6. cvičení z PSt — 25.–29.3.2024

Poznávačka náhodných veličin

1. Pravděpodobnost, že do našeho serveru pronikne hacker je během každého dne 0.01, nezávisle pro každý den. Označme T počet dnů do prvního průniku. Jaké je rozdělení T, $\mathbb{E}(T)$, $\mathrm{var}(T)$? Jaká je pravděpodobnost, že server zůstane bezpečný po celý rok?

2. Každý test programu může skončit buď nalezením chyby (úspěch) nebo ne (neúspěch). Předpokládáme, že pravděpodobnost nalezení chyby při jednom testu je 0.05 a vývojář provede 20 nezávislých testů, označíme X počet nalezených chyb. Jaké je rozdělení X, $\mathbb{E}(X)$, $\mathrm{var}(X)$? Jaká je pravděpodobnost, že nalezne právě tři chyby?

3. Historická data ukazují, že náš server obdrží průměrně 30 žádostí za minutu. Použijte Poissonovo rozdělení k určení pravděpodobnosti, že server obdrží přesně 40 žádostí v následující minutě.

Rozptyl

Připomenutí:

- definice $\mathrm{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$
- věta $\mathrm{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- odvozené veličiny: směrodatná odchylka $\sigma_X = \sqrt{\mathrm{var}(X)}$
- variační koeficient $CV_X = \sigma_X / \mathbb{E}(X)$ (pokud $\mathbb{E}(X) > 0$)

4. Předpokládejme, že vyřešení jednoho příkladu trvá X minut, kde $X = 1, 2, \dots,$ nebo 5. Doba trvání je náhodná (závislá na počasí) a pravděpodobnostní funkce je $p_X(1) = p_X(2) = 0.1$, $p_X(3) = p_X(4) = 0.2$, $p_X(5) = 0.4$. Z minula víme, že $\mathbb{E}(X) = 3.7$. Určete $\mathrm{var}(X)$ a σ_X.

5. Pro náhodné veličiny X, Y platí $Y = aX + b$ (kde a, b jsou reálná čísla).
(a) Vyjádřete $\mathrm{var}(Y)$ pomocí $\mathrm{var}(X)$.
(b) Zopakujte pro σ_X a CV_X.

6. Dávají smysl následující „definice alternativního rozptylu“? Tj., dozvíte se o náhodné veličině X něco zajímavého, pokud zjistíte
(a) $\mathbb{E}(X - \mathbb{E}(X))$?
(b) $\mathbb{E}(|X - \mathbb{E}(X)|)$?
(c) $\mathbb{E}((X - \mathbb{E}(X))^k)$ pro $k = 3, 4, \dots$.

7. Nechť $X \sim Bin(100, 0.5)$ a $Y \sim 10Bin(100, .05)$ (tedy $Y/10$ má binomické rozdělení $Bin(100, .05)$). Spočtěte (využitím vzorců z přednášky) $\mathbb{E}(X)$, $\mathrm{var}(X)$, σ_X, CV_X a totéž pro Y.

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 $p_{X,Y}$.

8. Označme X, Y výsledky dvou nezávislých hodů čtyřstěnnou kostkou (s čísly 1, …, 4).
(a) Jaká je pravděpodobnostní funkce $Z_1 = \max(X, Y)$?
(b) Jaká je pravděpodobnostní funkce $Z_2 = XY$?

[Nápověda: Jakých hodnot nabývá vektor (X, Y), pokud $\max(X, Y) = k$? Resp. v druhé části, pokud $XY = k$?]

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

(a) Najděte marginální rozdělení X i Y. Spočtěte $\mathbb{E}(X)$, $\mathbb{E}(Y)$.

(b) Jsou X a Y nezávislé? Neboli: platí $p_{X,Y}(x,y) = p_X(x)p_Y(y)$?

(c) Popište rozdělení $X+Y$ – tj. nalezněte pravděpodobnostní funkci n.v. $X+Y$. Spočtěte odsud $\mathbb{E}(X+Y)$ – ověřte, zda se to rovná $\mathbb{E}(X)+\mathbb{E}(Y)$.

(d) Popište rozdělení $X \cdot Y$. Spočtěte odsud $\mathbb{E}(XY)$ – ověřte, zda se to rovná $\mathbb{E}(X)\mathbb{E}(Y)$.

$x \backslash y$	0	1	2
0	1/4	1/6	1/12
1	1/6	1/4	1/12

K procvičení

10. 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.

11. Nechť $X \sim Poi(\lambda)$. Odvoďte vztahy $\mathbb{E}(X) = \lambda$ a $\operatorname{var}(X) = \lambda$.

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 $0 \le A \le 2m$ a $0 \le B \le m$. Pro každé $a = 0, \dots, 2m$ 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. $X_1, \dots, X_m$. (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 $Ber(p)$.