Zkouškový test z PMS

na vypracování máte 90 min

31.1.2017

.

T1 O jevech A_1, A_2 víme: $P(A_1 \cup A_2) = 1$, $P(A_1) + P(A_2) = 1$, $P(A_i) > 0$, $i \in \{1, 2\}$. Dokažte či vyvraťte následující výroky:

a) Pravděpodobnost, že jevy A_1, A_2 nastanou zároveň, je nulová.

b) Jevy A_1, A_2 jsou nezávislé.

c) Jevy $A_1, A_1 \cup A_2$ jsou nezávislé.

T2 V sáčku je 10 bílých a 5 černých koulí. Vytáhnu náhodně kouli a pokud je černá, vytáhnu náhodně ještě jednu. Pokud je i druhá koule černá, hra končí. Pokud vytáhnu jako druhou kouli bílou, obě vrátím zpět a celý proces opakuji. Pokud vytáhnu už jako první bílou kouli, vrátím ji zpět a celý proces opakuji. Jedno kolo označuje proces vytáhnutí jedné či dvou koulí (pokud první byla černá) a případné jejich navrácení zpět do sáčku.

a) Jaká je pravděpodobnost, že hra skončí již v prvním kole?

b) Jaká je pravděpodobnost, že hra bude mít alespoň čtyři kola?

c) V jakém kole hra průměrně skončí?

T3 Hodnoty pravděpodobnostní funkce náhodného vektoru (X, Y) jsou v tabulce.

X \ Y	0	2	4
-1	0.3	0.1	0.2
1	0.2	0.1	0.1

Určete $\text{cov}(X, Y)$ a rozdělení veličiny $Z = 2X + Y$.

T4 Realizovali jsme náhodný výběr z rozdělení daném hustotou

$$f_p(u) = \begin{cases} pe^{-p(u-7)}, & u > 7, \\ 0, & \text{jinak} \end{cases}$$

s výsledky: 15, 10.5, 13, 19, 20.5, 12. Metodou maximální věrohodnosti odhadněte parametr p.

T5 V daném městě jsme vybrali náhodně 50 mužů a 40 žen. Z nich se 18, resp. 12 nehlásí k žádné víře. Testujte, zda existuje významný rozdíl mezi religiozitou mužů a žen v daném městě. Jako hladinu spolehlivosti použijte 0.05.

Řešení

T1

a) Z principu inkluze a exkluze platí

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

Protože $P(A_1 \cup A_2) = 1$, $P(A_1) + P(A_2) = 1$, musí být $P(A_1 \cap A_2) = 0$, tedy tvrzení platí.

b) Protože $P(A_i) > 0$, $i \in \{1, 2\}$, je $P(A_1)P(A_2) > 0$, avšak $P(A_1 \cap A_2) = 0$, proto jevy A_1, A_2 nejsou nezávislé.

c) Je $P(A_1 \cap (A_1 \cup A_2)) = P(A_1) = P(A_1) \cdot 1 = P(A_1)P(A_1 \cup A_2)$, tedy jevy A_1 a $A_1 \cup A_2$ jsou nezávislé.

T2 Pravděpodobnost, že v jednom kole vytáhnu za sebou dvě černé koule je rovna $\frac{5}{15} \cdot \frac{4}{14} = \frac{2}{21} = p$.

a) Pravděpodobnost, že hra skončí v prvním kole je $p = \frac{2}{21}$.

b) Pravděpodobnost, že má hra alespoň čtyři kola je rovna

$$\begin{aligned} 1 - P(\text{hra má právě 1, 2 nebo 3 kola}) &= 1 - \big(p + (1-p)p + (1-p)^2 p\big) \\ &= 1 - \frac{2}{21}\left(1 + \frac{19}{21} + \frac{361}{441}\right) \doteq 74\% \end{aligned}$$

c) Veličina X určující počet kol hry má geometrické rozdělení s parametrem p, tedy průměrný počet kol hry je $\mathbb{E}X = \frac{1}{p} = 11.5$.

T3 Marginální pravděpodobnostní funkce veličin X, Y jsou rovny

$$p_X(-1) = 0.6, \quad p_X(1) = 0.4, \qquad p_Y(0) = 0.5, \quad p_Y(2) = 0.2, \quad p_Y(4) = 0.3.$$

Z toho již snadno určíme střední hodnoty

$$\begin{aligned} \mathbb{E}X &= \sum_{X=a} a \cdot p_X(a) = -1 \cdot 0.6 + 1 \cdot 0.4 = -0.2 \\ \mathbb{E}Y &= \sum_{Y=a} a \cdot p_Y(a) = 0 \cdot 0.5 + 2 \cdot 0.2 + 4 \cdot 0.3 = 1.6 \end{aligned}$$

Využitím sdružené pravděpodobnostní funkce spočítáme střední hodnotu součinu X, Y:

$$\mathbb{E}XY = \sum_{X=a, Y=b} ab \cdot p_{X,Y}(a, b) = -2 \cdot 0.1 + 2 \cdot 0.1 - 4 \cdot 0.2 + 4 \cdot 0.1 = -0.4$$

Tedy $\operatorname{cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = -0.4 - (-0.2 \cdot 1.6) = -0.08$.

Veličina Z může nabývat pouze hodnot $\{-2, 0, 2, 4, 6\}$ a to s pravděpodobnostmi

$$\begin{aligned} &P(Z = -2) = P(X = -1, Y = 0) = p_{X,Y}(-1, 0) = 0.3, \\ &P(Z = 0) = P(X = -1, Y = 2) = p_{X,Y}(-1, 2) = 0.1, \\ &P(Z = 2) = P(X = -1, Y = 4 \vee X = 1, Y = 0) = p_{X,Y}(-1, 4) + p_{X,Y}(1, 0) = 0.2 + 0.2 = 0.4, \\ &P(Z = 4) = P(X = 1, Y = 2) = p_{X,Y}(1, 2) = 0.1, \\ &P(Z = 6) = P(X = 1, Y = 4) = p_{X,Y}(1, 4) = 0.1. \end{aligned}$$

T4 Označme $(x_1, ..., x_6)$ realizace daného náhodného výběru. Protože jest $x_i > 7$ pro každé $i \in \{1, ..., 6\}$, pro věrohodnostní funkci L daného výběru platí

$$\begin{aligned}
L(p) &= \prod_{i=1}^{6} f_p(x_i) = \prod_{i=1}^{6} pe^{-p(x_i-7)} \\
\log L(p) &= \sum_{i=1}^{6} \left(\log p + \log e^{-p(x_i-7)} \right) = 6 \log p - p \sum_{i=1}^{6} (x_i - 7) \\
\frac{\partial}{\partial p} \log L(p) &= \frac{6}{p} - \sum_{i=1}^{6} (x_i - 7)
\end{aligned}$$

Nutná podmínka existence maxima nám říká $\frac{\partial}{\partial p} \log L(p) = 0$ nebo $\frac{\partial}{\partial p} \log L(p)$ v bodě p neexistuje. Protože L je hladce diferencovatelná ve všech bodech $p > 7$, dostáváme

$$\begin{aligned}
0 &= \frac{6}{p} - \sum_{i=1}^{6} (x_i - 7) \\
p &= \frac{6}{\sum_{i=1}^{6} (x_i - 7)} = \frac{6}{48} = \frac{1}{8}.
\end{aligned}$$

T5 Jde o test homogenity dvou binomických rozdělení. Nechť p_1, resp. p_2 je podíl mužů, resp. žen v daném městě, kteří se nehlásí k žádné víře. Nechť náhodné výběry 50, resp. 40 žen jsou reprezentovány veličinami $X_i \sim \mathrm{Alt}(p_1)$, $i \in \{1, ..., 50\}$, resp. $Y_i \sim \mathrm{Alt}(p_2)$. Označme $X = \sum_{i=1}^{50} X_i$ a $Y = \sum_{i=1}^{40}$. Potom $X \sim \mathrm{Bi}(50, p_1)$ a $Y \sim \mathrm{Bi}(p_2)$. Chceme testovat hypotézu $H_0 : \ p_1 = p_2$ proti $H_1 : \ p_1 \neq p_2$. Označme $x = \frac{X}{50}$ a $y = \frac{Y}{40}$ a $z = \frac{X+Y}{50+40}$. Za platnosti H_0 má testovací statistika

$$T = \frac{x - y}{\sqrt{z(1-z) \left(\frac{1}{50} + \frac{1}{40} \right)}}$$

rozdělení blízké $\mathrm{N}(0, 1)$. Dosazením $X = 18$, $Y = 12$ dostáváme hodnotu

$$T = \frac{\frac{18}{50} - \frac{12}{40}}{\sqrt{\frac{30}{90} \frac{60}{90} \left(\frac{1}{50} + \frac{1}{40} \right)}} = \frac{0.06}{1} = 0.06$$

Jedná se o oboustranný test, proto hodnotu $|T|$ porovnáváme s kvantilem $u_{0.975} = 1.96$, z čehož je jasně vidět $|T| \leq u_{0.975}$, proto na hladině 0.05 nelze prokázat významný rozdíl mezi religiozitou mužů a žen v daném městě.

Na úlohu lze také nahlížet jako na test homogenity multinomických rozdělení ve čtyřpolní tabulce. Uvádíme pozorované četnosti (E_j)

pohlaví	nevěří	věří	celkem
muži	18	32	50
ženy	12	28	40
celkem	30	60	90

Za předpokladu platnosti H_0 (viz výše) lze odhadnout hodnotu parametru $p = p_1 = p_2$ (kde $X \sim \mathrm{Mul}(50, p_1, 1 - p_1)$ a $Y \sim \mathrm{Mul}(40, p_2, 1 - p_2)$) jako $p' = \frac{90}{30} = \frac{1}{3}$. Potom tabulka očekávaných četností O_j je rovna

pohlaví	nevěří	věří	celkem
muži	16.67	33.33	50
ženy	13.33	26.67	40
celkem	30	60	90

Statistika χ^2 potom vychází

$$T = \sum_{j=1}^{4} \frac{(O_j - E_j)^2}{E_j} = \frac{4^2}{3^2} \left(\frac{3}{50} + \frac{3}{100} + \frac{3}{40} + \frac{3}{80} \right) = 0.36.$$

Porovnáním hodnoty T s kvantilem $q_{\chi_1^2}(0.95) = 3.8 > T$ dostáváme, že na hladině 0.05 nelze prokázat významný rozdíl mezi religiozitou mužů a žen v daném městě.