3. zkoušková písemka NMAI059 Pravd. a Stat. 1 – řešení – 30.6.2021

1. (10 bodů) Rozhodněte, zda existují náhodné veličiny X, Y na tomtéž pravděpodobnostním prostoru, takové, že

(a) $X \sim Pois(1/10)$, $Y \sim Bin(100, 1/10)$ a přitom $P(X \leq Y) = 1$.
(b) $X \sim Pois(1/10)$, $Y \sim Bin(100, 1/10)$ a přitom $P(X \geq Y) = 1$.
(c) $X \sim Bin(100, 1/2)$, $Y \sim Bin(100, 1/10)$ a přitom $P(X \geq Y) = 1$.
(d) $X \sim Bin(100, 1/2)$, $Y \sim Bin(100, 1/10)$ a přitom $P(X \leq Y) = 1$.

Řešení:

(a) Ne: vždy platí $Y \leq 100$, ale $P(X = 101)$ je kladná.

(b) Ne: Podle vlastností Poissonova a binomického rozdělení je $\mathbb{E}(X) = 1/10$, $\mathbb{E}(Y) = 10$. Proto $\mathbb{E}(X - Y) < 0$ a tudíž musí s kladnou pravděpodobností být $X - Y$ záporné.

Varianta: platí, že $P(X = 0) > P(Y = 0)$, tudíž s kladnou pravděpodobností je $X = 0$ ale $Y > 0$. Bez kalkulačky je to asi trochu horší ukázat, ale rozdíl je veliký, takže i to jde:

$$P(X = 0) = e^{-1/10}$$
$$P(Y = 0) = \binom{100}{0}\Big(1 - \frac{1}{10}\Big)^{100} \doteq e^{-100 \cdot \frac{1}{10}} = e^{-10}$$

Používáme užitečný vzorec $e^{-x} \doteq 1 - x$ (který se někdy hodí použít pro aproximaci exponenciální funkce, ale tady zrovna naopak, pro aproximaci $1 - x$). Přesné hodnoty jsou $P(X = 0) \doteq 0.9048$, $P(Y = 0) \doteq 2.6 \cdot 10^{-5}$. Navíc i intuitivně by mělo být jasné, že když $\mathbb{E}(X) = 1/10$, tak $P(X = 0)$ je dost velká, zatímco $\mathbb{E}(Y) = 10$, takže $P(Y = 0)$ je spíše malá.

(c) Ano: uvažme na příklad n hodů desetistěnnou kostkou. Označíme X počet hodů, kde padlo $1, \dots, 5$ a Y počet takových, kde padla jednička. Pak X i Y mají požadovaná rozdělení a $X \geq Y$ platí vždy.

(d) Ne: stejné zdůvodnění jako v části (b).

2. (10 bodů) Nechť $X \sim N(0, 1)$ a $Y = |X|$.

(a) Jaká je distribuční funkce Y?
(b) Jaká je hustota Y?
(c) Spočtěte $\mathbb{E}(Y)$.
(d) Určete medián Y.

Řešení:

(a) Podle definice je

$$F_Y(y) = P(Y \leq y) = P(-y \leq X \leq y).$$

To je jistě nula pro $y \leq 0$. Pro kladná y dostáváme $F_Y(y) = \Phi(y) - \Phi(-y)$. Využitím symetrie normálního rozdělení, tj. faktu $\Phi(y) + \Phi(-y) = 1$ můžeme ještě upravit

$$F_Y(y) = 2\Phi(y) - 1.$$

(b) Derivací vzorce z předchozí části: $f_Y(y) = 2\varphi(y) = \frac{2}{\sqrt{2\pi}} e^{-y^2/2}$ pro kladná y, $f_Y(y) = 0$ pro $y \le 0$. Ale je to vidět i přímo: díky sudosti funkce φ je každé kladné číslo u Y dvakrát pravděpodobnější než pro X.

(c) Podle definice a části (b) je

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \cdot f_Y(y)dy = \int_{0}^{\infty} y \cdot 2\varphi(y)dy = \int_{0}^{\infty} y \cdot \frac{2}{\sqrt{2\pi}} e^{-y^2/2}dy = \Big[\frac{-2}{\sqrt{2\pi}} e^{-y^2/2}\Big]_0^{\infty} = \frac{2}{\sqrt{2\pi}}$$

(d) Podle definice je medián $F_Y^{-1}(1/2)$. Hledáme tedy y, aby $2\Phi(y) - 1 = 1/2$, tudíž $y = \Phi^{-1}(3/4)$.

3. (10 bodů) Petr má (šestistěnnou) kostku, na které padá šestka s pravděpodobností p. Čísla 1, …, 5 padají všechna se stejnou pravděpodobností.

(a) V závislosti na p určete, jaká je střední hodnota čísla, které Petrovi padne.

(b) Najděte odhad $\hat{p}$ momentovou metodou. Jaké $\hat{p}$ stanovíte, pokud padla čísla 2, 6, 3?

(c) Určete odhad $\hat{p}$ pomocí metody maximální věrohodnosti. Opět určete výsledek pro čísla 2, 6, 3.

Řešení:

(a) Označme K výsledek hodu Petrovou kostkou. Podle definice je

$$\mathbb{E}(K) = 6 \cdot p + (1 + 2 + \cdots + 5) \cdot \left(\frac{1-p}{5}\right) = 3 + 3p.$$

(b) Označme μ výběrový průměr hozených čísel. Momentová metoda nám říká, že $3+3\hat{p} = \mu$, neboli $\hat{p} = \mu/3 - 1$. (S ohledem na význam čísla p bychom mohli říci, že pokud $\mu/3 - 1$ vyjde záporné, položíme $\hat{p} = 0$, ale momentová metoda jako taková to neříká.) Pro hozená čísla 2,6,3 je $\mu = 11/3$, tedy $\hat{p} = 2/9$.

(c) Označme x počet šestek a y počet ostatních hozených čísel. Pravděpodobnost takového výsledku je

$$p^X \left(\frac{1-p}{5}\right)^y = p^x (1-p)^y \cdot 5^{-y}.$$

Minimalizujeme stejně jako na přednášce (úloha s leváky) a vyjde nám, že maxima se nabývá v bodě

$$\hat{p} = \frac{x}{x+y}.$$

Pro čísla 2, 6, 3 nám tedy vychází $\hat{p} = 1/3$.

4. (10 bodů) (a) Definujte pojem korelace (dvou) náhodných veličin. Jaká je korelace $\varrho(X, X)$ pro $X \sim U(0,1)$?

(b) Definujte pojem nezávislé náhodné veličiny (spojitý případ, dvě veličiny). Uveďte obě ekvivalentní formulace. Rozhodněte, zda existují nezávislé X, Y takové, že $X \sim U(0,1)$ a $Y \sim U(0,1)$?

Řešení:

(a) Pro veličiny X, Y definujeme jejich kovarianci jako $\operatorname{cov}(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot E(Y)$ a jejich korelaci jako

$$\varrho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(Y)}}.$$

Dosazením $Y = X$ získáme $\varrho(X,X) = \frac{\operatorname{cov}(X,X)}{\operatorname{var}(X)} = 1$ (pro jakoukoli X, kde je $\operatorname{var}(X)$ definovaný a nenulový.

(b) Náhodné veličiny X, Y jsou nevávislé, pokud pro každé reálné x, y platí

$$F_{X,Y}(x,y) = F_X(x)F_Y(y),$$

neboli $P(X \le x \,\&\, Y \le y) = P(X \le x)P(Y \le y)$. Pro spojité náhodné veličiny je ekvivalentní požadovat $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

5. (10 bodů) Vysvětlete, jak se provádí test dobré shody.

Řešení: Opakujeme n-krát pokus, který může dopadnout jedním z k výsledků, jednotlivá opakování jsou nezávislé. Testujeme nulovou hypotézu, která říká, že i-tý výsledek nastává s pravděpodobností p_i. Pro $i = 1, \dots, k$ označme X_i počet pokusů s i-tým výsledkem a $E_i = p_i n$ střední hodnotu X_i. Dále spočtěme tzv. Pearsonovu statistiku

$$T = \sum_{i=1}^{k} \frac{(E_i - X_i)^2}{E_i}.$$

Konečně kritická hodnota γ je kvantilová funkce χ^2 rozdělení s $k-1$ stupni volnosti v bodě $1-\alpha$, tj. $\gamma = Q(1-\alpha)$, kde Q je kvantilová funkce pro χ^2_{k-1}.

Nulovou hypotézu (tj. to, že máme správné pravděpodobnosti) zamítneme, pokud $T > \gamma$. Při takové volbě je pravděpodobnost chyby prvního druhu přibližně rovna α.

6. (10 bodů) Vyslovte a dokažte větu o úplné střední hodnotě (neboli, o výpočtu střední hodnoty rozborem případů). Stačí varianta pro diskrétní náhodné veličiny.

Řešení: Věta:

Nechť X je náhodná veličina na pravděpodobnostním prostoru $(\Omega, \mathcal{F}, P)$. Pokud $B_1, B_2, \dots$ je rozklad Ω a $A \in \mathcal{F}$, tak

$$\mathbb{E}(X) = \sum_i P(B_i) \cdot \mathbb{E}(X \mid B_i),$$

kdykoliv má součet smysl. (Sčítance s $P(B_i) = 0$ považujeme za 0.)

Důkaz: Podle věty o úplné pravděpodobnosti je

$$P(X = x) = \sum_i P(B_i) \cdot P(X = x \mid B_i).$$

Vynásobíme tuto rovnost x a sečteme pro všechna $x \in ImX$. Na levé straně dostaneme $\sum_{x \in ImX} x \cdot P(X = x)$ což je definice $\mathbb{E}(X)$. Na pravé straně po prohození pořadí sčítání dostaneme $\sum_i \sum_{x \in ImX} P(B_i) \cdot x \cdot P(X = x \mid B_i) = \sum_i P(B_i) \cdot \mathbb{E}(X_i \mid B_i)$.