1. (15 bodů) Dvojrozměrný náhodný vektor (X, Y) má pravděpodobnosti hodnot dané tabulkou:

$X \backslash Y$	-1	0	1
-1	1/8	1/8	3/16
0	1/16	1/8	1/16
1	3/16	1/16	1/16

a) Určete marginální pravděpodobnostní funkce p_X a p_Y. Jsou náhodné veličiny X a Y závislé či nezávislé? Zdůvodněte.

b) Pro veličinu $W = X \cdot Y$ určete pravděpodobnostní funkci p_W a střední hodnotu $\mathrm{E}W$.

c) Spočítejte kovarianci $\mathrm{cov}(X, Y)$.

Řešení:

a) Marginální pravděpodobnostní funkce p_X a p_Y jsou

$X \backslash Y$	-1	0	1	p_X
-1	1/8	1/8	3/16	7/16
0	1/16	1/8	1/16	4/16
1	3/16	1/16	1/16	5/16
p_Y	6/16	5/16	5/16	

Protože je např. $p_{X,Y}(-1,-1) = \frac{1}{8} \neq \frac{6}{16} \cdot \frac{7}{16} = p_X(-1)p_Y(-1)$, jsou náhodné veličiny X a Y závislé.

b) Hodnoty veličiny $W = X \cdot Y$:

$X \backslash Y$	-1	0	1
-1	1	0	-1
0	0	0	0
1	-1	0	1

Její pravděpodobnostní funkce je

$$p_W(i) = \begin{cases} 3/16 + 3/16 = 3/8\,, & i = -1\,, \\ 1/8 + 1/16 + 1/8 + 1/16 + 1/16 = 7/16\,, & i = 0\,, \\ 1/8 + 1/16 = 3/16\,, & i = 1\,, \\ 0\,, & \text{jinak.} \end{cases}$$

Střední hodnotu $\mathrm{E}W$ vypočteme jako

$$\mathrm{E}W = (-1) \cdot 3/8 + 0 \cdot 7/16 + 1 \cdot 3/16 = -3/16\,.$$

c)

$$\begin{aligned}
\mathrm{E}X &= (-1)\cdot\frac{7}{16} + 0\cdot\frac{4}{16} + 1\cdot\frac{5}{16} = -\frac{2}{16}\,,\\
\mathrm{E}Y &= (-1)\cdot\frac{6}{16} + 0\cdot\frac{5}{16} + 1\cdot\frac{5}{16} = -\frac{1}{16}\,,\\
\mathrm{cov}(X,Y) &= \mathrm{E}W - \mathrm{E}X\cdot\mathrm{E}Y = -\frac{3}{16} - \left(-\frac{2}{16}\right)\cdot\left(-\frac{1}{16}\right) = -\frac{50}{256}\,.
\end{aligned}$$

Závislost X a Y plyne také z nenulové kovariance.

2. (15 bodů) U náhodně vybraných 100 osob, které navštěvují poradenské centrum pro zdravou výživu, byl sledován jejich BMI (body mass index = hmotnost[kg]/výška[m]^2) v závislosti na pohlaví. Pro jednoduchost byly zaznamenány pouze tři kategorie, a to podváha (BMI < 20), normální hmotnost (BMI v rozmezí 20 - 25) a nadváha (BMI > 25). Naměřeny byly následující četnosti:

	muži	ženy
podváha	10	10
normální hmotnost	20	20
nadváha	10	30

Statisticky otestujte na hladině

a) $\alpha = 5$ %, zda je hmotnost klientů poradenského centra závislá na pohlaví;

b) $\alpha = 1$ %, zda jsou mezi klienty poradenského centra počty mužů a žen přibližně stejné.

Řešení:

Nejprve doplníme tabulku o marginální četnosti

	muži	ženy	marginální četnosti
podváha	10	10	$n_{1.} = 20$
normální hmotnost	20	20	$n_{2.} = 40$
nadváha	10	30	$n_{3.} = 40$
marginální četnosti	$n_{.1} = 40$	$n_{.2} = 60$	$n = 100$

a) Za nulovou, resp. alternativní hypotézu si zvolíme

H_0 : hmotnost a pohlaví lidí v centru jsou nezávislé,
H_1 : hmotnost a pohlaví lidí v centru nejsou nezávislé.

Za platnosti H_0 má testová statistika

$$T = \sum_{i=1}^{3}\sum_{j=1}^{2} \frac{\left(n_{ij} - \frac{n_{i.}\cdot n_{.j}}{n}\right)^2}{\frac{n_{i.}\cdot n_{.j}}{n}}$$

rozdělení přibližně $\chi^2((3-1)(2-1)) = \chi^2(2)$. Vypočteme tedy hodnotu testové statistiky z dat, tj.

$$T = \frac{(10 - \frac{20\cdot 40}{100})^2}{\frac{20\cdot 40}{100}} + \frac{(10 - \frac{20\cdot 60}{100})^2}{\frac{20\cdot 60}{100}} + \frac{(20 - \frac{40\cdot 40}{100})^2}{\frac{40\cdot 40}{100}} + \frac{(20 - \frac{40\cdot 60}{100})^2}{\frac{40\cdot 60}{100}} + \frac{(10 - \frac{40\cdot 40}{100})^2}{\frac{40\cdot 40}{100}} + \\ + \frac{(30 - \frac{40\cdot 60}{100})^2}{\frac{40\cdot 60}{100}} = 6.25\,.$$

Jelikož $T > q_{\chi^2(2)}(0.95) \doteq 5.99$, zamítáme na hladině $\alpha = 5$ % hypotézu H_0 (nezávislost) ve prospěch H_1 (závislost).

b) Za nulovou, resp. alternativní hypotézu si zvolíme

$H_0 : p_1 = p_2 = 0.5$,
$H_1 : p_1 \neq p_2$,
kde p_1 a p_2 jsou pravděpodobnosti, že náhodně vybraný klient poradenského centra je muž, resp. žena.

Za platnosti H_0 má testová statistika

$$T = \sum_{i=1}^{2} \frac{(n_{i.} - np_i)^2}{np_i}$$

rozdělení $\chi^2(1)$. Vypočteme tedy hodnotu testové statistiky z dat, tj.

$$T = \frac{(40 - 100\cdot 0.5)^2}{100\cdot 0.5} + \frac{(60 - 100\cdot 0.5)^2}{100\cdot 0.5} = 4.$$

Jelikož $T < q_{\chi^2(1)}(0.99) \doteq 6.63$, nezamítáme na hladině $\alpha = 1$ % hypotézu H_0 (přibližně stejný počet mužů a žen) ve prospěch H_1 (významně různý počet mužů a žen).

3. (15 bodů) Alice a Bob opakovaně házejí jednou kostkou. Hra končí výhrou Boba, pokud padne šestka, výhrou Alice, pokud padnou dvě lichá čísla za sebou, jinak se hází dál. Jaké šance na výhru hráči mají?

Řešení:

Označme absorpční stavy:
1 ... vyhrál Bob,
2 ... vyhrála Alice.
Pokud v žádném z nich nejsme, rozlišujeme dva přechodné stavy:
3 ... ještě nic nepadlo (tj. začátek) nebo poslední číslo bylo sudé různé od 6, tj. 2 nebo 4,
4 ... poslední číslo bylo liché.

$$\boldsymbol{P} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1/6 & 0 & 1/3 & 1/2 \\ 1/6 & 1/2 & 1/3 & 0 \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{I}_2 & \boldsymbol{0} \\ \hline \boldsymbol{R} & \boldsymbol{Q} \end{array}\right),$$

$$\boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{R} = \begin{pmatrix} 1/6 & 0 \\ 1/6 & 1/2 \end{pmatrix}, \qquad \boldsymbol{Q} = \begin{pmatrix} 1/3 & 1/2 \\ 1/3 & 0 \end{pmatrix},$$

$$\boldsymbol{F} = (\boldsymbol{I}_2 - \boldsymbol{Q})^{-1} = \begin{pmatrix} 2/3 & -1/2 \\ -1/3 & 1 \end{pmatrix}^{-1} = 2 \begin{pmatrix} 1 & 1/2 \\ 1/3 & 2/3 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2/3 & 4/3 \end{pmatrix},$$

$$\boldsymbol{F}\,\boldsymbol{R} = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix}.$$

Po prvním hodu vyhrává Bob s pravděpodobností 1/6, ve zbývajících případech je v jednom z přechodných stavů, jejich pravděpodobnosti po prvním hodu jsou $(1/3, 1/2)$ a výsledky mají nepodmíněné pravděpodobnosti

$$(1/3, 1/2) \cdot \boldsymbol{F}\,\boldsymbol{R} = (1/3, 1/2),$$

takže hra je spravedlivá a Bob i Alice mají pravděpodobnost výhry 1/2.

4. (5 bodů) Mám pocit, že vždy stojím v těch nejdelších frontách, neboli: fronty, ve kterých stojím, jsou delší než průměrné. Co k tomu říká teorie pravděpodobnosti?

Řešení:

Něco na tom je. Počítáme-li průměrnou délku front, započítáme každou frontu jen jednou. Pokud počítáme, v jak dlouhé frontě stojí náhodně vybraný člověk, každá fronta se započítá tolikrát, kolik je v ní lidí. Dostaneme tedy vždy větší průměr; stačí k tomu, že fronty nejsou stejně dlouhé.

Člověk sám přispívá k délce fronty; ta je nutně nenulová pro frontu, ve které stojí.

Kromě toho přispívá výběrový efekt: pokud dlouhou frontu nestojíme, rychle na to zapomeneme. Když stojíme v dlouhé frontě, máme dost času na takovéto úvahy.

Pro útěchu lze říci, že není důvod se domnívat, že jiní lidé stráví ve frontách více či méně času.