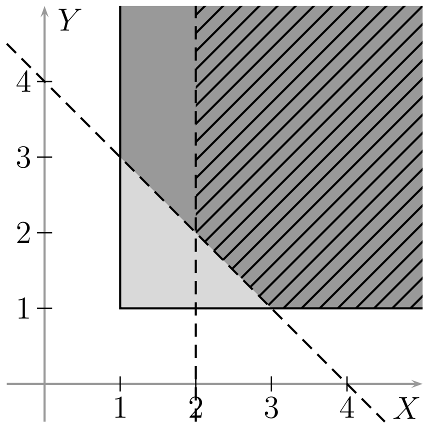
Rešitve izpita iz verjetnosti in statistike z dne 16. 2. 2011

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:

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

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

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:

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

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).