6. cvičení z PSt — 20.–24.3.2023

Poznávačka náhodných veličin

1. Na kroužku máme pět klíčů, jeden z nich je správný, ale my nevíme jaký. Zkoušíme otevřít dveře.

(a) Po každém pokusu se nám kroužek vysmekne, a vybíráme vždy znovu náhodně.

(b) Vybíráme v náhodném pořadí, ale každý klíč jenom jednou (můžeme si je poznačit).

V obou případech zkoumáme, kolikátým pokusem dveře otevřeme. Pojmenujte, jaké je rozdělení této náhodné veličiny a jaká je její střední hodnota.

(c) Jako část (a), ale správné jsou dva klíče z deseti?

(d) Jako část (b), ale správné jsou dva klíče z deseti? (Zde je určení střední hodnoty trochu těžší, stačí když určíte pravděpodobnostní funkci.)

2. Jste ve skupině s dalšími 500 lidmi. Jaké je pravděpodobnost, že právě jeden z nich má narozeniny ve stejný den jako vy? (Ignorujte přestupné roky a to, že ne ve všechny dny se rodí stejně dětí.)

Neboli: označme X počet lidí, se stejnými narozeninami, a spočtěte $p_X(1)$. Jaké je rozdělení X? Jaká je střední hodnota? Aproximujte $p_X(1)$ pomocí Poissonova rozdělení.

Pokud Y je obdobný počet pro vašeho souseda – jaký je vztah X a Y?

[Nápověda: můžete začít s $p_X(0)$, to je o něco snazší.]

3. X je uniformně náhodná mocnina dvojky mezi $\{2^a, 2^{a+1}, \dots, 2^b\}$.

(a) Vyjádřete X pomocí náhodné veličiny U, která je uniformně rozdělená na množině $\{a, a+1, \dots, b\}$.

(b) Určete $\mathbb{E}(X)$ a $\mathrm{var}(X)$.

Náhodné vektory

Sdružená pravděpodobnostní funkce je definována vztahem $p_{X,Y}(x,y) = P(X = x \,\&\, Y = y)$. „Jednorozměrné funkce“ p_X, p_Y se v tomto kontextu nazývají marginální pravděpodobnostní funkce. Připomeňte si, jak je zjistit z $p_{X,Y}$.

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 $Z = \max(X_1, X_2, X_3)$?

[Nápověda 1: Jakých hodnot nabývá vektor (X_1, X_2), pokud $Y = k$?
nebo jinak –
Nápověda 2: Určete napřed $P(Y \le k)$, $P(Z \le k)$?]

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

[Nápověda: použijte konvoluční vzorec podobně jako na přednášce.]

6. 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 funkci $p_{X|Y}$, tj. čísla $P(X = x \mid Y = y)$ pro všechny hodnoty x, y.

Spojité náhodné veličiny

Připomeňte si, že distribuční funkce F_X je definována vztahem

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

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

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

8. Vyřešte předchozí část znovu, pro n.v. X s hustotou f_X.

Bonusy

9. 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.

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.

K procvičení

11. 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 $Bern(p)$.