Opravný zápočtový test PST, 17.5.2021

T1 Náhodný vektor (X, Y) má rozdělení dané tabulkou:

$X \backslash Y$	-1	0	1
0	0.1	0.5	0.1
1	0.1	0.1	0.1

a) Určete rozdělení veličin X a Y, určete jejich kovariant.

b) Jsou veličiny X, Y nezávislé? Dokažte.

c) Určete pravděpodobnosti: $P(X \leq Y)$, $P(X + Y < 1)$, $P(X \neq 0 | Y \neq 0)$.

T2 Náhodná veličina X má rozdělení dané hustotou $f(x) = \frac{\alpha}{2} e^{-\alpha|x|}$, $x \in \mathbb{R}$, kde $\alpha > 0$ je parametr. Pozorovali jsme těchto 9 hodnot:

$$0.213,\ -2.932,\ 6.664,\ 0.182,\ -2.426,\ -0.195,\ -2.301,\ 0.334,\ 4.371.$$

Metodou maximální věrohodnosti odhadněte hodnotu parametru α.

T3 Z informačního zdroje přichází nezávisle znaky. Znak A má 3x větší pravděpodobnost výskytu než znak M. Na zdroj jsou paralelně zapojena dvě rozpoznávací zařízení. Obě jsou stejná a pracují nezávisle. Pravděpodobnost, že (kterékoliv jedno) zařízení přečte znak A jako M je 0.1 a pravděpodobnost, že přečte znak M jako A je 0.15. V ostatních případech písmena rozpoznají správně.

a) Na vstupu byl jeden ze znaků A,M. Jaká je pravděpodobnost, že první zařízení zapíše na výstup písmeno A?

b) Na vstupu byl jeden ze znaků A,M. Jaká je pravděpodobnost, že obě zařízení zapíší na výstup písmeno A?

c) Na výstupu zapsala obě zařízení písmeno A. Jaká je pravděpodobnost, že bylo písmeno A na vstupu? Výstupu A lze dosáhnout pouze písmenem A nebo M na vstupu.

Řešení

T1

a) Sečtením pravděpodobností v tabulce přes řádky resp. sloupce dostáváme marginální rozdělení veličiny X resp. Y: $P(X = 0) = 0.7$, $P(X = 1) = 0.3$, tedy $X \sim \mathrm{Alt}(0.3)$ a máme $\mathbb{E}X = 0.3$. Pro Y máme $P(Y = -1) = P(Y = 1) = 0.2$ a $P(Y = 0) = 0.6$. Ze symetrie Y okolo nuly plyne $\mathbb{E}Y = 0$. Pro výpočet kovariantu je třeba zjistit hodnotu

$$\mathbb{E}XY = \sum_{a,b \in \mathcal{H}_{(X,Y)}} a \cdot b \cdot p_{X,Y}(a, b) = 0.1 + 0 + 0 + 0 + 0 - 0.1 = 0.$$

Celkově tedy $\mathrm{cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0 - 0.3 \cdot 0 = 0$.

b) Jelikož je $p_{X,Y}(0, 0) = 0.5 \neq 0.42 = 0.6 \cdot 0.7 = p_X(0)p_Y(0)$, jsou veličiny X a Y závislé. (I když je jejich kovariant roven nule!)

c) $P(X \leq Y) = P(X=0, Y=0) + P(X=0, Y=1) + P(X=1, Y=1) = 0.7.$
$P(X+Y<1) = P(X=0, Y=-1) + P(X=0, Y=0) + P(X=1, Y=-1) = 0.7.$
$P(X \neq 0 | Y \neq 0) = \frac{P(X \neq 0 \cap Y \neq 0)}{P(Y \neq 0)} = \frac{P(X=1, Y=-1) + P(X=1, Y=1)}{P(Y \neq 0)} = \frac{0.2}{0.4} = 0.5.$

T2 Nejprve sestavíme věrohodnostní funkci jako součin hodnot hustoty v realizacích výběru:

$$L(\alpha) = \prod_{i=1}^{9} f(x_i) = \prod_{i=1}^{9} \frac{\alpha}{2} e^{-\alpha |x_i|} = \left(\frac{\alpha}{2}\right)^9 \exp\left(-\alpha \sum_{i=1}^{9} |x_i|\right).$$

Protože budeme hledat maximum L na množině $(0, \infty)$, funkci zlogaritmujeme a zderivujeme.

$$\log(L(\alpha)) = 9 \log\left(\frac{\alpha}{2}\right) + \log \exp\left(-\alpha \sum_{i=1}^{9} |x_i|\right) = 9 \log(\alpha) - 9 \log(2) - \alpha \sum_{i=1}^{9} |x_i|,$$

$$\frac{\partial}{\partial \alpha} \log(L(\alpha)) = \frac{9}{\alpha} - \sum_{i=1}^{9} |x_i|.$$

Položíme-li derivaci rovnou nule, dostáváme rovnici

$$\frac{9}{\alpha} = \sum_{i=1}^{9} |x_i|,$$

$$\alpha = \frac{9}{\sum_{i=1}^{9} |x_i|}.$$

Dosazením hodnot realizací máme $\sum_{i=1}^{9} |x_i| = 19.618$ a tedy $\alpha = 0.4588$.

T3 Označme jevy, že na vstupu byl znak A, resp. M jako A, resp. M. Označme jev, že na výstupu prvního přístroje bude A jako A_1. Ze zadání máme $P(A) = 3P(M)$ a omezíme-li se pouze na znaky A a M, máme $P(A) + P(M) = 1$, tedy $P(A) = 0.75$, $P(M) = 0.25$. Dále je zadáno nebo lze snadno spočítat $P(A_1|A) = 0.9$, $P(A_1|M) = 0.15$.

a) Věta o úplné pravděpodobnosti dává

$$P(A_1) = P(A_1|A) \cdot P(A) + P(A_1|M) \cdot P(M) = 0.9 \cdot 0.75 + 0.15 \cdot 0.25 = 0.7125.$$

b) Je potřeba si uvědomit, že jevy „první zařízení zapíše A“ a „druhé zařízení zapíše A“ nejsou nezávislé. Ukažme si to detailněji. Označme jevy A_1, resp. A_2, že bude A na výstupu prvního, resp. druhého přístroje. Dle a) jsme dostali $P(A_2) = P(A_1) = 0.7125$. Podíváme se na podmíněnou pravděpodobnost $P(A_2|A_1)$. Pokud tedy víme, že na výstupu z prvního zařízení byl znak A, lze pomocí Bayesova obratu spočítat pravděpodobnost, že bylo A na vstupu:

$$P(A|A_1) = \frac{P(A \cap A_1)}{P(A_1)} = \frac{P(A_1|A) \cdot P(A)}{P(A_1)} = \frac{0.9 \cdot 0.75}{0.7125} = 0.9474.$$

Potom ale máme

$$\begin{aligned} P(A_2|A_1) &= P(A_2|A_1 \cap A) \cdot P(A|A_1) + P(A_2|A_1 \cap M) \cdot P(M|A_1) \\ &= 0.9 \cdot 0.9474 + 0.15 \cdot (1 - 0.9474) \\ &= 0.8605. \end{aligned}$$

Tedy $P(A_2|A_1) = 0.8605 \neq 0.7125 = P(A_2)$, z čehož vyplývá, že A_1 a A_2 nejsou nezávislé. Tento postup by šlo popsat slovy následovně: Pokud víme, že první zařízení zapsalo na výstup A, znamená to (vzhledem k přesnosti zařízení), že je pravděpodobnější, že na vstupu byl znak A, než kdybychom informaci o A_1 neměli (konkrétně je tato pravděpodobnost rovna 0.9474 - tedy opravdu více než původních 0.75). To však opět ovlivňuje pravděpodobnost, se kterou druhé zařízení zapíše na výstup A - je to 0.8605 oproti původním 0.7125. Tedy informace o tom, že první zařízení zapsalo na výstup A zvyšuje pravděpodobnost, že druhé zařízení zapsalo na výstup také A - i když zařízení pracují nezávisle.

Jevy A_1 a A_2 jsou však podmíněně nezávislé - za podmínky, že na vstupu je A, případně za podmínky, že na vstupu je M. To tedy znamená, že $P(A_1 \cap A_2|A) = P(A_1|A) \cdot P(A_2|A)$ a obdobně, $P(A_1 \cap A_2|M) = P(A_1|M) \cdot P(A_2|M)$. To nám dává jeden ze způsobů výpočtu $P(A_1 \cap A_2)$:

$$\begin{aligned}
P(A_1 \cap A_2) &= P(A_1 \cap A_2|A) \cdot P(A) + P(A_1 \cap A_2|M) \cdot P(M) \\
&= P(A_1|A) \cdot P(A_2|A) \cdot P(A) + P(A_1|M) \cdot P(A_2|M) \cdot P(M) \\
&= P(A_1|A)^2 \cdot P(A) + P(A_1|M)^2 \cdot P(M) \\
&= 0.9^2 \cdot 0.75 + 0.15^2 \cdot 0.25 = 0.6131.
\end{aligned}$$

Druhý, složitější způsob výpočtu je pomocí výše spočítané podmíněné pravděpodobnosti $P(A_2|A_1)$. Tedy

$$P(A_1 \cap A_2) = P(A_2|A_1) \cdot P(A_1) = 0.8605 \cdot 0.7125 = 0.6131.$$

Výsledky tedy souhlasí.

c) Víme, že nastal jev $A_1 \cap A_2$. Podle Bayesovy věty je nyní

$$P(A|A_1 \cap A_2) = \frac{P(A_1 \cap A_2|A) \cdot P(A)}{P(A_1 \cap A_2)} = \frac{0.9^2 \cdot 0.75}{0.6131} = 0.9908.$$

Tedy pokud obě zařízení zapíší na výstupu znak A, potom je pravděpodobnost, že znak A byl na vstupu přibližně 99%.