, a) = \frac{a^\lambda}{\Gamma(\lambda)}\,x^{\lambda-1}e^{-ax} = \frac{a^\lambda}{\Gamma(\lambda)}\,e^{(\lambda-1)\ln x - ax}\,,$$

torej gre za eksponentno družino z naravnima parametroma $\lambda - 1$ in a ter pripadajočo naravno zadostno statistiko $(\ln X, X)$. Zaradi linearne neodvisnosti je potem minimalna zadostna statistika za vzorec lahko:

$$\left(\ln X_1 + \ln X_2 + \cdots + \ln X_n,\ X_1 + X_2 + \cdots + X_n\right),$$

lahko pa recimo tudi:

$$\left(X_1 X_2 \cdots X_n,\ X_1 + X_2 + \cdots + X_n\right).$$

10. To je recimo $|X_1 + X_2 + \cdots + X_n|$.

11. Pišemo lahko $X_i = \mu + \sigma Z_i$, kjer so $Z_1, \dots Z_n$ neodvisne in porazdeljene standardno normalno (torej poznamo porazdelitev slučajnega vektorja $(Z_1, Z_2, \dots Z_n)$).

a) Vzemimo:

$$\big(X_1 - \bar{X}, X_2 - \bar{X}, \dots X_n - \bar{X}\big) = \sigma\big(Z_1 - \bar{Z}, Z_2 - \bar{Z}, \dots Z_n - \bar{Z}\big)\,,$$

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 $(x_1, x_2, \dots x_n)$ preslika v $(x_1 - \bar{x}, x_2 - \bar{x}, \dots x_n - \bar{x})$, 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 $(x_1, x_2, \dots x_n)$, za katere je $\bar{x} = 0$.

b) Vzemimo:

$$\left(\frac{X_1 - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}, \frac{X_2 - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}, \dots \frac{X_n - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}\right) = \\ = \left(\frac{Z_1}{\sqrt{\sum_{i=1}^n Z_i^2}}, \frac{Z_2}{\sqrt{\sum_{i=1}^n Z_i^2}} \dots \frac{Z_n}{\sqrt{\sum_{i=1}^n Z_i^2}}\right).$$

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 $(Z_1, \dots Z_n)$ se vidi, da je naša statistika porazdeljena enakomerno na enotski sferi.

c) Vzemimo:

$$\left(\frac{X_1 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \frac{X_2 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}, \dots \frac{X_n - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}\right) = \\ = \left(\frac{Z_1 - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}, \frac{Z_2 - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}} \dots \frac{Z_n - \bar{Z}}{\sqrt{\sum_{i=1}^n (Z_i - \bar{Z})^2}}\right).$$

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.

12. Če so $X_1, \dots X_n$ neodvisne in porazdeljene normalno $\mathrm{N}(\mu, \sigma)$,

a) sta $\bar{X}$ in $(X_1 - \bar{X}, \dots X_n - \bar{X})$ neodvisna;

b) 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;

c) 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.

13. Velja:

$$\sum_{i=1}^n (X_i-\mu)^2 = \sum_{i=1}^n (X_i-\bar{X})^2 + n(\bar{X}-\mu)^2 .$$

Iz 12. naloge vemo, da sta statistiki $\sum_{i=1}^n (X_i-\bar{X})^2$ in $n(\bar{X}-\mu)^2$ neodvisni. Poleg tega iz 37. naloge iz 4. razdelka in 22. naloge iz 5. 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:

$$\sum_{i=1}^n (X_i-\bar{X})^2 \sim \text{Gama}\left(\frac{n-1}{2}, \frac{1}{2\sigma^2}\right) \quad \text{oziroma} \quad \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\bar{X})^2 \sim \chi^2(n-1) .$$

14. 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 13. 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 24. nalogo iz 5. razdelka, ki pravi, da mora imeti zato T Studentovo porazdelitev z $n-1$ prostostnimi stopnjami.

11. Točkasto ocenjevanje in vzorčenje

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:

$$\begin{aligned}D(\bar{X}) &= \frac{1}{n^2}\,D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K(X_i, X_j) = \\ &= \frac{1}{n^2}\left[\sum_{i=1}^{n} D(X_i) - \sum_{i,j;i\ne j} K(X_i, X_j)\right].\end{aligned}$$

Velja $D(X_i) = D(X) = \sigma^2$, za $i \ne j$ pa je:

$$\begin{aligned}E(X_iX_j) &= \frac{1}{N(N-1)}\sum_{k,l;k\ne l} x_k x_l = \\ &= \frac{1}{N(N-1)}\left[\sum_{k=1}^{N}\sum_{l=1}^{N} x_k x_l - \sum_{k=1}^{N} x_k^2\right] = \\ &= \frac{1}{N(N-1)}\big(N^2\mu^2 - N(\mu^2+\sigma^2)\big) = \\ &= \mu^2 - \frac{\sigma^2}{N-1}\,,\end{aligned}$$

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) Iz 23. naloge v 5. razdelku sledi, da je $X_{(k)} \sim \mathrm{Beta}(k, n-k+1)$. Torej je:

$$E\big[X_{(k)}\big] = \frac{n!}{(k-1)!\,(n-k)!} \int_0^1 t^k (1-t)^{n-k}\,\mathrm{d}t = \frac{k}{n+1}\,.$$

b) Če je $X \sim \mathrm{E_z}(a,b)$, iz točke a) sledi:

$$E\big(X_{(k)}\big) = \frac{n+1-k}{n+1}\,a + \frac{k}{n+1}\,b\,.$$

Za cenilko $\hat{q}_p$ iskanega kvantila $q_p = (1-p)a + pb$ nastavimo linearno kombinacijo $\lambda X_{(i)} + \mu X_{(j)}$. Iz nepristranskosti sledi, da mora za poljubna a in b veljati:

$$\frac{\lambda(n+1-i) + \mu(n+1-j)}{n+1}\,a + \frac{\lambda i + \mu j}{n+1}\,b = (1-p)a + pb\,,$$

od koder dobimo sistem:

$$\begin{aligned} \lambda\left(1 - \frac{i}{n+1}\right) + \mu\left(1 - \frac{j}{n+1}\right) &= 1-p \\ \lambda\,\frac{i}{n+1} + \mu\,\frac{j}{n+1} &= p\,, \end{aligned}$$

ki ima rešitev:

$$\lambda = \frac{j-(n+1)p}{j-i}\,, \qquad \mu = \frac{i-(n+1)p}{i-j}\,.$$

c) Če je $p = i/(n+1)$, je $\hat{q}_p = X_{(i)}$; podobno, če je $p = j/(n+1)$, je $\hat{q}_p = X_{(j)}$. Opazimo, da sta $X_{(i)}$ in $X_{(j)}$ tudi vzorčna kvantila za verjetnosti $i/(n+1)$ in $j/(n+1)$.

Cenilko za kvantil za dano verjetnost p lahko torej opišemo takole: če je $p = i/(n+1)$ ali $p = j/(n+1)$, je ocena enaka kar vzorčnemu kvantilu za verjetnost p. Sicer pa med tema dvema točkama linearno interpoliramo oz. ekstrapoliramo.

d) Če le gre, je smiselno izbrati tista i in j, pri katerih interpoliramo. Brž ko je $1/(n+1) \le p \le n/(n+1)$, lahko i izberemo tako, da je $i/(n+1) \le p \le (i+1)/(n+1)$, in postavimo $j = i+1$. Med tema dvema točkama linearno interpoliramo.

e) Velja $2/9 \le 1/4 \le 3/9$, torej bo ocena enaka:

$$\hat{q}_{1/4} = \left(3 - \frac{9}{4}\right)\cdot 6 + \left(\frac{9}{4} - 2\right)\cdot 14 = 8\,.$$

3. a) Naj bo N moč celotne statistične množice, $N_1 = Np_1, \ldots N_r = Np_r$ pa moči posameznih stratumov (torej je $N = \sum_{i=1}^r N_i$). Nadalje naj bo x_{ij}, $i = 1, \ldots r$, $j = 1, \ldots N_i$, vrednost naše statistične spremenljivke na j-ti enoti i-tega stratuma. Tedaj velja:

$$E\big[f(X_i)\big] = \frac{1}{N_i}\sum_{j=1}^{N_i} f(x_{ij}) = \frac{1}{Np_i}\sum_{j=1}^{N_i} f(x_{ij})\,.$$

Sledi:

$$\sum_{i=1}^{r} p_i\, E\big[f(X_i)\big] = \sum_{i=1}^{r} \frac{1}{N} \sum_{j=1}^{N_i} f(x_{ij}) = E\big[f(X)\big]\,.$$

b) Po prejšnjem je:

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

Nadalje je:

$$\sigma^2 = E\big[(X-\mu)^2\big] = \sum_{i=1}^{r} p_i\, E\big[(X-\mu)^2\big] = \sum_{i=1}^{r} p_i\big(\sigma_i^2 + (\mu_i-\mu)^2\big)\,.$$

Pišemo lahko:

$$\sigma^2 = \sigma_{\mathrm{W}}^2 + \sigma_{\mathrm{B}}^2\,,$$

kjer je:

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

disperzija *med stratumi* ali *pojasnjena* disperzija (pojasnjena z različnostjo stratumov),

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

pa je disperzija *znotraj stratumov* ali *nepojasnjena* disperzija.

c) Iskana nepristranska cenilka vzorčnega povprečja je:

$$\bar{X}^{(\mathrm{s})} = p_1\bar{X}_1 + p_2\bar{X}_2 + \cdots + p_r\bar{X}_r\,,$$

njena disperzija pa je enaka:

$$D\big(\bar{X}^{(\mathrm{s})}\big) = \frac{p_1^2\sigma_1^2}{n_1} + \frac{p_2^2\sigma_2^2}{n_2} + \cdots + \frac{p_r^2\sigma_r^2}{n_r}\,.$$

d) Ko vstavimo $n_i = np_i$, dobimo $D\big(\bar{X}^{(\mathrm{s})}\big) = \dfrac{\sigma_{\mathrm{W}}^2}{n} \le \dfrac{\sigma^2}{n} = D\big(\bar{X}^{(\mathrm{e})}\big)$, kjer je $\bar{X}^{(\mathrm{e})}$ vzorčno povprečje na nestratificiranem enostavnem slučajnem vzorcu.

e) Poiskati moramo ekstrem disperzije $D\big(\bar{X}^{(\mathrm{s})}\big)$, 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:

$$D\big(\bar{X}^{(\mathrm{s})}\big) = \frac{\big(p_1\sigma_1 + p_2\sigma_2 + \cdots + p_r\sigma_r\big)^2}{n}$$

in iz Jensenove neenakosti sledi, da je to res manjše ali enako σ_{W}^2/n.

4. a) Če enote populacije označimo kar z $1, 2, \ldots N$, lahko cenilko zapišemo v obliki $\sum_{i=1}^N a_i x_i \,\mathbf{1}(i \in S)$. Njeno matematično upanje je enako $\sum_{i=1}^N a_i x_i \pi_i$ in mora biti enako populacijskemu povprečju $\frac{1}{N}\sum_{i=1}^N x_i$ ne glede na vrednosti $x_1, \ldots x_N$ spremenljivke X. To pa je možno le pri $a_i = 1/(N\pi_i)$. Horvitz–Thompsonova cenilka obstaja natanko tedaj, ko je $\pi_i > 0$ za vse i.

b) Pri enostavnem slučajnem vzorcu velikosti n brez ponavljanja je $\pi_i = n/N$, torej je Horvitz–Thompsonova cenilka enaka $\frac{1}{n}\sum_{i\in S} x_i$, torej gre kar za vzorčno povprečje.

c) Pri vzorcu velikosti n s ponavljanjem, zajetem iz populacije velikosti N, je $\pi_i = 1 - \big(1 - \frac{1}{N}\big)^n$. Horvitz–Thompsonova cenilka je torej enaka:

$$\frac{1}{N\Big[1 - \big(1 - \frac{1}{N}\big)^n\Big]}\sum_{i\in S} x_i$$

oziroma za $N = 10$ in $n = 3$:

$$\frac{100}{271}\sum_{i\in S} x_i \,.$$

Vrednosti na naših vzorcih: $2{\cdot}21, 1{\cdot}85, 2{\cdot}21, 24{\cdot}35$.

Opazimo, da ocena povprečja pri vzorcu $(21, 22, 23)$ presega maksimum, kar ni smiselno! To je zato, ker Horvitz–Thompsonova cenilka ne da nujno konveksne kombinacije vrednosti spremenljivke in zato ne komutira s translacijami.

d) Horvitz–Thompsonova cenilka komutira s translacijami natanko tedaj, ko je $\sum_{i\in S}(1/\pi_i) = N$ za vse vzorce S. Primer takšnega vzorčnega načrta za populacijo iz treh enot, kjer verjetnosti, da je posamezna enota v vzorcu, niso vse enake:

$$\begin{pmatrix} \{1\} & \{2,3\} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Lahko pa Horvitz–Thompsonovo cenilko tudi normaliziramo – vzamemo:

$$\frac{\sum_{i\in S}\frac{x_i}{\pi_i}}{\sum_{i\in S}\frac{1}{\pi_i}}\,.$$

Ta cenilka je robustnejša, ni pa več linearna.

e) Enote populacije indeksirajmo z dvojnimi indeksi (i, j), kar pomeni j-ta enota v i-tem stratumu. Tedaj velja:

$$\pi_{(i,j)} = \frac{k}{K}\,\frac{n_i}{N_i}\,,$$

Če z I_i označimo indikator dogodka, da je i-ti stratum izbran, velja:

$$\sum_{(i,j)\in S} \frac{1}{\pi_{i,j}} = \frac{K}{k} \sum_{i=1}^{K} N_i I_i \,.$$

To je za vse vzorce enako N natanko tedaj, ko so stratumi bodisi vsi izbrani bodisi vsi enako veliki, t. j. $N_i = N/K$.

f) Če s T označimo Horvitz–Thompsonovo cenilko, velja:

$$\begin{aligned}
T^2 &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\mathbf{1}(x_i \in S, x_j \in S)}{\pi_i \pi_j} x_i x_j \,, \\
E(T^2) &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} x_i x_j \,, \\
\bigl(E(T)\bigr)^2 &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \,, \\
D(T) &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) x_i x_j \,.
\end{aligned}$$

Če pa za vsak S velja $\sum_{i\in S}(1/\pi) = N$, tudi za vsak u velja:

$$T = u + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\pi_i}(x_i - u) \,,$$

torej je tudi:

$$D(T) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) (x_i - u)(x_j - u) \,.$$

Opomba. Disperzija je zagotovo enaka nič, če je $\pi_{ij} = \pi_i \pi_j$ za vse i in j. To pa pomeni, da sta poljubna dogodka $\{i \in S\}$ in $\{j \in S\}$ neodvisna. Za različne i in j to lahko dosežemo: za vsako enoto se neodvisno odločimo, ali jo vzamemo v vzorec ali ne. Za $i = j$ pa je to res le, če je $\pi_i = 0$ (kar za Horvitz–Thompsonovo cenilko ni smiselno) ali $\pi_i = 1$, kar pomeni, da v vzorec zajamemo vso populacijo.

5. Ne, ker se porazdelitev vzorčnega povprečja ujema s porazdelitvijo populacije (glej 33. nalogo iz 8. razdelka) in je zato verjetnost v limiti pri pogoju za doslednost konstantna.

6. a) Iz $z_1 = E(X) = a/2$ dobimo $\hat{a} = 2\bar{X}$.

b) $\hat{a} = 6{\cdot}8$, kar je nesmiselno glede na to, da je v vzorcu tudi vrednost 10.

c) Če vzamemo četrti moment, iz $z_4 = a^4/5$ dobimo $\hat{a} = \sqrt[4]{5\hat{z}_4} = \sqrt[4]{10099} \doteq 10{\cdot}025$, kar je smiselno.

7. $\hat{\alpha} = \bar{X} = (X_1 + X_2 + \cdots + X_n)/n$. Cenilka je nepristranska. Je tudi dosledna, ker je dobljena po metodi momentov.

8. $\hat{a} = (\hat{z}_2 - 1)/2, \quad \hat{b} = (3 + \hat{z}_1 - 2\hat{z}_2)/2$.
Na našem konkretnem vzorcu iz $\hat{z}_1 = 1/2$ in $\hat{z}_2 = 13/10$ dobimo $\hat{a} = 3/20$ in $\hat{b} = 9/20$.

9. Velja $E(X) = 0$ (neodvisno od α), zato iz prvega momenta ne dobimo ničesar. Iz drugega momenta dobimo $\hat{\alpha} = \sqrt{\hat{z}_2/2}$.

10. 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 $s_0^2 = \hat{z}_2 - \bar{X}^2$. Z nekaj računanja dobimo, da lahko to zapišemo tudi v obliki:

$$s_0^2 = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

Označimo $Y := X - \mu$. Tedaj velja:

$$\begin{aligned} s_0^2 &= \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \\ &= \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 = \\ &= \frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Y_i Y_j \\ &= \frac{n-1}{n^2} \sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2} \sum\sum_{\substack{1 \le i,j \le n \\ i \ne j}} Y_i Y_j \end{aligned}$$

Ker je $E(Y) = 0$, je očitno:

$$E(s_0^2) = \frac{n-1}{n} \, \sigma^2 \,,$$

torej je s_0^2 pristranska cenilka za σ^2 (je pa *asimptotično nepristranska*). Če postavimo $k' := n/(n-1)$, dobimo, da je:

$$s^2 = \frac{n}{n-1} \, s_0^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

nepristranska cenilka za σ^2.

Za izračun srednje kvadratne napake pišimo:

$$
\begin{aligned}
s_0^4 = {} & \frac{(n-1)^2}{n^4} \sum_{i=1}^n \sum_{j=1}^n Y_i^2 Y_j^2 - \frac{2(n-1)}{n^4} \sum_{i=1}^n \mathop{\sum\sum}_{\substack{1 \le j,k \le n \\ j \ne k}} Y_i^2 Y_j Y_k + \\
& + \frac{1}{n^4} \mathop{\sum\sum}_{\substack{1 \le j,k \le n \\ i \ne j}} \mathop{\sum\sum}_{\substack{1 \le k,l \le n \\ k \ne l}} Y_i Y_j Y_k Y_l = \\
= {} & \frac{(n-1)^2}{n^4} \sum_{i=1}^n Y_i^4 + \frac{(n-1)^2}{n^4} \mathop{\sum\sum}_{\substack{1 \le i,j \le n \\ i \ne j}} Y_i^2 Y_j^2 - \frac{2(n-1)}{n^4} \sum_{i=1}^n \mathop{\sum\sum}_{\substack{1 \le j,k \le n \\ j \ne k}} Y_i^2 Y_j Y_k + \\
& + \frac{1}{n^4} \mathop{\sum\sum}_{\substack{1 \le i,j \le n \\ i \ne j}} \mathop{\sum\sum}_{\substack{1 \le k,l \le n \\ k \ne l}} Y_i Y_j Y_k Y_l \,.
\end{aligned}
$$

Matematično upanje prvih dveh členov v zadnjem izrazu je očitno. Pri tretjem členu opazimo, da je, če je $j \ne k$, $E(Y_i^2 Y_j Y_k)$ 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 $i \ne j$ in $k \ne l$, 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:

$$
\begin{aligned}
E(s_0^4) &= \frac{(n-1)^2}{n^3}\,\kappa^4 + \frac{(n-1)^3}{n^3}\,\sigma^4 + \frac{2(n-1)}{n^3}\,\sigma^4 = \\
&= \frac{(n-1)^2}{n^3}\,\kappa^4 + \frac{(n-1)(n^2-2n+3)}{n^3}\,\sigma^4 \,,
\end{aligned}
$$

kjer je κ^4 četrti centralni moment. Od tod dobimo:

$$
q(ks_0^2) = \left(\frac{(n-1)^2\kappa^4}{n^3} + \frac{(n-1)(n^2-2n+3)\sigma^4}{n^3} \right) k^2 - \frac{2(n-1)\sigma^4}{n}\,k + \sigma^4
$$

Za $k = 1$ dobimo:

$$
q(s_0^2) = \frac{(n-1)^2\kappa^4 + (-n^2+5n-3)\sigma^4}{n^3}
$$

(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 $\kappa^4 = 3\sigma^4$ in velja:

$$
q(ks_0^2) = \left(\frac{n^2-1}{n^2}\,k^2 - \frac{2(n-1)}{n}\,k + 1 \right) \sigma^4
$$

kar je minimalno pri $k = n/(n+1)$. Najučinkovitejša izmed cenilk oblike ks_0^2 je torej:

$$(s_0^*)^2 = \frac{1}{n+1} \sum_{i=1}^{n} (X_i - \bar{X})^2 .$$

Srednje kvadratične napake cenilk pa so:

$$q(s_0^2) = \frac{2n-1}{n^2}\,\sigma^4 , \quad q(s^2) = \frac{2}{n-1}\,\sigma^4 , \quad q\big((s_0^*)^2\big) = \frac{2}{n+1}\,\sigma^4 .$$

Ni težko preveriti, da je s_0^2 vselej učinkovitejša od s^2 in da je $(s_0^*)^2$ vselej učinkovitejša od s_0^2.

11. 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 8. 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 8. 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 $\bar{X}$.

12. Naj bo A dogodek, da prvi poskus uspe, drugi ne uspe, tretji pa spet uspe. Tedaj je:

$$P_\theta(A) = \theta \left(1 - \frac{1+\theta}{2}\right) \frac{\theta}{2} = \frac{\theta^2(1-\theta)}{4} .$$

Iščemo maksimum tega izraza za $\theta \in [0,1]$. Iz $P_0(A) = P_1(A) = 0$ in $\frac{\mathrm{d}}{\mathrm{d}\theta}\,P_\theta(A) = \frac{1}{4}(2\theta - 3\theta^2)$, kar je enako nič pri $\theta = 2/3$, dobimo, da je lahko maksimum dosežen kvečjemu v prej omenjenih točkah. Ker je edino $P_{2/3}(A) > 0$, se to zgodi pri $\theta = \hat{\theta} := 2/3$, kar je tudi ocena po metodi največjega verjetja.

13. Tu dobimo isto cenilko kot pri metodi momentov, t. j. $\hat{\alpha} = \bar{X}$.

14. a) Porazdelitev ima smisel pri $a > 1$, ko velja $c = a - 1$.

b) $\hat{a} = \dfrac{n}{\ln X_1 + \ln X_2 + \cdots + \ln X_n} + 1.$

15. $\hat{a} = 1/4$, $\hat{b} = 3/10$, kar se ne ujema z ocenama po metodi momentov.

16. Pri $X > 0$ je cenilka edina rešitev enačbe $-\frac{\ln \hat{\lambda}}{\hat{\lambda}} = X$.
Pri $X = 0$ cenilka ni natančno določena (funkcija verjetja je enaka za vse λ).
Pri $-e^{-1} \le X < 0$ je cenilka *večja* rešitev enačbe $-\frac{\ln \hat{\lambda}}{\hat{\lambda}} = X$.
Opažanje $X < -e^{-1}$ je v nasprotju z našim statističnim modelom.

17. 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 $M := \max\{X_1, \dots X_n\}$. Za $x \in [0, a]$ je kumulativna porazdelitvena funkcija te slučajne spremenljivke enaka:

$$F_M(x) = P(M < x) = P(X_1 < x, X_2 < x, \dots X_n < x) = \left(\frac{x}{a}\right)^n ,$$

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 $n \ge 3$ je M učinkovitejša od A in M' učinkovitejša od M.
Vselej pa je M^* učinkovitejša od vseh ostalih cenilk.

18. Pripadajoča zadostna statistika je število uspelih poskusov, ki ga označimo z S. Za iskano cenilko $h(S)$ bo torej moralo veljati:

$$E\big[h(S)\big] = p^2 .$$

Ker je S porazdeljena binomsko $\mathrm{Bin}(3, p)$, to pomeni:

$$h(0)\,(1-p)^3 + 3\,h(1)\,p(1-p)^2 + 3\,h(2)\,p^2(1-p) + h(3)\,p^3 = p^2$$

za vse $p \in (0, 1)$. S primerjavo koeficientov dobimo, da bo to natanko tedaj, ko bo:

$$h(0) = 0\,, \quad h(1) = 0\,, \quad h(2) = \tfrac{1}{3}\,, \quad h(3) = 1\,.$$

Za posplošitev cenilke na n poskusov le-to iščemo kot polinom pripadajoče zadostne statistike S. Iz $E(S) = np$ in $E(S^2) = np + (n^2 - n)p^2$ dobimo $E(S^2 - S) = (n^2 - n)p^2$, torej bo iskana cenilka enaka:

$$\frac{S^2 - S}{n^2 - n}$$

in zlahka se lahko prepričamo, da se za $n = 3$ ujema s prej dobljeno cenilko.

Opomba. Ker je $E(S^k)$ polinom stopnje k v θ in ker je vsako funkcijo na $\{0, 1, \ldots n\}$ možno zapisati kot polinom stopnje največ n, nepristranske cenilke obstajajo le za karakteristike, ki so polinomi parametra θ stopnje največ n.

19. a) Iz generičnega eksponentnega zapisa verjetnostne funkcije:

$$f_X(x) =$$
$$= \exp\left(\rho_0(x) \ln \frac{1}{1 + 3\theta + 2\theta^2} + \rho_1(x) \ln \frac{\theta}{1 + 3\theta + 2\theta^2} + \rho_2(x) \ln \frac{\theta^2}{1 + 3\theta + 2\theta^2}\right)$$

dobimo triparametrični zapis s pripadajočo zadostno statistiko $\big(\rho_0(X), \rho_1(X), \rho_2(X)\big)$. Iz zapisa:

$$f_X(x) = \frac{\rho_0(x) + 3\,\rho_1(x) + 2\,\rho_2(x)}{1 + 3\theta + 2\theta^2}\, e^{x \ln \theta}$$

pa dobimo enoparametričen zapis s pripadajočo zadostno statistiko X.

b) Cenilka $\varphi(X)$ bo nepristranska natanko tedaj, ko bo za vsak θ veljalo:

$$E\big[\varphi(X)\big] = \frac{\varphi(0) + 3\,\varphi(1)\,\theta + 2\,\varphi(2)\,\theta^2}{1 + 3\theta + 2\theta^2} = \frac{1}{1 + \theta}\,,$$

kar je ekvivalentno:

$$\varphi(0) + 3\,\varphi(1)\,\theta + 2\,\varphi(2)\,\theta^2 = 1 + 2\theta\,.$$

Cenilka bo torej nepristranska natanko tedaj, ko bo $\varphi(0) = 1$, $\varphi(1) = 2/3$ in $\varphi(2) = 0$. Ker je ena sama, ima seveda tudi najmanjšo možno disperzijo.

20. a) Seveda gre za eksponentno družino z naravnim parametrom λ in pripadajočo zadostno statististiko $S = X_1 + \cdots + X_n$. Iščemo funkcijo φ, za katero je $E\big[\varphi(S)\big] = \lambda$ za vse $\lambda > 0$. Statistika S ima porazdelitev Gama(n, λ), torej je:

$$E\big[\varphi(S)\big] = \frac{\lambda^n}{\Gamma(n)} \int_0^\infty \varphi(x)\, x^{n-1}\, e^{-\lambda x}\, \mathrm{d}x\,.$$

Veljati mora torej:

$$\int_0^\infty \varphi(x)\, x^{n-1}\, e^{-\lambda x}\, \mathrm{d}x = \frac{\Gamma(n)}{\lambda^{n-1}}\,.$$

Opazimo, da je izraz pod integralom Laplaceova transformiranka funkcije $x \mapsto x^{n-1}\varphi(x)$. Iz tabele Laplaceovih transformirank po nekaj računanja razberemo, da danemu pogoju ustreza funkcija $\varphi(x) = (n-1)/x$, torej bo iskana cenilka $(n-1)/S$.

b) Krajši račun pokaže:

$$q\left(\frac{a}{S}\right) = \lambda^2 \left(\frac{a^2}{(n-1)(n-2)} - \frac{2a}{n-1} + 1\right).$$

od koder dobimo, da je srednja kvadratična napaka minimalna pri $a = n - 2$.

21. Seveda gre za eksponentno družino s pripadajočo zadostno statistiko $S = X_1 + \cdots + X_n$. Razmeroma lahko je opaziti, da je statistika $\mathbf{1}(X_1 = 1)$ nepristranska cenilka za p. Za $n = 1$ je to tudi iskana statistika. Pri $n > 1$ pa iz:

$$P(X_1 = 1 \mid S = k) = \frac{n-1}{k-1}$$

razberemo iskano cenilko $\dfrac{n-1}{S-1}$.

22. a) Funkcija verjetja:

$$f_X(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

nam da:

$$L = \frac{\lambda^X e^{-\lambda}}{X!}, \quad \frac{\mathrm{d}L}{\mathrm{d}\lambda} = \frac{(X-\lambda)\lambda^{X-1} e^{-\lambda}}{X!},$$

od koder dobimo cenilko $\hat{\lambda} = X$, cenilka za λ^2 pa je X^2. Le-ta je pristranska, saj je $E(X^2) = \lambda^2 + \lambda$.

b) Ker gre za eksponentno družino, katere naravni parametrični prostor ima neprazno notranjost, bo nepristranska cenilka imela enakomerno najmanjšo možno disperzijo, brž ko bo funkcija minimalne zadostne statistike X. Iščemo torej tako funkcijo φ, da bo $E_\lambda\big(\varphi(X)\big) = \lambda^2$ za vse λ. To lahko naredimo z nastavkom:

$$\sum_{k=0}^{\infty} \frac{\varphi(k)\, \lambda^k e^{-\lambda}}{k!} = \lambda^2$$

oziroma:

$$\sum_{k=0}^{\infty} \frac{\varphi(k)\, \lambda^k}{k!} = \lambda^2 e^{\lambda},$$

nakar z razvojem:

$$\lambda^2 e^{\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^{n+2}}{n!} = \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} = \sum_{k=2}^{\infty} \frac{k(k-1)\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{k(k-1)\lambda^k}{k!}$$

in primerjavo koeficientov dobimo, da mora biti $\varphi(k) = k(k-1)$, torej je iskana statistika $X(X-1)$. Le-to lahko tudi kar uganemo iz prvih dveh momentov Poissonove porazdelitve.

c) Velja:

$$\begin{aligned}q(X^2 - aX) &= E\big[(X^2 - aX - \lambda^2)^2\big] = \\ &= E(X^4) + a^2\,E(X^2) + \lambda^4 - 2a\,E(X^3) - 2\lambda^2\,E(X^2) + 2a\lambda^2\,E(X) = \\ &= 4\lambda^3 + (7 - 6a + a^2)\lambda^2 + (1 - 2a + a^2)\lambda\,.\end{aligned}$$

Najnatančnejša cenilka bi bila tista, ki bi imela pri *vseh* λ najmanjšo srednjo kvadratično napako; smiselna pa je tista, za katero ne obstaja nobena druga vrednost a, pri kateri bi bila srednja kvadratična napaka strogo manjša za vse λ. Z drugimi besedami, cenilke delno uredimo, in sicer naj bo $\hat{\lambda}_1 \preceq \hat{\lambda}_2$, če je $q(\hat{\lambda}_1) \le q(\hat{\lambda}_2)$ za vse λ. Najboljša cenilka je tista, ki je glede na to ureditev najmanjši element, smiselna pa je taka, ki je minimalni element.

Delno urejenost cenilk v izbrani obliki lahko izrazimo s koeficientoma $c_1(a) := 1 - 2a + a^2$ in $c_2(a) = 7 - 6a + a^2$: za $\hat{\lambda}_1 = X^2 - a_1 X$ in $\hat{\lambda}_2 = X^2 - a_2 X$ je namreč $\hat{\lambda}_1 \preceq \hat{\lambda}_2$ natanko tedaj, ko je $c_1(a_1) \le c_1(a_2)$ in $c_2(a_1) \le c_2(a_2)$. Za $a < 1$ sta oba koeficienta strogo padajoča, za $a > 3$ sta oba strogo naraščajoča, za $1 \le a \le 3$ pa je c_1 naraščajoč, c_2 pa padajoč. Od tod sledi, da najmanjšega elementa iz te družine ni, minimalne pa dobimo za $1 \le a \le 3$. To so torej smiselne cenilke, najboljše pa ni.

23. a) Če pišemo $a = e^{-3\lambda}$ in $\hat{a} = h(X)$, mora za vsak λ veljati:

$$E\big[h(X)\big] = \sum_{k=0}^{\infty} f(k)\,\frac{\lambda^k\,e^{-\lambda}}{k!} = e^{-3\lambda}\,,$$

torej:

$$\sum_{k=0}^{\infty} f(k)\,\frac{\lambda^k}{k!} = e^{-2\lambda} = \sum_{k=0}^{\infty} \frac{(-2)^k \lambda^k}{k!}\,.$$

Iz enoličnosti potenčnih vrst sledi $f(k) = (-2)^k$, torej je *edina možna* nepristranska cenilka $\hat{a} = (-2)^X$. Toda ta cenilka lahko zavzame zelo velike vrednosti, čeprav je karakteristika, ki jo ocenjuje, strogo pozitivna.

b) Cenilka po metodi največjega verjetja je e^{-3X}. Iz:

$$q\big[a^X\big] = e^{(a^2-1)\lambda} - 2\,e^{(a-4)\lambda} + e^{-6\lambda}$$

dobimo:

$$q\big[(-2)^X\big] - q\big[e^{-3\lambda}\big] = e^{3\lambda} - e^{(e^{-6}-1)\lambda} + 2\,e^{(e^{-3}-4)\lambda} - 2\,e^{-6\lambda} > 0$$

za vse $\lambda > 0$.

c) Cenilko iščemo kot funkcijo minimalne zadostne statistike, ki je v našem primeru recimo $S_n = X_1 + X_2 + \cdots + X_n$. Ker gre za eksponentno družino, katere parametrični prostor vsebuje odprto množico, bo imela taka cenilka tudi enakomerno najmanjšo možno disperzijo med vsemi nepristranskimi.

Nastavimo torej $\hat{a} = h_n(S_n)$. Ker je $S_n \sim \mathrm{Poi}(n\lambda)$, bo moralo veljati:

$$E\big[h_n(S_n)\big] = \sum_{k=0}^{\infty} h_n(k)\,\frac{n^k \lambda^k\, e^{-n\lambda}}{k!} = e^{-3\lambda}\,,$$

torej:

$$\sum_{k=0}^{\infty} f(k)\,\frac{n^k \lambda^k}{k!} = e^{(n-3)\lambda} = \sum_{k=0}^{\infty} \frac{(n-3)^k \lambda^k}{k!}\,.$$

Dobimo cenilko $\hat{a} = \left(\dfrac{n-3}{n}\right)^S$, kar se za velike n bliža $e^{-3S/n}$, to pa je ravno cenilka po metodi največjega verjetja.

12. Intervali zaupanja

1. Obravnavajmo najprej primer, ko je $n = 1$. Najprej opazimo, da mora interval za $S = 0$ vsebovati določen interval oblike $[0, \varepsilon)$, $\varepsilon > 0$, sicer je lahko verjetnost pokritosti poljubno majhna. Torej mora biti $[0, b_0)$ ali $[0, b_0]$, kjer je $b_0 > 0$. Zaradi simetrije mora biti potem interval za opažanje $S = 1$ enak $(1 - b_0, 1]$ oz. $[1 - b_0, 1]$. Brž ko je $b_0 < 1/2$ ali pa je $b_0 = 1/2$ in je interval za $S = 0$ odprt pri b_0, interval pri $\theta = 1/2$ ni pokrit in je torej verjetnost pokritosti enaka nič. Pri $b_0 = 1/2$ in zaprti različici je verjetnost pokritosti enaka:

$$\begin{cases} 1-\theta & ; 0 \le \theta \le 1/2 \\ \theta & ; 1/2 \le \theta \le 1 \end{cases},$$

torej je minimalna verjetnost pokritosti enaka $1/2$. Pri $b_0 > 1/2$ in odprti različici pa je verjetnost pokritja enaka:

$$\begin{cases} 1-\theta & ; 0 \le \theta \le 1 - b_0 \\ 1 & ; 1 - b_0 < \theta < b_0 \\ \theta & ; b_0 \le \theta \le 1 \end{cases},$$

torej je minimalna verjetnost pokritosti enaka b_0. Pri $\beta > 1/2$ bo torej iskani interval oblike:

$$\begin{cases} [0, \beta) & ; S = 0 \\ (1-\beta, 1] & ; S = 1 \end{cases},$$

pri $\beta \le 1/2$ pa bo oblike:

$$\begin{cases} [0, 1/2] & ; S = 0 \\ [1/2, 1] & ; S = 1 \end{cases}.$$

Da se dokazati, da je to edina rešitev, ki izpolnjuje pogoje naloge.

Oglejmo si zdaj še primer, ko je $n = 2$. Podobno kot prej ugotovimo, da mora biti interval pri opažanju $S = 0$ oblike $[0, b_0)$ ali $[0, b_0]$ $(b > 0)$, kar za $S = 2$ da $(1 - b_0, 1]$ oziroma $[1 - b_1, 1]$. Pri $S = 1$ pa je interval lahko oblike $(1 - b_1, b_1)$ ali $[1 - b_1, b_1]$ (za $b_1 > 1/2$) ali pa tudi kar $\{1/2\}$. Zaradi monotonosti mora biti $b_1 \ge b_0$, poleg tega pa mora biti tudi $b_0 + b_1 \ge 1$ (oz. $b_0 \ge 1/2$, če je interval za $S = 1$ kar $\{1/2\}$), sicer za $b_0 < \theta < 1 - b_1$ (oz. za $b_0 < \theta < 1/2$) interval ni pokrit. Naj bo najprej $b_0 > 1/2$ in naj bo za $S = 0$ interval oblike $[0, b_0)$, za $S = 1$ pa oblike $(1 - b_1, b_1)$. Tedaj je verjetnost pokritosti enaka:

$$\begin{cases} (1-\theta)^2 & ; 0 \le \theta \le 1 - b_0 \\ 1-\theta^2 & ; 1 - b_1 < \theta \le 1 - b_0 \\ 1 & ; 1 - b_0 < \theta < b_0 \\ 2\theta - \theta^2 & ; b_0 \le \theta < b_1 \\ \theta^2 & ; \theta \ge b_1 \end{cases}$$

in minimalna verjetnost pokritosti bo $\min\{2b_0 - b_0^2, b_1^2\}$. Za $\beta > 3/4$ je možno nastaviti $2b_0 - b_0^2 = b_1^2 = \beta$: postavimo namreč $b_0 := 1 - \sqrt{1-\beta}$ in $b_1 := \sqrt{\beta}$.

Minimalni interval zaupanja je torej:

$$\begin{cases} [0, 1-\sqrt{1-\beta}) & ; S=0 \\ (1-\sqrt{\beta}, \sqrt{\beta}) & ; S=1 \\ (\sqrt{1-\beta}, 1] & ; S=2 \end{cases}.$$

in da se dokazati, da je edini minimalni interval zaupanja te oblike.

Pri $1/2 < \beta \le 3/4$ bo še vedno $b_1 = \sqrt{\beta} > 1/2$, medtem ko bo $b_0 = 1 - \sqrt{1-\beta} \le 1/2$. Toda za $b_0 \le 1/2$ bo verjetnost pokritosti pri prejšnjem intervalu zaupanja enaka:

$$\begin{cases} (1-\theta)^2 & ; 0 \le \theta \le 1-b_0 \\ 1-\theta^2 & ; 1-b_1 < \theta < b_0 \\ 2\theta(1-\theta) & ; b_0 \le \theta \le 1-b_0 \\ 2\theta-\theta^2 & ; b_0 \le \theta < b_1 \\ \theta^2 & ; \theta \ge b_1 \end{cases}$$

in minimalna verjetnost pokritosti bo $\min\{2b_0 - 2b_0^2, b_1^2\} \le 1/2$, kar ne bo v redu. Od tod dobimo, da je minimalni interval zaupanja pri $1/2 < \beta \le 3/4$ oblike:

$$\begin{cases} [0, 1/2] & ; S=0 \\ \bigl(1-\sqrt{\beta}, \sqrt{\beta}\bigr) & ; S=1 \\ [1/2, 1] & ; S=2 \end{cases}.$$

Sistem enačb $2b_0 - b_0^2 = b_1^2 = \beta$ ima v okviru naših pogojev rešitev $b_0 = \frac{1}{2}\bigl(1 - \sqrt{1-2\beta}\bigr)$, $b_1 = \sqrt{\beta}$. Toda zdaj moramo paziti na pogoj $b_0 + b_1 \ge 1$. Krajši račun pokaže, da gre to le pri $\beta \ge 4/9$; pri $\beta = 4/9$ dobimo $b_0 + b_1 = 1$, kjer moramo bolj paziti, zato bomo ta primer obravnavali kasneje. Torej bo minimalni interval zaupanja pri $4/9 < \beta \le 1/2$ enak:

$$\begin{cases} \Bigl[0, \frac{1}{2}\bigl(1-\sqrt{1-2\beta}\bigr)\Bigr) & ; S=0 \\ \bigl(1-\sqrt{\beta}, \sqrt{\beta}\bigr) & ; S=1 \\ \Bigl(\frac{1}{2}\bigl(1+\sqrt{1-2\beta}\bigr), 1\Bigr] & ; S=2 \end{cases}.$$

Za $\beta \le 4/9$ pa gremo lahko v dve smeri: lahko določimo $b_0 = \frac{1}{2}\bigl(1 - \sqrt{1-2\beta}\bigr)$ in $b_1 = 1 - b_0$ ali pa $b_1 = \sqrt{\beta}$ in $b_0 = 1 - b_1$. To nas pripelje do naslednjih družin intervalov zaupanja:

$$\begin{cases} [0, 1-b_1) & ; S=0 \\ [1-b_1, b_1] & ; S=1 \\ (b_1, 1] & ; S=2 \end{cases}, \quad \begin{cases} [0, 1-b_1] & ; S=0 \\ (1-b_1, b_1) & ; S=1 \\ [b_1, 1] & ; S=2 \end{cases} \quad \text{in} \quad \begin{cases} \bigl[0, \frac{1}{2}\bigr) & ; S=0 \\ \bigl\{\frac{1}{2}\bigr\} & ; S=1 \\ \bigl(\frac{1}{2}, 1\bigr] & ; S=2 \end{cases}$$

Vsak interval zaupanja iz te družine je minimalen, pogoje pa izpolnjuje, brž ko je njegova verjetnost pokritja vsaj β: za prvi dve družini to pomeni $b_1^2 \le \beta$ in $2b_1(1-b_1) \ge \beta$ (rešitev tega sistema neenačb na b_1 je prepuščena bralcu), tretji interval zaupanja pa izpolnjuje pogoje, brž ko je $\beta \le 1/4$. Pri $\beta < 4/9$ torej dobimo bistveno različne minimalne intervale zaupanja.

2. Označimo dano statistično spremenljivko z X. Najprej opazimo, da je X pri $\theta = 1$ z verjetnostjo 1 enak 0, kar pomeni, da mora interval zaupanja pri opažanju $X = 0$ obvezno vsebovati tudi 1. Zaradi predpisane oblike mora biti torej to kar interval $[0, 1]$. Preostala dva parametra, b_1 in b_2, lahko nastavimo strogo manjša od 1.

Če nastavimo $b_1 > b_2$, velja:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ P_\theta(X = 0) + P_\theta(X = 1) & ; b_2 \le \theta < b_1 \\ P_\theta(X = 0) & ; b_1 \le \theta \le 1 \end{cases}$$

oziroma:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ \dfrac{\theta}{\theta^2 - \theta + 1} & ; b_2 \le \theta < b_1 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & ; b_1 \le \theta \le 1 \end{cases} .$$

Ni se težko prepričati, da sta funkciji $\theta \mapsto \frac{\theta}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0, 1]$ obe naraščajoči, torej je:

$$\inf_{0\le\theta\le1} P_\theta(\theta \in I) = \min\left\{\frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_1^2 - b_1 + 1}\right\} .$$

Interval bo minimalen, če bo kar:

$$\frac{b_2}{b_2^2 - b_2 + 1} = \frac{b_1^2}{b_1^2 - b_1 + 1} = \beta .$$

Ker sta funkciji $\theta \mapsto \frac{\theta}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0, 1]$ obe naraščajoči in ker za $\theta \in (0, 1)$ velja $\frac{\theta}{\theta^2-\theta+1} > \frac{\theta^2}{\theta^2-\theta+1}$, je pri zgornji izbiri tudi $b_1 > b_2$. Velja:

$$b_1 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)} , \quad b_2 = \frac{1 + \beta - \sqrt{(1 + 3\beta)(1 - \beta)}}{2\beta} .$$

Pri $\beta = 0{\cdot}95$ dobimo $b_1 \doteq 0{\cdot}9523$ in $b_2 \doteq 0{\cdot}7954$ (obakrat smo zaokrožili navzgor). Ni pa to edina možnost. Če postavimo $b_1 < b_2$, dobimo:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_2 \\ P_\theta(X = 0) + P_\theta(X = 2) & ; b_1 \le \theta < b_2 \\ P_\theta(X = 0) & ; b_2 \le \theta \le 1 \end{cases}$$

oziroma:

$$P_\theta(\theta \in I) = \begin{cases} 1 & ; \theta < b_1 \\ \dfrac{2\theta^2 - 2\theta + 1}{\theta^2 - \theta + 1} & ; b_1 \le \theta < b_2 \\ \dfrac{\theta^2}{\theta^2 - \theta + 1} & ; b_2 \le \theta \le 1 \end{cases} .$$

Kot smo že omenili, je funkcija $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $[0, 1]$ naraščajoča. Funkcija $\theta \mapsto \frac{2\theta^2-2\theta+1}{\theta^2-\theta+1}$ pa je na intervalu $\left[0, \frac{1}{2}\right]$ padajoča, na intervalu $\left[\frac{1}{2}, 1\right]$ pa naraščajoča; v $\frac{1}{2}$ ima minimum $\frac{2}{3}$. Zato mora biti $b_1 \geq \frac{1}{2}$, brž ko je $\beta \geq \frac{1}{2}$. V tem primeru velja:

$$\inf_{0\leq\theta\leq 1} P_\theta(\theta \in I) = \min\left\{\frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_2^2 - b_2 + 1}\right\}.$$

Interval bo minimalen, če bo kar:

$$\frac{2b_1^2 - 2b_1 + 1}{b_1^2 - b_1 + 1} = \frac{b_2^2}{b_2^2 - b_2 + 1} = \beta\,.$$

Spet ker sta funkciji $\theta \mapsto \frac{2\theta^2-2\theta+1}{\theta^2-\theta+1}$ in $\theta \mapsto \frac{\theta^2}{\theta^2-\theta+1}$ na intervalu $\left[\frac{1}{2}, 1\right]$ obe naraščajoči in ker za $\theta \in \left(\frac{1}{2}, 1\right)$ velja $\frac{2\theta^2-2\theta+1}{\theta^2-\theta+1} > \frac{\theta^2}{\theta^2-\theta+1}$, je pri zgornji izbiri tudi $b_1 < b_2$. Velja:

$$b_1 = \frac{2 - \beta + \sqrt{(3\beta - 2)(2 - \beta)}}{2(2 - \beta)}, \quad b_2 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)}.$$

Pri $\beta = 0˙95$ dobimo $b_1 \doteq 0˙9499$ in $b_2 \doteq 0˙9523$ (obakrat smo zaokrožili navzgor). Ta izbira je torej slabša glede na seštevek dolžin pri vseh opazovanjih. Spet je podobno kot prej dovolj, da je velja le ena enakost, drugi parameter pa se lahko poveča (a potem dobimo še daljši interval zaupanja za enako stopnjo zaupanja).

Za $b_1 = b_2$ pa bi moralo veljati:

$$b_1 = b_2 = \frac{-\beta + \sqrt{\beta(5\beta - 4)}}{2(1 - \beta)}$$

(torej $0˙9523$ pri $\beta = 0˙95$), kar ni minimalno, saj ta interval strogo vsebuje katero koli izmed prejšnjih dveh možnosti.

3. Očitno mora biti $b_0 > 0$, sicer dobimo poljubno majhno ali celo ničelno verjetnost pokritosti. Poleg tega mora biti interval zaupanja pri opažanju $S = n$ kar $[0, 1]$, sicer pri $\theta = 1$ nimamo pokritosti. Za $k = 1, 2, \ldots$ pa lahko pri $S = k$ nastavimo kar $[0, b_k)$, saj za zaprto različico dobimo enako minimalno verjetnost pokritosti.

Če fiksiramo $b_k < \theta < b_{k+1}$, smo torej v intervalu zaupanja natanko tedaj, ko je $S > k$. Pri vseh $\theta \in (b_k, b_{k+1})$ mora biti torej $P(S > k) \geq \beta$, kar drži natanko tedaj, ko je to res pri $\theta = b_k$. Minimalnost dobimo natanko tedaj, ko za b_k postavimo tisti θ, pri katerem je $P(S > k) = \beta$.

Zadevo pa si lahko predstavljamo tudi v duhu 23. naloge iz 5. razdelka (glej tudi 2. nalogo iz 11. razdelka): naj torej n gostov pride na zabavo, vsak z zamudo, porazdeljeno enakomerno na $[0, 1]$, neodvisno od drugih gostov. Če z S_θ označimo število gostov, ki pridejo na zabavo do časa θ, je $S_\theta \sim \mathrm{Bin}(n, \theta)$. V tem duhu mora torej veljati $P(S_{b_k} > k) = \beta$. Toda če s T_k označimo prihod gosta, ki pride k-ti po vrsti, je $\{S_\theta > k\} = \{T_{k+1} \leq \theta\}$, torej mora veljati $P(T_{k+1} \leq b_k) = \beta$. Zgornja meja b_k mora biti torej natanko kvantil porazdelitve $\mathrm{Beta}(k + 1, n - k)$ za verjetnost β.

Podobno dobimo tudi nasprotni enostranski interval zaupanja, t. j. $(a_k, 1]$ pri $S = k$, $k = 1, 2, \ldots n$. Lahko postavimo $a_k = 1 - b_{n-k}$, kar je kvantil porazdelitve $\mathrm{Beta}(k, n - k + 1)$ za verjetnost $1 - \beta$.

4. Dovolj je izračunati zgornje krajišče, in sicer le za primer, ko ne uspe noben poskus ali pa uspe en poskus. Če ne uspe noben poskus, je zgornje krajišče pri Clopper–Pearsonovem intervalu kvantil porazdelitve $\mathrm{Beta}(2,1)$ za verjetnost $(1+\beta)/2$. Ta porazdelitev ima gostoto $p(t) = 2(1-t)$ in kumulativno porazdelitveno funkcijo $F(t) = 1-(1-t)^2$, kar nam da krajišče $1-\sqrt{(1-\beta)/2}$. Če pa uspe natanko en poskus, dobimo kvantil porazdelitve $\mathrm{Beta}(2,1)$; le-ta ima gostoto $p(t) = 2t$ in kumulativno porazdelitveno funkcijo $F(t) = t^2$, kar nam da krajišče $\sqrt{(1+\beta)/2}$. Popoln zapis je torej:

$$\begin{cases} \left[0, 1-\sqrt{\frac{1-\beta}{2}}\right) & ; S=0 \\ \left(1-\sqrt{\frac{1+\beta}{2}}, \sqrt{\frac{1+\beta}{2}}\right) & ; S=1 \\ \left(\sqrt{\frac{1-\beta}{2}}, 1\right] & ; S=2 \end{cases},$$

kar je pri vseh izidih strogo več kot pri intervalu zaupanja iz 1. naloge.

5. Za mejo $a_{\max}$ lahko brez škode za splošnost postavimo funkcijo minimalne zadostne statistike. Iz gostote porazdelitve je razvidno, da gre za eksponentno družino z minimalno zadostno statistiko $S_n := X_1 + \cdots + X_n$. Iz centralnega limitnega izreka pa sledi, da je le-ta porazdeljena približno normalno $\mathrm{N}\big(n\mu, \sqrt{n}\,\sigma\big)$, kjer je:

$$\mu = E(X_i) = a\,, \qquad \sigma = \sqrt{D(X_i)} = a\sqrt{3}/2\,.$$

Postavimo torej $a_{\max} := h(S_n)$ in poskusimo, ali konstrukcija intervala zaupanja deluje, če je h strogo naraščajoča funkcija. Tedaj je namreč:

$$P\big(a \in [0, a_{\max}]\big) = P\big(a \le h(S_n)\big) = P\big(S_n \ge h^{-1}(a)\big) \approx \frac{1}{2} - \Phi\left(\frac{h^{-1}(a) - na}{a\sqrt{3n}/2}\right).$$

Desna stran bo enaka $\beta = 0{\cdot}95$, če bo:

$$\frac{h^{-1}(a) - na}{a\sqrt{3n}/2} = -z\,,$$

kjer je $z := \Phi^{-1}\big(\beta/2\big) \doteq 1{\cdot}645$. Sledi:

$$h^{-1}(a) = \left(n - z\frac{\sqrt{3n}}{2}\right) a$$

oziroma:

$$h(s) = \frac{s}{n - z\sqrt{3n}/2}\,.$$

Za velike n je to dobro definirana strogo naraščajoča funkcija, torej lahko postavimo:

$$a_{\max} = \frac{S_n}{n - z\sqrt{3n}/2}\,.$$

Za funkcijo h pa ne bi mogli vzeti strogo padajoče funkcije. V tem primeru bi namreč veljalo:

$$P\big(a \in [0, a_{\max}]\big) = P\big(a \le h(S_n)\big) = P\big(S_n \le h^{-1}(a)\big) \approx \frac{1}{2} + \Phi\left(\frac{h^{-1}(a) - na}{a\sqrt{3n}/2}\right)$$

in desna stran bi bila enaka β za

$$\frac{h^{-1}(a) - na}{a\sqrt{3n}/2} = z\,,$$

od koder bi sledilo $h^{-1}(a) = \big(n + z\sqrt{3n}/2\big)a$ oziroma $h(s) = s/\big(n + z\sqrt{3n}/2\big)$. Ta funkcija pa je spet strogo naraščajoča, kar je v protislovju s prvotno zahtevo.

6. $\bar{X} = 97$, $\Delta \doteq 3{\cdot}27$ (zaokroženo navzgor),
$93{\cdot}73 < \mu < 100{\cdot}27$.

7. $s = 5$, $t_{0.975}(8) \doteq 2{\cdot}31$, $\Delta \doteq 3{\cdot}85$ (zaokroženo navzgor),
$93{\cdot}15 < \mu < 100{\cdot}85$.

8. $\bar{X} \doteq 1{\cdot}55$, $s_0 \doteq s \doteq 1{\cdot}30$. Interval zaupanja za μ: $1{\cdot}46 < \mu < 1{\cdot}64$.

9. Definirajmo slučajne spremenljivke $X_1, \ldots X_n$, in sicer naj bo $X_i = 1$, če je i-ti poskus uspel, sicer pa 0. Tedaj je $\bar{X} = S/n =: \hat{\theta}$. Nadalje velja:

$$s_0^2 = \frac{S(1-\hat{\theta})^2 + (n-S)\hat{\theta}^2}{n} = \hat{\theta}(1-\hat{\theta})\,.$$

Dobimo *Waldov interval zaupanja*:

$$\hat{\theta} - c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\,,$$

kjer je $c = z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$.

Je lahko izračunljiv, a ne dosega nominalne verjetnosti pokritosti niti v povprečju. Dejanska verjetnost pokritosti je često daleč pod nominalno. Res pa je, da se dejanska minimalna verjetnost pokritosti, ko θ preteče interval $[a, b]$ $(0 < a \le b < 1)$, bliža nominalni β, ko gre n proti neskončno.

10.

Clopper–Pearsonov interval:	$(0{\cdot}0713, 0{\cdot}4011)$.
Waldov interval:	$(0{\cdot}0528, 0{\cdot}3472)$.
Wilsonov interval:	$(0{\cdot}0806, 0{\cdot}4161)$.
Agresti–Coullov interval:	$(0{\cdot}0749, 0{\cdot}4218)$.
Wilsonov interval s popravkom za zveznost:	$(0{\cdot}0661, 0{\cdot}4427)$.
Agresti–Coullov interval s popravkom za zveznost:	$(0{\cdot}0499, 0{\cdot}4468)$.

Spodnja krajišča so zaokrožena navzdol, zgornja pa navzgor.

11. Clopper–Pearsonov interval: $0{.}1266, 0{.}2919$.
Waldov interval: $(0{.}1216, 0{.}2784)$.
Wilsonov interval: $(0{.}1333, 0{.}2889)$.
Agresti–Coullov interval: $(0{.}1326, 0{.}2896)$.
Wilsonov interval s popravkom za zveznost: $(0{.}1292, 0{.}2944)$.
Agresti–Coullov interval s popravkom za zveznost: $(0{.}1276, 0{.}2946)$.
Spodnja krajišča so zaokrožena navzdol, zgornja pa navzgor.

12. Clopper–Pearsonov interval: $(0{.}496525849771,\ 0{.}5130720831478)$.
Waldov interval: $(0{.}49657611084,\ 0{.}51302388916)$.
Wilsonov interval: $(0{.}496575924714,\ 0{.}513021478666)$.
Agresti–Coullov interval: $(0{.}496575924612,\ 0{.}513021478769)$.
Wilsonov interval s popravkom za zveznost: $(0{.}496525930387,\ 0{.}513071457209)$.
Agresti–Coullov interval s pop. za zveznost: $(0{.}496525924612,\ 0{.}513071478769)$.
Spodnja krajišča so zaokrožena navzdol, zgornja pa navzgor.

Agresti–Coullov interval je sicer nasploh odličen približek Wilsonovega, a tu je razlika še posebej majhna, ker je opaženi delež grbov blizu 1/2. Clopper–Pearsonov interval je potrebno določiti z računalnikom.

13. $\chi^2_{0.025}(8) \doteq 2{.}180$, $\chi^2_{0.975}(8) \doteq 17{.}53$,
$3{.}37 < \sigma < 9{.}58$ (spodnja meja zaokrožena navzdol, zgornja pa navzgor).

14. $\bar{X} = 125$, $s \doteq 2{.}646$, $\chi^2_{0.05}(4) \doteq 0{.}711$, $\chi^2_{0.95}(4) \doteq 9{.}49$, $1{.}72 < \sigma < 6{.}28$.

15. $\bar{X} \doteq 45{.}51$, $s \doteq 3{.}71$,
$t_{0.995}(74) \doteq 2{.}64$, $\Delta \doteq 1{.}14$ (zaokroženo navzgor), $44{.}37 < \mu < 46{.}65$,
$\chi^2_{0.005}(74) \doteq 46{.}4$, $\chi^2_{0.995}(74) \doteq 109{.}1$, $3{.}05 < \sigma < 4{.}69$.
Pri obeh intervalih zaupanja je spodnja meja zaokrožena navzdol, zgornja pa navzgor.

16. Če s κ^4 ($\kappa \ge 0$) označimo četrti centralni moment spremenljivke X, po centralnem limitnem izreku približno velja:

$$\sum_{i=1}^{n} (X_i - \mu)^2 \sim \mathrm{N}\big(n\sigma^2, \sqrt{n(\kappa^4 - \sigma^4)}\big)$$

Iz Jensenove neenakosti sledi $\kappa \ge \sigma$, pri čemer enakost velja natanko tedaj, ko je spremenljivka X bodisi skoncentrirana v eni točki bodisi skoncentrirana v dveh točkah, pri čemer ima v vsaki verjetnost 1/2. Privzemimo najprej, da to ni res, torej da je $\kappa > \sigma$. V tem primeru lahko zapišemo:

$$\frac{1}{\sqrt{n(\kappa^4 - \sigma^4)}} \left[\sum_{i=1}^{n} (X_i - \mu)^2 - n\sigma^2 \right] \xrightarrow[n\to\infty]{\mathrm{d}} \mathrm{N}(0,1)\,.$$

V nadaljevanju moramo iz spremenljivke na levi odstraniti vse neopazljive parametre razen σ, pri tem pa moramo paziti, da se ohrani asimptotična normalnost. Najprej velja:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\bar{X})^2-\sqrt{n}(\bar{X}-\mu)^2\,.$$

Ker gre $D\big[\sqrt[4]{n}(\bar{X}-\mu)\big]=\sigma^2/\sqrt{n}\xrightarrow[n\to\infty]{}0$, gre po neenačbi Čebiševa tudi $\sqrt[4]{n}(\bar{X}-\mu)\xrightarrow[n\to\infty]{P}0$ in zato tudi $\sqrt{n}(\bar{X}-\mu)^2\xrightarrow[n\to\infty]{P}0$. Po izreku Sluckega potem velja tudi:

$$\frac{1}{\sqrt{n(\kappa^4-\sigma^4)}}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2-n\sigma^2\right]\xrightarrow[n\to\infty]{\mathrm{d}}\mathrm{N}(0,1)\,.$$

Naj bosta $\hat{\sigma}$ in $\hat{\kappa}$ cenilki za κ in σ po metodi momentov, t. j.:

$$\hat{\sigma}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}\,,\quad \hat{\kappa}=\sqrt[4]{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^4}\,.$$

Ker so cenilke po metodi momentov dosledne, gre $\hat{\sigma}\xrightarrow[n\to\infty]{P}\sigma$ in $\hat{\kappa}\xrightarrow[n\to\infty]{P}\kappa$. Iz izreka Sluckega za deljenje sedaj sledi:

$$\frac{1}{\sqrt{n(\hat{\kappa}^4-\hat{\sigma}^4)}}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2-n\sigma^2\right]\xrightarrow[n\to\infty]{\mathrm{d}}\mathrm{N}(0,1)$$

oziroma:

$$\frac{\sqrt{n}}{\sqrt{\hat{\kappa}^4-\hat{\sigma}^4}}(\hat{\sigma}^2-\sigma^2)\xrightarrow[n\to\infty]{\mathrm{d}}\mathrm{N}(0,1)\,.$$

Seveda se lahko zgodi, da pride do deljenja z nič, a asimptotično prihaja do tega vse redkeje. V primeru deljenja z nič lahko rezultat poljubno nastavimo (lahko tudi na vrednost "nedefinirano") in bo konvergenca v porazdelitvi še vedno veljala.

Sedaj si oglejmo še primer, ko je $\kappa=\sigma$, torej ko je X bodisi skoncentrirana v eni točki bodisi v dveh točkah z enakima verjetnostma. Če je X skoncentrirana v eni točki, vemo, da je $\sigma=0$, velja pa tudi $\hat{\sigma}=\hat{\kappa}=0$. Torej zgornja konstrukcija še vedno velja, če pri neenačajih vključimo še enakost. Preostane še primer, ko je X skoncentrirana v dveh enako verjetnih točkah. Zaradi invariantnosti leve strani za translacije in raztege lahko brez škode za splošnost privzamemo, da je $X\sim\mathrm{Ber}(1/2)$. Če spet s $\hat{\theta}$ označimo delež enic, velja:

$$\begin{aligned}
\hat{\sigma}^2&=\hat{\theta}(1-\hat{\theta})\,,\\
\hat{\kappa}^4&=\hat{\theta}(1-\hat{\theta})(1-3\hat{\theta}+3\hat{\theta}^2)\,,\\
\hat{\kappa}^4-\hat{\sigma}^4&=\hat{\theta}(1-\hat{\theta})(1-2\hat{\theta}^2)\,,\\
\frac{\sqrt{n}}{\sqrt{\hat{\kappa}^4-\hat{\sigma}^4}}(\hat{\sigma}^2-\sigma^2)&=\sqrt{n}\,\frac{\hat{\theta}(1-\hat{\theta})-\frac{1}{4}}{\sqrt{\hat{\theta}(1-\hat{\theta})(1-2\hat{\theta})^2}}=\sqrt{n}\,\frac{2\hat{\theta}-1}{4\sqrt{\hat{\theta}(1-\hat{\theta})}}\,.
\end{aligned}$$

Po Laplaceovi integralski formuli gre $\sqrt{n}\,(2\hat{\theta} - 1) \xrightarrow[n\to\infty]{P} \mathrm{N}(0,1)$. Ker gre $\hat{\theta} \xrightarrow[n\to\infty]{P} \frac{1}{2}$, spet iz izreka Sluckega za deljenje dobimo:

$$\frac{\sqrt{n}}{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}(\hat{\sigma}^2 - \sigma^2) \xrightarrow[n\to\infty]{P} \mathrm{N}\big(0, \tfrac{1}{2}\big)\,,$$

od koder dobimo naslednjo konstrukcijo intervala zaupanja:

$$\hat{\sigma}^2 - c\,\frac{\hat{\kappa}^4 - \hat{\sigma}^4}{2\sqrt{n}} < \sigma^2 < \hat{\sigma}^2 + c\,\frac{\hat{\kappa}^4 - \hat{\sigma}^4}{2\sqrt{n}}\,,$$

kar pomeni, da prvotna konstrukcija za predpisano stopnjo zaupanja še vedno velja, pravkar napisana pa tudi, a je natančnejša. Toda to pomeni le, da nam pri konstrukciji intervalov zaupanja za standardni odklon ni treba skrbeti za primer, ko je $\kappa = \sigma$: če to vemo, lahko preprosto zatrdimo, da je kar $\sigma = (\max_i X_i - \min_i X_i)/2$. Če bo dobljena vrednost večja od nič, bo ta izjava z verjetnostjo ena pravilna, sicer pa bo pravilna z verjetnostjo β, brž ko bo vzorec dovolj velik.

17. $\bar{X} \doteq 1\text{˙}55$, $\quad \hat{\sigma} \doteq 1\text{˙}30$, $\quad \hat{\kappa} \doteq 2\text{˙}09$.
Interval zaupanja za σ: $1\text{˙}19 < \sigma < 1\text{˙}41$.

Opomba. Zanimiva je tudi ocena za sploščenost (kurtozis):

$$\frac{\hat{\kappa}^4}{\hat{\sigma}^4} - 3 \doteq 3\text{˙}56\,,$$

kar namiguje na to, da je dejanska porazdelitev daleč od normalne. Vendar pa največji relativni delež v κ^4 prispeva edina ženska z desetimi otroki. Le-ta nam torej znatno poveča širino intervala zaupanja.

13. Testi značilnosti

1. Če z D označimo število dobitnih srečk med kupljenimi, je $D \sim \mathrm{Bin}(n, \theta)$, kjer je $n = 8$, θ pa delež dobitnih med vsemi srečkami v seriji. Verjetnostna funkcija je torej enaka:

$$P(D = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} ,$$

kar nam da verjetje $L = \binom{n}{D} p^D (1 - \theta)^{n-D}$. Ničelna hipoteza trdi, da je $\theta \geq 1/2$, alternativna pa, da je $\theta < 1/2$. Torej je $\Theta = [0, 1]$ in $\Theta_{H_0} = [0, 1/2]$. Po krajšem računu dobimo:

$$\sup_{\theta \in \Theta} L = \binom{n}{D} \left(\frac{D}{n}\right)^D \left(1 - \frac{D}{n}\right)^{n-D}$$

Za $D \geq n/2$ velja tudi:

$$\sup_{\theta \in \Theta_{H_0}} L = \binom{n}{D} \left(\frac{D}{n}\right)^D \left(1 - \frac{D}{n}\right)^{n-D} ,$$

medtem ko za $D \leq n/2$ velja:

$$\sup_{\theta \in \Theta_{H_0}} L = \binom{n}{D} \left(\frac{1}{2}\right)^n .$$

Torej velja:

$$\Lambda = \begin{cases} 2^{-n} \left(\frac{D}{n}\right)^{-D} \left(1 - \frac{D}{n}\right)^{-(n-D)} & ; D \leq n/2 \\ 1 & ; D \geq n/2 \end{cases} .$$

Za $D \leq n/2$ velja $\ln \Lambda = g(D/n)$, kjer je $g(x) = -n\big(\ln 2 + x \ln x + (1-x)\ln(1-x)\big)$. Z odvajanjem dobimo, da je funkcija g strogo naraščajoča na $(0, 1/2]$. Torej je vzorec toliko ustreznejši, kolikor večji je D. Naj bo D' kopija statistike D. Ker je:

$$\sup_{\theta \in \Theta_{H_0}} P(D' \leq 2) \doteq 0{\cdot}145 , \qquad \sup_{\theta \in \Theta_{H_0}} P(D' < 2) \doteq 0{\cdot}035 ,$$

brez randomizacije hipoteze ne moremo zavrniti. Sicer pa jo zavrnemo z verjetnostjo 0˙135 (zaokrožili smo jo navzdol).

2. Označimo s θ verjetnost po modelu, da pade grb, z S pa število opaženih grbov (za naš konkretni primer bo torej $S = 5$, a za konstrukcijo testa potrebujemo slučajno spremenljivko). Označimo še $n = 20$. Iz:

$$P_\theta(S = k) = \binom{n}{k} \theta^k (1 - \theta)^{n-k} ,$$

$$P_{1/2}(S = k) = \binom{n}{k} 2^{-n} ,$$

$$\sup_{0 \leq \theta \leq 1} P_\theta(S = k) = \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k}$$

dobimo razmerje verjetij:
$$\Lambda = \frac{n^n}{2^n\, S^S\, (n-S)^{n-S}}\,.$$
Velja $\ln\Lambda = g(S)$, kjer je $g(s) = n\ln\frac{n}{2} - s\ln s - (n-s)\ln(n-s)$. Z odvajanjem dobimo, da je g strogo naraščajoča na $(0, n/2]$. Brž ko je torej $S \le n/2$, je opažanje toliko ustreznejše za ničelno hipotezo, kolikor večji je S. Toda ker se razmerje verjetij ohrani, če S zamenjamo z $n - S$, velja, da je opažanje toliko ustreznejše, kolikor bližje je S številu $n/2$. Ničelno hipotezo bomo torej zavrnili, če bo vrednost $\left|S - \frac{n}{2}\right|$ prevelika.

Naj bo S' kopija statistike S. Za naše opažanje dobimo:

$$P_{1/2}(S' \le 5 \text{ ali } S' \ge 15) \doteq 0\text{'}041\,,$$
$$P_{1/2}(S' < 5 \text{ ali } S' > 15) \doteq 0\text{'}012\,.$$

Pri $\alpha = 0\text{'}05$ bomo torej našo hipotezo zavrnili, pri $\alpha = 0\text{'}01$ pa ne. Do randomizacije tukaj ne pride.

3. Ničelno hipotezo lahko interpretiramo kot to, da je verjetnost, da je posameznik, ki dobi nagrado, ženska z verjetnostjo $\theta_0 = 0\text{'}7$. Pri alternativni hipotezi pa je posameznik, ki dobi nagrado, ženska z verjetnostjo $\theta \ne 0\text{'}7$.

V splošnem je torej verjetnost, da nagrado dobi k žensk in $n - k$ moških, enaka:

$$f_n(k \mid \theta) = \binom{n}{k}\theta^k(1-\theta)^{n-k}\,.$$

Če velja ničelna hipoteza, v ta izraz samo vstavimo $\theta = \theta_0$. V splošnem pa to zoptimiziramo po θ. Velja:

$$\frac{\partial \ln f(k \mid \theta)}{\partial\theta} = \frac{k}{\theta} - \frac{n-k}{1-\theta}$$

in maksimum je dosežen pri $\theta = k/n$. Če n fiksiramo, število žensk, ki dobijo nagrado (doslej k), pa naredimo za slučajno spremenljivko S, je torej razmerje verjetij enako:

$$\Lambda = \left(\theta_0\,\frac{n}{S}\right)^S \left((1-\theta_0)\,\frac{n}{n-S}\right)^{n-S}\,.$$

Ničelno hipotezo bomo zavrnili, če bo to razmerje premajhno, natančneje, če bo verjetnost, da bo tako majhno ali še manjše, kot smo ga opazili, največ $0\text{'}05$. Za $n = 12$ in $\theta_0 = 0\text{'}7$ tabelirajmo:

S	Λ
0	$5{\cdot}314410 \cdot 10^{-7}$
1	$3{\cdot}875139 \cdot 10^{-5}$
2	$6{\cdot}449463 \cdot 10^{-4}$
3	0˙005754585
4	0˙03270222
5	0˙1273437
6	0˙3512980
7	0˙6933160
8	0˙9693585
9	0˙9286985
10	0˙5666800
11	0˙1853767
12	0˙01384129

Razmerje verjetij je manjše ali enako opaženemu pri $S \in \{0,1,2,3,4,5,12\}$. Če je torej S' porazdeljena tako kot S pri H_0, torej $S' \sim \mathrm{Bin}(12, 0{\cdot}7)$, je p-vrednost našega opažanja enaka:

$$p(5) = P\big(S' \in \{0,1,2,3,4,5,12\}\big) \doteq 0{\cdot}0524\,.$$

Ker je to večje od α, izračunamo še:

$$p^-(5) = P\big(S' \in \{0,1,2,3,4,12\}\big) \doteq 0{\cdot}0233\,.$$

Ker je to manjše od α, ničelno hipotezo zavrnemo z verjetnostjo 0˙916.

4. Izvedemo dvostranski test. Če je $S' \sim \mathrm{Bin}(12, 0{\cdot}7)$, je $P(S' \le 5) \doteq 0{\cdot}0386$ in $P(S' \ge 5) \doteq 0{\cdot}9905$, torej je $p(5) \doteq 0{\cdot}0772$. Hipoteze ne moremo zavrniti. Zavrnili pa bi jo, če bi za alternativno hipotezo postavili, da so izjemni študijski dosežki pristranski v korist moških.

Enostranska testa se ujemata s testom na podlagi razmerja verjetij. Če npr. testiramo ničelno hipotezo, da je $\theta = \theta_0$, proti alternativni, da je $\theta < \theta_0$, je za $S \le n\theta_0$ razmerje verjetij enako:

$$\Lambda = n^n \left(\frac{\theta_0}{S}\right)^S \left(\frac{1-\theta_0}{n-S}\right)^{n-S}\,.$$

Z odvajanjem po S dobimo, da je razmerje verjetij naraščajoča funkcija spremenljivke S, kar pomeni, da sta razmerje verjetij in S ekvivalentni testni statistiki.

Dvostranski test pa ni primerljiv s testom na podlagi razmerja verjetij. Lahko pa rečemo, da je pretirano konservativen. Za $k \le n\theta_0$ je namreč $\rho(k) = 2\,P(S' \le k)$, medtem ko je verjetnost značilnosti za opažanje $S = k$ glede na test, ki temelji na testni statistiki ρ, enaka $p(k) = P(S' \le k) + P(S' \ge l)$, kjer je l najmanjše naravno število, za katero je $P(S' \ge l) \le P(S' \le k)$. Seveda je $P(S' \ge l) \le P(S' \le k)$, a tipično $P(S' \ge l) < P(S' \le k)$. Torej je $p(k) \le \rho(k)$, a tipično $p(k) < \rho(k)$. Enako velja tudi za primer, ko je $k \ge n\theta_0$. Test, ki zavrne ničelno hipotezo na podlagi testne statistike ρ in verjetnosti značilnosti, še vedno ustreza predpisani stopnji značilnosti, a ničelno hipotezo zavrne večkrat. Njegova računska zahtevnost pa je primerljiva s testom na podlagi razmerja verjetij.

5. Označimo s θ verjetnost po modelu, da bo dobitek izžreban. Označimo $\theta_0 = 1/50$ in naj bo N število iger, po katerih je dobitek prvič izžreban. Tedaj velja $P_\theta(N = k) = \theta(1-\theta)^{k-1}$, torej je $L = \theta(1-\theta)^{N-1}$ in:

$$\sup_{\theta \le \theta_0} L = \begin{cases} \theta_0(1-\theta_0)^{N-1} & ;\ N \le 1/\theta_0 \\ \frac{1}{N}\left(1 - \frac{1}{N}\right)^{N-1} & ;\ N \ge 1/\theta_0 \end{cases} .$$

Razmerje verjetij je zato enako:

$$\Lambda = \begin{cases} 1 & ;\ N \le 1/\theta_0 \\ \theta_0 N \left(\dfrac{1-\theta_0}{1-\frac{1}{N}}\right)^{N-1} & ;\ N \ge 1/\theta_0 \end{cases} .$$

Za $N > 1/\theta_0$ velja $\ln\Lambda = g(N)$, kjer je $g(x) = \ln\theta_0 + \ln x + (x-1)\ln(1-\theta_0) - (x-1)\ln\left(1 - \frac{1}{x}\right)$. Z odvajanjem dobimo, da je g na intervalu $[1/\theta_0, \infty)$ strogo padajoča funkcija. Test bo torej potekal tako, da bomo ničelno hipotezo zavrnili, če bo testna statistika $\max\{\frac{1}{\theta_0}, N\}$ prevelika.

Poiskati moramo torej najmanjši tak k, za katerega bo $P_{\theta_0}\left(\max\left\{\frac{1}{\theta_0}, N\right\} > k\right) \le \alpha$. Velja $P_{\theta_0}(N > k) = (1-\theta_0)^k$. Ker je $P_{1/50}(\max\{50, N\} > 50) = P_{1/50}(N > 50) \doteq 0{\cdot}364$, bo to najmanjši k, za katerega bo $(49/50)^k < 0{\cdot}1$. Iz $\log_{49/50} 0{\cdot}1 \doteq 113{\cdot}97$ dobimo, da je ta k enak 114 – velja:

$$P_{1/50}\big(\max\{50, N\} > 114\big) = P_{1/50}(N > 114) \doteq 0{\cdot}09994766\,,$$
$$P_{1/50}\big(\max\{50, N\} \ge 114\big) = P_{1/50}(N \ge 114) \doteq 0{\cdot}1019874\,.$$

Test bo torej potekal tako, da bomo igro igrali, dokler ne bo izžreban dobitek, vendar pa bomo odigrali največ 114 iger. Če bo dobitek izžreban pred zadnjo, 114. igro, hipoteze ne bomo zavrnili. Če dobitek tudi po 114 igrah še ne bo izžreban, bomo hipotezo zavrnili. Če pa bo dobitek (prvič) izžreban v zadnji, 114. igri, bomo hipotezo zavrnili z verjetnostjo:

$$\frac{0{\cdot}1 - 0{\cdot}09994766}{0{\cdot}1019874 - 0{\cdot}09994766} \doteq 0{\cdot}025$$

(zaokrožili smo navzdol).

6. Glede na to, da sta ničelna in alternativna hipoteza obe enostavni, je test, ki temelji na razmerju verjetij, najmočnejši. Če z L_0 označimo verjetje pri ničelni, z L_1 pa pri alternativni hipotezi, velja:

$$\frac{L_0}{L_1} = \frac{\frac{1}{500}\,e^{-X/500}}{\frac{1}{100}\,e^{-X/100}} = \frac{1}{5}\,e^{4X/500}\,,$$

kjer je X opažena življenjska doba žarnice. Ničelno hipotezo bomo torej zavrnili, če bo testna statistika $\frac{1}{5}\,e^{4X/500}$ premajhna, to pa bo tedaj, ko bo življenjska doba X premajhna. Ker je X pri ničelni hipotezi porazdeljena zvezno, lahko H_0 zavrnemo, brž ko je $X \le c$, kjer je $P_{H_0}(X \le c) = \alpha$. Iz:

$$P_{H_0}(X \le c) = \int_0^c \frac{1}{500}\,e^{-x/500}\,\mathrm{d}x = 1 - e^{-c/500}$$

dobimo $c = -500\ln(0\text{'}95) \doteq 25\text{'}64$ (zaokroženo navzdol). Moč tega testa je:

$$P_{H_1}(X \le c) = \int_0^c \frac{1}{100}\,e^{-x/100}\,\mathrm{d}x = 1 - e^{-c/100} \doteq 0\text{'}226\,.$$

7. a) Tinetovo povprečje je porazdeljeno normalno $\mathrm{N}(0, 1/10)$, torej ima gostoto:

$$p_{\mathrm{Ti}}(x) = \frac{10}{\sqrt{2\pi}}\,e^{-50x^2}$$

Tonetovo povprečje pa je porazdeljeno normalno $\mathrm{N}(0, 1/100)$, torej ima gostoto:

$$p_{\mathrm{To}}(x) = \frac{100}{\sqrt{2\pi}}\,e^{-5000x^2}\,.$$

Ker sta ničelna in alternativna hipoteza obe enostavni, je test, ki temelji na razmerju verjetij, najmočnejši. Če torej z X označimo opaženo povprečje, ničelno hipotezo torej zavrnemo, če je razmerje:

$$\frac{p_{\mathrm{To}}(X)}{p_{\mathrm{Ti}}(X)} = 10\,e^{-4950X^2}$$

premajhno, to pa je takrat, ko je vrednost $|X|$ prevelika. Če je P_{To} verjetnost pri Tonetovem povprečju, za $c \ge 0$ velja:

$$P_{\mathrm{To}}(|X| \ge c) = 1 - 2\Phi(100c)$$

Vrednost c moramo nastaviti tako, da bo to enako α, torej:

$$c = \frac{\Phi^{-1}\left(\frac{1-\alpha}{2}\right)}{100} \doteq 0\text{'}0196\,.$$

Moč testa pa je:

$$P_{\mathrm{Ti}}(|X| \ge c) = 1 - 2\Phi(10c) \doteq 0\text{'}845\,.$$

8. Iz zapisa gostote:

$$p_{X_1,\dots X_n}(x_1, \dots x_n \mid \mu, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right]\,.$$

je očitno, da gre za eksponentno družino, in ni težko preveriti pogojev zgornje trditve (za karakteristiko lahko postavimo npr. $\mu - \sigma^2$ ali μ/σ^2). Maksimum verjetja je pri polnem modelu dosežen pri:

$$\mu = \hat{\mu}_1 := \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad \sigma^2 = \hat{\sigma}_1^2 := \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 .$$

Če se omejimo na ničelno hipotezo $\mu = \sigma^2$, pa je:

$$L = p_{X_1,\dots X_n}(X_1, \dots X_n \mid \sigma^2, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} X_i - \frac{n\sigma^2}{2}\right]$$

in iz:

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}\ln L = \frac{1}{\sigma^3}\sum_{i=1}^{n} x_i^2 - n\sigma - \frac{n}{\sigma}$$

dobimo, da je maksimum dosežen pri:

$$\sigma^2 = \hat{\sigma}_2^2 := \hat{\mu}_2 := \frac{1}{2}\left[\sqrt{1 + \frac{4}{n}\sum_{i=1}^{n} X_i^2} - 1\right] .$$

Če z Λ označimo razmerje verjetij:

$$\Lambda = \frac{\sup_{\Theta_{H_0}} L}{\sup_{\Theta} L} = \frac{p_{X_1\dots X_n}(X_1, \dots X_n \mid \hat{\mu}_2, \hat{\sigma}_2)}{p_{X_1,\dots X_n}(X_1, \dots X_n \mid \hat{\mu}_1, \hat{\sigma}_1)}\,,$$

ima $-2\ln\Lambda$ približno porazdelitev hi kvadrat z eno prostostno stopnjo. Za naše opažanje dobimo $-2\ln\Lambda \doteq 5{\cdot}61$, kritična vrednost pa je $\chi^2_{0.99} \doteq 6{\cdot}63$. Hipoteze torej ne moremo zavrniti.

9. $n = 100, S = 24, \alpha = 0{\cdot}05$: $0{\cdot}875 < 1{\cdot}645$, hipoteze ne moremo zavrniti.
$n = 500, S = 120, \alpha = 0{\cdot}05$: $2{\cdot}18 \geq 1{\cdot}645$, hipotezo zavrnemo.
$n = 500, S = 120, \alpha = 0{\cdot}01$: $2{\cdot}18 < 2{\cdot}33$, hipoteze ne moremo zavrniti.

10. $n = 10000$, $S = 5090$, $H_0\colon\ \theta = 0{\cdot}5 : H_1\colon\ \theta \neq 0{\cdot}5$: $1{\cdot}79 < 1{\cdot}96$, hipoteze ne moremo zavrniti.
$n = 10000$, $S = 5090$, $H_0\colon\ \theta = 0{\cdot}5 : H_1\colon\ \theta > 0{\cdot}5$: $1{\cdot}79 > 1{\cdot}645$, hipotezo zavrnemo.

11. $\bar{X} = 97$, $Z = -1{\cdot}8$.
Pri alternativni hipotezi $\mu \neq 100$ upoštevamo $z_{0.975} \doteq 1{\cdot}96$ in hipoteze ne moremo zavrniti.
Pri alternativni hipotezi $\mu < 100$ upoštevamo $z_{0.95} \doteq 1{\cdot}645$ in hipotezo zavrnemo.
Pri alternativni hipotezi $\mu > 100$ tudi upoštevamo $z_{0.95} \doteq 1{\cdot}645$, a hipoteze ne zavrnemo.

12. $\bar{X} = 107, \quad s = 10, \quad T = 2\text{˙}1, \quad t_{0.975}(8) \doteq 2\text{˙}31.$
Hipoteze ne moremo zavrniti.
Če bi vedeli, da je $\sigma = 10$, pa bi upoštevali $z_{0.975} \doteq 1\text{˙}96$ in hipotezo bi zavrnili.

13. $\bar{X} = 48, \quad s \doteq 2\text{˙}58, \quad T = -2\text{˙}45, \quad df = 9.$
Če za H_1 vzamemo, da je $\mu \neq 50$, upoštevamo $t_{0.975}(9) \doteq 2\text{˙}26$ in hipotezo zavrnemo.
Če za H_1 vzamemo, da je $\mu < 50$, pa upoštevamo $t_{0.95}(9) \doteq 1\text{˙}83$ in hipotezo prav tako zavrnemo.

14. 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 $\bar{\Delta} \doteq -5\text{˙}6, \quad s \doteq 6\text{˙}433, \quad T \doteq -2\text{˙}75, \quad t_{0.95}(9) \doteq 1\text{˙}83.$
Ničelno hipotezo zavrnemo, torej sprejmemo hipotezo, da dieta deluje.

15. $\bar{X} = 20, \quad \bar{Y} = 22, \quad s = 2\text{˙}65, \quad T = -1\text{˙}5, \quad t_{0.975}(14) \doteq 2\text{˙}14.$
Hipoteze ne moremo zavrniti.

16. $\bar{X} = 100, \quad \bar{Y} = 96, \quad s = 2\text{˙}5, \quad T = 3\text{˙}2, \quad t_{0.99}(16) \doteq 2\text{˙}58.$
Hipotezo zavrnemo.

17. $\bar{X}_1 = 4, \quad \bar{X}_2 = 3, \quad \bar{X}_3 \doteq 1\text{˙}667,$
$S_{\mathrm{B}}^2 \doteq 9\text{˙}333, \quad S_{\mathrm{W}}^2 \doteq 8\text{˙}667, \quad F \doteq 4\text{˙}31, \quad F_{0.95}(2, 8) \doteq 4\text{˙}46.$
Hipoteze ne moremo zavrniti.

18. $s = 7\text{˙}45.$

a) $\chi^2 = 20, \quad \chi^2_{0.025}(9) \doteq 2\text{˙}70, \quad \chi^2_{0.975}(9) \doteq 19\text{˙}0,$ hipotezo zavrnemo.

b) $\chi^2 = 5, \quad \chi^2_{0.05}(9) \doteq 3\text{˙}33,$ hipoteze ne moremo zavrniti.

19. Pri frekvencah 21, 42, 77 in 116 je: $\chi^2 \doteq 81\text{˙}3 \quad \chi^2_{0.99}(3) \doteq 11\text{˙}3.$
Hipotezo zavrnemo.
Če so frekvence 21, 37, 53 in 25, pa je $\chi^2 \doteq 18\text{˙}2$ in hipotezo še vedno zavrnemo.

20. Če so frekvence 2, 5 in 4, je: $\chi^2 = 1, \quad \chi^2_{0.95}(2) \doteq 5\text{˙}99.$
Hipoteze ne moremo zavrniti.
Pri frekvencah 20, 50 in 40 pa je $\chi^2 = 10$. Ker je kritično območje isto, hipotezo zdaj zavrnemo.

21. Diskretizirana porazdelitev iz ničelne hipoteze:

pod -1	od -1 do 0	od 0 do 1	nad 1
0˙1839	0˙3161	0˙3161	0˙1839

$\chi^2 \doteq 21\text{˙}9, \quad \chi^2_{0.95}(3) \doteq 7\text{˙}81.$
Hipotezo zavrnemo.

22. 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

$\chi^2 \doteq 11˙8, \quad df = 4, \quad \chi^2_{0.95}(4) \doteq 9˙49.$
Hipotezo zavrnemo.

23. Iz metode največjega verjetja dobimo oceno $\hat{\theta} = 0˙65$ in iz nje ocenjeno porazdelitev:

$$\begin{pmatrix} RR & Rr & rr \\ 0˙1225 & 0˙455 & 0˙4225 \end{pmatrix}.$$

To nam da $\chi^2 = 11˙60$, kar primerjamo s $\chi^2_{0.95}(1) \doteq 3˙84$ in hipotezo zavrnemo.

24. Gre za isti statistični model kot v 15. nalogi iz 11. razdelka, torej dobimo oceni $\hat{a} = 1/4$ in $\hat{b} = 3/10$, od koder dobimo $\chi^2 = 2$. To spet primerjamo s $\chi^2_{0.95}(1) \doteq 3˙84$ in dobimo, da hipoteze ne moremo zavrniti.

25. Porazdelitev najprej diskretiziramo:

pod -1	od -1 do 0	od 0 do 1	nad 1
$\theta/2$	$(1-\theta)/2$	$(1-\theta)/2$	$\theta/2$

kjer je $\theta = e^{-\lambda}$. Iz vzorca dobimo $\hat{\theta} = 0˙59$ in pripadajočo ocenjeno diskretizirano porazdelitev:

pod -1	od -1 do 0	od 0 do 1	nad 1
0˙295	0˙205	0˙205	0˙295

kar nam da $\chi^2 \doteq 0˙448$. Ker je $\chi^2_{0.95}(2) \doteq 5˙99$, hipoteze ne zavrnemo.

26. Velja $S^+ = 2$ in $S^- = 8$. Ker za slučajno spremenljivko $S' \sim \mathrm{Bin}(10, 1/2)$ velja $P(S' \le 2) \doteq 0˙0547$, ničelne hipoteze o enakosti porazdelitev s testom z znaki pri tej stopnji značilnosti ne moremo zavrniti.

Pač pa lahko zavrnemo hipotezo, da imata enaki povprečji:
$\bar{X} = 17˙4, \quad \bar{Y} = 27˙1, \quad s \doteq 11˙27, \quad T \doteq -2˙72, \quad t_{0.95}(9) \doteq 1˙83.$

27. Velja $S^+ = 10$ in $S^- = 2$. Ker za slučajno spremenljivko $S' \sim \mathrm{Bin}(12, 1/2)$ velja $P(S' \le 2) \doteq 0˙0193$, ničelno hipotezo o enakosti porazdelitev zavrnemo: sprejmemo hipotezo, da je X stohastično večja in vsaj kdaj tudi strogo večja od Y.

Pač pa ne moremo zavrniti hipoteze, da imata X in Y enaki povprečji:
$\bar{X} = 27˙25, \quad \bar{Y} = 26˙5, \quad s \doteq 4˙575, \quad T \doteq 0˙57, \quad t_{0.95}(11) \doteq 1˙80.$

28. Velja $S^+ = 12$ in $S^- = 4$. Ker za slučajno spremenljivko $S' \sim \text{Bin}(16, 1/2)$ velja $P(S' \le 4) \doteq 0{\cdot}0193$, ničelno hipotezo o enakosti porazdelitev zavrnemo, če jo testiramo proti hipotezi, da je X stohastično večja in vsaj kdaj strogo večja od Y.

Oglejmo pa si zdaj, kaj pravita ustrezna testa povprečij. Iz:

$$\bar{X} = 25{\cdot}5\,, \quad \bar{Y} = 48{\cdot}375\,, \quad s \doteq 44{\cdot}36\,, \quad T \doteq 2{\cdot}06\,, \quad t_{0.95}(15) \doteq 1{\cdot}75$$

dobimo, da lahko ničelno hipotezo zavrnemo v korist hipoteze, da ima Y ves čas večje povprečje od X. Tako smo pri testu povprečij zavrnili ničelno hipotezo ravno v nasprotno smer kot pri testu z znaki.

Opomba. Pri tej in tudi pri prejšnji nalogi smo pri obeh testih dobili kontradiktorne rezultate. To kaže na to, da je nekaj narobe z našim modelom, recimo s tem, da je X ves čas stohastično večja od Y.

29. Ko preštejemo, dobimo, da se je 12 ljudi pred ogledom počutilo boljše kot po ogledu, 25 ljudi pa po ogledu boljše kot pred ogledom; 13 ljudi se je pred in po ogledu počutilo enako. Ker je:

$$Z = \frac{|12 - 25| - 1}{\sqrt{37}} \doteq 1{\cdot}97 > 1{\cdot}96\,,$$

hipotezo zavrnemo.

30. $m = 9$, $n = 11$, $\sum_{i=1}^{9} R_i = 70$, $Z \doteq -1{\cdot}82$, $z_{0.975} \doteq 1{\cdot}96$.
Hipoteze ne moremo zavrniti.

31. Pripravljalo se jih je 12, 8 pa se jih ni pripravljalo. Vsota rangov tistih, ki se niso pripravljali, je 110. Testna statistika v korist teh rangov pride $Z \doteq 1{\cdot}97$, kar je več od kritične vrednosti $1{\cdot}645$. Ničelno hipotezo tako zavrnemo in sprejmemo hipotezo, da so tisti, ki so se pripravljali, tekli bolje od tistih, ki se niso.

Pri T-testu pa povprečje tistih, ki so se pripravljali, pride $8{\cdot}083$, povprečje tistih, ki se niso pripravljali, pa $8{\cdot}362$. Velja še $s \doteq 0{\cdot}4066$, testna statistika pa pride $T \doteq 1{\cdot}50$, kar primerjamo s kritično vrednostjo $t_{0.95}(18) \doteq 1{\cdot}73$. Ničelne hipoteze tokrat ne moremo zavrniti.

32. Iz tabele:

	1	2	3	4	5	Skupaj
da	0	2	1	5	2	10
ne	0	5	2	0	0	7
Skupaj	0	7	3	5	2	17
Kumulativno	0	7	10	15	17	
Vezani rang	–	4	9	13	16·5	

razberemo vsoto rangov tistih, ki so odgovorili z da:

$$2 \cdot 4 + 1 \cdot 9 + 5 \cdot 13 + 2 \cdot 16{\cdot}5 = 115\,,$$

Testna statistika pride 2˙39, kar presega kritično vrednost $z_{0.99} \doteq 2{\cdot}33$. Torej zavrnemo hipotezo, da so oboji študenti enako pod stresom, in sprejmemo hipotezo, da so tisti, ki so se bolj posvečali študiju, bolj pod stresom kot tisti, ki se niso. Enostranska odstopanja so prišla zelo značilna.

14. 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. $T \doteq 0{\cdot}825$, $df = 13$, $t_{0.95}(13) \doteq 1{\cdot}77$.
Hipoteze ne moremo zavrniti.

3. Teoretične frekence:

oči\lasje	rdeči, blond	rjavi, črni	Skupaj
modre	6˙29	6˙71	13
zelene	11˙12	11˙87	23
rjave	12˙58	13˙42	26
Skupaj	30	32	62

so vse višje od 5. Velja $\chi^2 \doteq 22{\cdot}8$, kar primerjamo z $\chi^2_{0.99}(2) \doteq 9{\cdot}21$.
Hipotezo o neodvisnosti zavrnemo: barva las in barva oči sta bili statistično zelo značilno povezani.

4. Najprej opazimo, da so glede na to, da so vse opažene frekvence vsaj 5, tudi vse teoretične frekvence vsaj 5. Nadalje pride $\chi^2 \doteq 0{\cdot}0077$, kar primerjemo s $\chi^2_{0.95}(1) \doteq 3{\cdot}84$.
Hipoteze o neodvisnosti ne moremo zavrniti.

5. Vse opažene in zato tudi vse teoretične frekvence so enake vsaj 5.
$\chi^2 \doteq 169$, $\chi^2_{0.95}(8) \doteq 15{\cdot}5$. Hipotezo zavrnemo.

6. $\bar{X} = 3$, $\bar{Y} = 7$, $C_x^2 = 10$, $C_{xy} = 20$.
Regresijska premica: $y = 2x + 1$.
Pri $X = 10$ je $\hat{Y} = 21$.
$t_{0.975}(3) \doteq 3{\cdot}18$, $S \doteq 1{\cdot}155$, $\Delta \doteq 9{\cdot}07$.
Interval zaupanja: $11{\cdot}93 < Y < 30{\cdot}07$.