4. zkoušková písemka NMAI059 Pravd. a Stat. 1 – řešení – 21.7.2021

1.

$x \backslash y$	0	1	2
1	a	1/8	b
0	1/8	3/8	c

V tabulce je sdružená pravděpodobnostní funkce náhodných veličin X, Y. Jiné než vyznačené hodnoty tyto veličiny nenabývají.

(a) Rozhodněte, zda je možné tabulku doplnit (tj. zvolit a, b, c) tak, aby byly X a Y nezávislé.

(b) Doplňte tabulku tak, aby byla střední hodnota $\mathbb{E}(X)$ co největší.

(c) Doplňte tabulku tak, aby byla střední hodnota $\mathbb{E}(Y)$ co největší.

(d) Doplňte tabulku tak, aby byl rozptyl $\mathrm{var}(X)$ co nejmenší.

Řešení:

(a) Musí být $p_Y(1) = P(Y = 1) = 1/8 + 3/8 = 1/2$ a pokud jsou X, Y nezávislé, tak $p_X(0) = 3/4$ a $p_X(1) = 1/4$. Odsud $p_Y(0) = 1/6$ a tedy $p_Y(2) = 1 - 1/2 - 1/6 = 1/3$. Známe tedy pravděpodobnostní funkce X i Y a snadno dopočítáme jediné řešení $a = p_X(1)p_Y(0) = 1/24$, $b = p_X(1)p_Y(2) = 1/12$ a $c = p_X(0)p_Y(2) = 1/4$.

Mnozí tvrdili, že řešení neexistuje, protože ve druhém sloupečku jsou dvě různá čísla. Ujasněte si rozdíl mezi nezávislostí dvou veličin a tím, že jednotlivé veličiny jsou uniformní!

(b) Potřebujeme co největší $a+b$, protože ty připívají $\mathbb{E}(X)$, zatímco c nepřispívá. Např. tedy $a = c = 0$, $b = 3/8$.

(c) Potřebujeme co největší $b+c$, protože ty připívají $\mathbb{E}(X)$, zatímco a nepřispívá. Např. tedy $a = c = 0$, $b = 3/8$.

(d) X nabývá jen hodnoty 0, 1, má tedy Bernoulliho rozdělení s parametrem $p = a + 1/8 + b$. Víme, že rozptyl takové veličiny je $p(1-p)$. Rozptyl bude tedy nejmenší, pokud bude $a = b = 0$ a $c = 3/8$. (Rozptyl se bude rovnat $\frac{1}{8} \cdot \frac{7}{8} = \frac{7}{64}$.)

2. Opakovaně házíme spravedlivou mincí. Označme X_k číslo hodu, v němž poprvé padne k-krát po sobě panna. Např. pro posloupnost POPPP je $X_1 = 1$, $X_2 = 4$ a $X_3 = 5$.

(a) Určete $\mathbb{E}(X_1)$.

(b) Určete $\mathbb{E}(X_2)$. K tomu účelu určete $\mathbb{E}(X_2 \mid \text{první hod byl O})$ pomocí $\mathbb{E}(X_2)$. A také $\mathbb{E}(X_2 \mid \text{první dva hody byly PO})$ pomocí $\mathbb{E}(X_2)$.

(c) Určete $\mathbb{E}(X_3)$.

Řešení:

(a) Jedná se o gemetrické rozdělení – čekání na úspěch, jehož pravděpodobnost je 1/2. Je tedy $\mathbb{E}(X_1) = 1/(1/2) = 2$.

(b) Označme $\mathbb{E}(X_2) = x$. Pokud první hod byl orel, tak jsme ve stejné pozici jako na začátku, jenom jsme jeden hod „promarnili“. Je tedy $\mathbb{E}(X_2 \mid \text{první hod byl O}) = 1 + x$. Obdobně $\mathbb{E}(X_2 \mid \text{první dva hody byly PO}) = 2 + x$. Pokud budou první dva hody PP, tak $X_2 = 2$. Nyní využijeme větu o úplné střední hodnotě (o rozboru případů):

$$
\begin{aligned}
x = \mathbb{E}(X_2) &= P(O) \cdot \mathbb{E}(X_2 \mid O) + P(PO) \cdot \mathbb{E}(X_2 \mid PO) + P(PP) \cdot \mathbb{E}(X_2 \mid PP) = \\
&= \frac{1}{2} \cdot (1 + x) + \frac{1}{4} \cdot (2 + x) + \frac{1}{4} \cdot 2 \\
x &= \frac{3}{4}x + \frac{3}{2} \\
x &= 6
\end{aligned}
$$

(c) Označme $\mathbb{E}(X_3) = y$. Budeme postupovat podobně jako v části (b), rozlišíme začátky O, PO, PPO, PPP. Dostáváme

$$
\begin{aligned}
y = \mathbb{E}(X_3) =& P(O) \cdot \mathbb{E}(X_3 \mid O) + P(PO) \cdot \mathbb{E}(X_3 \mid PO) + \\
& + P(PPO) \cdot \mathbb{E}(X_3 \mid PPO) + P(PPP) \cdot \mathbb{E}(X_3 \mid PPP) \\
=& \frac{1}{2} \cdot (1 + y) + \frac{1}{4} \cdot (2 + y) + \frac{1}{8} \cdot (3 + y) + \frac{1}{8} \cdot 3 \\
y =& \frac{7}{8}y + \frac{14}{8} \\
y =& 14
\end{aligned}
$$

3. Buď X náhodná veličina s hustotou $f_X(t) = 1/t^2$ pro $t \geq 1$ a $f_X(t) = 0$ jinak.

(a) Ověřte, že se jedná o hustotu.
(b) Určete $\mathbb{E}(X)$.
(c) Spočtěte distribuční funkci, F_X.
(d) Buď $Y = 1/X$. Jaká je distribuční funkce náhodné veličiny Y?
(e) Určete hustotu náhodné veličiny Y. Pojmenujte jejjí rozdělení.

Řešení:

(a) Daná funkce je nezáporná a platí pro ni

$$\int_{-\infty}^{\infty} f_X(t)dt = \int_1^{\infty} 1/t^2 dt = \left[-1/t\right]_1^{\infty} = 0 - (-1) = 1.$$

Je to tedy hustota.

(b) Počítáme podle definice.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)dt = \int_1^{\infty} 1/t dt = \left[\log t\right]_1^{\infty} = \infty - 0 = \infty.$$

Takže střední hodnota existuje, ale je nekonečná.

(c) Pro $x \leq 1$ je jistě $F_X(x) = 0$. Pro $x \geq 1$ integrujeme

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_1^{x} 1/t^2 dt = \left[-1/t\right]_1^{x} = -1/x - (-1) = 1 - 1/x.$$

Několika zkoušeným vyšla distribuční funkce záporná, nebo větší než 1. To by mělo být jasné varovné znamení, že je něco špatně!

(d) Náhodná veličina X nabývá jen hodnoty z $[1, \infty)$, proto Y bude z intervalu $(0, 1]$. Proto pro $y \leq 0$ je $F_Y(y) = 0$ a pro $y \geq 1$ je $F_Y(y) = 1$. Pro $y \in (0, 1]$ máme

$$F_Y(y) = P(Y \leq y) = P(X \geq 1/y) = 1 - P(X < 1/y) = 1 - \left(1 - \frac{1}{1/y}\right) = y.$$

(Je totiž $1/y \geq 1$, takže pro ně funguje vzorec získaný integrováním.)

(e) Z předchozí části vidíme, že se jedná o uniformní rozdělení $U(0, 1)$ s hustotou $f_Y(y) = 1$ na intervalu $[0, 1]$ a 0 jinde. (Hodnota v krajních bodech intervalu není jednoznačně určena.)

4. (a) Definujte pojem distribuční funkce náhodné veličiny. Definujte, co znamená, že nějaký jev platí skoro jistě.

Rozhodněte zda některá z následujících implikací platí pro libovolnou dvojici náhodných veličin X, Y:

1. $X \leq Y$ s.j. $\Rightarrow F_X(t) \leq F_Y(t)$
2. $X \leq Y$ s.j. $\Rightarrow F_Y(t) \leq F_X(t)$
3. $F_X(t) \geq F_Y(t) \Rightarrow X \leq Y$ s.j.

(b) Definujte pojem rozptyl náhodné veličiny. Kdy je rozptyl roven 0?

Řešení:

(a) Distribuční funkce náhodné veličiny X je definována předpisem $F_X(x) = P(X \leq x)$. Jev A platí skoro jistě znamená, že $P(A) = 1$.

Pro každé dvě náhodné veličiny platí 2. Je totiž

$$F_Y(y) = P(Y \leq y) = P(X \leq Y \leq y) \leq P(X \leq y) = F_X(y)$$

Bod 1 už platit nemůže, to by musely být obě distribuční funkce stejné. Bod 3 nemůže platit vždy, protože X, Y můžou být např. nezávislé se stejným rozdělením, řekněme $Ber(1/2)$. Pak mají stejnou distribuční funkci, ale neplatí vždy $X \leq Y$.

(b) Rozptyl náhodné veličiny X se střední hodnotou $\mu = \mathbb{E}(X)$ je definován jako

$$\mathrm{var}(X) = \mathbb{E}((X - \mu)^2).$$

(Pokud $\mathbb{E}(X) = \infty$ nebo $\mathbb{E}(X)$ neexistuje, tak rozptyl nedefinujeme.) Často ho počítáme podle vzorce

$$\mathrm{var}(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2.$$

Z definice plyne (s ohledem na to, že druhá mocnina je vždy kladná), že $\mathrm{var}(X) = 0$ právě tehdy, když je $X = \mu$ s.j., neboli když je X skoro jistě konstantní.

5. Popište metody, jak generovat náhodnou veličinu s danou distribuční funkcí. Zejména základní metodu (inverse transformation), případně též zamítací metodu (rejection sampling).

Řešení: Na přednášce jsme si uváděli na příklad následující dvě metody (pro plný počet bodů stačila ta první). A také speciální případ té první metody, kdy generujeme diskrétní náhodnout veličinu.

Inverse transformation

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.

Zamítací metoda (rejection sampling)

- Chceme vygenerovat n.v. s hustotou f.
- Umíme vygenerovat n.v. s hustotou g (která je „podobná“).
- $\frac{f(y)}{g(y)} \leq c$ pro nějakou konstantu c.
- Postup
  1. Vygenerujeme Y s hustotou g, a $U \sim U(0,1)$.
  2. Pokud $U \leq \frac{f(Y)}{cg(Y)}$, tak $X := Y$.
  3. Jinak hodnotu Y, U zamítneme a opakujeme od bodu 1.
- Zdůvodnění: vygenerovat náhodnou hodnotu X s hustotou f je totéž, jako vygenerovat náhodný bod pod grafem funkce f, jehož vodorovná (x-ová) souřadnice je X (a svislá je uniformně náhodná mezi 0 a X).

6. Vyslovte a dokažte větu o střední hodnotě součinu nezávislých náhodných veličin. (Stačí varianta pro diskrétní náhodné veličiny.)

Řešení:

Věta:

Pro diskrétní nezávislé náhodné veličiny platí $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, má-li pravá strana smysl.

Důkaz: Využijeme pravidla LOTUS pro funkci $g(x, y) = xy$. Podle něj je

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(g(X, Y)) && \text{definice } g \\
&= \sum_{x \in Im(X), y \in Im(Y)} g(x, y) \cdot P(X = x \,\&\, Y = y) && \text{LOTUS} \\
&= \sum_{x \in Im(X), y \in Im(Y)} xy \cdot P(X = x) P(Y = y) && \text{defce } g \text{ a nezávislost } X,\, Y \\
&= \Big(\sum_{x \in Im(X)} x \cdot P(X = x) \Big) \cdot \Big(\sum_{y \in Im(Y)} y \cdot P(Y = y) \Big) \\
&= \mathbb{E}(X) \cdot \mathbb{E}(Y) && \text{definice } \mathbb{E}(X),\, \mathbb{E}(Y)
\end{aligned}$$