REŠITVE NOVEJŠIH KOLOKVIJEV IN IZPITOV IZ VERJETNOSTI IN STATISTIKE

Matematika – univerzitetni študij

Zbral: Martin Raič

2012/13

Rešitve kolokvija iz verjetnosti in statistike z dne 15. 11. 2012

Matematika – univerzitetni študij

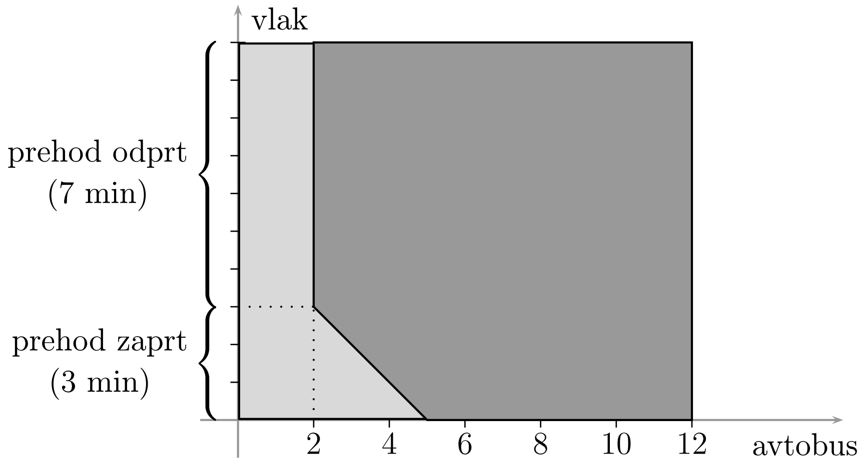
1. Dogodek, da Albert dobi dvoboj, ustreza naslednjim potekom partij:

$$AA\,,\ ABAA\,,\ ABABAA\,,\ldots \quad \text{in} \quad BAA\,,\ BABAA\,,\ BABABAA\,,\ldots$$

Verjetnost iskanega dogodka je torej enaka:

$$\left[1+\frac{2}{3}\right]\left[1+\frac{1}{3}\cdot\frac{2}{3}+\left(\frac{1}{3}\cdot\frac{2}{3}\right)^2+\left(\frac{1}{3}\cdot\frac{2}{3}\right)^4+\cdots\right]\left(\frac{1}{3}\right)^2=\frac{5}{21}\doteq 0\text{˙}238\,.$$

2. Marina bo prišla še pravočasno, če ji bo čakanje na avtobus skupaj s čakanjem na vlak vzelo največ 10 minut. Verjetnostni prostor lahko predstavimo s pravokotnikom na ravnini, kjer abscisa predstavlja čas, ki je minil od odhoda zadnjega avtobusa, ordinata pa zaprtost prehoda oz. koliko časa je prehod že odprt:



Verjetnost, da Marina pride še pravočasno, je torej enaka $\dfrac{191}{240}\doteq 0\text{˙}796.$

3. a) $\dfrac{1}{5}\cdot\dfrac{\binom{30}{1}\binom{20}{4}}{\binom{50}{5}}+\dfrac{2}{5}\cdot\dfrac{\binom{30}{2}\binom{20}{3}}{\binom{50}{5}}+\dfrac{3}{5}\cdot\dfrac{\binom{30}{3}\binom{20}{2}}{\binom{50}{5}}+\dfrac{4}{5}\cdot\dfrac{\binom{30}{4}\binom{20}{1}}{\binom{50}{5}}+\dfrac{5}{5}\cdot\dfrac{\binom{30}{5}\binom{20}{0}}{\binom{50}{5}}=\dfrac{3}{5}=0\text{˙}6.$

b) $\dfrac{5}{3}\left[\dfrac{1}{5}\cdot\dfrac{\binom{30}{1}\binom{20}{4}}{\binom{50}{5}}+\dfrac{2}{5}\cdot\dfrac{\binom{30}{2}\binom{20}{3}}{\binom{50}{5}}\right]=\dfrac{5415}{30268}\doteq 0\text{˙}179.$

4. Označimo z A_i dogodek, da ima i-ti igralec vse karte v barvi pika s samimi različnimi vrednostmi. Očitno se lahko zgodita največ dva taka dogodka. Glede na to, da je možnih 10 pikovih lestvic petih kart (od A,2,3,4,5 do 10,J,Q,K,A), za vsak i velja:

$$p_1 := P(A_i) = \frac{10}{\binom{52}{5}} \doteq 3\text{˙}847693 \cdot 10^{-6}\,.$$

Za dva igralca pa je število razporeditev kart prikazano v naslednjih dveh tabelah – v prvi vrstici so prikazane karte prvega igralca, v drugi pa število možnih razporeditev za drugega:

A,2,3,4,5	2,3,4,5,6	3,4,5,6,7	4,5,6,7,8	5,6,7,8,9
4	4	3	2	1

6,7,8,9,10	7,8,9,10,J	8,9,10,J,Q	9,10,J,Q,K	10,J,Q,K,A
1	2	3	4	4

Torej za poljubna različna i in j velja:

$$p_2 := P(A_i \cap A_j) = \frac{28}{\binom{52}{5}\binom{47}{5}} \doteq 7\text{˙}023448 \cdot 10^{-12}\,.$$

Verjetnost iskanega dogodka je torej enaka $10p_1 - \binom{10}{2}p_2 \doteq 3\text{˙}847661 \cdot 10^{-5}$.

4P. Naj bo K_i dogodek, da je bila karta, ki je bila pred mešanjem na i-tem položaju v kupu, tam tudi po mešanju. Dogovorimo se, da indeksi $i = 1, 2, 3$ označujejo spodnje tri položaje. Nadalje naj bo A dogodek, da je as po mešanju ostal na istem mestu (to bi bil lahko dogodek K_4). Izračunati moramo pogojno verjetnost dogodka $K_1 \cup K_2 \cup K_3$ glede na $\bar{A}$.

Pogojno verjetnost dogodka K_i izračunamo tako, da še nadalje pogojujemo glede na to, kje po mešanju pristane as (ne sme pristati na i-tem mestu). Dobimo:

$$P(K_i \mid \bar{A}) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}\,.$$

Nadalje opazimo, da se lahko zgodita največ dva dogodka K_i (če se zgodijo vsi trije, je tudi as na istem mestu). Za $i \neq j$ velja:

$$P(K_i \cap K_j \mid \bar{A}) = \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18}\,.$$

Iskana verjetnost je:

$$\begin{aligned} P(K_1 \cup K_2 \cup K_3 \mid \bar{A}) = {} & P(K_1 \mid \bar{A}) + P(K_2 \mid \bar{A}) + P(K_3 \mid \bar{A}) - \\ & - P(K_1 \cap K_2 \mid \bar{A}) + P(K_1 \cap K_3 \mid \bar{A}) + \\ & + P(K_2 \cap K_3 \mid \bar{A}) = \\ = {} & \frac{1}{2}\,. \end{aligned}$$

Rešitve kolokvija iz verjetnosti in statistike z dne 5. 2. 2013

Matematika – univerzitetni študij

1. Če z S označimo število dobitkov v prvih 100 stavah, je $S \sim \mathrm{b}(100, 18/37)$. Po teh 100 stavah ima Renato $10S$ žetonov. Če dobi zadnjo stavo (ko stavi desetino žetonov), se mu število žetonov poveča za 10%, če zadnjo stavo izgubi, pa se mu zmanjša za 10%. Iskana verjetnost je tako enaka:

$$
\begin{aligned}
&\frac{18}{37}\,P(11S > 500) + \frac{19}{37}\,P(9S > 500) = \\
&\quad= \frac{18}{37}\,P\big(S > 45\tfrac{5}{11}\big) + \frac{19}{37}\,P\big(S > 55\tfrac{5}{9}\big) = \\
&\quad= \frac{18}{37}\,P(S > 45{\cdot}5) + \frac{19}{37}\,P(S > 55{\cdot}5) \approx \\
&\quad\approx \frac{18}{37}\left[\frac{1}{2} - \Phi\left(\frac{45{\cdot}5 - 100\cdot\frac{18}{37}}{\sqrt{100\cdot\frac{18}{37}\cdot\frac{19}{37}}}\right)\right] + \frac{19}{37}\left[\frac{1}{2} - \Phi\left(\frac{55{\cdot}5 - 100\cdot\frac{18}{37}}{\sqrt{100\cdot\frac{18}{37}\cdot\frac{19}{37}}}\right)\right] \doteq \\
&\quad\doteq 0{\cdot}402.
\end{aligned}
$$

2. Najprej iz:

$$1 = \iint_{x>y>1} \frac{c}{x^3}\,\mathrm{d}x\,\mathrm{d}y = \int_1^\infty \int_y^\infty \frac{c}{x^3}\,\mathrm{d}x\,\mathrm{d}y = \frac{c}{2}$$

ali iz:

$$1 = \iint_{x>y>1} \frac{c}{x^3}\,\mathrm{d}x\,\mathrm{d}y = \int_1^\infty \int_1^x \frac{c}{x^3}\,\mathrm{d}y\,\mathrm{d}x = \frac{c}{2}$$

izračunamo $c = 2$. Nato izračunamo kumulativno porazdelitveno funkcijo:

$$F_Z(z) = \iint_{\substack{x-2y<z\\ x>y>1}} \frac{2}{x^3}\,\mathrm{d}x\,\mathrm{d}y\,.$$

Račun lahko dokončamo na vsaj dva načina.

Prvi način:

$$F_Z(z) = \int_{\max\{1,-z\}}^\infty \int_y^{z+2y} \frac{2}{x^3}\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{\max\{-z,1\}} - \frac{1}{2(z+2\max\{-z,1\})}$$

oziroma:

$$F_Z(z) = \begin{cases} -\dfrac{1}{2z} & ;\ z \le -1 \\ 1 - \dfrac{1}{2(z+2)} & ;\ z \ge -1 \end{cases}.$$

Drugi način: za $z \le -1$ velja:

$$F_Z(z) = \int_{-z}^\infty \int_{(x-z)/2}^x \frac{2}{x^3}\,\mathrm{d}y\,\mathrm{d}x = -\frac{1}{2z}\,,$$

za $z \geq -1$ pa velja:

$$F_Z(z) = \int_1^{z+2} \int_1^x \frac{2}{x^3}\,\mathrm{d}y\,\mathrm{d}x + \int_{z+2}^{\infty} \int_{(x-z)/2}^x \frac{2}{x^3}\,\mathrm{d}y\,\mathrm{d}x = 1 - \frac{1}{2(z+2)}\,.$$

Seveda dobimo isto kot prej. Ker je F_Z zvezna in odsekoma zvezno odvedljiva, ima Z gostoto:

$$f_Z(z) = \begin{cases} \dfrac{1}{2z^2} & ;\ z < -1 \\ \dfrac{1}{2(z+2)^2} & ;\ z > -1 \end{cases}.$$

3. Naj bo T čas okvare. Tedaj je čas od okvare do kontrole (glede na besedilo naloge) enak $\max\{t_0 - T, 0\}$. Pričakovana vrednost je enaka:

$$E\big[\max\{t_0 - T, 0\}\big] = \int_0^{t_0} (t_0 - t)\,\lambda\, e^{-\lambda t}\,\mathrm{d}t = t_0 - \frac{1 - e^{-\lambda t_0}}{\lambda}\,.$$

4. *Prvi način.* Za $n = 1, 2, 3, \ldots$ označimo:

$$\begin{aligned} A_{ni} &:= \{\text{v prvih } n-1 \text{ metih nikoli ne pade } i \text{ pik}\}\,, \\ B_{ni} &:= \{\text{v } n\text{-tem metu pade } i \text{ pik}\}\,. \end{aligned}$$

Tedaj je dogodek $\{N = n\}$ unija nezdružljivih dogodkov $(A_{n1} \setminus A_{n6}) \cap B_{n1}$ in $(A_{n6} \setminus A_{n1}) \cap B_{n6}$. Očitno je:

$$P(A_{ni}) = \left(\frac{5}{6}\right)^{n-1}, \qquad P(B_{ni}) = \frac{1}{6}\,.$$

Nadalje sta dogodka $(A_{ni} \setminus A_{nj})$ in B_{nk} za poljubne i, j in k neodvisna. Dogodka A_{ni} in A_{nj} pa sta odvisna, a za $i \neq j$ velja:

$$P(A_{ni} \cap A_{nj}) = \left(\frac{4}{6}\right)^{n-1}.$$

Sledi:

$$P(A_{ni} \setminus A_{nj}) = P(A_{ni}) - P\big(A_{ni} \setminus (A_{ni} \cap A_{nj})\big) = \left(\frac{5}{6}\right)^{n-1} - \left(\frac{4}{6}\right)^{n-1}.$$

Torej velja:

$$P(N = n) = 2 \cdot \left[\left(\frac{5}{6}\right)^{n-1} - \left(\frac{4}{6}\right)^{n-1}\right] \cdot \frac{1}{6} = \frac{1}{3}\left[\left(\frac{5}{6}\right)^{n-1} - \left(\frac{2}{3}\right)^{n-1}\right].$$

Drugi način: izračunamo $P(N > n)$, t. j. verjetnost dogodka, da po n metih še ni padla ena pika ali pa še ni padlo šest pik. Po formuli za verjetnost unije dogodkov za $n = 0, 1, 2, \ldots$ velja:

$$P(N > n) = 2\left(\frac{5}{6}\right)^n - \left(\frac{4}{6}\right)^n.$$

Torej za $n = 1, 2, 3, \ldots$ velja:

$$P(N = n) = P(N > n - 1) - P(N > n) = \frac{1}{3}\left[\left(\frac{5}{6}\right)^{n-1} - \left(\frac{2}{3}\right)^{n-1}\right],$$

kar je isto kot prej.

Opomba. Jasno je, da lahko slučajna spremenljivka N zavzame le vrednosti $2, 3, 4, \ldots$, a zgornjo formulo lahko uporabimo tudi za $n = 1$: pravilno dobimo $P(N = 1) = 0$.

4P. Za verjetnostni prostor lahko postavimo kar $\binom{4}{2} = 6$ enako verjetnih razporeditev asa in fanta v kup štirih kart: ni potrebno ločiti, katera karta je as in katera fant. Očitno lahko slučajna spremenljivka N zavzame le vrednosti 2, 3, ali 4; za n izmed teh vrednosti dogodek $\{N = n\}$ zajema $n-1$ izidov, saj mora biti ena izmed izbranih kart na n-tem položaju od zgoraj, druga pa nad njo. Sklep:

$$N \sim \begin{pmatrix} 2 & 3 & 4 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}.$$

Rešitve kolokvija iz verjetnosti in statistike z dne 4. 4. 2013

Matematika – univerzitetni študij

1. 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 $\arccos R$, kjer je $R = \sqrt{X^2 + Y^2}$. Od tod naprej gre na vsaj dva načina.

Prvi način. Pričakovani kot je 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:

$$\bar{\alpha} = 2\int_0^{\pi/2} t\cos t\sin t\,\mathrm{d}t = \int_0^{\pi/2} t\sin(2t)\,\mathrm{d}t = -\left.\frac{t\cos(2t)}{2}\right|_0^{\pi/2} + \frac{1}{2}\int_0^{\pi/2}\cos(2t)\,\mathrm{d}t = \\ = \frac{\pi}{4}\,.$$

Drugi način. Označimo z α kot, pod katerim meteorit pade na planet. To je zdaj slučajna spremenljivka. Izračunamo najprej njeno kumulativno porazdelitveno funkcijo. Za $0 < t < \pi/2$ velja:

$$F_\alpha(t) = P(\alpha \le t) = P(\arccos R \le t) = P(R \ge \cos t) = 1 - \cos^2 t\,.$$

Od tod dobimo porazdelitveno gostoto:

$$f_\alpha(t) = 2\cos t\sin t\,,$$

iz nje pa pričakovano vrednost:

$$E(\alpha) = 2\int_0^{\pi/2} t\cos t\sin t\,\mathrm{d}t = \frac{\pi}{4}\,.$$

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 (glej drugi način).

2. Velja:

$$r(X, X + aY) = \frac{4 - a}{2\sqrt{4 - 2a + 9a^2}}\,,$$

torej mora biti:

$$a \le 4\,, \qquad \frac{(4 - a)^2}{4(4 - 2a + 9a^2)} = \frac{1}{36}\,,$$

od koder dobimo $a = 2$.

3. Iz:

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

dobimo:

$$P(Y = k) = E\left[\frac{X^k\, e^{-X}}{k!}\right] = \frac{\lambda}{k!}\int_0^\infty x^k\, e^x\, e^{-\lambda x}\,\mathrm{d}x = \frac{\lambda}{(\lambda + 1)^{k+1}}\,.$$

Od tod dobimo:

$$P(Y + 1 = l) = \frac{\lambda}{(\lambda + 1)^l} = \frac{\lambda}{\lambda + 1}\left(1 - \frac{\lambda}{\lambda + 1}\right)^{l-1}\,,$$

torej ima $Y + 1$ geometrijsko porazdelitev $\mathrm{Geom}\big(\frac{\lambda}{\lambda+1}\big)$.

4. a) Označimo $p_k = P(A_k)$ in naj bo B_n dogodek, da se zgodi vsaj eden izmed dogodkov $A_1, \ldots A_n$. Tedaj velja:

$$P(B_n) = 1 - (1 - p_1)(1 - p_2)\cdots(1 - p_n) \ge 1 - e^{-(p_1 + p_2 + \ldots + p_n)}\,.$$

Dogodki $B_1, B_2, B_3, \ldots$ tvorijo naraščajoče zaporedje, čigar unija je dogodek, da se zgodi vsaj eden izmed dogodkov dogodkov $A_1, A_2, A_3, \ldots$; verjetnost tega dogodka je $\lim_{n\to\infty} P(B_n)$. Ker vrsta $\sum_{i=1}^\infty p_i$ divergira, je ta limita večja ali enaka 1, torej je enaka ena.

b) Videli smo, da se skoraj gotovo zgodi vsaj eden izmed dogodkov $A_1, A_2, \ldots$. Naj bo I_1 prvi indeks i, za katerega se zgodi A_i. Pogojno na I_1 so dogodki A_{I_1+1}, A_{I_1+2} spet neodvisni z verjetnostmi $p_{I_1+1}, p_{I_2+1}, \ldots$. Ker vsota teh verjetnosti spet divergira, se skoraj gotovo zgodi vsaj eden izmed teh dogodkov. Torej se skoraj gotovo zgodita vsaj dva izmed dogodkov $A_1, A_2, \ldots$. Isti sklep lahko naredimo tudi na I_2, drugem indeksu, za katerega se zgodi A_i. Tako dobimo, da se skoraj gotovo zgodijo vsaj trije dogodki. Ko nadaljujemo, dobimo, da se za vsak m skoraj gotovo zgodi vsaj m teh dogodkov. Od tod pa sledi, da se skoraj gotovo zgodi neskončno mnogo teh dogodkov.

4P. Iz tabele verjetnosti:

k	$P(X = k)$	$P(X < k)$	$P(X \leq k)$
0	0˙00673	0	0˙00673
1	0˙03369	0˙00673	0˙04043
2	0˙08422	0˙04043	0˙12465
3	0˙14037	0˙12465	0˙26503
4	0˙17547	0˙26503	0˙44049
5	0˙17547	0˙44049	0˙61596
6	0˙14622	0˙61596	0˙76218

odčitamo $q_{1/4} = 3$, $q_{1/2} = 5$ in $q_{3/4} = 6$. Vsi trije kvartili so natančno določeni.

Rešitve kolokvija iz verjetnosti in statistike z dne 6. 6. 2013

Matematika – univerzitetni študij

1. Najprej izračunajmo:

$$E(X_i) = a\,, \qquad D(X_i) = E\big[(X_i - a)^2\big] = \frac{a^2+2}{3}\,.$$

Če z $\bar{X}$ označimo aritmetično sredino, torej velja $E(\bar{X}) = a$ in $D(\bar{X}) = (a^2+2)/300$. Po centralnem limitnem izreku je približno $\bar{X} \sim \mathrm{N}\big(a, \sqrt{(a^2+2)/300}\big)$, torej je:

$$P\left(\bar{X} > \frac{11a}{10}\right) \approx \frac{1}{2} - \Phi\left(\frac{\frac{a}{10}}{\sqrt{\frac{a^2+2}{300}}}\right) = \frac{1}{2} - \Phi\left(\frac{a\sqrt{3}}{\sqrt{a^2+2}}\right)\,.$$

Torej mora biti približno $\Phi\big(a\sqrt{3}/\sqrt{a^2+2}\big) = 0\text{˙}45$ oziroma približno $a\sqrt{3}/\sqrt{a^2+2} = z_{0.95} \doteq 1\text{˙}645$, kar je res pri $a = z_{0.95}\sqrt{2/(3 - z_{0.95}^2)} \doteq 4\text{˙}3$.

2. Iz gostote dobimo logaritem verjetja:

$$\ln L = n \ln \frac{\sin(\pi a)}{\pi} + (a-1)\sum_{i=1}^n \ln X_i - a \sum_{i=1}^n \ln(1 - X_i)$$

Ta funkcija gre proti minus nekončno, ko gre a proti 0 ali proti 1. Torej bo maksimum zagotovo dosežen tam, kjer je odvod:

$$\frac{\mathrm{d}\ln L}{\mathrm{d}a} = n\pi \operatorname{ctg}(\pi a) + \sum_{i=1}^n \ln \frac{X_i}{1-X_i}$$

enak nič, to pa se zgodi, ko je a enak:

$$\begin{aligned}
\hat{a} &= \frac{1}{\pi}\operatorname{arcctg}\left(-\frac{1}{n\pi}\sum_{i=1}^n \ln\frac{X_i}{1-X_i}\right) = \\
&= \frac{1}{\pi}\operatorname{arcctg}\left(\frac{1}{n\pi}\sum_{i=1}^n \ln\frac{1-X_i}{X_i}\right) = \\
&= \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg}\left(\frac{1}{n\pi}\sum_{i=1}^n \ln\frac{X_i}{1-X_i}\right)\,.
\end{aligned}$$

Izražava:

$$\hat{a} = \frac{1}{\pi}\operatorname{arctg}\left(-\frac{n\pi}{\sum_{i=1}^n \ln\frac{X_i}{1-X_i}}\right) = \frac{1}{\pi}\operatorname{arctg}\left(\frac{n\pi}{\sum_{i=1}^n \ln\frac{1-X_i}{X_i}}\right)$$

pa ne da vedno pravilnega rezultata, saj je tedaj $-\frac{1}{2} < a < \frac{1}{2}$, medtem ko mora biti v našem statističnem modelu $0 < a < 1$.

3. a) Imamo podatke:

$$m = 17\,, \quad \bar{X} = 45\text{˙}76\,, \quad S_{\mathrm{p},X} = 15\text{˙}61\,, n = 12\,, \quad \bar{Y} = 31\text{˙}17\,, \quad S_{\mathrm{p},Y} = 20\text{˙}31\,.$$

Popravljeni vzorčni standardni odklon za vse študente znaša:

$$S_{\mathrm{p}} = \sqrt{\frac{(m-1)S_{\mathrm{p},X}^2 + (n-1)S_{\mathrm{p},Y}^2}{m+n-2}} \doteq 17\text{˙}68\,.$$

b) Po primerjavi testne statistike $T \doteq 2\text{˙}19$ s kritičnima vrednostma $t_{0.975}(27) \doteq 2\text{˙}05$ in $t_{0.995}(27) \doteq 2\text{˙}77$ dobimo, da so odstopanja statistično značilna, niso pa zelo značilna.

4. Nastavimo neznano funkcijo h, za katero bo veljalo $E\big[h(X)\big] = \lambda\, e^{\lambda}$. Iz Poissonove porazdelitve dobimo:

$$E\big[h(X)\big] = \sum_{k=0}^{\infty} h(k)\,\frac{\lambda^k\, e^{-\lambda}}{k!}\,,$$

Torej mora veljati:

$$\sum_{k=0}^{\infty} \frac{h(k)}{k!}\,\lambda^k = \lambda\, e^{2\lambda} = \sum_{k=1}^{\infty} \frac{2^{k-1}}{(k-1)!}\,\lambda^k$$

in to velja za $h(k) = k\,2^{k-1}$. To pomeni, da je $X\,2^{X-1}$ nepristranska cenilka za $\lambda\, e^{\lambda}$.

4P. S prvim momentom:

$$m_1 = E(X) = \frac{1}{2a}\int_{-a}^{a} x\,\mathrm{d}x = 0$$

si ne moremo nič pomagati, zato pa si lahko pomagamo z drugim:

$$m_2 = E(X^2) = \frac{1}{2a}\int_{-a}^{a} x^2\,\mathrm{d}x = \frac{a^2}{3}\,.$$

Cenilka za a je torej:

$$\hat{a} = \sqrt{3m_2} = \sqrt{\frac{3}{n}(X_1^2 + X_2^2 + \cdots + X_n^2)}\,.$$

Rešitve izpita iz verjetnosti in statistike z dne 28. 6. 2013

Matematika – univerzitetni študij

1. Označimo s H_i dogodek, da smo iz vreče vzeli natanko i lešnikov dobavitelja B, Z_2 pa dogodek, da smo dobili natanko dva žarka lešnika. Očitno dogodki H_0, H_1 in H_2 tvorijo popoln sistem dogodkov. Velja:

$$P(H_0) = \frac{\binom{8}{3}\binom{2}{0}}{\binom{10}{3}} = \frac{7}{15}\,, \quad P(H_1) = \frac{\binom{8}{2}\binom{2}{1}}{\binom{10}{3}} = \frac{7}{15}\,, \quad P(H_2) = \frac{\binom{8}{1}\binom{2}{2}}{\binom{10}{3}} = \frac{1}{15}\,,$$

in:

$$\begin{aligned} P(Z_2 \mid H_0) &= 3 \cdot 0\text{'}05^2 \cdot 0\text{'}95 = 0\text{'}007125\,, \\ P(Z_2 \mid H_1) &= 0\text{'}05^2 \cdot 0\text{'}85 + 2 \cdot 0\text{'}05 \cdot 0\text{'}95 \cdot 0\text{'}15 = 0\text{'}016375\,, \\ P(Z_2 \mid H_2) &= 0\text{'}05 \cdot 2 \cdot 0\text{'}15 \cdot 0\text{'}85 + 0\text{'}95 \cdot 0\text{'}15^2 = 0\text{'}034125\,. \end{aligned}$$

Pri točki a) je zahtevana verjetnost enaka:

$$P(Z_2) = \frac{7}{15} \cdot 0\text{'}007125 + \frac{7}{15} \cdot 0\text{'}016375 + \frac{1}{15} \cdot 0\text{'}034125 \doteq 0\text{'}0132\,,$$

pri točki b) pa je enaka:

$$\frac{1}{P(Z_2)} \cdot \frac{1}{15} \cdot 0\text{'}95 \cdot 0\text{'}15^2 \doteq 0\text{'}108\,.$$

2. Uporabimo centralni limitni izrek, za kar pa moramo poznati matematično upanje in disperzijo slučajne spremenljivke S. Velja:

$$\begin{gathered} E(X_iY_i) = E(X_i)\,E(Y_i) = -1\,, \\ E(X_i^2) = D(X_i) + \big(E(X_i)\big)^2 = 5\,, \\ E(Y_i^2) = D(Y_i) + \big(E(Y_i)\big)^2 = 10\,, \\ E(X_i^2Y_i^2) = E(X_i^2)\,E(Y_i^2) = 50\,, \\ D(X_iY_i) = E(X_i^2Y_i^2) - \big(E(X_iY_i)\big)^2 = 49\,, \\ E(S) = -100\,, \qquad D(S) = 4900\,. \end{gathered}$$

Sledi:

$$P(S > 0) \approx \frac{1}{2} - \Phi\left(\frac{100}{\sqrt{4900}}\right) = \frac{1}{2} - \Phi\left(\frac{10}{7}\right) \doteq 0\text{'}077\,.$$

3. Iz logaritma gostote:

$$\ln f(x) = \frac{3}{2}\ln a - \ln\sqrt{2\pi} + 2\ln x - ax^2$$

dobimo logaritem verjetja:

$$\ln L = \frac{3n}{2} \ln a - n \ln \sqrt{2\pi} + 2\big(\ln X_1 + \cdots + \ln X_n\big) - a(X_1^2 + \cdots + X_n^2)\,.$$

Opazimo, da gre $\ln L$ pri vsakem naboru opažanj proti minus neskončno, ko gre a bodisi proti nič bodisi proti neskončno. Zato bo maksimum nujno dosežen v stacionarni točki. Velja:

$$\frac{\mathrm{d} \ln L}{\mathrm{d} a} = \frac{3n}{2a} - (X_1^2 + \cdots + X_n^2)\,,$$

kar je enako nič, če je a enak:

$$\hat{a} = \frac{3n}{2(X_1^2 + X_2^2 + \cdots + X_n^2)}\,.$$

4. Velja:

$$P(X = k) = \frac{1}{a} \left(1 - \frac{1}{a}\right)^{k-1} \quad ; \quad k = 1, 2, 3, \ldots$$

in:

$$f_{Y|X}(y) = \sqrt{\frac{X}{2\pi}}\, e^{-Xy^2/2}\,.$$

Brezpogojna porazdelitev slučajne spremenljivke Y je torej zvezna z gostoto:

$$f_Y(y) = E\big[f_{Y|X}(y)\big] = \sum_{k=1}^{\infty} \frac{1}{a} \left(1 - \frac{1}{a}\right)^{k-1} \sqrt{\frac{k}{2\pi}}\, e^{-ky^2/2} =$$

$$= \frac{1}{(a-1)\sqrt{2\pi}} \sum_{k=1}^{\infty} \sqrt{k} \left(\frac{a-1}{a}\, e^{-y^2/2}\right)^k\,.$$

Vrsta konvergira za vse y in vse $a > 1$, za njeno vsoto pa ne kaže, da bi se dala zapisati v elementarni sklenjeni obliki.

4P. Iz:

$$\int_{-\infty}^{\infty} f_X(x)\, \mathrm{d}x = \int_1^{\infty} \big(cx^{-2} + x^{-4}\big)\, \mathrm{d}x = c + \frac{1}{3}$$

dobimo, da mora biti $c = 2/3$. Kumulativna porazdelitvena funkcija je enaka:

$$F_X(x) = P(X < x) = \begin{cases} 0 & ; x \le 1 \\ 1 - \dfrac{2}{3x} - \dfrac{1}{3x^3} & ; x \ge 1 \end{cases}\,.$$

Za $y < 0$ je očitno $F_Y(y) = 0$. Za $y \ge 0$ velja:

$$F_Y(y) = P(Y < y) = P\big(2 - \sqrt{y} < X < 2 + \sqrt{y}\big) = F_X\big(2 + \sqrt{y}\big) - F_X\big(2 + \sqrt{y}\big)\,.$$

Za $0 \leq y \leq 1$ velja:

$$F_Y(y) = \frac{2}{3\big(2-\sqrt{y}\big)} + \frac{1}{3\big(2-\sqrt{y}\big)^3} - \frac{2}{3\big(2+\sqrt{y}\big)} - \frac{1}{3\big(2+\sqrt{y}\big)^3}\,,$$

za $y \geq 1$ pa velja:

$$F_Y(y) = 1 - \frac{2}{3\big(2+\sqrt{y}\big)} - \frac{1}{3\big(2+\sqrt{y}\big)^3}\,.$$

Od tod sledi, da je porazdelitev slučajne spremenljivke Y zvezna z gostoto:

$$f_Y(y) = \begin{cases} 0 & ;\ y < 0 \\ \frac{1}{3\sqrt{y}(2-\sqrt{y})^2} + \frac{1}{2\sqrt{y}(2-\sqrt{y})^4} + \frac{1}{3\sqrt{y}(2+\sqrt{y})^2} + \frac{1}{2\sqrt{y}(2+\sqrt{y})^4} & ;\ 0 < y < 1 \\ \frac{1}{3\sqrt{y}(2+\sqrt{y})^2} + \frac{1}{2\sqrt{y}(2+\sqrt{y})^4} & ;\ y > 1 \end{cases}\,.$$

Rešitve izpita iz verjetnosti in statistike z dne 23. 8. 2013

Matematika – univerzitetni študij

1. a) Slučajna spremenljivka N lahko zavzame celoštevilske vrednosti od 2 do $2n - 2k + 2$. Za i izmed teh števil se dogodek $\{N = i\}$ zgodi, če je med prvimi $i - 1$ nogavicami natanko ena karirasta in če je i-ta karirasta. To pomeni, da je med zadnjimi $2n - i$ nogavicami natanko $2k - 2$ karirastih. Sledi:

$$P(N = i) = \frac{(i-1)\binom{2n-i}{2k-2}}{\binom{2n}{2k}}\,.$$

b) Za $n = 4$ in $k = 2$ dobimo:

$$N \sim \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ \frac{15}{70} & \frac{20}{70} & \frac{18}{70} & \frac{12}{70} & \frac{5}{70} \end{pmatrix},$$

od koder izračunamo:

$$E(N) = \frac{18}{5}\,, \quad E(N^2) = \frac{72}{5}\,, \quad D(N) = \frac{36}{25}\,.$$

2. Ker gre za 100 neodvisnih seštevancev, je smiselno uporabiti približek na podlagi centralnega limitnega izreka. Velja:

$$E(X_i Y_i) = E(X_i)\,E(Y_i) = 3\,, \quad E(S) = 300\,,$$
$$E(X_i^2) = D(X_i) + \big(E(X_i)\big)^2 = 9\,, \quad E(Y_i^2) = D(Y_i) + \big(E(Y_i)\big)^2 = 10\,,$$
$$E(X_i^2 Y_i^2) = E(X_i^2)\,E(Y_i^2) = 90\,, \quad D(X_i Y_i) = 81\,, \quad D(S) = 8100\,,$$

torej je približno:

$$P(S < 240) \approx \Phi\left(\frac{240 - 300}{90}\right) + \frac{1}{2} = \frac{1}{2} - \Phi\left(\frac{2}{3}\right) \approx 0{\cdot}252\,.$$

3. Označimo ustrezni aritmetični sredini z $\mu_{\text{Ž}}$ in μ_{M} ter ustrezna popravljena standardna odklona z $S_{\text{p,Ž}}$ in $S_{\text{p,M}}$. Naj bo v vzorcu n moških in n žensk. Ničelno hipotezo zavrnemo v korist alternativne, če je:

$$\frac{\mu_{\text{M}} - \mu_{\text{Ž}}}{\sqrt{\dfrac{S_{\text{p,M}}^2 + S_{\text{p,Ž}}^2}{2}}}\sqrt{\frac{n}{2}} > t_{0.99}(2n-1)$$

ali ekvivalentno:

$$\frac{t_{0.99}(2n-1)}{\sqrt{n}} < \frac{\mu_{\text{M}} - \mu_{\text{Ž}}}{\sqrt{S_{\text{p,M}}^2 + S_{\text{p,Ž}}^2}} \doteq 0{\cdot}18034\,,$$

kar velja za $n \geq 169$ (če kvantil Studentove porazdelitve aproksimiramo s kvantilom standardne normalne porazdelitve, pa ustrezno neenakost dobimo že za $n \geq 162$).

4. Označimo z R razdaljo od izhodišča do najbližje točke (če točk sploh ni, pa naj bo $R = \infty$). Najlažje je računati komplemenarno kumulativno porazdelitveno funkcijo, torej $P(R > r)$. To je verjetnost dogodka, da v krogu okoli izhodišča s polmerom r ni nobene točke.

Kvadrat oddaljenosti posamezne točke od izhodišča ima porazdelitev hi kvadrat z dvema prostostnima stopnjama, ki ima gostoto:

$$g(x) = \begin{cases} \frac{1}{2}\,e^{-x/2} & ; \; x > 0 \\ 0 & ; \; \text{sicer} \end{cases} .$$

Verjetnost, da je posamezna točka od izhodišča oddaljena za več kot r, je enaka:

$$\int_{r^2}^{\infty} g(x)\,\mathrm{d}x = e^{-r^2/2} .$$

Od tod sledi:

$$P(R > r \mid N) = e^{-Nr^2/2}$$

in:

$$P(R > r) = \sum_{n=0}^{\infty} \frac{\lambda^n\,e^{-\lambda}}{n!}\,e^{-nr^2/2} = e^{\lambda(e^{-r^2/2}-1)} .$$

Kumulativna porazdelitvena funkcija je torej enaka:

$$F_R(r) = \begin{cases} 0 & ; \; r \le 0 \\ 1 - e^{\lambda(e^{-r^2/2}-1)} & ; \; r \ge 0 \end{cases} .$$

Opazimo, da njena limita, ko gre r proti neskončno, ni 1, temveč $1-e^{-\lambda}$. Ta limita je verjetnost, da je slučajna spremenljivka R končna ali, ekvivalentno, da je v ravnini vsaj ena točka.

4P. Najprej določimo kumulativno porazdelitveno funkcijo $F_Y(y) = P(Y < y)$. Ker lahko Y zavzame vrednosti od 0 do 4, je za $y \le 0$ očitno $F_Y(y) = 0$, za $y \ge 4$ pa $F_Y(y) = 1$. Za $0 \le y \le 1$ velja:

$$F_Y(y) = P(X^2 < y) = P\left(-\sqrt{y} < X < \sqrt{y}\right) = \frac{2}{3}\sqrt{y}\,,$$

za $1 \le y \le 4$ pa velja:

$$F_Y(y) = P(X^2 < y) = P\left(-\sqrt{y} < X < \sqrt{y}\right) = P\left(-1 < X < \sqrt{y}\right) = \frac{\sqrt{y}+1}{3} .$$

2011/12

Rešitve kolokvija iz verjetnosti in statistike z dne 17. 11. 2011

Matematika – univerzitetni študij

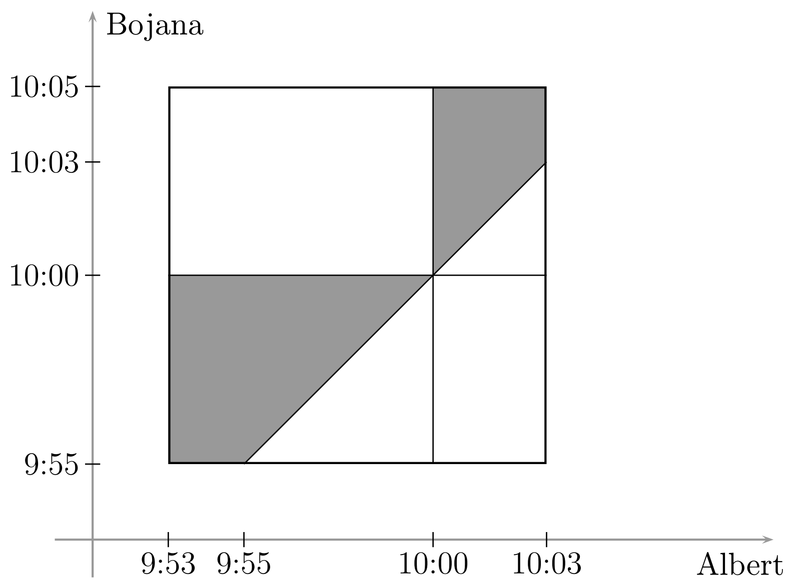
1. a) Igralca najprej dobivala igre izmenoma, brž ko nekdo dvakrat zmaga, pa končata. Če z J_2 označimo dogodek, da je zadnji dve rundi dobil Janez, velja:

$$P(J_2) = \sum_{k=0}^{\infty} \left[\left(\frac{2}{3} \cdot \frac{1}{3}\right)^k + \frac{1}{3} \left(\frac{2}{3} \cdot \frac{1}{3}\right)^k \right] \cdot \left(\frac{2}{3}\right)^2 = \frac{16}{21} \doteq 0\text{˙}762\,.$$

b) Naj bo A dogodek, da sta odigrala več kot tri runde. Izračunamo lahko:

$$P(A \mid J_2) = 1 - P(A^{\mathrm{c}} \mid J_2) = 1 - \frac{1}{P(J_2)} \left[\left(\frac{2}{3}\right)^2 + \frac{1}{3} \left(\frac{2}{3}\right)^2 \right] = \frac{2}{9} \doteq 0\text{˙}222\,.$$

2. Če postavimo, da je izid v verjetnostnem prostoru določen z Albertovim in Bojaninim prihodom na postajo, je tak izid izbran na slepo v okviru predpisanih meja za prihode (okvir na sliki). S sivo barvo je označen dogodek, da se Albert in Bojana srečata:



Verjetnost iskanega dogodka je razmerje ploščin, ki je enako 0˙33.

3. Označimo z n število metov, z S pa število šestic. Tedaj je $S \sim \mathrm{b}(n, 1/6)$. Veljati mora:

$$P(S \geq 0{\cdot}15\, n) \geq 0{\cdot}95\,.$$

Po Laplaceovi integralski formuli je:

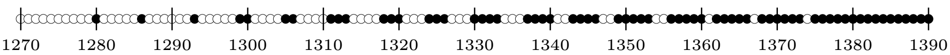
$$P(S \geq 0{\cdot}15\, n) \approx \frac{1}{2} - \Phi\left(\frac{0{\cdot}15\, n - n/6}{\sqrt{n \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right) = \frac{1}{2} + \Phi\left(\sqrt{\frac{n}{500}}\right)\,.$$

Torej bo število metov ustrezalo približno tedaj, ko bo $\sqrt{n/500} \geq \Phi^{-1}(0{\cdot}45)$ oziroma $n \geq 500\bigl(\Phi^{-1}(0{\cdot}45)\bigr)^2 \doteq 1352{\cdot}77$, torej $n \geq 1353$.

V resnici je najmanjše možno število metov, ki ustrezajo zahtevi, že 1280. Ne ustreza pa *vsako* število izdelkov, ki je večje ali enako 1280: prvo število, od katerega naprej vsako število naročenih ustreza, je 1375. Nekaj točnih verjetnosti:

$$\begin{aligned} n = 1280: &\quad P(S \geq 192) \doteq 0{\cdot}9507721 \\ n = 1281: &\quad P(S \geq 193) \doteq 0{\cdot}9437958 \\ n = 1374: &\quad P(S \geq 207) \doteq 0{\cdot}9497926 \\ n = 1375: &\quad P(S \geq 207) \doteq 0{\cdot}9510002 \end{aligned}$$

Na naslednji sliki so prikazana števila naročenih izdelkov, ki ustrezajo (polni krogci) in števila, ki ne ustrezajo (prazni krogci).



4. 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:

$$\begin{aligned} P(H_0) &= \frac{10}{45}, && P(U = 0 \mid H_0) = 1\,, && P(U = 1 \mid H_0) = 0\,, && P(U = 2 \mid H_0) = 0\,, \\ P(H_1) &= \frac{25}{45}, && P(U = 0 \mid H_1) = 0{\cdot}8\,, && P(U = 1 \mid H_1) = 0{\cdot}2\,, && P(U = 2 \mid H_1) = 0 \\ P(H_2) &= \frac{10}{45}, && P(U = 0 \mid H_2) = 0{\cdot}64\,, && P(U = 1 \mid H_2) = 0{\cdot}32\,, && P(U = 2 \mid H_2) = 0{\cdot}04\,. \end{aligned}$$

Torej je:

$$
\begin{aligned}
P(U=0) &= \frac{10}{45}\cdot 1 + \frac{25}{45}\cdot 0\text{˙}8 + \frac{10}{45}\cdot 0\text{˙}64 \doteq 0\text{˙}8089\,,\\
P(U=1) &= \frac{10}{45}\cdot 0 + \frac{25}{45}\cdot 0\text{˙}2 + \frac{10}{45}\cdot 0\text{˙}32 \doteq 0\text{˙}1822\,,\\
P(U=2) &= \frac{10}{45}\cdot 0 + \frac{25}{45}\cdot 0 + \frac{10}{45}\cdot 0\text{˙}04 \doteq 0\text{˙}0089\,,
\end{aligned}
$$

oziroma približno:

$$
U \sim \begin{pmatrix} 0 & 1 & 2 \\ 0\text{˙}8089 & 0\text{˙}1822 & 0\text{˙}0089 \end{pmatrix}.
$$

Opomba: seveda bi bilo dovolj izračunati $P(U=1)$ in $P(U=2)$ in nato upoštevati, da je vsota verjetnosti enaka 1.

Rešitve kolokvija iz verjetnosti in statistike z dne 12. 1. 2012

Matematika – univerzitetni študij

1. Za $k = 1, 2, \ldots 9$ velja:

$$\begin{aligned} P(D = k) &= \sum_{n=0}^{\infty} P\left(k \cdot 10^n \le \frac{1}{U} \le (k+1) \cdot 10^n\right) = \\ &= \sum_{n=0}^{\infty} P\left(\frac{1}{(k+1) \cdot 10^n} \le U \le \frac{1}{k \cdot 10^n}\right) = \\ &= \sum_{n=0}^{\infty} 10^{-n}\left(\frac{1}{k} - \frac{1}{k+1}\right) = \\ &= \frac{10}{9k(k+1)} \,. \end{aligned}$$

2. Iz $Y = X\left(\dfrac{1}{Z} - 2\right)$ dobimo:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x)\, f\left(x\left(\frac{1}{z} - 2\right)\right) \left|\frac{x}{z^2}\right| \mathrm{d}x \,.$$

Očitno Z skoraj gotovo zavzame vrednosti iz $(0, 1/2)$. Za z iz tega intervala po krajšem računu dobimo:

$$f_Z(z) = \frac{\lambda^2}{z^2} \int_0^{\infty} e^{-\lambda(1/z-1)x}\, x \,\mathrm{d}x = \frac{1}{(1-z)^2}$$

(neodvisno od λ – zakaj?) Sledi:

$$f_Z(z) = \begin{cases} \frac{1}{(1-z)^2} & ;\ 0 < z < \frac{1}{2} \\ 0 & ;\ \text{sicer} \end{cases} \,.$$

3. *Prvi način.* Obstaja $\binom{9}{3} = 84$ razporeditev treh križcev v 9 polj. Preštejmo razporeditve pri posameznih vrednostih slučajne spremenljivke S:

- $S = 0$: vse vrstice in vsi stolpci so zapolnjeni, kar ustreza razporeditvam, pri katerih v vsako vrstico damo po en križec, obenem pa jih damo tudi v same različne stolpce: 6 razporeditev.
- $S = 1$: lahko imamo nič praznih vrstic in en prazen stolpec ali pa obratno. Za vsako možnost imamo po $3 \cdot 2 \cdot 3 = 18$ razporeditev, skupaj 36.
- $S = 2$: lahko imamo dve prazni vrstici in nič praznih stolpcev (3 razporeditve), obratno (spet 3 razporeditve) ali pa eno prazno vrstico in en prazen stolpec ($3 \cdot 3 \cdot 4 = 36$ razporeditev). Skupaj 42 razporeditev.

Dovolj je sicer prešteti razporeditve za samo dve vrednosti slučajne spremenljivke S. Torej je:

$$S \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{6}{84} & \frac{36}{84} & \frac{42}{84} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{14} & \frac{6}{14} & \frac{7}{14} \end{pmatrix}$$

in $E(S) = \dfrac{20}{14} = \dfrac{10}{7}$.

Drugi način. Označimo z X število vrstic, z Y pa število stolpcev, ki ostanejo prazni. Tedaj je $S = X + Y$. Spet preštevamo razporeditve:

- $X = 0$: v vsako vrstico damo po en križec: $3^3 = 27$ razporeditev.
- $X = 1$: v eno vrstico gresta dva križca, v eno en križec, v eno pa noben križec. Dobimo $3 \cdot 2 \cdot 3 \cdot 3 = 54$ razporeditev.
- $S = 2$: v eno vrstico gredo vsi trije križci: 3 razporeditve.

Torej sta slučajni spremenljivki X in Y obe porazdeljeni diskretno s shemo:

$$\begin{pmatrix} 0 & 1 & 2 \\ \frac{27}{84} & \frac{54}{84} & \frac{3}{84} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 \\ \frac{9}{28} & \frac{18}{28} & \frac{1}{28} \end{pmatrix}$$

iz katere dobimo $E(X) = E(Y) = \dfrac{20}{28} = \dfrac{5}{7}$ in nazadnje še $E(S) = E(X) + E(Y) = \dfrac{10}{7}$.

Tretji način. Pišemo $S = X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3$, kjer je:

$$X_i = \begin{cases} 1 & ; i\text{-ta vrstica je prazna} \\ 0 & ; i\text{-ta vrstica ni prazna} \end{cases}, \qquad Y_j = \begin{cases} 1 & ; j\text{-ta vrstica je prazna} \\ 0 & ; j\text{-ta vrstica ni prazna} \end{cases}.$$

Velja $E(X_i) = P(i\text{-ta vrstica je prazna}) = \dfrac{\binom{6}{3}}{\binom{9}{3}} = \dfrac{20}{84} = \dfrac{5}{21}$ in podobno tudi $E(Y_j) = \dfrac{5}{21}$. Sledi $E(S) = 6 \cdot \dfrac{5}{21} = \dfrac{10}{7}$.

4. Najprej izračunamo gostoto: $f_X(x) = F_X'(x) = e^{-e^{-x}}\, e^{-x}$. Sledi:

$$E\big(e^{X/2}\big) = \int_{-\infty}^{\infty} e^{e^{-x}}\, e^{-x/2}\,\mathrm{d}x\,.$$

S substitucijo $t = e^{-x/2}$ dobimo:

$$E\big(e^{X/2}\big) = 2\int_0^\infty e^{-t^2}\,\mathrm{d}t \tag{$*$}$$

Od tod naprej gre na vsaj dva načina.

Prvi način: zaradi sodosti lahko $(*)$ pišemo tudi v obliki:

$$E\big(e^{X/2}\big) = \int_{-\infty}^{\infty} e^{-t^2}\,\mathrm{d}t$$

in s pomočjo gostote normalne porazdelitve $\mathrm{N}(0, \sigma)$:

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-t^2/(2\sigma^2)}$$

za $\sigma = 1/\sqrt{2}$ dobimo:

$$E\big(e^{X/2}\big) = \int_{-\infty}^{\infty} e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}\,.$$

Drugi način: z nadaljnjo substitucijo $s = t^2$ dobimo:

$$E\big(e^{X/2}\big) = \int_0^{\infty} \frac{e^{-s}}{\sqrt{s}}\,\mathrm{d}s = \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\,.$$

Rešitve kolokvija iz verjetnosti in statistike z dne 12. 4. 2012

Matematika – univerzitetni študij

1. Ker mora biti vsota vseh verjetnosti enaka 1, velja $p = 1/12$. Nadalje je:

$$E(X) = \frac{2}{3}\,, \quad E(Y) = \frac{1}{3} + \frac{a}{12}\,, \quad E(XY) = \frac{1}{12} + \frac{a}{12}\,.$$

Slučajni spremenljivki X in Y sta torej nekorelirani natanko tedaj, ko velja $\dfrac{1}{12} + \dfrac{a}{12} = \dfrac{2}{3}\left(\dfrac{1}{3} + \dfrac{a}{12}\right)$, torej $a = 5$.

2. a) *Prvi način.* Iz rodovne funkcije binomske porazdelitve dobimo pogojno rodovno funkcijo slučajne spremenljivke S:

$$G_{S|N}(z) = \left(\frac{1+z}{2}\right)^{2N}\,,$$

brezpogojna rodovna funkcija pa je njeno matematično upanje:

$$G_S(z) = E\big[G_{S|N}(z)\big] = \sum_{n=0}^{\infty} \frac{\lambda^n\,e^{-\lambda}}{n!}\left(\frac{1+z}{2}\right)^{2n} = \exp\left[\left(\frac{z+1}{2}\right)^2 - \lambda\right] =$$
$$= e^{\lambda(z^2+2z-3)/4}\,.$$

Drugi način. Uporabimo, da je binomska porazdelitev $\mathrm{b}(2n, 1/2)$ dobljena kot porazdelitev vsote $X_1 + X_2 + \cdots + X_n$, kjer so $X_1, \ldots X_n$ neodvisne in porazdeljene binomsko $\mathrm{b}(2, 1/2)$. Rodovna funkcija te porazdelitve je $G_X(z) = \left(\dfrac{1+z}{2}\right)^2$. Ker se porazdelitev slučajne spremenljivke S ujema s porazdelitvijo slučajne vsote $X_1 + X_2 + \cdots + X_N$, mora biti njena rodovna funkcija enaka $G_S(z) = G_N(G_X(z))$. Iz $G_N(z) = e^{\lambda(z-1)}$ dobimo:

$$G_S(z) = \exp\left[\lambda\left(\left(\frac{z+1}{2}\right)^2 - 1\right)\right]\,,$$

kar je isto kot prej.

b) *Prvi način.* Z odvajanjem rodovne funkcije dobimo:

$$G_S'(z) = \frac{\lambda(z+1)}{2}\,e^{\lambda(\lambda^2+2z-3)/4}\,,$$

torej je $P(S = 1) = \dfrac{G'(0)}{1!} = \dfrac{\lambda}{2}\,e^{-3\lambda/4}$.

Drugi način. Iz razvoja rodovne funkcije v eksponentno vrsto dobimo:

$$G_S(z) = e^{-3\lambda/4}\,e^{\lambda(z^2+2z)/4} = e^{-3\lambda/4}\left(1 + \lambda\,\frac{z^2+2z}{4} + \frac{\lambda^2}{2}\left(\frac{z^2+2z}{4}\right)^2 + \cdots\right)\,,$$

od koder sledi $P(S = 1) = \dfrac{\lambda}{2}\, e^{-3\lambda/4}$.

Tretji način. Ker pogojno glede na N velja $S \sim \mathrm{b}(2N, 1/2)$, je tudi:

$$P(S = 1 \mid N) = \binom{2N}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{2N-1} = \frac{N}{2^{2N-1}}$$

in posledično:

$$P(S = 1) = E\left[\frac{N}{2^{2N-1}}\right] = \sum_{n=1}^{\infty} \frac{n}{2^{2n-1}} \, \frac{\lambda^n\, e^{-\lambda}}{n!} = \sum_{m=0}^{\infty} \frac{\lambda^{m+1}\, e^{-\lambda}}{2^{2m+1}\, m!} = \frac{\lambda}{2}\, e^{-3\lambda/4}\,.$$

3. Velja $E(X_i) = 0$ in $D(X_i) = 1/6$, torej $E(S) = 0$ in $D(S) = 25$. Sledi:

$$P(S \ge x) \approx \frac{1}{2} - \Phi\left(\frac{x}{5}\right)\,,$$

torej mora biti $x \approx 5\Phi^{-1}(0\text{˙}45) \doteq 8\text{˙}22427$.

Točen rezultat: 0˙224672.

4. a) Če z m označimo mediano, mora veljati:

$$\int_0^m 2u \, \mathrm{d}u = u^2 = \frac{1}{2}\,,$$

torej je $m = \sqrt{2}/2$.

b) *Prvi način.* Za $0 \le x \le 1$ velja:

$$\begin{aligned} F_{U_{(1)}}(x) &= P(U_1 \le x \text{ ali } U_2 \le x) = P(U_1 \le x) + P(U_2 \le x) - P(U_1 \le x, U_2 \le x) = \\ &= 2x^2 - x^4 \\ f_{U_{(1)}}(x) &= 4x - 4x^3\,, \end{aligned}$$

torej je $E\big(U_{(1)}\big) = \displaystyle\int_0^1 (4x^2 - 4x^4)\, \mathrm{d}x = \frac{8}{15}$. Nadalje za $0 \le x \le 1$ velja:

$$\begin{aligned} F_{U_{(2)}}(x) &= P(U_1 \le x, U_2 \le x) = P(U_1 \le x, U_2 \le x) = x^4 \\ f_{U_{(2)}}(x) &= 4x^3\,, \end{aligned}$$

torej je $E\big(U_{(2)}\big) = \displaystyle\int_0^1 4x^4 \, \mathrm{d}x = \frac{4}{5}$.

Drugi način. Velja:

$$\begin{aligned} E(U_{(1)}) &= 4 \iint_{[0,1]^2} \min\{x, y\}\, xy \, \mathrm{d}x \, \mathrm{d}y = \\ &= 4 \int_0^1 \int_0^x xy^2 \, \mathrm{d}y \, \mathrm{d}x + 4 \int_0^1 \int_x^1 x^2 y \, \mathrm{d}y \, \mathrm{d}x = \\ &= \frac{4}{3} \int_0^1 x^4 \, \mathrm{d}x + 2 \int_0^1 x^2 (1 - x^2) \, \mathrm{d}x = \\ &= \frac{8}{15} \end{aligned}$$

in:

$$
\begin{aligned}
E(U_{(2)}) &= 4 \iint_{[0,1]^2} \max\{x, y\}\, xy \, \mathrm{d}x \, \mathrm{d}y = \\
&= 4 \int_0^1 \int_0^x x^2 y \, \mathrm{d}y \, \mathrm{d}x + 4 \int_0^1 \int_x^1 xy^2 \, \mathrm{d}y \, \mathrm{d}x = \\
&= 2 \int_0^1 x^4 \, \mathrm{d}x + \frac{4}{3} \int_0^1 x(1 - x^3) \, \mathrm{d}x = \\
&= \frac{4}{5} \, .
\end{aligned}
$$

c) Cenilko za mediano slučajne spremenljivke X, ki je enaka $a\,\frac{\sqrt{2}}{2} + b$, nastavimo v obliki $\lambda X_{(1)} + \mu X_{(2)}$. Njeno matematično upanje je enako $\lambda\left(\frac{8}{15}a + b\right) + \mu\left(\frac{4}{5}a + b\right)$. Cenilka bo nepristranska natanko tedaj, ko se bo to dvoje ujemalo pri vseh a in b, to pa bo tedaj, ko bo:

$$
\begin{aligned}
\frac{8}{15}\,\lambda + \frac{4}{5}\,\mu &= \frac{\sqrt{2}}{2} \, , \\
\lambda + \mu &= 1 \, .
\end{aligned}
$$

torej ko bo $\lambda = 3 - \frac{15\sqrt{2}}{8}$ in $\mu = \frac{15\sqrt{2}}{8} - 2$.

Rešitve kolokvija iz verjetnosti in statistike z dne 7. 6. 2012

Matematika – univerzitetni študij

1. Dana porazdelitev ima gostoto:

$$f(x) = \frac{1}{a\sqrt{2\pi}}\, e^{-(x-a)^2/(2a^2)} = \frac{1}{a\sqrt{2\pi}} \exp\left(-\frac{x^2}{2a^2} + \frac{x}{a} - \frac{1}{2}\right),$$

kar nam da verjetje:

$$L = \frac{1}{a^n (2\pi)^{n/2}} \exp\left(-\frac{1}{2a^2}\sum_{i=1}^n X_i^2 + \frac{1}{a}\sum_{i=1}^n X_i - \frac{n}{2}\right).$$

Le-to je ves čas strogo pozitivno in gre proti nič, brž ko gre a proti nič ali proti neskončno. Ugodno je gledati njegov logaritem:

$$\ln L = -\frac{n}{2}\ln(2\pi) - n\ln a - \frac{1}{2a^2}\sum_{i=1}^n X_i^2 + \frac{1}{a}\sum_{i=1}^n X_i - \frac{n}{2},$$

ki ima odvod:

$$\frac{\mathrm{d}\ln L}{\mathrm{d}a} = -\frac{n}{a} + \frac{1}{a^3}\sum_{i=1}^n X_i^2 - \frac{1}{a^2}\sum_{i=1}^n X_i\,.$$

Edina ničla odvoda na intervalu $(0,\infty)$ je:

$$\hat{a} = -\frac{1}{2n}\sum_{n=1}^{\infty} X_i + \sqrt{\left(\frac{1}{2n}\sum_{n=1}^{\infty} X_i\right)^2 + \frac{1}{n}\sum_{i=1}^{\infty} X_i^2}$$

in to je iskana cenilka za a.

2. Iz zapisa gostote:

$$f(x) = \frac{1}{x\,\mathrm{B}(a,1-a)} \exp\left(a\ln\frac{x}{1-x}\right)$$

(za $0 < x < 1$) razberemo, da gre za eksponentno družino, katere parametrični prostor je odprta množica $(0,1)$ (za vse a iz te množice ima porazdelitev smisel). Nepristranska cenilka z najmanjšo možno disperzijo bo torej vsaka funkcija pripadajoče zadostne statistike $\ln\frac{X}{1-X}$, ki bo nepristranska cenilka za a. Ker je funkcija $x \mapsto \ln\frac{x}{1-x}$ strogo naraščajoča, je dovolj iskati kar funkcije opažanja X. Iz:

$$\begin{aligned} E(X) &= \frac{1}{\mathrm{B}(a,1-a)}\int_0^1 x^a(1-x)^{-a}\,\mathrm{d}x = \frac{\mathrm{B}(a+1,1-a)}{\mathrm{B}(a,1-a)} = \\ &= \frac{\Gamma(a+1)\,\Gamma(1-a)}{\Gamma(2)}\,\frac{\Gamma(1)}{\Gamma(a)\,\Gamma(1-a)} = \frac{\Gamma(a+1)}{\Gamma(a)} = a \end{aligned}$$

sledi, da je iskana cenilka kar opažanje samo, t. j. X.

3. 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:

$$\inf_{0 \le \theta \le 1} P_\theta(\theta \in I) = \min\left\{ \frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_1^2 - b_1 + 1} \right\} .$$

Pogoj $\inf_{0 \le \theta \le 1} P_\theta(\theta \in I) = \beta$ bo zagotovo izpolnjen, če bo:

$$\frac{b_2}{b_2^2 - b_2 + 1} = \frac{b_1^2}{b_1^2 - b_1 + 1} = \beta .$$

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). No, dovolj je, da je velja le ena enakost, drugi parameter pa se lahko poveča (a potem dobimo daljši interval zaupanja za enako stopnjo zaupanja).

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:

$$\inf_{0\leq\theta\leq1} P_\theta(\theta \in I) = \min\left\{\frac{b_2}{b_2^2 - b_2 + 1}, \frac{b_1^2}{b_2^2 - b_2 + 1}\right\}.$$

Pogoj $\inf_{0\leq\theta\leq1} P_\theta(\theta \in I) = \beta$ bo zagotovo izpolnjen, če bo:

$$\frac{2b_1^2 - 2b_1 + 1}{b_1^2 - b_1 + 1} = \frac{b_2^2}{b_2^2 - b_2 + 1} = \beta\,.$$

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 $\beta = 0{\cdot}95$ dobimo $b_1 \doteq 0{\cdot}9499$ in $b_2 \doteq 0{\cdot}9523$ (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).

Možno pa je izbrati tudi $b_1 = b_2$. V tem primeru dobimo:

$$b_1 = b_2 = \frac{-\beta + \sqrt{\beta(5\beta-4)}}{2(1-\beta)}$$

(torej $0{\cdot}9523$ pri $\beta = 0{\cdot}95$), kar je kombinacija prejšnjih dveh možnosti.

4. Označimo naše opažanje, t. j. število metov, pri katerih pade šestica, z X. Velja $X \sim \mathrm{b}(13, \theta)$, kar je enoparametrična eksponentna družina, katere naravni parameter je naraščajoča funkcija parametra θ, pripadajoča zadostna statistika pa je natančno X. Torej ničelno hipotezo zavrnemo, če je opažanje X preveliko. Ker je $P_{1/6}(X \geq 5) \doteq 0{\cdot}0512$, hipoteze ne moremo zavrniti – vsaj če testa ne randomiziramo.

Če pa dopustimo randomizacijo, nam verjetnost $P_{1/6}(X > 5) \doteq 0{\cdot}0127$ pove, da lahko hipotezo zavrnemo z verjetnostjo:

$$\frac{0{\cdot}05 - P_{1/6}(X > 5)}{P_{1/6}(X \geq 5) - P_{1/6}(X > 5)} = \frac{0{\cdot}05 - P_{1/6}(X > 5)}{P_{1/6}(X = 5)} \doteq 0{\cdot}97\,.$$

Rešitve izpita iz verjetnosti in statistike z dne 18. 6. 2012

Matematika – univerzitetni študij

1. 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:

$$P(J_3 \mid N_5) = \frac{\frac{1}{3\cdot 4} \cdot \frac{1}{4}}{\frac{1}{3\cdot 4} \cdot \frac{1}{4} + \frac{1}{4\cdot 5} \cdot \frac{1}{5} + \frac{1}{5\cdot 6} \cdot \frac{1}{6}} = \frac{75}{131} \doteq 0\text{'}573\,.$$

2. *Prvi način*: s pomočjo kumulativne porazdelitvene funkcije. Če označimo $Y = X - \lfloor X \rfloor$ in je $0 \le y \le 1$, je $Y \le y$ natanko tedaj, ko je $n \le X \le n + y$ za neki $n \in \mathbb{N}$. Sledi:

$$\begin{aligned} F_Y(y) &= \sum_{n=1}^{\infty} P(n \le X \le n+y) = \frac{1}{\ln 2} \sum_{n=1}^{\infty} \int_n^{n+y} \frac{1}{x(x+1)}\,\mathrm{d}x = \\ &= \frac{1}{\ln 2} \sum_{n=1}^{\infty} \ln \left(\frac{n+y}{n+y+1}\, \frac{n+1}{n} \right)\,. \end{aligned}$$

Iz delne vsote:

$$\sum_{n=1}^{m} \ln\left(\frac{n+y}{n+y+1}\,\frac{n+1}{n}\right) = \ln\frac{(1+m)(1+y)}{1+y+m}$$

po limitiranju dobimo $F_Y(y) = \dfrac{\ln(1+y)}{\ln 2}$. Natančneje, kumulativna porazdelitvena funkcija je enaka:

$$F_Y(y) = \begin{cases} 0 & ;\ y \le 0 \\ \dfrac{\ln(1+y)}{\ln 2} & ;\ 0 \le y \le 1 \\ 1 & ;\ y \ge 1 \end{cases}.$$

To tudi pomeni, da je slučajna spremenljivka Y zvezna z gostoto:

$$f_Y(y) = \begin{cases} \dfrac{1}{(1+y)\ln 2} & ;\ 0 < y < 1 \\ 0 & ;\ \text{sicer} \end{cases}.$$

Drugi način: izračunamo neposredno gostoto. Funkcija $h(x) = x - \lfloor x \rfloor$ je namreč v slučajni točki X z verjetnostjo 1 zvezno odvedljiva, njen odvod pa je enak 1. Sledi:

$$f_Y(y) = \sum_{x;h(x)=y} f_X(x)\,.$$

Za $0 < y < 1$ to pomeni:

$$f_Y(y) = \sum_{n=1}^{\infty} f_X(y+n) = \frac{1}{\ln 2}\sum_{n=1}^{\infty}\frac{1}{(n+y)(n+y+1)} = $$
$$= \frac{1}{\ln 2}\sum_{n=1}^{\infty}\left(\frac{1}{(n+y)} - \frac{1}{(n+y+1)}\right).$$

Iz delne vsote:

$$\sum_{n=1}^{m}\left(\frac{1}{(n+y)} - \frac{1}{(n+y+1)}\right) = \frac{1}{1+y} - \frac{1}{1+y+m}$$

po limitiranju dobimo $f_Y(y) = 1/(1+y)$, kar je enako kot prej.

3. a) Iz:

$$E\big(e^{-X}\big) = \lambda\int_0^{\infty} e^{-x}\,e^{-\lambda x}\,\mathrm{d}x = \frac{\lambda}{\lambda+1}$$

dobimo:

$$\frac{1}{\lambda+1} = 1 - \frac{\lambda}{\lambda+1} = 1 - E\big(e^{-X}\big) = E\big(1-e^{-X}\big)\,,$$

torej je $1 - e^{-X}$ nepristranska cenilka za $1/(\lambda+1)$. Ker gre za eksponentno družino s parametričnim prostorom, ki je odprta množica, ima ta cenilka tudi enakomerno najmanjšo disperzijo.

b) Za eno opažanje X je X tudi statistika, ki pripada ustreznemu zapisu eksponentne družine. Za dve neodvisni opažanji X_1 in X_2 je pripadajoča statistika vsota X_1+X_2, ki ima porazdelitev Gama$(2,\lambda)$, torej zvezno z gostoto:

$$f(x) = \lambda^2\, x\, e^{-\lambda x}\,.$$

Iz prejšnje izpeljave cenilke dobimo:

$$\frac{1}{\lambda+1} = \lambda \int_0^\infty \bigl(1-e^{-x}\bigr) e^{-\lambda x}\,\mathrm{d}x\,,$$

torej je:

$$\frac{\lambda}{\lambda+1} = \lambda^2 \int_0^\infty \frac{1-e^{-x}}{x}\, x\, e^{-\lambda x}\,\mathrm{d}x = E\left(\frac{1-e^{-(X_1+X_2)}}{X_1+X_2}\right)\,,$$

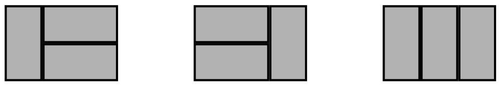
od koder končno dobimo iskano cenilko $1-\dfrac{1-e^{-(X_1+X_2)}}{X_1+X_2}$.

4. $\bar{X} = 1{\cdot}69$, $\quad S_{\mathrm{p}} \doteq 0{\cdot}940$, $\quad t_{0.975}(99) \doteq 1{\cdot}98$.
Interval zaupanja: $1{\cdot}50 < X < 1{\cdot}88$ (spodnja meja je zaokrožena navzdol, zgornja pa navzgor).

Rešitve izpita iz verjetnosti in statistike z dne 20. 8. 2012

Matematika – univerzitetni študij

1. a) Sprejeta je lahko ena, dve ali pa tri palice. Tri palice so lahko sprejete, če padejo na enega od naslednjih treh načinov:



Verjetnost tega dogodka je $\dfrac{3 \cdot 3!}{7^3} = \dfrac{18}{343}$.

Oglejmo si zdaj dogodek, da je sprejeta natanko ena palica. Očitno je to prva palica, ki pade, drugi dve pa morata pasti tako, da se z njo prekrivata. Če pade palica vodoravno (4 možnosti), se lahko naslednja z njo prekriva na 4 načine. Če palica pade navpično ob strani (2 možnosti), se lahko naslednja z njo prekriva na 3 načine, če pa pade navpično na sredino (1 možnost), pa se lahko naslednja z njo prekriva na 5 načinov. Torej je verjetnost, da je sprejeta ena sama palica, enaka:

$$\frac{4 \cdot 4^2 + 2 \cdot 3^2 + 1 \cdot 5^2}{7^3} = \frac{107}{343}\,.$$

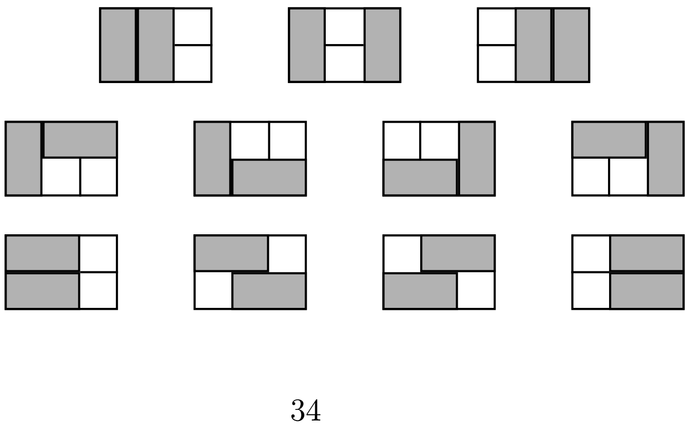
Verjetnost, da sta sprejeti dve palici, je enaka $1 - \dfrac{18}{343} - \dfrac{107}{343} = \dfrac{218}{343}$. Če torej z X označimo število sprejetih palic, velja:

$$X \sim \begin{pmatrix} 1 & 2 & 3 \\ \frac{107}{343} & \frac{218}{343} & \frac{18}{343} \end{pmatrix} \doteq \begin{pmatrix} 1 & 2 & 3 \\ 0{\cdot}312 & 0{\cdot}636 & 0{\cdot}052 \end{pmatrix}\,.$$

b) Označimo z A_1, A_2 in A_3 dogodke, da je bila prva, druga oz. tretja palica sprejeta. Izračunati moramo:

$$P(A_1 \cap A_2 \mid X = 2) = \frac{P(A_1 \cap A_2 \cap A_3^c)}{P(X = 2)} = \frac{P(A_1 \cap A_2) - P(X = 3)}{P(X = 2)}\,.$$

Dogodek $A_1 \cap A_2$ se lahko zgodi na naslednjih 11 načinov:



in ima verjetnost $\dfrac{11 \cdot 2!}{49} = \dfrac{22}{49}$. Sledi:

$$P(A_1 \cap A_2 \mid X = 2) = \frac{\frac{22}{49} - \frac{18}{343}}{\frac{218}{343}} = \frac{34}{53} \doteq 0{\cdot}642\,.$$

2. Nalogo najlažje rešimo s pomočjo rodovnih funkcij: če je G_1 rodovna funkcija slučajnih spremenljivk X_i, G_2 pa rodovna funkcija slučajne spremenljivke $2N$, je rodovna funkcija iskane vsote S kompozitum $G_2 \circ G_1$. 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 pa je funkcija:

$$s \mapsto \frac{\frac{3}{4}s}{1 - \frac{1}{4}s} = \frac{3s}{4-s}\,,$$

torej je $G_2(s) = \dfrac{3s^2}{4-s^2}$. Rodovna funkcija slučajne vsote S je tako funkcija:

$$s \mapsto \frac{3\left(\frac{2s}{3-s}\right)^2}{4 - \left(\frac{2s}{3-s}\right)^2} = \frac{s^2}{3-2s} = \frac{\frac{1}{3}s^2}{1 - \frac{2}{3}s}\,,$$

torej ima S geometrijsko porazdelitev $\mathrm{Geom}(1/3)$, pomaknjeno za 1 v desno. Z drugimi besedami, $S - 1 \sim \mathrm{Geom}(1/3)$.

3. Iz zapisa gostote:

$$f(x) = \frac{1}{x\,\mathrm{B}(a, 1-a)} \exp\left(a \ln \frac{x}{1-x}\right)$$

(za $0 < x < 1$) razberemo, da gre za eksponentno družino, katere parametrični prostor je odprta množica $(0,1)$ (za vse a iz te množice ima porazdelitev smisel). Nepristranska cenilka z najmanjšo možno disperzijo bo torej vsaka funkcija pripadajoče zadostne statistike $\ln \frac{X}{1-X}$, ki bo nepristranska cenilka za a^2. Ker je funkcija $x \mapsto \ln \frac{x}{1-x}$ strogo naraščajoča, je dovolj iskati kar funkcije opažanja X. Iz:

$$\begin{aligned}
E(X) &= \frac{1}{\mathrm{B}(a,1-a)} \int_0^1 x^a (1-x)^{-a}\,\mathrm{d}x = \frac{\mathrm{B}(a+1,1-a)}{\mathrm{B}(a,1-a)} = \\
&= \frac{\Gamma(a+1)\,\Gamma(1-a)}{\Gamma(2)} \frac{\Gamma(1)}{\Gamma(a)\,\Gamma(1-a)} = \frac{\Gamma(a+1)}{\Gamma(a)} = a\,, \\
E(X^2) &= \frac{1}{\mathrm{B}(a,1-a)} \int_0^1 x^{a+1} (1-x)^{-a}\,\mathrm{d}x = \frac{\mathrm{B}(a+2,1-a)}{\mathrm{B}(a,1-a)} = \\
&= \frac{\Gamma(a+2)\,\Gamma(1-a)}{\Gamma(3)} \frac{\Gamma(1)}{\Gamma(a)\,\Gamma(1-a)} = \frac{\Gamma(a+2)}{2\,\Gamma(a)} = \frac{a(a+1)}{2}
\end{aligned}$$

sledi, da je iskana cenilka $2X^2 - X$.

4. $\bar{X} = 88\dot{\ }3, \quad S_{\mathrm{p}} \doteq 18\dot{\ }54, \quad t_{0.995}(9) \doteq 3\dot{\ }25.$
Interval zaupanja: $69\dot{\ }2 < X < 107\dot{\ }4$ (spodnja meja je zaokrožena navzdol, zgornja pa navzgor).

Rešitve izpita iz verjetnosti in statistike z dne 3. 9. 2012

Matematika – univerzitetni študij

1. a) Iskana verjetnost je enaka pogojni verjetnosti $P(A \mid B)$, kjer je A dogodek, da Rajko v posameznem strelu zadene 10 krogov, B pa dogodek, da zadene več kot 7 krogov (glej pojasnilo spodaj). Ker je $A \subseteq B$, je:

$$P(A \mid B) = \frac{P(A)}{P(B)} = \frac{\frac{1}{66}}{\frac{1}{11}} = \frac{1}{6}\,.$$

Pojasnilo. Denimo, da izvajamo zaporedje neodvisnih poskusov in da lahko pri vsakem pride do opažanja A, prav tako pa tudi do opažanja B. Možne so vse kombinacije ($\bar{A} \cap \bar{B}$, $A \setminus B$, $B \setminus A$ in $A \cap B$) in vse se pri posameznem poskusu zgodijo s fiksnimi verjetnostmi, ki jih označimo s $P(\bar{A} \cap \bar{B})$, $P(A \setminus B)$, $P(B \setminus A)$ in $P(A \cap B)$. To so verjetnosti v verjetnostnem prostoru, ki se nanaša na *posamezen* poskus. Privzemimo, da je $P(B) > 0$.

Poleg prej omenjenega pa gledamo še verjetnostni prostor, ki se nanaša na celotno zaporedje poskusov. Namesto A in B imamo na njem dogodke A_n in B_n: to sta dogodka, da v n-tem poskusu pride do opažanja A oziroma B. Verjetnost na tem prostoru označimo s $\tilde{P}$ (to je torej produkt števno neskončno verjetnostnih mer P). Zanima nas $\tilde{P}(C_0)$, kjer je C_0 dogodek, da pri prvem poskusu, pri katerem pride do opažanja B, pride tudi do opažanja A. Dokazali bomo dokaj intuitivno trditev:

$$\tilde{P}(C_0) = P(A \mid B)\,. \tag{$*$}$$

To lahko storimo na vsaj dva načina.

Prvi način: dogodek C_0 zapišemo kot disjunktno unijo dogodkov $\bar{B}_1 \cap \cdots \cap \bar{B}_{n-1} \cap A_n \cap B_n$, kjer je $n = 1, 2, 3, \ldots$. Sledi:

$$\tilde{P}(C_0) = \sum_{n=1}^{\infty} \tilde{P}(\bar{B}_1 \cap \cdots \cap \bar{B}_{n-1} \cap A_n \cap B_n) = \sum_{n=1}^{\infty} (1 - P(B))^{n-1}\, P(A \cap B) = \frac{P(A \cap B)}{P(B)} = P(A \mid B)\,.$$

Drugi način: s pomočjo rekurzivne zveze. Dogodek C_0 zapišemo kot disjunktno unijo dogodkov $A_1 \cap B_1$ in $\bar{B}_1 \cap C_0^{\rightarrow}$, kjer je $C_0^{\rightarrow}$ dogodek, da pri prvem poskusu *od drugega naprej*, pri katerem pride do opažanja B, pride tudi do opažanja A. Tedaj velja $\tilde{P}(C_0^{\rightarrow} \mid \bar{B}_1) = P(C_0)$ (intuitivno je to zelo jasno, teoretična utemeljitev pa sloni na t. i. *Dynkinovi lemi*, znani tudi kot *izrek* π–λ). sledi:

$$\tilde{P}(C_0) = \tilde{P}(A_1 \cap B_1) + \tilde{P}(\bar{B}_1)\tilde{P}(C_0) = P(A \cap B) + (1 - P(B))\tilde{P}(C_0)\,,$$

od koder dobimo:

$$\tilde{P}(C_0) = \frac{P(A \cap B)}{P(B)} = P(A \mid B)\,.$$

b) Slučajna spremenljivka R lahko zavzame vrednosti 1, 2 ali 3. Dogodke, povezane z njenimi vrednostmi, lahko prikažemo z naslednjimi vzorci rekordov s pripadajočimi verjetnostmi:

$$\begin{array}{lll}
R = 1: & 7, \ldots 10, \ldots & \frac{1}{6} \\
R = 2: & 7, \ldots 8, \ldots 10, \ldots & \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6} \\
 & 7, \ldots 9, \ldots 10, \ldots & \frac{1}{3} \cdot 1 = \frac{1}{3} \\
R = 3: & 7, \ldots 8, \ldots 9, \ldots 10, \ldots & \frac{1}{2} \cdot \frac{2}{3} \cdot 1 = \frac{1}{3}
\end{array}$$

(dogodek $\{R = 1\}$ je natanko dogodek iz prejšnje točke). Torej je:

$$R \sim \begin{pmatrix} 1 & 2 & 3 \\ 1/6 & 1/2 & 1/3 \end{pmatrix}.$$

c) Iskana verjetnost je kvocient verjetnosti vzorca $7, 8, \ldots 10, \ldots$ in verjetnosti dogodka, da je $R = 2$, torej je enaka:

$$\frac{\frac{1}{22} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{33}\,.$$

2. *Prvi način.* Za $x > 0$ velja:

$$f_{U|X}(u \mid x) = \begin{cases} \frac{1}{x} & ; x \le u \le 2x \\ 0 & ; \text{sicer} \end{cases} .$$

Od tod se vidi, da je U skoncentrirana na intervalu $(0, \infty)$, saj so tam skoncentrirane vse pogojne gostote. Za $u > 0$ velja:

$$f_U(u) = \int_0^\infty f_{U|X}(u \mid x) f_X(x) \, \mathrm{d}x = \int_{u/2}^u \frac{1}{x} \cdot x \, e^{-x} \, \mathrm{d}x = e^{-u/2} - e^{-u} .$$

Torej je:

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

Matematično upanje lahko izračunamo bodisi kot:

$$E(U) = \int_{-\infty}^\infty u f_U(u) \, \mathrm{d}u = \int_0^\infty u \big(e^{-u/2} - e^{-u} \big) \, \mathrm{d}u = 3$$

bodisi iz matematičnega upanja enakomerne porazdelitve:

$$E(U \mid X) = \frac{3X}{2} , \qquad E(U) = \frac{3}{2} \, E(X) = 3 .$$

Drugi način. Opazimo, da je slučajna spremenljivka $V := U/X$ pogojno na X porazdeljena enakomerno na intervalu $(1, 2)$, neodvisno od X. Torej je V neodvisna od X. Iščemo porazdelitev produkta $U = VX$. Za $u > 0$ velja:

$$f_U(u) = \int_{-\infty}^\infty f_X(x) f_V\left(\frac{u}{x}\right) \left| \frac{1}{x} \right| \mathrm{d}x = \int_{u/2}^u x \, e^{-x} \, \frac{1}{x} \, \mathrm{d}x = e^{-u/2} - e^{-u}$$

ali tudi:

$$f_U(u) = \int_{-\infty}^\infty f_V(v) f_X\left(\frac{u}{v}\right) \left| \frac{1}{v} \right| \mathrm{d}v = \int_1^2 \frac{u}{v^2} \, e^{-u/v} \, \mathrm{d}v = u \int_{1/2}^1 e^{-tu} \, \mathrm{d}t = e^{-u/2} - e^{-u} .$$

Matematično upanje je seveda enako:

$$E(U) = E(X) \, E(V) = \frac{3}{2} \cdot 2 = 3 .$$

3. Iz zapisa gostote:

$$f(x) = a \, e^{(a-1) \ln x}$$

(za $0 < x < 1$) razberemo, da gre za eksponentno družino, katere parametrični prostor je odprta množica $(0, \infty)$ (za vse a iz te množice ima porazdelitev smisel). Nepristranska cenilka z najmanjšo možno disperzijo bo torej vsaka funkcija pripadajoče zadostne statistike $\ln X$, ki bo nepristranska cenilka za a^2. Ker je funkcija $x \mapsto \ln x$ strogo naraščajoča, je dovolj iskati kar funkcije opažanja X. Iz:

$$E(X^2) = a \int_0^1 x^{a+1} \, \mathrm{d}x = \frac{a}{a+2} = 1 - \frac{2}{a+2}$$

sledi, da je iskana cenilka $(1 - X^2)/2$.

4. Ker ima spremenljivka X disperzijo $1/12$ *ne glede na* a in ker je vzorec velik, se lahko naslonimo na konstrukcijo intervala zaupanja za matematično upanje normalne porazdelitve z znano disperzijo:

$$\bar{X} - \frac{c\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{c\sigma}{\sqrt{n}}\,,$$

kjer je μ matematično upanje, c pa je kvantil standardne normalne porazdelitve za verjetnost $(1+\beta)/2 = 0^{\cdot}975$, torej $c = z_{0.975} \doteq 1^{\cdot}96$. Parameter σ je standardni odklon, ki je, kot smo že omenili, enak $1/\sqrt{12}$, matematično upanje μ pa je enako $a + \frac{1}{2}$. Iskani asimptotični interval zaupanja je torej:

$$\bar{X} - \frac{1}{2} - \frac{z_{0.975}}{\sqrt{12n}} < a < \bar{X} - \frac{1}{2} + \frac{z_{0.975}}{\sqrt{12n}}$$

oziroma približno:

$$\bar{X} - 0^{\cdot}5 - \frac{0^{\cdot}57}{\sqrt{n}} < a < \bar{X} - 0^{\cdot}5 + \frac{0^{\cdot}57}{\sqrt{n}}\,.$$

Rešitve izpita iz verjetnosti in statistike z dne 24. 1. 2013

Matematika – univerzitetni študij

1. Označimo z N_i, C_i, R_i in K_i dogodke, da bo navadni, čokoladni, rdeči oz. karamelni krof pošel kot i-ti. Nadalje naj bo še A dogodek, da je čokoladni krof pošel pred karamelnim. Tedaj velja:

$$\begin{aligned}
P(A) &= P(C_1) + P(N_1 \cap C_2) + P(R_1 \cap C_2) + \\
&\qquad + P(N_1 \cap R_2 \cap C_3) + P(R_1 \cap N_2 \cap C_3) = \\
&= \frac{3}{10} + \frac{4}{10} \cdot \frac{3}{6} + \frac{2}{10} \cdot \frac{3}{8} + \frac{4}{10} \cdot \frac{2}{6} \cdot \frac{3}{4} + \frac{2}{10} \cdot \frac{4}{8} \cdot \frac{3}{4} = \\
&= \frac{3}{4} = 0{\cdot}75.
\end{aligned}$$

Nadalje velja:

$$P(C_2 \cap A) = P(N_1 \cap C_2) + P(R_1 \cap C_2) = \frac{4}{10} \cdot \frac{3}{6} + \frac{2}{10} \cdot \frac{3}{8} = \frac{11}{40}\,.$$

Sledi:

$$P(C_2 \mid A) = \frac{P(C_2 \cap A)}{P(A)} = \frac{11}{30} \doteq 0{\cdot}367\,.$$

2. Slučajna spremenljivka $W = X + Y + 1$ je porazdeljena normalno $\mathrm{N}\big(1, \sqrt{2}\big)$, torej ima gostoto:

$$f_W(w) = \frac{1}{2\sqrt{2\pi}}\, e^{-(w-1)^2/4}\,.$$

Slučajna spremenljivka $Z = W^2$ pa ima za $z > 0$ gostoto:

$$f_Z(z) = \frac{f_W\big(\sqrt{z}\big) + f_W\big(-\sqrt{z}\big)}{2\sqrt{z}} = \frac{e^{-(\sqrt{z}-1)^2/4} + e^{-(-\sqrt{z}-1)^2/4}}{4\sqrt{2\pi z}}\,.$$

Sklep:

$$f_Z(z) = \begin{cases} \dfrac{e^{-(\sqrt{z}-1)^2/4} + e^{-(\sqrt{z}+1)^2/4}}{4\sqrt{2\pi z}} & ;\ z > 0 \\ 0 & ;\ \text{sicer} \end{cases}\,.$$

3. Iz:

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = c(\lambda) \int_0^{\infty} x\, e^{-\lambda x}\,\mathrm{d}x = c(\lambda)\, \frac{\lambda+1}{\lambda^2}\, e^{-\lambda}$$

dobimo $c(\lambda) = \dfrac{\lambda^2 e^{\lambda}}{\lambda + 1}$. Logaritem verjetja je torej enak:

$$\ln L = n\big(\lambda + 2\ln\lambda - \ln(\lambda+1)\big) + \ln X_1 + \cdots + \ln X_n - \lambda(X_1 + \cdots + X_n)\,.$$

Po odvajanju dobimo:

$$\frac{\mathrm{d} \ln L}{\mathrm{d}\lambda} = n\left(1 + \frac{2}{\lambda} - \frac{1}{\lambda+1}\right) - \lambda(X_1 + \cdots + X_n) .$$

Cenilka po metodi največjega verjetja je tisti λ, za katerega je zgornji izraz enak nič. Po krajšem računu in upoštevanju pogoja, da je $\lambda > 0$, dobimo cenilko:

$$\hat{\lambda} = \frac{-\bar{X} + 2 + \sqrt{\bar{X}^2 + 4\bar{X} - 4}}{2(\bar{X} - 1)} ,$$

kjer kot ponavadi označimo $\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$.

4. 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 $z := \Phi^{-1}\big((1+\beta)/2\big) \doteq 1{\cdot}645$. 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.

2010/11

Rešitve kolokvija iz verjetnosti in statistike z dne 15. 11. 2010

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Za izbranega posameznika je verjetnost, da ne spozna nikogar, enaka $\left(\frac{2}{3}\right)^3$.

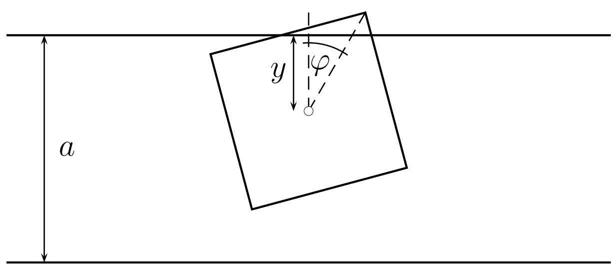
Za dva izbrana posameznika je verjetnost, da ne spoznata nikogar, enaka $\left(\frac{2}{3}\right)^5$.

Za tri ali štiri izbrane posameznike je verjetnost, da ne spoznajo nikogar, enaka $\left(\frac{2}{3}\right)^6$.

Po načelu vključitev in izključitev je verjetnost iskanega dogodka enaka:

$$4 \cdot \left(\frac{2}{3}\right)^3 - \binom{4}{2}\left(\frac{2}{3}\right)^5 + \binom{4}{3}\left(\frac{2}{3}\right)^6 - \binom{4}{4}\left(\frac{2}{3}\right)^6 = \frac{160}{243} \doteq 0{\cdot}658\,.$$

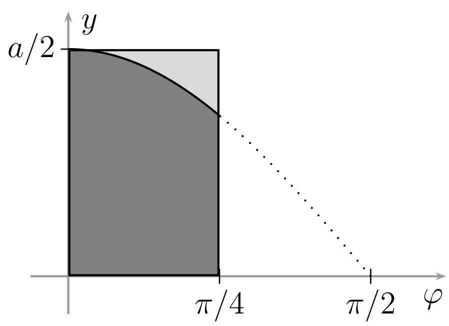
2. Označimo z y oddaljenost središča kvadrata od najbližje črte, s φ pa najmanjši kot med diagonalo in pravokotnico na črte (glej sliko):



Če je kvadrat vržen na slepo, to pomeni, da je oddaljenost y izbrana na slepo iz intervala $[0, a/2]$, kot φ pa je izbran na slepo iz intervala $[0, \pi/4]$, in sicer neodvisno od y. Nadalje kvadrat seka diagonalo natanko tedaj, ko velja:

$$y \le \frac{a}{2}\cos\varphi\,.$$

Verjetnostni prostor in naš dogodek sta prikazana na spodnji sliki:



in iskana verjetnost je enaka:

$$\frac{\displaystyle\int_0^{\pi/4} \frac{a}{2} \cos\varphi \, \mathrm{d}\varphi}{\displaystyle\frac{a}{2} \cdot \frac{\pi}{4}} = \frac{4}{\pi} \int_0^{\pi/4} \cos\varphi \, \mathrm{d}\varphi = \frac{2\sqrt{2}}{\pi} \doteq 0\text{˙}900\,.$$

3. Označimo naslednje dogodke:

$$\begin{aligned} H_1 &= \{\text{Ula izbere prvo posodo}\}\,, & H_2 &= \{\text{Ula izbere drugo posodo}\}\,, \\ U &= \{\text{Ula izvleče belo kroglico}\}\,, & V &= \{\text{Vid izvleče belo kroglico}\}\,. \end{aligned}$$

Velja:

$$\begin{aligned} P(H_1) &= 0\text{˙}3\,, \quad & P(H_2) &= 0\text{˙}7\,, \\ P(U \mid H_1) &= \frac{2}{3}\,, \quad & P(U \mid H_2) &= \frac{1}{3}\,, \end{aligned}$$

zaradi simetrije pa tudi:

$$P(V \mid H_1) = \frac{2}{3}\,, \quad P(V \mid H_2) = \frac{1}{3}\,.$$

Tako lahko po izreku o popolni verjetnosti izpeljemo:

$$P(V) = P(H_1)\, P(V \mid H_1) + P(H_2)\, P(V \mid H_2) \doteq 0\text{˙}433\,.$$

Nadalje velja:

$$P(U \mid V) = \frac{P(U \cap V)}{P(V)} = \frac{P(H_1 \cap U \cap V)}{P(V)} = \frac{1}{P(V)} \cdot 0\text{˙}3 \cdot \frac{2}{3} \cdot \frac{1}{2} \doteq 0\text{˙}231\,.$$

4. Laplaceova integralska formula nam da:

$$P(X = k) \approx \frac{1}{\sqrt{2\pi \cdot 50 \cdot 0\text{˙}4 \cdot 0\text{˙}4}} \exp\left[-\frac{(k-20)^2}{2 \cdot 50 \cdot 0\text{˙}4 \cdot 0\text{˙}6}\right] = \frac{1}{24\pi} \exp\left[-\frac{(k-20)^2}{24}\right]\,,$$

od koder dobimo, da bo $P(X = k) > 10^{-3}$ približno takrat, ko bo:

$$|k - 20| < \sqrt{24 \ln \frac{1000}{\sqrt{24\pi}}} \doteq 10\text{˙}67\,.$$

Kaže torej, da je največje število k, za katero je $P(X = k) > 10^{-3}$, enako 30. Tudi v resnici je tako, saj je:

$$\begin{aligned} P(X = 30) &= \binom{50}{30} \cdot 0\text{˙}4^{30} \cdot 0\text{˙}6^{20} \doteq 0\text{˙}001987\,, \\ P(X = 31) &= \binom{50}{31} \cdot 0\text{˙}4^{31} \cdot 0\text{˙}6^{19} \doteq 0\text{˙}0008545\,. \end{aligned}$$

Verjetnost $P(X \geq 30)$ lahko ocenimo navzdol z vsoto nekaj verjetnosti:

$$P(X \geq 30) \geq P(X = 30) + P(X = 31) + P(X = 32) + P(X = 33) > 0\text{˙}00330\,.$$

Za oceno navzgor pa si pomagamo s kvocientom:

$$\frac{P(X = l + 1)}{P(X = l)} = \frac{50 - l}{l + 1} \cdot \frac{0\text{˙}4}{0\text{˙}6}\,,$$

ki je padajoča funkcija spremenljivke l, torej za $l \geq 30$ velja:

$$\frac{P(X = l + 1)}{P(X = l)} \leq \frac{P(X = 31)}{P(X = 30)} \doteq 0\text{˙}4301$$

in:

$$P(X = l) \leq P(X = 30) \left(\frac{P(X = 31)}{P(X = 30)} \right)^{l-30}$$

Torej je:

$$P(X \geq 30) \leq P(X = 30) \sum_{l=30}^{\infty} \left(\frac{P(X = 31)}{P(X = 30)} \right)^{l-30} = \frac{P(X = 30)}{1 - \frac{P(X=31)}{P(X=30)}} < 0\text{˙}00349\,.$$

Laplaceova integralska formula:

$$P(X \geq 30) \approx \frac{1}{2} - \Phi\left(\frac{50 - 29\text{˙}5}{\sqrt{50 \cdot 0\text{˙}4 \cdot 0\text{˙}6}} \right) \doteq 0\text{˙}00305\,.$$

nam da rezultat izven našega intervala $(0\text{˙}00330, 0\text{˙}00349)$.
Točen rezultat: 0˙003360382.

Rešitve kolokvija iz verjetnosti in statistike z dne 10. 1. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Označimo z A zamudo Ahačičevih, z B pa zamudo Berkopčevih. Tedaj je $T = \min\{A, B\}$. Sledi:

$$
\begin{aligned}
F_T(t) &= P(T \le t) = P\big(\{A \le t\} \cup \{B \le t\}\big) = \\
&= P(A \le t) + P(B \le t) - P(A \le t)\, P(B \le t)\,.
\end{aligned}
$$

Za $t \le 0$ je gotovo $F_T(t) = 0$ (ker še nihče ne pride), za $t \ge 10$ pa velja $F_T(t) = 1$ (ker Ahačičevi zagotovo pridejo). Za $0 < t < 10$ pa velja $P(A \le t) = t/10$ in $P(B \le t) = t/20$, od koder sledi:

$$
F_T(t) = \frac{3t}{20} - \frac{t^2}{200}\,, \quad f_T(t) = \frac{3}{20} - \frac{t}{100}\,.
$$

Torej je:

$$
f_T(t) = \begin{cases} \frac{3}{20} - \frac{t}{100} & ;\ 0 < t < 10 \\ 0 & ;\ \text{sicer} \end{cases}\,.
$$

2. Velja:

$$
\begin{aligned}
P(Y = 0) &= P(X = 1) + P(X = 2) = \frac{1}{3} + \frac{2}{9} = \frac{5}{9}\,, \\
P(Y = k) &= P(X = 3k) + P(X = 3k + 1) + P(X = 3k + 2) = \\
&= \frac{1}{3}\left[\left(\frac{2}{3}\right)^{3k-1} + \left(\frac{2}{3}\right)^{3k} + \left(\frac{2}{3}\right)^{3k+1}\right] = 19 \cdot \frac{2^{3k-1}}{3^{3k+2}}\,; \quad k \in \mathbb{N}\,.
\end{aligned}
$$

3. a) *Prvi način.* Velja:

$$
\begin{aligned}
f_Z(z) &= \frac{1}{2} \int_{-\infty}^{\infty} f(x)\, f\left(\frac{z + x}{2}\right) \mathrm{d}x = \\
&= \frac{\lambda^2}{2} \int_{\substack{x \ge 0 \\ (z+x)/2 \ge 0}} e^{-\lambda x}\, e^{-\lambda(z+x)/2}\, \mathrm{d}x = \\
&= \frac{\lambda^2}{2}\, e^{-\lambda z/2} \int_{\max\{-z,0\}}^{\infty} e^{-3\lambda x/2}\, \mathrm{d}x\,.
\end{aligned}
$$

Za $z \ge 0$ je torej:

$$
f_Z(z) = \frac{\lambda^2}{2}\, e^{-\lambda z/2} \int_0^{\infty} e^{-3\lambda x/2}\, \mathrm{d}x = \frac{\lambda}{3}\, e^{-\lambda z/2}\,,
$$

za $z \le 0$ pa velja:

$$
f_Z(z) = \frac{\lambda^2}{2}\, e^{-\lambda z/2} \int_{-z}^{\infty} e^{-3\lambda x/2}\, \mathrm{d}x = \frac{\lambda}{3}\, e^{\lambda z}\,.
$$

Drugi način. Velja:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f(2y-z)\, f(y)\, \mathrm{d}y = \\
&= \lambda^2 \int_{\substack{2y-z\geq 0\\ y\geq 0}} e^{-\lambda(2y-z)}\, e^{-\lambda y}\, \mathrm{d}y = \\
&= \lambda^2\, e^{\lambda z} \int_{\max\{z/2,0\}}^{\infty} e^{-3\lambda y}\, \mathrm{d}y\,.
\end{aligned}
$$

Za $z \geq 0$ je torej:

$$
f_Z(z) = \lambda^2\, e^{\lambda z} \int_{z/2}^{\infty} e^{-3\lambda y}\, \mathrm{d}y = \frac{\lambda}{3}\, e^{-\lambda z/2}\,,
$$

za $z \leq 0$ pa velja:

$$
f_Z(z) = \lambda^2\, e^{\lambda z} \int_{0}^{\infty} e^{-3\lambda y}\, \mathrm{d}y = \frac{\lambda}{3}\, e^{\lambda z}\,.
$$

b) Velja $E(Z) = 2\,E(X) - E(Y) = \dfrac{2}{\lambda} - \dfrac{1}{\lambda} = \dfrac{1}{\lambda}$ (ni potrebno integrirati $z\, f_Z(z)$).

4. Velja:

$$
\begin{aligned}
E\big(|Z|\big) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |z|\, e^{-z^2/2}\, \mathrm{d}z = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z\, e^{-z^2/2}\, \mathrm{d}z = \frac{2}{\sqrt{2\pi}} \doteq 0\text{˙}798\,, \\
D(|Z|) &= E\big(|Z|^2\big) - \Big(E\big(|Z|\big)\Big)^2 = E\big(Z^2\big) - \Big(E\big(|Z|\big)\Big)^2 = \\
&= D(Z) - \Big(E\big(|Z|\big)\Big)^2 = 1 - \frac{2}{\pi} \doteq 0\text{˙}363\,.
\end{aligned}
$$

V splošnem za vsako slučajno spremenljivko velja:

$$
\begin{aligned}
D\big(|X|\big) &= E\big(|X|^2\big) - \Big(E\big(|X|\big)\Big)^2 = E(X^2) - \Big(E\big(|X|\big)\Big)^2 \leq E(X^2) - \big(E(X)\big)^2 = \\
&= D(X)\,.
\end{aligned}
$$

Zadnja neenakost velja zato, ker je $\big|E(X)\big| \leq E\big(|X|\big)$ in zato $\big(E(X)\big)^2 \leq \Big(E\big(|X|\big)\Big)^2$.

Rešitve kolokvija iz verjetnosti in statistike z dne 4. 4. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Za $0 < x \le 1$ velja:

$$f_{Y|X}(y \mid x) = \frac{\sqrt{x}}{\sqrt{2\pi}}\, e^{-xy^2/2}\,, \quad f_{X,Y}(x,y) = f_X(x)\, f_{Y|X}(y \mid x) = \frac{1}{2\sqrt{2\pi}}\, e^{-xy^2/2}$$

Sledi:

$$f_Y(y) = \frac{1}{2\sqrt{2\pi}} \int_0^1 e^{-xy^2/2}\, \mathrm{d}x = \frac{1 - e^{-y^2/2}}{y^2\sqrt{2\pi}}\,.$$

2. a) Če označimo $M(t) := E\big[e^{tS}\big]$, velja:

$$M(t) = E\big[e^{tX_1}\big]\, E\big[e^{tX_2}\big] \cdots E\big[e^{tX_{100}}\big] = \Big(E\big[e^{tX_1}\big]\Big)^{100}$$

Nadalje je:

$$E\big[e^{tX_1}\big] = \sum_{k=1}^{100} \frac{e^{kt}}{2^k} = \frac{e^t}{2 - e^t}\,.$$

Količina obstaja za $t < \ln 2$, sicer geometrijska vrsta divergira. Torej je:

$$M(t) = \left(\frac{e^t}{2 - e^t}\right)^{100}$$

in funkcija je definirana za $t < \ln 2$.
b) Iz neenačbe Markova dobimo:

$$P(S \ge 300) = P\big(e^{tS} \ge e^{300t}\big) \le e^{-300t}\, E\big[e^{tS}\big] = \frac{e^{-200t}}{\big(2 - e^t\big)^{100}} =: g(t)\,.$$

Iz:

$$g'(t) = \left[-200 + \frac{100\, e^t}{2 - e^t}\right] g(t)$$

dobimo, da je minimum dosežen pri $t = \ln(4/3)$, od koder dobimo:

$$P(S \ge 300) \le \frac{\left(\frac{4}{3}\right)^{-200}}{\left(\frac{2}{3}\right)^{100}} \doteq 4{\cdot}18 \cdot 10^{-8}\,.$$

Točen rezultat: $2{\cdot}65 \cdot 10^{-9}$.

Opomba: normalna aproksimacija (centralni limitni izrek) nam tu da premajhne vrednosti:

- $7{\cdot}69 \cdot 10^{-13}$ za interpretacijo s $P(S \ge 300)$;
- $1{\cdot}28 \cdot 10^{-12}$ za interpretacijo s $P(S > 299)$;

- $9{\cdot}91 \cdot 10^{-13}$ za interpretacijo s $P(S > 299{\cdot}5)$.

3. a) Če $X_1, X_2, \ldots$ v porazdelitvi konvergirajo proti X, $Y_1, Y_2, \ldots$ pa proti c, slučajne spremenljivke $\ln X_1, \ln X_2, \ldots$ v porazdelitvi konvergirajo proti $\ln X$, $\ln Y_1, \ln Y_2, \ldots$ pa proti $\ln c$. Torej slučajne spremenljivke $\ln(X_1Y_1) = \ln X_1 + \ln Y_1, \ln(X_2Y_2) = \ln X_2 + \ln Y_2, \ldots$ konvergirajo proti $\ln X + \ln c = \ln(cX)$, se pravi, da $X_1Y_1, X_2Y_2, \ldots$ konvergirajo proti cX.

b) Recimo, da je $X_1, X_2, \ldots > M$. Tedaj je $X_n - M > 0$ za vsak n. Ker slučajne spremenljivke $X_1 - M, X_2 - M, \ldots$ konvergirajo proti $X - M$, po prejšnji točki slučajne spremenljivke $(X_1 - M)Y_1 = X_1Y_1 - MY_1, (X_2 - M)Y_2 = X_2Y_2 - MY_2, \ldots$ konvergirajo proti $c(X - M) = cX - cM$, potem pa morajo tudi slučajne spremenljivke $X_1Y_1, X_2Y_2, \ldots$ konvergirati proti cX.

Podobno, če je $X_1, X_2, \ldots < M$, je $M - X_n > 0$ za vsak n. Ker slučajne spremenljivke $M - X_1, M - X_2, \ldots$ konvergirajo proti $M - X$, po prejšnji točki slučajne spremenljivke $(M - X_1)Y_1 = MY_1 - X_1Y_1, (M - X_2)Y_2 = MY_2 - X_2Y_2, \ldots$ konvergirajo proti $c(M - X) = cM - cX$, potem pa morajo tudi slučajne spremenljivke $X_1Y_1, X_2Y_2, \ldots$ konvergirati proti cX.

Opomba: v resnici produkti $X_1Y_1, X_2Y_2, \ldots$ v porazdelitvi konvergirajo proti cX, brž ko $X_1, X_2, \ldots$ konvergirajo proti X, $Y_1, Y_2, \ldots$ pa proti c: nobene druge omejitve niso potrebne.

4. Če z $x_{11}, \ldots x_{1N}$ označimo vrednosti dane statistične spremenljivke na prvi, z $x_{21}, \ldots x_{2N}$ pa na drugi skupini, za $r = 1, 2$ velja:

$$\mu_r = \frac{1}{N}\sum_{s=1}^N x_{rs}\,, \quad \sigma_r^2 = \frac{1}{N}\sum_{s=1}^N (x_{rs} - \mu_r)^2 = \frac{1}{N}\sum_{s=1}^N x_{rs}^2 - \mu_r^2\,.$$

Označimo zdaj še z $X_1, \ldots X_n$ vrednosti statistične spremenljivke na enotah, ki naj bi bile iz prve, $Y_1, \ldots Y_n$ pa na enotah, ki naj bi bile iz druge populacije. Tedaj so vse te slučajne spremenljivke neodvisne in velja:

$$\begin{aligned} E(X_i) &= \frac{1-p}{N}\sum_{s=1}^N x_{1s} + \frac{p}{N}\sum_{s=1}^N x_{2s} = (1-p)\mu_1 + p\mu_2 = \mu_1 + p(\mu_2 - \mu_1)\,, \\ E(Y_i) &= \frac{1-p}{N}\sum_{s=1}^N x_{2s} + \frac{p}{N}\sum_{s=1}^N x_{1s} = (1-p)\mu_2 + p\mu_1 = \mu_2 - p(\mu_2 - \mu_1) \end{aligned}$$

ter še:

$$\begin{aligned} E(X_i^2) &= \frac{1-p}{N}\sum_{s=1}^N x_{1s}^2 + \frac{p}{N}\sum_{s=1}^N x_{2s}^2 = (1-p)(\mu_1^2 + \sigma^2) + p(\mu_2^2 + \sigma^2) = \\ &= \mu_1^2 + \sigma^2 + p(\mu_2^2 - \mu_1^2)\,, \\ E(Y_i^2) &= \frac{1-p}{N}\sum_{s=1}^N x_{2s}^2 + \frac{p}{N}\sum_{s=1}^N x_{1s}^2 = (1-p)(\mu_2^2 + \sigma^2) + p(\mu_1^2 + \sigma^2) = \\ &= \mu_2^2 + \sigma^2 - p(\mu_2^2 - \mu_1^2)\,. \end{aligned}$$

Po nekaj računanja dobimo:

$$D(X_i) = D(Y_i) = \sigma^2 + p(1-p)(\mu_2 - \mu_1)^2 \,.$$

Če torej z

$$\bar{Z} = \frac{X_1 + X_2 + \cdots + X_n + Y_1 + Y_2 + \cdots + Y_n}{2n}$$

označimo povprečje dane statistične spremenljivke na našem vzorcu, dobimo:

$$D(\bar{Z}) = \frac{\sigma^2 + p(1-p)(\mu_2 - \mu_1)^2}{2n} \,.$$

Rešitve kolokvija iz verjetnosti in statistike z dne 9. 6. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Če taka cenilka obstaja, mora biti funkcija pripadajoče zadostne statistike. Za eno samo opažanje X je le-ta enaka kar X, za neodvisna opažanja $X_1, \ldots X_n$ pa je enaka $S := X_1 + X_2 + \cdots + X_n$. Slučajna spremenljivka S ima porazdelitev Gama(n, λ). Izračunajmo:

$$E(S^2) = \int_0^\infty s^2 \, \frac{\lambda^n}{(n-1)!} s^{n-1} e^{-\lambda s} \, \mathrm{d}s = \frac{n(n+1)}{\lambda^2} \, .$$

To lahko izračunamo tudi s pomočjo disperzije:

$$E(S^2) = D(S) + \bigl(E(S)\bigr)^2 = \frac{n}{\lambda^2} + \left(\frac{n}{\lambda}\right)^2 = \frac{n(n+1)}{\lambda^2} \, .$$

Nepristranska cenilka za $\dfrac{1}{\lambda^2}$ bo torej $\dfrac{S^2}{n(n+1)}$.

2. 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.

3. $\bar{X} \doteq 1{\cdot}55, \quad S \doteq 1{\cdot}20, \quad c \doteq 1{\cdot}96, \quad \Delta \doteq 0{\cdot}21.$
Interval zaupanja: $1{\cdot}34 < \mu < 1{\cdot}76$.

4. Hipotezo bomo zavrnili, če bo premalo izvlečenih listkov dobitnih. Natančneje, če je D število dobitnih listkov med izvlečenimi, bomo za določen prag c ničelno hipotezo zavrnili, če bo $D < c$. Za $D > c$ hipoteze ne bomo zavrnili, za $D = c$ pa bomo *randomizirali*, t. j. hipotezo bomo zavrnili z neko verjetnostjo γ. Slučajna spremenljivka D je porazdeljena hipergeometrijsko, natančneje, velja:

$$P(D = k) = \frac{\binom{12}{k}\binom{8}{5-k}}{\binom{20}{5}} \, .$$

Izračunajmo:

$$\begin{aligned} P(D = 0) &\doteq 0{\cdot}0036 \, , \qquad & P(D \le 0) &\doteq 0{\cdot}0036 \\ P(D = 1) &\doteq 0{\cdot}0542 \, , \qquad & P(D \le 1) &\doteq 0{\cdot}0578 \, . \end{aligned}$$

Torej bomo postavili $c = 1$. Verjetnost γ bomo izračunali tako, da bo veljalo:

$$P(D = 0) + \gamma\, P(D = 1) = \alpha\,,$$

torej:

$$\gamma = \frac{\alpha - P(D = 0)}{P(D = 1)} \doteq 0{\dot{}}85$$

(zaokrožili smo navzdol).

Rešitve izpita iz verjetnosti in statistike z dne 30. 6. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. 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:

$$P(M^{\mathrm{c}} \mid F_k) = \left(\frac{9}{10}\right)^{k-1}\,,$$

torej je:

$$P(M^{\mathrm{c}} \mid F) = \frac{P(M^{\mathrm{c}} \cap F)}{P(F)} = \frac{1}{P(F)} \sum_{k=1}^{\infty} P(F_k)\, P(M^{\mathrm{c}} \mid F_k) = \frac{20}{21}\,.$$

iskana pogojna verjetnost pa je enaka:

$$P(M \mid F) = 1 - P(M^{\mathrm{c}} \mid F) = \frac{1}{21}\,.$$

2. Najprej opazimo, da Z skoraj gotovo zavzame vrednosti na intervalu $[0, 1]$ (celo le na $(0, 1)$). Iz izražave:

$$X_2 = g(X_1, Z)\,, \quad \text{kjer je} \quad g(x, z) = \frac{x}{z} - x\,, \quad \frac{\partial g}{\partial z}(x, z) = -\frac{x}{z^2}$$

po krajšem računu sledi, da za $0 < z < 1$ velja:

$$f_Z(z) = \lambda^3\, \frac{1 - z}{z^3} \int_0^{\infty} x^2 e^{-\lambda x/z}\, \mathrm{d}x = 2(1 - z)\,.$$

Torej lahko zapišemo:

$$f_Z(z) = \begin{cases} 2(1 - z) & ;\ 0 < z < 1 \\ 0 & ;\ \text{sicer} \end{cases}\,.$$

3. 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 (centralni limitni izrek se sicer da posplošiti tudi na vsote slučajnih spremenljivk z določeno vrsto odvisnosti, vendar pa 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:

$$\begin{aligned} P(S > 150) &= P(X = 1)\, P(S > 150 \mid X = 1) + P(X = 2)\, P(S > 150 \mid X = 2) = \\ &= \frac{2}{3}\, P(T > 150) + \frac{1}{3}\, P(T > 75)\,. \end{aligned}$$

Za slučajno spremenljivko T pa centralni limitni izrek velja: iz $E(T) = 100$ in $D(T) = 10000$ dobimo:

$$P(T > a) \approx \frac{1}{2} - \Phi\left(\frac{a - 100}{100}\right),$$

torej je:

$$P(S > 150) \approx \frac{1}{2} - \frac{2}{3}\,\Phi\left(\frac{1}{2}\right) + \frac{1}{3}\,\Phi\left(\frac{1}{4}\right) \doteq 0'405\,.$$

4. 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 $\mathrm{b}(3, p)$, 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:

$$h(0) = 0\,, \quad h(1) = 0\,, \quad h(2) = \tfrac{1}{3}\,, \quad h(3) = 1\,.$$

Cenilko lahko posplošimo tudi na primer, ko izvedemo n poskusov. V tem primeru se jo splača iskati 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.

Rešitve izpita iz verjetnosti in statistike z dne 26. 8. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Označimo z N_i dogodek, da je v i-ti škatli nagrada, z V_2 pa dogodek, da je vodja igre odprl drugo škatlo. Tedaj velja:

$$\begin{aligned} P(V_2) &= P(N_1)\,P(V_2 \mid N_1) + P(N_3)\,P(V_2 \mid N_3) + P(N_4)\,P(V_2 \mid N_4) = \\ &= \frac{1}{4}\left(\frac{1}{6} + \frac{1}{4} + \frac{1}{3}\right) = \\ &= \frac{3}{16} \end{aligned}$$

in nadalje:

$$\begin{aligned} P(N_1 \mid V_2) &= \frac{P(N_1)\,P(V_2 \mid N_1)}{P(V_2)} = \frac{2}{9}\,, \\ P(N_3 \mid V_2) &= \frac{P(N_1)\,P(V_2 \mid N_1)}{P(V_2)} = \frac{1}{3}\,, \\ P(N_4 \mid V_2) &= \frac{P(N_1)\,P(V_2 \mid N_1)}{P(V_2)} = \frac{4}{9}\,. \end{aligned}$$

Igralcu se torej najbolj splača odpreti četrto škatlo. Nagrada je notri s pogojno verjetnostjo 4/9.

2. Velja:

$$E\big(e^{-X}\big) = \int_0^\infty e^{-2x}\,\mathrm{d}x = \frac{1}{2}\,, \quad E\Big[\big(e^{-X}\big)^2\Big] = \int_0^\infty e^{-3x}\,\mathrm{d}x = \frac{1}{3}\,, \quad D\big(e^{-X}\big) = \frac{1}{12}$$

in nadalje:

$$\begin{aligned} E\big(e^{-X-Y}\big) &= \int_0^\infty e^{-2x}\,\mathrm{d}x \int_0^\infty e^{-2y}\,\mathrm{d}y = \frac{1}{4}\,, \\ E\Big[\big(e^{-X-Y}\big)^2\Big] &= \int_0^\infty e^{-2x}\,\mathrm{d}x \int_0^\infty e^{-2y}\,\mathrm{d}y = \frac{1}{9}\,, \\ D\big(e^{-X}\big) &= \frac{7}{144}\,. \end{aligned}$$

Končno je:

$$E\big(e^{-X}e^{-X-Y}\big) = \int_0^\infty e^{-3x}\,\mathrm{d}x \int_0^\infty e^{-2y}\,\mathrm{d}y = \frac{1}{6}\,,$$

$$K\big(e^{-X}, e^{-X-Y}\big) = \frac{1}{24}\,, \quad r\big(e^{-X}, e^{-X-Y}\big) = \sqrt{\frac{3}{7}} \doteq 0{\cdot}655.$$

3. a) Iz zapisa:

$$f_X(x) = \frac{\rho_0(x) + 3\,\rho_1(x) + 2\,\rho_2(x)}{1 + 3\theta + 2\theta^2}\, e^{x \ln \theta}$$

takoj razberemo, da gre za eksponentno družino s pripadajočo zadostno statistiko X.

b) Cenilka $h(X)$ bo nepristranska natanko tedaj, ko bo za vsak θ veljalo:

$$E\big[h(X)\big] = \frac{h(0) + 3\,h(1)\,\theta + 2\,h(2)\,\theta^2}{1 + 3\theta + 2\theta^2} = \frac{1}{1+\theta}\,,$$

kar je ekvivalentno:

$$h(0) + 3\,h(1)\,\theta + 2\,h(2)\,\theta^2 = 1 + 2\theta\,.$$

Cenilka bo torej nepristranska natanko tedaj, ko bo $h(0) = 1$, $h(1) = 2/3$ in $h(2) = 0$. Ker se deterministično izraža z X, ima tudi najmanjšo možno disperzijo.

4. Vzorčno povprečje: $\bar{X} = 29\text{˙}85$.
Popravljeni vzorčni standardni odklon: $S_\mathrm{p} \doteq 2\text{˙}695$.
Kvantila porazdelitve hi kvadrat pri *df* $= 9$: $\chi^2_{0.025} \doteq 2\text{˙}700$, $\chi^2_{0.975} \doteq 19\text{˙}02$.
Interval zaupanja (ustrezno zaokrožen): $1\text{˙}85 < \sigma < 4\text{˙}92$.

Rešitve izpita iz verjetnosti in statistike z dne 5. 9. 2011

Matematika – bolonjski univerzitetni študij in pedagoška matematika

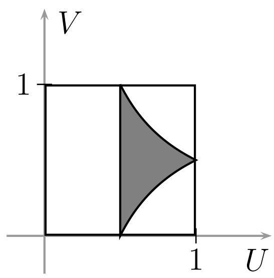
1. Brez škode za splošnost lahko privzamemo, da v drugo prelomimo npr. levi konec palice. Naj bo U dolžina levega konca prvič prelomljene palice, deljena z dolžino celotne palice, V pa naj bo dolžina levega konca drugič prelomljene palice, deljena z dolžino dela palice, ki smo ga drugič lomili (t. j. levega konca prvič prelomljene palice). Tedaj sta slučajni spremenljivki U in V neodvisni in porazdeljeni enakomerno na $(0,1)$, dolžine končnih delov palice pa so si v razmerju $UV : U(1-V) : 1-U$. Trikotnik lahko torej sestavimo, če velja:

$$\begin{aligned} UV &\le U(1-V) + (1-U)\,, \\ U(1-V) &\le UV + (1-U)\,, \\ 1-U &\le UV + U(1-V)\,, \end{aligned}$$

kar je ekvivalentno:

$$U \ge \frac{1}{2}\,, \quad 1 - \frac{1}{2U} \le V \le \frac{1}{2U}\,.$$

Slika:



Verjetnost našega dogodka je enaka:

$$\int_{1/2}^{1} \left[\frac{1}{2u} - \left(1 - \frac{1}{2u}\right)\right] \mathrm{d}u = \int_{1/2}^{1} \frac{\mathrm{d}u}{u} = \ln 2 - \frac{1}{2} \doteq 0{\cdot}193\,.$$

2. Najprej izračunamo:

$$E(X) = E\big[E(X \mid Y)\big] = E(Y) = 2 \cdot \frac{1}{2} = 1\,.$$

Za disperzijo pa imamo dve možnosti. Lahko najprej izračunamo:

$$E(X^2 \mid Y) = D(X \mid Y) + \big(E(X \mid Y)\big)^2 = Y^2 + Y + 1$$

ter nato:

$$\begin{aligned} E(X^2) &= E\big[E(X^2 \mid Y)\big] = E(Y^2) + E(Y) + 1 = D(Y) + \big(E(Y)\big)^2 + E(Y) + 1 = \\ &= 2 \cdot \frac{1}{2} \cdot \frac{1}{2} + 3 = \frac{7}{2}\,, \\ D(X) &= E(X^2) - \big(E(X)\big)^2 = \frac{5}{2}\,. \end{aligned}$$

Disperzijo pa lahko izračunamo tudi s pomočjo razbitja na pojasnjeno in nepojasnjeno disperzijo:

$$D(X) = D\big(E(X \mid Y)\big) + E\big(D(X \mid Y)\big) = D(Y) + E(Y+1) = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} + 1 = \frac{5}{2}\,.$$

3. Iz porazdelitve posameznega opažanja $\begin{pmatrix} 0 & 1 & 2 \\ (1-p)/2 & 1/2 & p/2 \end{pmatrix}$ in dejanskih opažanj dobimo opaženo funkcijo verjetja:

$$L = \left(\frac{1-p}{2}\right)^{15} \left(\frac{1}{2}\right)^{55} \left(\frac{p}{2}\right)^{30}.$$

Prikladneje je delati z logaritmom:

$$\ln L = -100 \ln 2 + 15 \ln(1-p) + 30 \ln p\,.$$

Po odvajanju dobimo:

$$\frac{\partial \ln L}{\partial p} = \frac{30}{p} - \frac{15}{1-p}\,,$$

kar je enako nič pri $p = 2/3$ in to je naša ocena.

4. $\bar{X} = 1\dot{}449$, $\quad S_{\mathrm{p}} \doteq 1\dot{}189$, $\quad df = 365$,
$t_{0.995} \doteq 2\dot{}59$ (pri $df = \infty$ pa pride $2\dot{}58$, kar je dober približek),
$\Delta \doteq 0\dot{}161$.

Interval zaupanja (ustrezno zaokrožen): $(1\dot{}28, 1\dot{}62)$.

Rešitve izpita iz verjetnosti in statistike z dne 1. 2. 2012

Matematika – bolonjski univerzitetni študij in pedagoška matematika

1. Označimo z A_n dogodek, da Petrček n-tič zasitnari in pri tem uspe, z N_n pa dogodek, da n-tič zasitnari in pri tem ne uspe (t. j. da še po n-tem sitnarjenju ni dobil avtomobilčka). Naj bo še N_0 gotovi dogodek. Tedaj je:

$$P\big(A_n \mid N_{n-1}\big) = \frac{2^{-n-2}}{1 + 2^{-n-1}}\,, \qquad P\big(N_n \mid N_{n-1}\big) = 1 - \frac{2^{-n-2}}{1 + 2^{-n-1}} = \frac{1 + 2^{-n-2}}{1 + 2^{-n-1}}$$

(dogodka A_n in N_n za $n > 1$ nista nasprotna, sta pa *pogojno nasprotna* glede na N_{n-1}, t. j. $A_n = N_{n-1} \setminus N_n$). Velja tudi $N_0 \supseteq N_1 \supseteq N_2 \cdots$. Sledi:

$$P(N_n) = P(N_0)\, P(N_1 \mid N_0)\, P(N_2 \mid N_1) \cdots P(N_n \mid N_{n-1}) = \frac{1 + 2^{-n-2}}{1 + 2^{-2}} = \\ = \frac{4}{5}\big(1 + 2^{-n-2}\big)\,.$$

Dogodek, da Petrček dobi avtomobilček, lahko izrazimo kot:

$$A := A_1 \cup A_2 \cup \cdots = (N_1 \cap N_2 \cap \cdots)^{\mathrm{c}}$$

in njegova verjetnost je enaka:

$$1 - \lim_{n \to \infty} P(N_n) = \frac{1}{5}\,.$$

Označimo še z M dogodek, da Petrček dobi avtomobilček pri mami. Velja:

$$M = A_1 \cup A_3 \cup A_5 \cup \cdots$$

Dogodki $A_1, A_2, A_3, \ldots$ so nezdružljivi in velja:

$$P(A_n) = P(N_{n-1}) - P(N_n) = \frac{2^{-n}}{5}\,,$$

torej je:

$$P(M) = \frac{1}{5}\big(2^{-1} + 2^{-3} + \cdots\big) = \frac{2}{15}\,,$$

iskana pogojna verjetnost pa je enaka:

$$P(M \mid A) = \frac{P(M)}{P(A)} = \frac{2}{3}\,.$$

2. *Prvi način*: zapišemo vrednosti slučajne spremenljivke S za vse možne izide. To lahko naredimo recimo tako, da izberemo nekega moškega in gledamo, kako glede na njega sedita ostala dva moška. Dobimo:

Izid	S
M M M Ž Ž Ž	4
M M Ž M Ž Ž	4
M M Ž Ž M Ž	4
M M Ž Ž Ž M	4
M Ž M M Ž Ž	4
M Ž M Ž M Ž	0
M Ž M Ž Ž M	4
M Ž Ž M M Ž	4
M Ž Ž M Ž M	4
M Ž Ž Ž M M	4

Torej je $S \sim \begin{pmatrix} 0 & 4 \\ 1/10 & 9/10 \end{pmatrix}$, od koder dobimo $E(S) = \dfrac{18}{5}$ in $D(S) = \dfrac{36}{25}$.

Drugi način: slučajno spremenljivko S zapišemo kot vsoto. Za i-ti sedež definiramo slučajno spremenljivko:

$$X_i = \begin{cases} 1 & ; \text{ na sedežih, ki sta sosedna } i\text{-temu, sedita osebi nasprotnega spola} \\ 0 & ; \text{ sicer} \end{cases} .$$

Velja $S = X_1 + X_2 + X_3 + X_4 + X_5 + X_6$ in iz:

$$\begin{aligned} E(X_i) &= p_0 := \\ &:= P(\text{na sedežih, ki sta sosedna } i\text{-temu, sedita osebi nasprotnega spola}) = \\ &= \frac{3}{5} \end{aligned}$$

dobimo $E(S) = \dfrac{18}{5} \doteq 3{\dot{}}6$. Izračunajmo še disperzijo: $D(S) = E(S^2) - \big(E(S)\big)^2$. Velja:

$$E(S^2) = \sum_{i=1}^{6} \sum_{j=1}^{6} E(X_i X_j) = 6(p_0 + 2p_1 + 2p_2 + p_3)$$

kjer je p_k verjetnost, da za sedeža, oddaljena za k (t. j. med njima je $k-1$ sedežev) velja, da na sosednih sedežih sedita osebi nasprotnega spola. Verjetnost p_0 smo že izračunali, velja pa še:

$$p_1 = \frac{2}{5}, \qquad p_2 = \frac{3}{10}, \qquad p_3 = \frac{2}{5},$$

torej je:

$$D(S) = 6(p_0 + 2p_1 + 2p_2 + p_3) - (6p_0)^2 = \frac{36}{25}.$$

3. Iz $E(X_i) = 0$ in $D(X) = a^2$ ter posledično $E(S) = 0$ in $D(S) = 500a^2$ ter še centralnega limitnega izreka dobimo:

$$P(S > 1000) \approx \frac{1}{2} - \Phi\left(\frac{1000}{a\sqrt{500}}\right),$$

torej mora biti:
$$\frac{1000}{a\sqrt{500}} \approx \Phi^{-1}(0\dot{}45) \doteq 1\dot{}645$$
oziroma $a \approx \dfrac{1000}{\Phi^{-1}(0.45)\sqrt{500}} \doteq 27\dot{}2.$

V resnici je $P(S > 1000) < 0\dot{}05$, če je $a \leq 1000/37 \doteq 27\dot{}03$, in $P(S > 1000) > 0\dot{}05$, če je $a > 1000/37$.

4. a) Iz porazdelitvene gostote:
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/(2\sigma^2)}$$
je jasno, da gre za eksponentno družino s pripadajočo zadostno statistiko X^2. Nepristranska cenilka z najmanjšo možno disperzijo mora biti torej funkcija statistike X^2. Ker je $\sigma = \sqrt{\sigma^2}$, je smiselno domnevati, da bo iskana cenilka podobna statistiki $\sqrt{X^2} = |X|$. Iz:
$$E(|X|) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x|\, e^{-x^2/(2\sigma^2)} = \sqrt{\frac{2}{\pi}}\,\sigma$$
sledi, da je iskana cenilka enaka $\sqrt{\dfrac{\pi}{2}}\,|X|$.

2009/10

Rešitve kolokvija iz verjetnosti in statistike z dne 23. 11. 2009

Matematika – UNI-BOL

1. Označimo z A dogodek, da Goran in Hedvika prekrižata same različne številke. Naloga se da rešiti na vsaj dva načina.

Prvi način. Za $i = 0, 1, 2, 3$ označimo s H_i dogodek, da Goran prekriža natanko i številk izmed 19, 20 in 21. Tedaj velja:

$$P(H_i) = \frac{\binom{3}{i}\binom{18}{7-i}}{\binom{21}{7}}, \qquad P(A \mid H_i) = \frac{\binom{21-i}{7}}{\binom{21}{7}}$$

in iz izreka o polni verjetnosti dobimo:

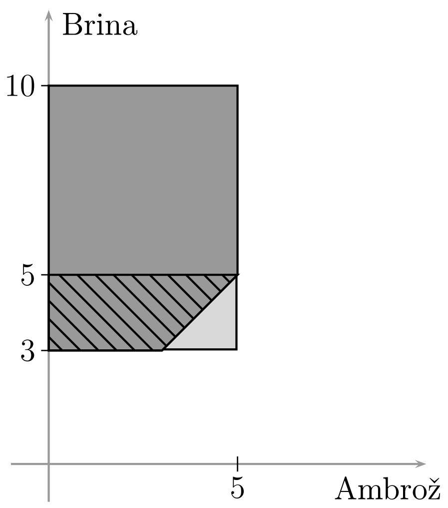
$$P(A) = \sum_{i=0}^{3} P(H_i)\ P(A \mid H_i) = \frac{\binom{18}{7}\binom{21}{7} + 3\cdot\binom{18}{6}\binom{20}{7} + 3\cdot\binom{18}{5}\binom{19}{7} + \binom{18}{4}\binom{18}{7}}{\binom{21}{7}^2} =$$

$$= \frac{37687}{54150} \doteq 0\text{˙}696\,.$$

Drugi način. Če z B_i označimo dogodek, da sta Goran in Hedvika oba prekrižala številko i, velja $A = (B_{19} \cup B_{20} \cup B_{21})^{\mathrm{c}}$. Po načelu vključitev in izključitev dobimo:

$$\begin{aligned} P(A) = 1 &- P(B_{19}) - P(B_{20}) - P(B_{21}) + \\ &+ P(B_{19} \cap B_{20}) + P(B_{19} \cap B_{21}) + P(B_{20} \cap B_{21}) - \\ &- P(B_{19} \cap B_{20} \cap B_{21}) = \\ = 1 &- 3\cdot\left(\frac{7}{21}\right)^2 + 3\cdot\left(\frac{7\cdot 6}{21\cdot 20}\right)^2 - \left(\frac{7\cdot 6\cdot 5}{21\cdot 20\cdot 19}\right)^2 = \\ = \frac{37687}{54150} &\doteq 0\text{˙}696\,. \end{aligned}$$

2. Iz skice:



in razmerja ploščin dobimo, da je verjetnost enaka $8/33 \doteq 0{\cdot}242$.

3. Označimo s p verjetnost, da je posamezen izdelek brezhiben. Iz Laplaceove integralske formule dobimo, da p zadošča zahtevi naloge približno takrat, ko velja:

$$\frac{1}{2} - \Phi\left(\frac{349{\cdot}5 - 400p}{\sqrt{400p(1-p)}}\right) \geq 0{\cdot}95\,,$$

kar je približno (v okviru zaokrožitvenih napak) ekvivalentno:

$$\frac{400p - 349{\cdot}5}{\sqrt{400p(1-p)}} \geq 1{\cdot}645\,.$$

Po množenju in kvadriranju dobimo, da je to nadalje ekvivalentno $p \geq \dfrac{349{\cdot}5}{400} = 0{\cdot}87375$ in še:

$$(400^2 + 400 \cdot 1{\cdot}645^2)p^2 - 400(2 \cdot 349{\cdot}5 + 1{\cdot}645^2)p + 349{\cdot}5^2 \geq 0\,.$$

Kvadratna neenačba je ekvivalentna pogoju $p \leq p_1$ ali $p \geq p_2$, kjer je:

$$\begin{aligned}
p_1 &= \frac{400(2 \cdot 349{\cdot}5 + 1{\cdot}645^2) - \sqrt{4 \cdot 400 \cdot 349{\cdot}5 \cdot 1{\cdot}645^2(400 - 349{\cdot}5) + 400^2 \cdot 1{\cdot}645^4}}{2(400^2 + 400 \cdot 1{\cdot}645^2)} \doteq \\
&\doteq 0{\cdot}844 \quad \text{(zaokroženo navzdol)} \\
p_2 &= \frac{400(2 \cdot 349{\cdot}5 + 1{\cdot}645^2) + \sqrt{4 \cdot 400 \cdot 349{\cdot}5 \cdot 1{\cdot}645^2(400 - 349{\cdot}5) + 400^2 \cdot 1{\cdot}645^4}}{2(400^2 + 400 \cdot 1{\cdot}645^2)} \doteq \\
&\doteq 0{\cdot}899 \quad \text{(zaokroženo navzgor)}\,.
\end{aligned}$$

Približno zadosten pogoj bo torej $p \geq 0{\cdot}899$.

V resnici je pogoj $p \geq 0{\cdot}899058$ zadosten, medtem ko $p = 0{\cdot}899057$ še ne zadošča zahtevi naloge.

4. Kumulativna porazdelitvena funkcija:

$$F_X(x) = \begin{cases} 0 & ;\ x \leq -1 \\ 1 - \dfrac{\arccos x}{\pi} & ;\ -1 \leq x \leq 1 \\ 1 & ;\ x \geq 1 \end{cases}$$

Porazdelitvena gostota:

$$f_X(x) = \begin{cases} \dfrac{1}{\pi\sqrt{1-x^2}} & ;\ -1 < x < 1 \\ 0 & ;\ \text{sicer} \end{cases}$$

Rešitve kolokvija iz verjetnosti in statistike z dne 18. 1. 2010

Matematika – UNI-BOL

1. Označimo $Z := XY$.

Prvi način. Iz formule za gostoto funkcije dveh slučajnih spremenljivk in neodvisnosti dobimo, da je slučajna spremenljivka Z porazdeljena zvezno z gostoto:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}\left(x, \frac{z}{x}\right) \frac{1}{|x|}\, \mathrm{d}x = \int_{-\infty}^{\infty} f_X(x)\, f_Y\left(\frac{z}{x}\right) \frac{1}{|x|}\, \mathrm{d}x\,.$$

Iz porazdelitve obeh slučajnih spremenljivk dobimo:

$$f_Z(z) = \frac{1}{3} \int_{\substack{1 \le x \le 2 \\ 1 \le z/x \le 4}} \frac{\mathrm{d}x}{x} = \int_{\substack{1 \le x \le 2 \\ z/4 \le x \le z}} \frac{\mathrm{d}x}{x}\,.$$

Za $1 \le x \le 2$ dobimo:

$$f_Z(z) = \frac{1}{3} \int_1^z \frac{\mathrm{d}x}{x} = \frac{\ln z}{3}\,,$$

za $2 \le z \le 4$ dobimo:

$$f_Z(z) = \frac{1}{3} \int_1^2 \frac{\mathrm{d}x}{x} = \frac{\ln 2}{3}\,,$$

za $4 \le z \le 8$ dobimo

$$f_Z(z) = \frac{1}{3} \int_{z/4}^2 \frac{\mathrm{d}x}{x} = \frac{\ln 8 - \ln z}{3}\,,$$

sicer je $f_Z(z) = 0$. Torej velja:

$$f_Z(z) = \begin{cases} \frac{1}{3} \ln z & ;\ 1 \le z \le 2 \\ \frac{1}{3} \ln 3 & ;\ 2 \le z \le 4 \\ \frac{1}{3}(\ln 8 - \ln z) & ;\ 4 \le z \le 8 \\ 0 & ;\ \text{sicer} \end{cases}\,.$$

Drugi način. Iz porazdelitve obeh slučajnih spremenljivk in neodvisnosti dobimo, da je kumulativna porazdelitvena funkcija slučajne spremenljivke Z enaka:

$$F_Z(z) = \frac{1}{3} \iint_{\substack{1 \le x \le 2 \\ 1 \le y \le 4 \\ xy < z}} \mathrm{d}x\, \mathrm{d}y = \frac{1}{3} \int_1^2 \int_{1 \le y \le \min\{z/x, 4\}} \mathrm{d}y\, \mathrm{d}x\,.$$

Za $1 \le x \le 2$ dobimo:

$$F_Z(z) = \frac{1}{3} \int_1^z \int_1^{z/x} \mathrm{d}y\, \mathrm{d}x = \frac{1}{3} \int_1^z \left(\frac{z}{x} - 1\right) \mathrm{d}x = \frac{z \ln z - z + 1}{3}\,,$$

za $2 \leq z \leq 4$ dobimo:

$$F_Z(z) = \frac{1}{3} \int_1^2 \int_1^{z/x} \mathrm{d}y\, \mathrm{d}x = \frac{1}{3} \int_1^2 \left(\frac{z}{x} - 1\right) \mathrm{d}x = \frac{z \ln 2 - 1}{3}\,,$$

za $4 \leq z \leq 8$ pa dobimo:

$$\begin{aligned} f_Z(z) &= \frac{1}{3} \int_1^{z/4} \int_1^4 \mathrm{d}y\, \mathrm{d}x + \frac{1}{3} \int_{z/4}^2 \int_1^{z/x} \mathrm{d}y\, \mathrm{d}x = \\ &= \frac{z}{4} - 1 + \frac{1}{3} \int_{z/4}^2 \left(\frac{z}{x} - 1\right) \mathrm{d}x = \\ &= \frac{z - 5 + z(\ln 8 - \ln z)}{3}\,. \end{aligned}$$

Torej velja:

$$F_Z(z) = \begin{cases} 0 & ;\ z \leq 1 \\ \frac{1}{3}\big(z \ln z - z + 1\big) & ;\ 1 \leq z \leq 2 \\ \frac{1}{3}\big(z \ln 2 - 1\big) & ;\ 2 \leq z \leq 4 \\ \frac{1}{3}\big(z - 5 + z(\ln 8 - \ln z)\big) & ;\ 4 \leq z \leq 8 \\ 1 & ;\ z \geq 8 \end{cases}.$$

Opomba. Brez težav preverimo, da za f_X iz prvega načina in F_X iz drugega načina velja $f_X = F_X'$, torej smo res obakrat dobili isto.

2. Za i-ti sedež definirajmo slučajno spremenljivko X_i, ki je enaka 1, če tisti, ki tam sedi, in njegov desni sosed oba dvigneta roko, sicer pa naj bo $X_i = 0$. Lahko pišemo tudi $X_i = Y_i Y_{i+1}$, kjer je $Y_i = 1$, če tisti, ki sedi na i-tem sedežu, dvigne roko, sicer pa je $Y_i = 0$. V tem primeru se dogovorimo, da sedeže označujemo z elementi iz $\mathbb{Z}_n$, in sicer od leve proti desni. Velja torej:

$$S = X_0 + X_1 + \ldots + X_{n-1} = Y_0 Y_1 + Y_1 Y_2 + \ldots Y_{n-2} Y_{n-1} + Y_{n-1} Y_0\,.$$

Brž ko je $n \geq 2$, velja $E\,X_i = 1/4$ in ker je matematično upanje aditivno, od tod sledi $E\,S = n/4$. Za $n = 1$ pa je $S = 0$ in zato $E\,S = 0$.

Za izračun disperzije je ugodno vpeljati razdaljo med sedežema: za $k \in \mathbb{Z}_n$ naj bo $|k|$ najmanjše število iz $\mathbb{N}_0$, ki v $\mathbb{Z}_n$ ustreza k ali pa $-k$ (recimo $|n-1| = 1$). Tedaj je razdalja med sedežema i in j enaka $|i-j|$ (ni se težko prepričati, da je to metrika na $\mathbb{Z}_n$). Od tod naprej lahko računamo na vsaj dva načina.

Prvi način. Najprej izračunamo:

$$E\,S^2 = \sum_{i \in \mathbb{Z}_n} \sum_{j \in \mathbb{Z}_n} E(X_i X_j)\,.$$

Če je $i = j$, je $E(X_i X_j) = 1/4$. Če je $|i-j| = 1$ in $n \geq 3$, je $E(X_i X_j) = 1/8$ (saj recimo za $j = i+1$ velja $X_i = Y_i Y_{i+1}^2 Y_{i+2} = Y_i Y_{i+1} Y_{i+2}$). Če pa je $|i-j| \geq 2$, velja

$E(X_iX_j) = 1/16$. Brž ko je $n \geq 3$, je za vsak $i \in \mathbb{Z}_n$ natanko en indeks $j \in \mathbb{Z}_n$ enak i, za natanko dva velja $|i - j| = 1$, za preostalih $n - 3$ indeksov pa velja $|i - j| \geq 2$. Sledi:

$$E\,S^2 = n\left(\frac{1}{4} + 2 \cdot \frac{1}{8} + (n-3) \cdot \frac{1}{16}\right) = \frac{n^2 + 5n}{16}$$

in končno:

$$D(S) = E\,S^2 - (E\,S)^2 = \frac{5n}{16}\,.$$

Drugi način. Nastavimo:

$$D(S) = \sum_{i \in \mathbb{Z}_n} \sum_{j \in \mathbb{Z}_n} K(X_i, X_j)\,.$$

Če je $i = j$, je $K(X_i, X_j) = D(X_i) = 3/16$. Če je $|i - j| = 1$ in $n \geq 3$, izračunamo $K(X_i, X_j) = E(X_iX_j) - (E\,X_i)(E\,X_j) = 1/16$. Za $|i - j| \geq 2$ pa sta slučajni spremenljivki X_i in X_j neodvisni, torej je $K(X_i, X_j) = 0$. Podobno kot pri prvem načinu seštejemo (za $n \geq 3$):

$$D(S) = n\left(\frac{3}{16} + 2 \cdot \frac{1}{16}\right) = \frac{5n}{16}\,.$$

Za $n = 1$ posebej dobimo, da je $D(S) = 0$, za $n = 2$ pa je

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

in zato $D(S) = 3/4$.

3. a) Zaradi linearnosti matematičnega upanja velja:

$$E\big[(X - X')(Y - Y')\big] = E(XY) - E(XY') - E(X'Y) + E(X'Y')\,.$$

Nadalje zaradi neodvisnosti velja:

$$E\big[(X - X')(Y - Y')\big] = E(XY) - (E\,X)(E\,Y') - (E\,X')(E\,Y) + E(X'Y')\,.$$

Ker je slučajni vektor (X', Y') porazdeljen enako kot (X, Y), končno velja:

$$E\big[(X-X')(Y-Y')\big] = E(XY) - (E\,X)(E\,Y) - (E\,X)(E\,Y) + E(XY) = 2\,K(X,Y)\,.$$

b) Iz prejšnje točke sledi:

$$K(X,Y) = \frac{1}{2}\,E\Big[\big(f(X) - f(X')\big)\big(g(X) - g(X')\big)\Big]\,. \qquad (*)$$

Če je $X \geq X'$, iz monotonosti funkcij f in g sledi $f(X) \geq f(X')$ in $g(X) \geq g(X')$, torej $\big(f(X) - f(X')\big)\big(g(X) - g(X')\big) \geq 0$. Če pa je $X \leq X'$, velja $f(X) \leq f(X')$ in $g(X) \leq g(X')$, torej prav tako $\big(f(X) - f(X')\big)\big(g(X) - g(X')\big) \geq 0$. Zato je desna stran v $(*)$ vedno nenegativna.

4. *Prvi način.* Iz rodovne funkcije:

$$G_X(s) = E(s^X) = \sum_{k=0}^{\infty} s^k \frac{\lambda^k e^{-\lambda}}{k!} = e^{\lambda(s-1)}$$

in njenih odvodov:

$$G_X^{(r)}(s) = \lambda^r e^{\lambda(s-1)}$$

dobimo:

$$E\big[X(X-1)(X-2)\cdots(X-r+1)\big] = \lambda^r\,.$$

Sledi:

$$\begin{aligned} E\,X &= \lambda \\ E\,X^2 &= E\big[X(X-1)\big] + E\,X = \lambda^2 + \lambda \\ E\,X^3 &= E\big[X(X-1)(X-2)\big] + 3\,E\,X^2 - 2\,E\,X = \lambda^3 + 3\lambda^2 + \lambda\,. \end{aligned}$$

Torej velja:

$$E\big[(X - E\,X)^3\big] = E\,X^3 - 3(E\,X^2)(E\,X) + 2(E\,X)^3 = \lambda\,.$$

Drugi način. Pomagamo si z momentno-rodovno funkcijo:

$$M_X(t) = G_X(e^t) = e^{\lambda(e^t-1)}$$

Z zaporednim odvajanjem dobimo:

$$\begin{aligned} M_X'(t) &= \lambda e^{\lambda(e^t-1)+t}\,, \\ M_X''(t) &= \lambda^2 e^{\lambda(e^t-1)+2t} + \lambda e^{\lambda(e^t-1)+t}\,, \\ M_X'''(t) &= \lambda^3 e^{\lambda(e^t-1)+3t} + 3\lambda^2 e^{\lambda(e^t-1)+2t} + \lambda e^{\lambda(e^t-1)+t}\,, \end{aligned}$$

od koder sledi:

$$E\,X^3 = M_X'''(0) = \lambda^3 + 3\lambda^2 + \lambda$$

in tako kot pri prvem načinu dobimo iskani rezultat.

Tretji način. Spomnimo se, da je tretji centralni moment enak tretji kumulanti, torej:

$$E\big[(X - E\,X)^3\big] = \kappa_3 = \frac{\mathrm{d}^3}{\mathrm{d}t^3}\bigg|_{t=0} \ln M_X(t) = \frac{\mathrm{d}^3}{\mathrm{d}t^3}\bigg|_{t=0} \big[\lambda(e^t-1)\big] = \lambda e^t\Big|_{t=0} = \lambda\,.$$

(vse kumulante Poissonove porazdelitve $\mathrm{P}(\lambda)$ so enake λ).

Rešitve kolokvija iz verjetnosti in statistike z dne 22. 4. 2010

Matematika – UNI-BOL

1. Uporabimo lahko centralni limitni izrek. Iz:

$$E(X_i) = 2\,, \quad D(X_i) = 0{\cdot}6\,, \quad E(S) = 800\,, \quad D(S) = 240\,,$$

kjer je $S = X_1 + X_2 + \cdots + X_{400}$, dobimo, da je približno $S \sim \mathrm{N}\big(800, \sqrt{240}\big)$. Če ustrezno verjetnost aproksimiramo tako, kot je napisana, dobimo:

$$P(S < 780) \approx \frac{1}{2} - \Phi\left(\frac{20}{\sqrt{240}}\right) \doteq 0{\cdot}5 - \Phi(1{\cdot}2910) \doteq 0{\cdot}0984\,.$$

Lahko pa upoštevamo, da ima S zalogo vrednosti na celoštevilski mreži, in aproksimiramo:

$$P(S < 779{\cdot}5) \approx \frac{1}{2} - \Phi\left(\frac{20{\cdot}5}{\sqrt{240}}\right) \doteq 0{\cdot}5 - \Phi(1{\cdot}3233) \doteq 0{\cdot}0929\,.$$

Točen rezultat: $0{\cdot}0914338$.

2. Iz navzkrižne gostote vzorca:

$$\begin{aligned} p(x_1, \ldots x_n) &= \frac{a^{2an}}{\big(\Gamma(2a)\big)^n}\, x_1^{2a-1} x_2^{2a-1} \cdots x_n^{2a-1}\, e^{-ax_1}\, e^{-ax_2} \cdots e^{-ax_n} = \\ &= \frac{a^{2an}}{\big(\Gamma(2a)\big)^n\, x_1\, x_2 \cdots x_n}\, e^{a(2\ln x_1 + 2\ln x_2 + \cdots + 2\ln x_n - x_1 - x_2 - \cdots - x_n)} \end{aligned}$$

razberemo, da je minimalna zadostna statistika recimo:

$$2(\ln X_1 + \ln X_2 + \cdots + \ln X_n) - X_1 - X_2 - \cdots - X_n\,.$$

3. a) Slučajne spremenljivke $Z_i = (X_i - \mu)/\sigma$ so neodvisne in porazdeljene standardno normalno, torej neodvisno od μ in σ. Ker je:

$$U = \frac{(Z_1 - \bar{Z})(Z_2 - \bar{Z}) \cdots (Z_n - \bar{Z})}{\big[(Z_1 - \bar{Z})^2 + (Z_2 - \bar{Z})^2 + \cdots + (Z_n - \bar{Z})^2\big]^{n/2}}$$

kjer je $\bar{Z} = (Z_1 + \cdots + Z_n)/n$, je tudi porazdelitev te statistike neodvisna od μ in σ, se pravi, da je statistika U postranska.

b) Znano je, da je normalna družina eksponentna družina, katere naravni parametrični prostor ima neprazno notranjost. V takih modelih so postranske statistike neodvisne od minimalne zadostne statistike, ki pa je v našem primeru $\big(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\big)$. Torej je U neodvisna od $\sum_{i=1}^n X_i^2$.

4. a) Iz funkcije verjetja in njenega odvoda:

$$L(k,\lambda) = \frac{\lambda^k\, e^{-\lambda}}{k!}\,, \qquad \frac{\partial L}{\partial \lambda} = \frac{(k-\lambda)\lambda^{k-1}\, e^{-\lambda}}{k!}$$

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 najmanjšo možno disperzijo, brž ko bo funkcija minimalne zadostne statistike X. Iščemo torej tako funkcijo h, da bo $E_\lambda\big(h(X)\big) = \lambda^2$ za vse λ. To lahko naredimo z nastavkom:

$$\sum_{k=0}^{\infty} \frac{h(k)\,\lambda^k\, e^{-\lambda}}{k!} = \lambda^2$$

oziroma:

$$\sum_{k=0}^{\infty} \frac{h(k)\,\lambda^k}{k!} = \lambda^2\, e^{\lambda}\,,$$

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 $h(k) = k(k-1)$, torej je iskana statistika $X(X-1)$. Le-to lahko tudi kar uganemo iz prvih dveh momentov Poissonove porazdelitve.

c) Velja:

$$\begin{aligned} q(X^2 - aX) &= E\big[(X^2 - aX - \lambda^2)^2\big] = \\ &= E(X^4) + a^2\, E(X^2) + \lambda^4 - 2a\, E(X^3) - 2\lambda^2\, E(X^2) + 2a\lambda^2\, E(X) = \\ &= 4\lambda^3 + (7 - 6a + a^2)\lambda^2 + (1 - 2a + a^2)\lambda\,. \end{aligned}$$

Smiselne izbire parametra a so tiste, pri katerih ne obstaja nobena druga vrednost, pri kateri bi bila srednja kvadratična napaka manjša za *vse* λ. To pa so tiste, pri katerih se koeficienta $c_1(a) := 1 - 2a + a^2$ in $c_2(a) = 7 - 6a + a^2$ ne moreta hkrati zmanjšati. Za $a < 1$ sta oba koeficienta strogo padajoča, za $a > 3$ pa oba strogo naraščajoča v a, torej ju lahko hkrati zmanjšamo in izbira v teh dveh območjih ni smiselna. Za $1 \le a \le 3$ pa je c_1 naraščajoč, c_2 pa padajoč v a, torej se ne moreta hkrati zmanjšati (poljubni dve različni izbiri iz $1 \le a \le 3$ sta neprimerljivi). Zato je iz intervala $[1, 3]$ smiselno izbirati.

Rešitve kolokvija iz verjetnosti in statistike z dne 3. 6. 2010

Matematika – UNI-BOL

1. a) *Prvi način*: parameter s stabilno disperzijo iščemo kot funkcijo naravnega parametra. Tu gre namreč za enoparametrično eksponentno družino, kjer verjetnostno funkcijo zapišemo v obliki:

$$p(x;\lambda) = \frac{e^{-\lambda}}{x!}\, e^{x\ln\lambda}\,,$$

torej je naravni parameter $\theta = \ln\lambda$. Parameter s stabilno disperzijo iščemo v obliki $g(\theta)$, kjer je $g'(\theta) = \sigma(X) = e^{\theta/2}$, kar bo res za $g(\theta) = 2\,e^{\theta/2} = 2\sqrt{\lambda}$.

Drugi način: parameter s stabilno disperzijo iščemo kot funkcijo matematičnega upanja, v tem primeru $E(X) = \lambda$. Za dovolj gladko naraščajočo funkcijo h bo približno veljalo:

$$\sigma\big(h(X)\big) \approx h'\big(E(X)\big)\sigma(X) = h'(\lambda)\sqrt{\lambda}\,.$$

Desna stran bo enaka 1 za $h'(\lambda)\sqrt{\lambda} = 1$, to pa velja za $h(\lambda) = 2\sqrt{\lambda}$. Dobimo isti parameter kot pri prvem načinu (v eksponentnih družinah to vedno velja).

b) Interval zaupanja konstruiramo na podlagi aproksimacije z normalno porazdelitvijo, ta pa velja tudi za eno samo opažanje, brž ko je λ dovolj velik. Za velike λ se namreč Poissonova porazdelitev bliža normalni, to pa zato, ker je to tudi porazdelitev velikega števila vsote neodvisnih Poissonovih slučajnih spremenljivk s parametrom blizu 1, kjer je Poissonova porazdelitev še dovolj "lepa". Glede na točko a), normalno aproksimacijo in dejstvo, da je cenilka po metodi največjega verjetja enaka $\hat{\lambda} = X$, je iskani aproksimativni interval zaupanja določen z neenačbo:

$$2\big|\sqrt{\lambda} - \sqrt{X}\big| < \Phi^{-1}\left(\frac{1-\alpha}{2}\right) \doteq 1{\cdot}96\,,$$

kar za naše opažanje (ustrezno zaokroženo) znese:

$$81{\cdot}3 < \lambda < 120{\cdot}6\,.$$

2. $\bar{X} = 19{\cdot}71$, $\quad S_\mathrm{p} \doteq 2{\cdot}925$, $\quad df = 9$, $\quad c = t_{0.975} \doteq 3{\cdot}25$, $\quad \Delta \doteq 3{\cdot}01$.
Interval zaupanja: $16{\cdot}70 < \mu < 22{\cdot}72$.

3. 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, za $x > 0$ velja:

$$\frac{L_0(x)}{L_1(x)} = \frac{\frac{1}{500}\,e^{-x/500}}{\frac{1}{100}\,e^{-x/100}} = \frac{1}{5}\,e^{4x/500}\,.$$

Če je torej X opažena življenjska doba žarnice, bomo ničelno hipotezo zavrnili, če bo testna statistika $\frac{1}{5}\,e^{4X/500}$ premajhna, to pa bo tedaj, ko bo življenjska doba X premajhna.

Isti sklep lahko dobimo tudi iz dejstva, da gre za enoparametrično eksponentno družino z naravno zadostno statistiko $-X$ glede na parametrizacijo z λ, in dejstva, da je vrednost parametra pri alternativni hipotezi večja od tiste pri ničelni hipotezi. Ker je X pri ničelni hipotezi porazdeljena zvezno, randomizacija ni potrebna in 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 $c = -500 \ln(0˙95) \doteq 25˙64$ (zaokroženo navzdol). Moč tega testa je:

$$P_{H_1}(X \le c) = \int_0^c \frac{1}{100}\, e^{-x/100}\, \mathrm{d}x = 1 - e^{-c/100} \doteq 0˙226\,.$$

4. Pri modelu, ki ga obravnava naloga, lahko funkcijo verjetja zapišemo v obliki:

$$L(x_1, \dots x_n; \mu, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2\right)\right]$$

in maksimum je pri celem 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:

$$L(x_1, \dots x_n; \sigma^2, \sigma) = \left(\sigma\sqrt{2\pi}\right)^{-n} \exp\left[-\frac{1}{\sigma^2}\sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i - \frac{n\sigma^2}{2}\right]$$

in iz:

$$\frac{\mathrm{d}}{\mathrm{d}\sigma} \ln L(x_1, \dots x_n; \sigma^2, \sigma) = \frac{1}{\sigma^3}\sum_{i=1}^n x_i^2 - n\sigma - \frac{n}{\sigma}$$

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:

$$\Lambda = \frac{\sup_{\Theta_{H_0}} L(X_1, \dots X_n; \mu, \sigma)}{\sup_{\Theta} L(X_1, \dots X_n; \mu, \sigma)} = \frac{L(X_1, \dots X_n; \hat{\mu}_2, \hat{\sigma}_2)}{L(X_1, \dots X_n; \hat{\mu}_1, \hat{\sigma}_1)}\,,$$

ima $-2\ln(\Lambda)$ približno porazdelitev hi kvadrat z eno prostostno stopnjo. Za naše opažanje dobimo $-2\ln(\Lambda) \doteq 5˙61$, kritična vrednost pa je $\chi^2_{0.99} \doteq 6˙63$. Hipoteze torej ne moremo zavrniti.

Rešitve izpita iz verjetnosti in statistike z dne 23. 6. 2010

Matematika – UNI-BOL

1. a) Označimo z A dogodek, da Tonetu toča obtolče avto, in s T_k dogodek, da toča prvič tolče k-tega junija. Tedaj velja:

$$P(A) = \sum_{k=1}^{\infty} P(A \cap T_k) = \sum_{k=1}^{\infty} P(T_k)\, P(A \mid T_k) = \sum_{k=1}^{\infty} \left(\frac{9}{10}\right)^{k-1} \cdot \frac{1}{10} \cdot P(A \mid T_k)\,.$$

Za $k = 1, 2, 3, 4$ je $P(A \mid T_k) = 1$, za $k > 4$ pa se $P(A \mid T_k)$ ujema z verjetnostjo dogodka, da krovci nadstreška niso končali prek k-tim junijem, le-ta pa je enaka $\sum_{r=k}^{\infty} 2^{3-r} = 2^{4-k}$. Torej je:

$$\begin{aligned}
P(A) &= \sum_{k=1}^{4} \left(\frac{9}{10}\right)^{k-1} \cdot \frac{1}{10} + \sum_{k=5}^{\infty} \left(\frac{9}{10}\right)^{k-1} \cdot \frac{1}{10} \cdot 2^{4-k} = \\
&= 1 - \left(\frac{9}{10}\right)^4 + \frac{1}{11} \cdot \left(\frac{9}{10}\right)^4 = \\
&= 1 - \frac{10}{11} \cdot \left(\frac{9}{10}\right)^4 = \\
&= \frac{4439}{11000} \doteq 0\text{˙}404\,.
\end{aligned}$$

b) Označimo z Z dogodek, da krovci zamudijo. Velja:

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

$$P(Z \cap A) = \sum_{k=1}^{\infty} P(T_k \cap A \cap Z) = \sum_{k=1}^{\infty} P(A \cap T_k)\, P(Z \mid A \cap T_k)\,.$$

Za $k = 1, 2, 3, 4$ je dogodek Z neodvisen od dogodka $A \cap T_k$, torej je $P(Z \mid A \cap T_k) = P(Z) = 1/2$. Za $k > 4$ pa je $P(Z \mid A \cap T_k) = 1$. Sledi:

$$\begin{aligned}
P(Z \cap A) &= \frac{1}{2} \sum_{k=1}^{4} \left(\frac{9}{10}\right)^{k-1} \cdot \frac{1}{10} + \sum_{k=5}^{\infty} \left(\frac{9}{10}\right)^{k-1} \cdot \frac{1}{10} \cdot 2^{4-k} = \\
&= \frac{1}{2} \left[1 - \left(\frac{9}{10}\right)^4\right] + \frac{1}{11} \cdot \left(\frac{9}{10}\right)^4 = \\
&= \frac{1}{2} \left[1 - \frac{9}{11} \cdot \left(\frac{9}{10}\right)^4\right] = \\
&= \frac{50951}{220000}
\end{aligned}$$

in od tod končno:

$$P(Z \mid A) = \frac{50951}{220000} \Bigg/ \frac{4439}{11000} = \frac{50951}{88780} \doteq 0\text{˙}574\,.$$

2. Najprej opazimo, da je:

$$\begin{aligned}
E\big[\operatorname{sgn}(X-Y)\operatorname{sgn}(X-Z)\big] &= P(X>Y, X>Z) + P(X<Y, X<Z) - \\
&\qquad - P(X>Y, X<Z) - P(X<Y, X>Z) = \\
&= P(Z<Y<X) + P(Y<Z<X) + \\
&\qquad + P(X<Y<Z) + P(X<Z<Y) - \\
&\qquad - P(Y<X<Z) - P(Z<X<Y)\,.
\end{aligned}$$

Zaradi simetrije je vseh šest možnih ureditev enako verjetnih in zaradi zveznosti je verjetnost, da sta kateri izmed danih slučajnih spremenljivk enaki, enaka nič. Torej je verjetnost posamezne stroge ureditve enaka $1/6$ in zato:

$$E\big[\operatorname{sgn}(X-Y)\operatorname{sgn}(X-Z)\big] = \frac{1}{3}\,.$$

Opomba. Seveda lahko nalogo rešimo tudi z integrali.

3. Najprej gostoto posamezne spremenljivke zapišemo v obliki, iz katere razberemo, da gre za eksponentno družino: za $1/e < x < 1$ velja:

$$p(x) = (1+\ln x)\frac{\alpha}{1-e^{-\alpha/e}}\,e^{\alpha x\ln x}\,,$$

funkcijo verjetja, t. j. navzkrižno verjetnostno gostoto, pa lahko za $1/e < x_1, \dots x_n < 1$ zapišemo v obliki:

$$p(x_1)\,p(x_2)\cdots p(x_n) = (1+\ln x_1)\cdots(1+\ln x_n)\left(\frac{\alpha}{1-e^{-\alpha/e}}\right)^n e^{\alpha(x_1\ln x_1+\cdots+x_n\ln x_n)}\,.$$

Ker gre za eksponentno družino, katere parametrični prostor vsebuje odprto množico, je minimalna zadostna statistika enaka:

$$X_1\ln X_1 + X_2\ln X_2 + \cdots + X_n\ln X_n\,.$$

4. a) Tinetovo povprečje je porazdeljeno normalno $\mathrm{N}(0, 1/10)$, torej ima gostoto:

$$f_{\mathrm{Ti}}(x) = \frac{10}{\sqrt{2\pi}}\,e^{-50x^2}$$

Tonetovo povprečje pa je porazdeljeno normalno $\mathrm{N}(0, 1/100)$, torej ima gostoto:

$$f_{\mathrm{To}}(x) = \frac{100}{\sqrt{2\pi}}\,e^{-5000x^2}\,.$$

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:

$$\frac{f_{\mathrm{To}}(X)}{f_{\mathrm{Ti}}(X)} = 10\,e^{-4950X^2}$$

premajhno, to pa je takrat, ko je vrednost $|X|$ prevelika. Če je P_{To} verjetnost pri Tonetovem povprečju, za $c \geq 0$ velja:

$$P_{\mathrm{To}}(|X| \geq c) = 1 - 2\Phi(100c)$$

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:

$$P_{\mathrm{Ti}}(|X| \geq c) = 1 - 2\Phi(10c) \doteq 0\text{·}845\,.$$

Rešitve izpita iz verjetnosti in statistike z dne 6. 9. 2010

Matematika – UNI-BOL

1. Računajmo:

$$
\begin{aligned}
P(Y > 1) &= \int_1^\infty e^{-y}\,\mathrm{d}y = \frac{1}{e}\,,\\
P(X > Y, Y > 1) &= \int_1^\infty \int_y^\infty e^{-x}\,e^{-y}\,\mathrm{d}x\,\mathrm{d}y = \int_1^\infty e^{-2y}\,\mathrm{d}y = \frac{1}{2e^2}\,,\\
P(X > Y \mid Y > 1) &= \frac{P(X > Y, Y > 1)}{(Y > 1)} = \frac{1}{2e}\,.
\end{aligned}
$$

2. Označimo z A_i dogodek, da i-ti otrok ne dobi nobene žoge, in pišimo:

$$
S = X_1 + X_2 + \cdots + X_n\,,\ \text{kjer je}\, X_i = \mathbf{1}_{A_i} = \begin{cases} 1 & ;\ i\text{-ti otrok ne dobi nobene žoge}\\ 0 & ;\ \text{sicer}\end{cases}
$$

Iz $E(X_i) = P(A_i) = \left(1 - \dfrac{1}{n-1}\right)^{n-1}$ dobimo $E(S) = n\left(1 - \dfrac{1}{n-1}\right)^{n-1}$. Za izračun disperzije pa pišimo:

$$
\begin{aligned}
D(S) &= E(S^2) - \big(E(S)\big)^2,\\
E(S^2) &= \sum_{i=1}^n E(X_i^2) + \sum_{\substack{1 \le i,j \le n\\ i \ne j}} E(X_i X_j) =\\
&= \sum_{i=1}^n P(A_i) + \sum_{\substack{1 \le i,j \le n\\ i \ne j}} P(A_i \cap A_j) =\\
&= n\left(1 - \frac{1}{n-1}\right)^{n-1} + n(n-1)\left(1 - \frac{1}{n-1}\right)^2\left(1 - \frac{2}{n-1}\right)^{n-2}.
\end{aligned}
$$

Torej je:

$$
\begin{aligned}
D(S) = n\left(1 - \frac{1}{n-1}\right)^{n-1} &+ n(n-1)\left(1 - \frac{1}{n-1}\right)^2\left(1 - \frac{2}{n-1}\right)^{n-2} -\\
&- n^2\left(1 - \frac{1}{n-1}\right)^{2n-2}.
\end{aligned}
$$

Opomba. Zanimivo je raziskati asimptotično obnašanje količin, ki jih računamo, ko gre n proti neskončno. Če z $a_n \sim b_n$ označimo, da je $\lim_{n\to\infty} a_n/b_n = 1$, se da s sredstvi iz elementarne analize dokazati, da je:

$$
E(S) \sim n/e \quad \text{in} \quad D(S) \sim n(e^{-1} - 2e^{-2})\,.
$$

Z bolj sofisticiranimi sredstvi iz verjetnostnega računa pa se da raziskati tudi asimptotično obnašanje cele porazdelitve slučajne spremenljivke S, in sicer se izkaže, da slučajne spremenljivke:

$$\frac{S - n/e}{\sqrt{n}}$$

šibko konvergirajo proti normalni porazdelitvi z matematičnim upanjem nič in disperzijo $e^{-1} - 2e^{-2}$.

3. Ker lahko zapišemo:

$$p(x) = \frac{a}{x}\, e^{a \ln x}$$

in za ustrezni vzorec:

$$p_{X_1,\ldots X_n}(x_1, \ldots x_n) = \frac{a^n}{x_1 x_2 \cdots x_n}\, e^{a(\ln x_1 + \cdots + \ln x_n)}\,,$$

gre za eksponentno družino, kjer ima naravni parametrični prostor neprazno notranjost. Torej je nepristranska cenilka za $1/a$ z najmanjšo možno disperzijo, če obstaja, funkcija minimalne zadostne statistike $\ln X_1 + \cdots + \ln X_n$. Iz:

$$E(\ln X_1) = a \int_0^1 x^{a-1} \ln x \,\mathrm{d}x = x^a \ln x \Big|_0^1 - \int_0^1 x^{a-1}\,\mathrm{d}x = -\frac{1}{a}$$

dobimo $E(\ln X_1 + \cdots + \ln X_n) = -n/a$. Iskana cenilka je torej:

$$-\frac{\ln X_1 + \ln X_2 + \cdots + \ln X_n}{n}\,.$$

To cenilko dobimo tudi po metodi največjega verjetja, a je potrebno preveriti, da je nepristranska.

4. Vzorčno povprečje: $\bar{X} = 24{\cdot}95$.
Popravljeni vzorčni standardni odklon: $S_{\mathrm{p}} \doteq 3{\cdot}776$.
Kvantila porazdelitve hi kvadrat pri $df = 9$: $\chi^2_{0.025} \doteq 2{\cdot}700$, $\chi^2_{0.975} \doteq 19{\cdot}02$.
Interval zaupanja (ustrezno zaokrožen): $2{\cdot}59 < \sigma < 6{\cdot}90$.

Rešitve izpita iz verjetnosti in statistike z dne 20. 9. 2010

Matematika – UNI-BOL

1. a) $\dfrac{1}{\binom{n}{2}}\left[\binom{k}{2}\cdot 2\cdot\dfrac{1}{2}\cdot\dfrac{1}{2}+k(n-k)\left(p\cdot\dfrac{1}{2}+(1-p)\cdot\dfrac{1}{2}\right)+\binom{n-k}{2}\cdot 2p(1-p)\right]=$

$=\dfrac{k(k-1)+2k(n-k)p+4(n-k)(n-k-1)p(1-p)}{2n(n-1)}$.

b) $\dfrac{2k(n-k)p+4(n-k)(n-k-1)p(1-p)}{k(k-1)+2k(n-k)p+4(n-k)(n-k-1)p(1-p)}$.

2. Iz:

$$T=\begin{cases} A+1 & ;\ A\le t_0\\ t_0+3 & ;\ A>t_0\end{cases}$$

dobimo:

$$E(T)=\int_0^{t_0}\frac{(t+1)}{(1+t)^2}\,\mathrm{d}t+(t_0+3)\int_{t_0}^{\infty}\frac{\mathrm{d}t}{(1+t)^2}=\ln(1+t_0)+\frac{t_0+3}{1+t_0}\,.$$

Z odvajanjem dobimo:

$$\frac{\mathrm{d}\,E(T)}{\mathrm{d}t_0}=\frac{1}{1+t_0}-\frac{2}{(1+t_0)^2}\,,$$

kar je minimalno pri $t_0=1$. Optimalna strategija za Toneta je torej, da se odpravi peš, če po eni uri še ni avtobusa.

3. a) Ker je $E(X)=a/2$ in $D(X)=a^2/12$, bo za velike n po centralnem limitnem izreku približno veljalo $\bar{X}\sim \mathrm{N}\left(\dfrac{a}{2},\dfrac{a}{\sqrt{12n}}\right)$. Po delta metodi bo potem za dovolj lepo funkcijo g približno:

$$g(\bar{X})\sim \mathrm{N}\left(g\left(\frac{a}{2}\right),\left|g'\left(\frac{a}{2}\right)\right|\frac{a}{\sqrt{12n}}\right)\,.$$

Približno (asimptotično) enakost $D\big[g(\bar{X})\big]=1$ bomo torej dosegli, če bo veljalo:

$$g'\left(\frac{a}{2}\right)\frac{a}{\sqrt{12n}}=1$$

oziroma:

$$g'(t)=\frac{\sqrt{3n}}{t}\,,$$

torej lahko postavimo $g(t)=\sqrt{3n}\ln t$.

b) Ker je približno $\sqrt{3n}\ln\bar{X} \sim \mathrm{N}\left(\sqrt{3n}\ln\frac{a}{2}, 1\right)$, bo približni interval zaupanja pri stopnji zaupanja β določen z zvezo:

$$\left|\sqrt{3n}\ln\bar{X} - \sqrt{3n}\ln\frac{a}{2}\right| < \Phi^{-1}\left(\frac{\beta}{2}\right)$$

oziroma:

$$2\bar{X}\exp\left[-\frac{1}{\sqrt{3n}}\Phi^{-1}\left(\frac{\beta}{2}\right)\right] < a < 2\bar{X}\exp\left[\frac{1}{\sqrt{3n}}\Phi^{-1}\left(\frac{\beta}{2}\right)\right].$$

Ko vstavimo $\bar{X} = 3$ in $\beta = 0\text{˙}99$, ob ustreznem zaokrožanju dobimo $5\text{˙}17 < a < 6\text{˙}97$.

4. Pri modelu, ki ga obravnava naloga, lahko funkcijo verjetja zapišemo v obliki:

$$L(x, y; \lambda, \mu) = \frac{\lambda^x \mu^y e^{-\lambda-\mu}}{x!y!}$$

in maksimum je pri celem modelu dosežen pri $\lambda = x$ in $\mu = y$. Če se omejimo na ničelno hipotezo $\lambda = \mu = 200$, pa je:

$$L(x, y; 200, 200) = \frac{\lambda^x \mu^y e^{-400}}{x!y!}$$

(ker je hipoteza enostavna, ni potrebno iskati maksimuma). Če z Λ označimo razmerje verjetij:

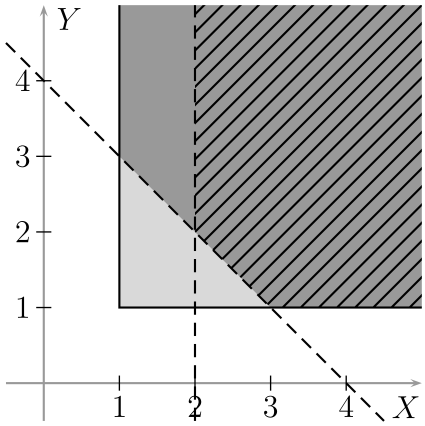
$$\Lambda = \frac{\sup_{\Theta_{H_0}} L(X, Y; \lambda, \mu)}{\sup_{\Theta} L(X, Y; \lambda, \mu)} = \frac{L(X, Y; 200, 200)}{L(X, Y; X, Y)} = \frac{200^{X+Y} e^{-400}}{X^X\, Y^Y e^{-X-Y}},$$

ima $-2\ln(\Lambda)$ približno porazdelitev hi kvadrat z dvema prostostnima stopnjama. Za naše opažanje dobimo $-2\ln(\Lambda) \doteq 6\text{˙}36$, kritična vrednost pa je $\chi^2_{0.95} \doteq 5\text{˙}99$. Hipotezo torej zavrnemo.

Rešitve izpita iz verjetnosti in statistike z dne 16. 2. 2011

Matematika – UNI-BOL

1. Na skici je svetlosivo označena zaloga vrednosti slučajnega vektorja (X, Y), sivo je označen dogodek $\{X + Y > 4\}$, šrafiran pa je presek $\{X > 2, X + Y > 4\}$:



Velja:

$$\begin{aligned} P(X + Y > 4) &= P(1 < X < 3, X + Y > 4) + P(X \geq 3) = \\ &= \int_1^3 \int_{4-x}^{\infty} \frac{1}{x^2 y^2} \, \mathrm{d}y \, \mathrm{d}x + \int_3^{\infty} \frac{1}{x^2} \, \mathrm{d}x = \\ &= \int_1^3 \frac{\mathrm{d}x}{x^2(4-x)} + \frac{1}{3} \,. \end{aligned}$$

Iz:

$$\int \frac{\mathrm{d}x}{x^2(4-x)} = \frac{1}{16} \int \left(\frac{1}{x} + \frac{4}{x^2} + \frac{1}{4-x} \right) \mathrm{d}x = \frac{1}{16} \ln \left| \frac{x}{4-x} \right| - \frac{1}{4x} + C$$

dobimo:

$$P(X + Y > 4) = \frac{\ln 3}{8} + \frac{1}{2} \,.$$

Nadalje je:

$$\begin{aligned} P(X > 2, X + Y > 4) &= P(2 < X < 3, X + Y > 4) + P(X \geq 3) = \\ &= \int_2^3 \frac{\mathrm{d}x}{x^2(4-x)} + \frac{1}{3} = \\ &= \frac{\ln 3}{16} + \frac{3}{8} \end{aligned}$$

in končno:

$$P(X > 2 \mid X + Y > 4) = \frac{P(X > 2, X + Y > 4)}{P(X + Y > 4)} = \frac{\ln 3 + 6}{2 \ln 3 + 8} \doteq 0{\cdot}696 \,.$$

2. Za $k = 2, 3, 4, 5$ je $\{X = k\}$ dogodek, da smo bodisi najprej izvlekli $k - 1$ rdečih kart, nato pa črno, bodisi da smo najprej izvlekli $k - 1$ črnih kart, nato pa rdečo. Torej je:

$$\begin{aligned}
P(X = 2) &= 2 \cdot \frac{4}{8} \cdot \frac{4}{7} = \frac{4}{7} \\
P(X = 3) &= 2 \cdot \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{4}{6} = \frac{2}{7} \\
P(X = 4) &= 2 \cdot \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{4}{5} = \frac{4}{35} \\
P(X = 5) &= 2 \cdot \frac{4}{8} \cdot \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{5} \cdot \frac{4}{4} = \frac{1}{35}
\end{aligned}$$

oziroma:

$$X \sim \begin{pmatrix} 2 & 3 & 4 & 5 \\ \frac{4}{7} & \frac{2}{7} & \frac{4}{35} & \frac{1}{35} \end{pmatrix},$$

od koder izračunamo $E(X) = \dfrac{13}{5} = 2\text{˙}6$ in $D(X) = \dfrac{16}{25} = 0\text{˙}64$.

3. Gostota normalne porazdelitve $\mathrm{N}\big(a/2, \sqrt{a}\big)$ je enaka:

$$\frac{1}{2\pi a} \exp\left(-\frac{\left(x - \frac{a}{2}\right)^2}{2a}\right) = \frac{1}{2\pi a} \exp\left(-\frac{x^2}{2a} + \frac{1}{2} - \frac{a}{8}\right),$$

torej je gostota porazdelitve našega opažanja enaka:

$$f_{X_1,\ldots X_n}(x_1, \ldots x_n) = (2\pi a)^{-n/2} \exp\left(-\frac{1}{2a}\sum_{i=1}^n x_i^2 + \frac{n}{2} - \frac{na}{8}\right).$$

Od tod razberemo, da gre za enoparametrično eksponentno družino z minimalno zadostno statistiko $\sum_{i=1}^n X_i^2$. Nepristranska cenilka ima v tem primeru najmanjšo možno disperzijo natanko tedaj, ko je funkcija minimalne zadostne statistke. Ker je $E(X_i^2) = D(X_i) + \big(E(X_i)\big)^2 = a + a^2/4$, bo cenilka:

$$\frac{4}{n}\sum_{i=1}^n X_i$$

nepristranska za $a^2 + 4a$ in bo imela najmanjšo možno disperzijo.

4. a) $f_{-X}(x) = \begin{cases} \frac{2}{3}\, e^x & ; x \le 0 \\ \frac{2}{3}\, e^{-2x} & ; x \ge 0 \end{cases}$.

b) Glede na to, da sta ničelna in alternativna hipoteza obe enostavni, je test, ki temelji na razmerju verjetij, najmočnejši. Če z Z označimo naše opažanje (X ali $-X$), bomo torej ničelno hipotezo zavrnili, če bo razmerje verjetij:

$$\frac{f_X(Z)}{f_{-X}(Z)} = e^Z$$

premajhno, to pa je natanko tedaj, ko je sama vrednost Z premajhna. Ker je Z pri ničelni hipotezi porazdeljena zvezno, randomizacija ni potrebna in lahko H_0 zavrnemo, brž ko je $Z \le c$, kjer je $P_{H_0}(Z \le c) = \alpha$. Za $Z \le 0$ velja:

$$P_{H_0}(Z \le c) = P(X \le c) = \frac{2}{3} \int_{-\infty}^{c} e^{2x} \, \mathrm{d}x = \frac{e^{2c}}{3} \, .$$

Če je iskana kritična vrednost c negativna, torej reši enačbo $e^{2c}/3 = \alpha$, torej $c = \frac{1}{2} \ln(3\alpha)$. Velja tudi obratno: brž ko je tako dobljena vrednost negativna, je to že iskana kritična vrednost. Za $\alpha = 0{\cdot}05$ dobimo $c = -0.949$ (zaokroženo navzdol).