Řešení vzorové zkouškové písemky NMAI059 Pravd. a Stat. 1 – 2020/21

1. (10 bodů) Následující tabulka popisuje sdruženou pravděpodobnostní funkci $p_{X,Y}(x,y)$ náhodného vektoru (X,Y). Jiné než vyznačené hodnoty tyto veličiny nenabývají. Víme, že $\mathbb{E}(X) = 1/2$.

$x \backslash y$	-1	0	1
0	a	$1/8$	$1/4$
1	$1/8$	b	$1/8$

(a) Určete hodnotu a a b.
(b) Zjistěte, zda jsou X a Y nezávislé.

Řešení:

(a) Víme, že

$$\mathbb{E}(X) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p_{X,Y}(1,-1) + p_{X,Y}(1,0) + p_{X,Y}(1,1) = 1/8 + b + 1/8,$$

tudíž $b = 1/4$. Součet všech čísel v tabulce musí být 1, tudíž $a = 1/8$.

(b) I bez řešení části (a) víme, že $p_X(0) = p_X(1) = 1/2$. Kdyby X a Y byly nezávislé, tak by v každém sloupci byla obě čísla stejná, a to není pravda např. v posledním sloupečku.

---

2. (10 bodů) Petr se opakovaně pokouší porazit silnějšího protivníka v šachu. Pokud Petr vyhraje, získá 10 bodů, jinak (i v případě remízy) ztratí 2 body. Petr vyhrává s pravděpodobností 1/4.

(a) V kolikátém kole Petr poprvé vyhraje (v průměru)?

(b) Má-li Petr na začátku 6 bodů, s jakou pravděpodobností bude nejpozději po pěti kolech na nule?

(c) Jaké je rozdělení Petrových bodů po pěti hrách, jestliže na začátku má nula bodů (záporné body jsou povoleny)? (Popište pravděpodobnostní funkci tohoto počtu bodů.)

(d) Je vyšší pravděpodobnost, že bude mít po deseti kolech alespoň dvacet bodů, nebo po sto kolech alespoň dvě stě bodů?

Řešení:

(a) Číslo kola, kdy Petr poprvé vyhraje, má rozdělení $Geom(1/4)$, tudíž zkoumaná střední hodnota je $\frac{1}{1/4} = 4$.

(b) Pokud by Petr vyhrál, má 16 bodů a ani čtyři prohry ho nedovedou k nule. Pokud má být po ≤ 5 kolech na nule, je to možné jen tak, že třikrát za sebou prohraje. To se stane s pravděpodobností $(3/4)^3$.

(c) Můžeme si představit, že v každém kole Petr ztratí dva body a pak má šanci získat 12 bodů, s pravděpodobností 1/4. Výsledný počet bodů bychom tedy mohli označit

$-10 + 12 \cdot Bin(5, 1/4)$. Platí tedy, že Petr bude mít $-10 + 12k$ bodů (pro $k = 0, 1, \ldots, 5$) s pravděpodobností danou binomickým rozdělením, tj. $\binom{5}{k}(1/4)^k(3/4)^{5-k}$.

(d) Střední hodnota zisku bodů za jedno kolo je $\mu = +10 \cdot \frac{1}{4} - 2 \cdot \frac{3}{4} = 1$. Rozptyl bychom mohli spočítat také, ale stačí nám, že je to nějaká konstanta σ^2.

Za deset kol bude mít průměrně deset bodů, za sto kol sto bodů. Dotazujeme se tedy na pravděpodobnost podstatně nadprůměrného zisku, navíc zisku „lineárně nadprůměrného“. Tato pravděpodobnost by měla pro větší n klesat, protože pro velká n se bude součet chovat „víc podle střední hodnoty“.

Teď pořádně: Podle Centrální limitní věty je rozdělení $(X_1 + \cdots + X_n - n\mu)/(\sigma\sqrt{n})$ přibližně normální. Při přechodu od $n = 10$ k $n = 100$ se čitatel zvýší desetkrát, jmenovatel jen $\sqrt{10}$-krát. Ptáme se tedy na pravděpodobnost, že ve standardním normálním rozdělení „padne“ mnohem vyšší číslo, tj. pravděpodobnost pro sto kol bude (mnohem) menší.

(Pro zajímavost: pro $n = 10$ je daná pravděpodobnost 0.224, pro $n = 100$ jen 0.027. To při písemce vyčíslovat nemusíte.)

---

3. (10 bodů) Pořádáte oslavu pro 100 hostů a přemýšlíte, kolik chlebíčků objednat. Ze zkušenosti víte, že počet chlebíčků snědených náhodným hostem je určen Poissonovým rozdělením se střední hodnotou 3. Kolik chlebíčků přibližně musíte objednat, abyste věděli, že s pravděpodobností 0.95 žádný host nebude hladový?

(Použijte vhodnou limitní větu.)

Řešení: Počet chlebíčků, které sní i-tý host, označíme X_i. Podle zadání je $X_i \sim Pois(3)$, podle vlastností Poissonova rozdělení odsud odvodíme, že $\mathbb{E}(X_i) = \mathrm{var}(X_i) = 3$. Podle Centrální limitní věty je rozdělení $\frac{X_1 + \cdots + X_{100} - 300}{\sqrt{3} \cdot \sqrt{100}}$ přibližně standardní normální. Proto

$$P(X_1 + \cdots + X_{100} \leq 300 + t\sqrt{300}) \doteq \Phi(t).$$

Aby tato pravděpodobnost byla 95 %, musíme nakoupit $300 + \sqrt{300} \cdot \Phi^{-1}(0.95)$ chlebíčků. (Po dosazení vyjde 328.5, ale to není během zkoušky třeba dopočítávat. V R-ku se můžeme např. zeptat i přímo na **qnorm**(0.95,300,**sqrt**(300)).)

Alternativní řešení bez CLV: víme, že součet Poissonových rozdělení je zase Poissonovo rozdělení. Tedy 100 hostů sní dohromady počet chlebíčků, jejichž rozdělení je $Pois(300)$. Zajímá nás tedy kvantilová funkce tohoto rozdělení v bodě 0.95. To nejde nijak hezky zjednodušit, ale R-ko to spočítá pomocí příkazu **qpois**(0.95,300). Vyjde 329.

---

4. (10 bodů) (a) Definujte pojem hustota náhodné veličiny X.

(b) Popište, jak pomocí hustoty určit $\mathrm{var}(X)$.

Řešení:

(a) Říkáme, že funkce $f : \mathbb{R} \to [0, \infty)$ je hustota náhodné veličiny X, pokud pro každé $x \in \mathbb{R}$ platí, že

$$P(X \leq x) = \int_{-\infty}^{x} f(t)dt.$$

(b) Podle pravidla LOTUS a pravidla pro výpočet rozptylu platí

$$\operatorname{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^{\infty} t^2 f(t)dt - \Big(\int_{-\infty}^{\infty} t f(t)dt\Big)^2$$

---

5. (10 bodů) Vyslovte větu o univerzalitě uniformního rozdělení. Vysvětlete, k čemu se hodí.

Řešení:

> Nechť F je funkce „typu distribuční funkce“: neklesající zprava spojitá funkce s $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$. Nechť Q je odpovídající kvantilová funkce.
>
> Nechť $U \sim U(0, 1)$ a $X = Q(U)$. Pak X má distribuční funkci F.

Věta je vhodná ke konstrukci náhodných veličin s danou distribucí – potřebujeme umět vyrobit uniformně náhodnou veličinu na intervalu $(0, 1)$ a spočítat kvantilovou funkci. Říkali jsme si ještě druhou větu podobného typu:

> Nechť X je n.v. s distribuční funkcí $F_X = F$, nechť F je spojitá a rostoucí. Pak $F(X) \sim U(0, 1)$.

Tuto větu bychom mohli použít třeba pro testování: pokud máme proceduru na otestování, že náhodná veličina je uniformní, můžeme ji použít na otestování libovolného jiného spojitého rozdělení.

---

6. (10 bodů) Vyslovte větu – slabý zákon velkých čísel. Dokažte ji.

Řešení: Nechť $X_1, X_2, \dots$ jsou n.n.v. se střední hodnotou μ a rozptylem σ^2. Definujme výběrový průměr $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Pak posloupnost $\bar{X}_n$ konverguje v pravděpodobnosti k μ (píšeme $\bar{X} \xrightarrow{P} \mu$), což značí, že pro každé $\varepsilon > 0$ platí

$$\lim_{n\to\infty} P(|X_n - \mu| > \varepsilon) = 0.$$

Větu dokážeme pomocí Čebyševovy nerovnosti. Spočteme napřed střední hodnotu a rozptyl $\bar{X}_n$:

$$\begin{aligned} \mathbb{E}(\bar{X}_n) &= \frac{1}{n}(\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)) = \frac{n\mu}{n} = \mu \\ \operatorname{var}(\bar{X}_n) &= \frac{1}{n^2}(\operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n)) = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

Podle Čebyševovy nerovnosti je

$$P(|\bar{X}_n - \mu| > t \cdot \sigma/\sqrt{n}) < \frac{1}{t^2}.$$

Pro $t\sigma/\sqrt{n} = \varepsilon$ nám vychází $1/t^2 = \frac{\sigma^2}{\varepsilon^2 n}$, což má limitu 0 pro $n \to \infty$.