Ukázková řešení domácích cvičení úterní skupinka

4. Označme X_1, X_2, X_3 výsledky tří nezávislých hodů čtyřstěnnou kostkou (s čísly 1, ..., 4).

(a) Jaká je pravděpodobnostní funkce $X = X_1$?

(b) Jaká je pravděpodobnostní funkce $Y = \max(X_1, X_2)$?

(c) Jaká je pravděpodobnostní funkce $Z = \max(X_1, X_2, X_3)$?

(d) O kolik se zvýší střední hodnota tím, že můžeme házet třikrát? Neboli, o kolik je vyšší $\mathbb{E}(Z)$ než $\mathbb{E}(X)$?

Řešení: Určíme napřed distribuční funkce. Pro $i \in \{1,2,3\}$ a pro $k \in \{0,1,2,3,4\}$ platí $F_{X_i}(k) = k/4$. Zjevně je $F_X = F_{X_1}$. Pro Y je nejsnazší si všimnout, že pro $k \in \{0,\ldots,4\}$

$$F_Y(k) = P(Y \le k) = P(X_1 \le k \,\&\, X_2 \le k) = P(X_1 \le k) \cdot P(X_2 \le k) = F_{X_1}(k) \cdot F_{X_2}(k) = (k/4)^2.$$

Analogicky pro Z dostáváme

$$F_Z(k) = P(Z \le k) = P(X_1 \le k \,\&\, X_2 \le k \,\&\, X_3 \le k) = F_{X_1}(k) \cdot F_{X_2}(k) \cdot F_{X_3}(k) = (k/4)^3.$$

Odsud snadno dostaneme pravděpodobnostní funkce pomocí vztahu $p(k) = F(k) - F(k-1)$. Pro $k \in \{1,2,3,4\}$ tedy máme

$$\begin{aligned} p_X(k) &= 1/4 \\ p_Y(k) &= \frac{1}{16}(k^2 - (k-1)^2) = \frac{2k-1}{16} \\ p_Z(k) &= \frac{1}{64}(k^3 - (k-1)^3) \end{aligned}$$

Mohli bychom i přímočaře rozebrat, pro jaké kombinace 16 možných hodů je Y rovno 1, 2, 3 nebo 4 – buď pomocí rozdílu čtverců, nebo i jakkoli jinak. Pro Z je takový postup už trochu náročný na představivost.

Nebo v tabulkovém zápisu:

k	1	2	3	4
$p_X(k)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$p_Y(k)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$
$p_Z(k)$	$\frac{1}{64}$	$\frac{7}{64}$	$\frac{19}{64}$	$\frac{37}{64}$

Střední hodnota X je $\mathbb{E}(X) = 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 5/2 = 2.5$.

Střední hodnota Z je rovna $\mathbb{E}(Z) = 1 \cdot \frac{1}{64} + 2 \cdot \frac{7}{64} + 3 \cdot \frac{19}{64} + 4 \cdot \frac{37}{64} = 220/64 \doteq 3.4$. Opakovaným házením si tedy polepšíme v průměru o 0.9.

U řešení na počítači samplováním vygenerujeme dostatečně mnoho hodů kostkou X_1, X_2, X_3 a pak spočítáme, kolikrát bylo příslušné maximum rovno 1, 2, 3, nebo 4 – a vydělíme počtem pokusů. Podrobněji v přiloženém R-kovém notebooku.

5.

Uvažme skupinu m manželských párů (tj. celkem $2m$ osob). Předpokládejme, že po deseti letech bude každý z těch $2m$ lidí stále naživu s pravděpodobností p, nezávisle na ostatních. Možnosti rozvodů apod. neuvažujeme, tj. páry jsou neměnné.

Označme L množinu lidí, kteří budou po deseti letech naživu a A jejich počet (tj. $A = |L|$). Dále buď B počet párů, kde budou naživu oba; tj. A, B jsou náhodné veličiny splňující $0 \le A \le 2m$ a $0 \le B \le m$. Pro každé $a = 0, \ldots, 2m$ chceme spočítat $\mathbb{E}(B \mid A = a)$.

(a) Uvážíme jednoho konkrétního člověka. Jaká je pravděpodobnost, že bude po deseti letech naživu, pokud víme, že $A = a$? Jinými slovy, pokud ten člověk je x, jaká je $P(x \in L \mid A = a)$?

(b) Uvážíme jeden konkrétní manželský pár. Jaká je pravděpodobnost, že budou oba naživu, pokud víme, že $A = a$?

(c) Vyjádřete B jako součet m vhodných indikátorových n.v.

(d) Linearita střední hodnoty platí i pro podmíněnou střední hodnotu, neboli

$$\mathbb{E}(\sum_{i=1}^{m} X_i \mid J) = \sum_{i=1}^{m} \mathbb{E}(X_i \mid J),$$

pro jakýkoliv jev J a n.v. $X_1, \ldots, X_m$. (To nemusíte dokazovat.) Využijte toho k vypočtení $\mathbb{E}(B \mid A = a)$.

(e) Jaké je rozdělení n.v. A? (Buď ho pojmenujte, nebo napište pravděpodobnostní funkci, tj. určete $P(A = a)$.)

(f) Pro pevně danou a-prvkovou množinu lidí M, jaká je pravděpodobnost, že je to přesně množina přeživších? Neboli, kolik je $P(L = M)$? A kolik $P(L = M \mid A = a)$? (L je náhodná množina, M daná množina).

(g) Pro $m = 10$ a $a = 4$ ověřte výsledek samplováním v libovolném programovacím jazyce. Budete-li používat R, doporučuji pozornosti příkaz rbinom(m,1,p) – vyrobí vektor s m čísly, každé z nich je rozděleno podle $Bin(1, p)$, neboli $Bern(p)$.

Řešení: Z cvičných důvodů si ukážeme několik různých postupů. Stačilo samozřejmě použít jen jeden.

(a) Výpočet podle definice – podmíněná pravděpodobnost, binomické rozdělení, pokud $x \in L$, tak ze zbylých lidí mám vybrat $a - 1$ přeživších (což se formálně zapíše jako podmíněná pravděpodobnost $P(A = a \mid x \in L)$).

$$P(x \in L \mid A = a) = \frac{P(x \in L \,\&\, A = a)}{P(A = a)} = \frac{P(x \in L)P(A = a \mid x \in L)}{P(A = a)} = \frac{p\binom{2m-1}{a-1}p^{a-1}(1-p)^{2m-a}}{\binom{2m}{a}p^a(1-p)^{2m-a}} = \frac{a}{2m}.$$

Výpočet pomocí „jiného pravděpodobnostního prostoru“: omezujeme se na případy, kdy $|L| = a$, každý z nich má stejnou pravděpodobnost $p^a(1-p)^{2m-a}$. Vybíráme tedy uniformně náhodně jednu z množin velikosti a.

$$P(x \in L \mid A = a) = \frac{\text{počet množin obsahujících } x}{\text{počet všech množin}} = \frac{\binom{2m-1}{a-1}}{\binom{2m}{a}} = \frac{a}{2m}.$$

Výpočet pomocí symetrie. Označme hledanou pravděpodobnost t. Střední hodnota počtu přeživších je tedy $2m \cdot t$. (Formální zdůvodnění: pro každého člověka zavedeme indikátorovou náhodnou veličinu (1 = přežil). Máme tedy veličiny $A_1, \ldots, A_m$, ze symetrie mají všechny střední hodnotu t. Linearita střední hodnoty nám dává

$$a = \mathbb{E}(A \mid A = a) = \mathbb{E}(A_1 + \cdots + A_{2m} \mid A = a) = \sum_{i=1}^{2m} \mathbb{E}(A_i \mid A = a) = 2m \cdot t.)$$

Odsud $t = a/(2m)$

(b) Manželé se budou jmenovat x a y. Výpočet podle definice – tak jako v části (a).

$$P(x, y \in L \mid A = a) = \frac{P(x, y \in L \,\&\, A = a)}{P(A = a)} = \frac{P(x, y \in L)P(A = a \mid x, y \in L)}{P(A = a)} = \frac{p^2\binom{2m-2}{a-2}p^{a-2}(1-p)^{2m-a}}{\binom{2m}{a}p^a(1-p)^{2m-a}} = \frac{a(a-1)}{2m(2m-1)}.$$

Výpočet pomocí „jiného pravděpodobnostního prostoru“: zase vybíráme uniformně náhodně jednu z množin velikosti a.

$$P(x, y \in L \mid A = a) = \frac{\text{počet množin obsahujících } x, y}{\text{počet všech množin}} = \frac{\binom{2m-2}{a-2}}{\binom{2m}{a}} = \frac{a(a-1)}{2m(2m-1)}.$$

(c) Označme B_i indikátor jevu „v i-tém manželském páru jsou oba naživu“. Jistě je $B = \sum_{i=1}^m B_i$.

(d) Podle části (b) je $\mathbb{E}(B_i \mid A = a) = \frac{a(a-1)}{2m(2m-1)}$. Máme tedy

$$\mathbb{E}(B \mid A = a) = \sum_{i=1}^{m} \mathbb{E}(B_i \mid A = a) = \frac{a(a-1)}{2(2m-1)}.$$

(e) $A \sim Bin(2m, p)$, jedná se o binomické rozdělení. Jinými slovy, $P(A = a) = \binom{2m}{a} p^a (1-p)^{2m-a}$.

(f) Pro danou množinu M velikosti a: má-li být $L = M$, tak musí každý člověk v M přežít (pravděpodobnost p^a) a každý mimo M umřít (pravděpodobnost $(1-p)^{2m-a}$. Celkem tedy

$$P(L = M) = p^a (1-p)^{2m-a}.$$

Pokud víme, že $A = a$, tak vybíráme množinu L jen z a-prvkových, tj.

$$P(L = M \mid A = a) = \frac{1}{\binom{2m}{a}}.$$

Odsud bychom mohli (spolu s částí (e)) spočítat jinak $P(L = M)$ jako součin $P(A = a)P(L = M \mid A = a)$.

(g) Viz přiložený soubor v R. Vysvětlení: principem samplování je ověřit teoretický výpočet jinak. Případně jej nahradit, pokud ho neumíme provést. Pokud použijeme počítač pro vyčíslení vzorce, který jsme spočítali, může to být často užitečné, ale neříká se tomu samplování (a nebylo to předmětem úlohy).

Smyslem úlohy tedy bylo skutečně nasimulovat posaný náhodný proces a změřit jednotlivé pravděpodobnosti, případně střední hodnoty. Šlo mi zejména o simulaci části (d), ale v přiloženém souboru počítáme i ostatní.

6. Nechť F_X je dána předpisem $F_X(x) = x/3$ pro $x \in [0, 3]$, $F_X(x) = 0$ pro $x < 0$ a $F_X(x) = 1$ pro $x > 3$. Nechť $Y = 1/X$ a $Z = X^2$. Spočtěte

(a) $P(1 \le X \le 2)$
(b) $P(X \le Y)$
(c) $P(X \le Z)$
(d) hustotní funkci f_X.
(e) distribuční funkce F_Y a F_Z.

Řešení:

(a) Podle definice distribuční funkce je $P(X \le 2) = F_X(2)$ a $P(X \le 1) = F_X(1)$. Je tedy

$$P(1 \le X \le 2) = P(1 < X \le 2) = P(X \le 2) - P(X \le 1) = F_X(2) - F_X(1) = 2/3 - 1/3 = 1/3,$$

kde první rovnost platí proto, že je spojitá náhodná veličina.

(b) Protože $X > 0$ s.j., nastává $X \le Y = 1/X$ právě tehdy, když $X^2 \le 1$, což je právě tehdy, když $X \le 1$. Čili

$$P(X \le Y) = P(X \le 1) = F_X(1) = 1/3.$$

(c) Protože $X > 0$ s.j., nastává $X \le Z = X^2$ právě tehdy, když $1 \le X$, neboli když není $X < 1$.

$$P(X \le Z) = 1 - P(X < 1) = 1 - F_X(1) = 2/3.$$

(d) Hustota $f_X(x) = 1/3$ pro $x \in [0, 3]$ a 0 jinde. (Jedná se o uniformní rozdělení na intervalu $[0, 3]$.)

(e) Opět stačí řešit jen kladné hodnoty. Potřebujeme prozkoumat jev $Y \le y$. Protože $Y = 1/X$ a $X > 0$, je to ekvivalentní $X \ge 1/y$. Neboli $F_Y(y) = 1 - F_X(1/y)$. To znamená, že

$$F_Y(y) = \begin{cases} 0 & \text{pro } y \le 1/3 \\ 1 - \frac{1}{3y} & \text{pro } y \ge 1/3. \end{cases}$$

F_Z nebyla součástí domácího úkolu.

V zadání domácího úkolu se tentokrát samplování nežádalo, nicméně pro ilustraci je též přiloženo.