5. cvičení z PSt — 30.3.2021

Rozdělení n.v.

1. (Newton-Pepy's problem) Označme A_1 jev „z šesti hodů kostkou padne aspoň jedna šestka". Dále A_2 jev „z dvanácti hodů kostkou padnou aspoň dvě šestky". Obecně: A_k jev „z $6k$ hodů kostkou padne aspoň k šestek".

(a) Vyjádřete $P(A_i)$ pomocí vzorečku s kombinačními čísly.
(b) Vyjádřete $P(A_i)$ distribuční funkce binomického rozdělení.
(c) Jaká je střední hodnota počtu šestek pro každou z variant?
(d) * Zapřemýšlejte, která z $P(A_k)$ by měla být největší bez použití vzorce.

2. (de Mèreho problém) (a) Jaká je pravděpodobnost, že padne ze čtyř hodů aspoň jedna šestka?
(b) Jaká je pravděpodobnost, že padne z 24 hodů dvojicí kostek aspoň jedna dvojitá šestka?
(c) De Mèreho problém spočíval v tom, jestli jsou odpovědi na části a,b stejné.
(d) Umíte úlohu interpretovat jako otázku o binomickém rozdělení? O geometrickém rozdělení?

Náhodné vektory

3. Nezávislé n.v. $X_1, \ldots, X_n$ mají geometrické rozdělení s parametry $p_1, \ldots, p_n$. Jaké je rozdělení $\min(X_1, \ldots, X_n)$?

4. Nechť $X \sim Pois(\lambda)$, $Y \sim Pois(\mu)$ jsou n.n.v. Pak $X + Y \sim Pois(\lambda + \mu)$.

5. Hodíme třikrát mincí. Označíme X počet rubů v prvních dvou hodech a Y počet líců v posledních dvou hodech.

(a) Určete sdruženou pravděpodobnostní funkci $p_{X,Y}$ a také marginální pravděpodobnostní funkce p_X, p_Y.
(b) Jsou X a Y nezávislé?
(c) Určete $P(X < Y)$.
(d) Určete podmíněnou pravděpodobnostní funkce $p_{X|Y}$.

6. Připomeňte si definici indikátorové náhodné veličiny I_A.

(a) Jaká je $\mathbb{E}(I_A)$?
(b) Nechť $A = A_1 \cup A_2 \cup \cdots \cup A_n$. Ověřte rovnost

$$1 - I_A = \prod_{i=1}^{n} (1 - I_{A_i}).$$

(c) Roznásobte a použijte větu o linearitě střední hodnoty, abyste získali princip inkluze a exkluze.

Spojité náhodné veličiny

Připomeňte si, že distribuční funkce F_X je definována vztahem

$$F_X(x) = P(X \leq x).$$

V některých případech je $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ pro vhodnou nezápornou funkci f_X (hustotu X). Pak je $P(X \in A) = \int_A f_X(t)dt$. Platí také $\mathbb{E}(X) = \int_{-\infty}^{\infty} t\ f_X(t)dt$ a obecněji

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t)\ f_X(t)dt.$$

Stejně jako pro diskrétní n.v. platí, že $var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

7. Pro n.v. X s distribuční funkci F_x vyjádřete (a) $P(X \in (0, 1])$ (b) $P(X > 0)$ (c) $P(X < 0)$ (d) $P(X \in [0, 1])$

8. Nechť X je spojitá náhodná veličina. Vyjádřete pomocí F_X distribuční funkci náhodných veličin
(a) $-X$. (b) $X^+ = \max(0, X)$, (c) $X^- = -\min(X, 0)$, (d) $|X| = X^+ + X^-$,

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

Bonusy

10. * Označme M počet emailů, které dostaneme za den, S počet spamů mezi nimi, H počet „hamů" – těch, co nejsou spamy. Předpokládejme, že $M \sim Pois(\lambda)$ a že každý email má nezávisle na ostatních pravděpodobnost p, že je to spam.
(a) Vyjádřete $P(S = k)$ (jako nekonečnou sumu) pomocí sdruženého rozdělení M a S.
(b) Odvoďte, že $S \sim Pois(p\lambda)$.
(c) Odvoďte, že $H \sim Pois((1-p)\lambda)$ a také, že H, S jsou nezávislé n.v.

11. * V tipovací hře má soutěžící na výběr n otázek, ze kterých si může postupně vybírat. U i-té otázky s pravděpodobností p_i odpoví správně, získá za to h_i korun a právo dalšího výběru. Pokud neodpoví správně, končí. Předpokládejme, že cílem je maximalizovat střední hodnotu zisku. Ukažte, že toho docílí, bude-li vybírat otázky seřazené podle hodnoty $\frac{p_i h_i}{1-p_i}$.

K procvičení

12. 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 $0 \leq A \leq 2m$ a $0 \leq B \leq m$. 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 zvolenou a-prvkovou množinu lidí M, jaká je pravděpodobnost, že je to přesně množina přeživších? Neboli, kolik je $P(L = M)$? A kolik $P(L = M \mid A = a)$?

(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 $Bern(p)$.