6. cvičení

Poznávačka náhodných veličin

Úloha 1 (Hacker)
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?

Řešení
$T \sim \mathrm{Geom}(0.01)$, $\mathbb{E}(T) = 100, \mathrm{var}(T) = \frac{1-p}{p^2} = \frac{0.99}{0.0001} = 9900$, $P[365 \text{ dní bez incidentu}] = 0.99^{365} \approx 0.025518$

Úloha 2 (Test-driven probability)
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?

Řešení
$X \sim \mathrm{Bin}(20, 0.05), \mathbb{E}[X] = 1, \mathrm{var}(X) = np(1-p) = 0.95$, $P[X = 3] = \binom{20}{3} 0.05^3 \cdot 0.95^{17} \approx 0.059582$

Úloha 3 (Vyřizování žádostí)
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.

Řešení
Použijeme tedy $Z \sim \mathrm{Poi}(30) \rightsquigarrow P(Z = 40) = \frac{30^{40}}{40!} \cdot e^{-30} = \frac{5845177845843136310577392578125 00}{392276786508598684 5929e^{30}} \approx 0.0139435$

Rozptyl

Úloha 4 (Here we go again)
Předpokládejme, že vyřešení jednoho příkladu trvá X minut, kde $X = 1, 2, \ldots,$ 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.

Řešení
Spočteme $\mathbb{E}(X^2) = 0.1 + 4 \cdot 0.1 + 9 \cdot 0.2 + 16 \cdot 0.2 + 25 \cdot 0.4 = 0.5 + 5 + 10 = 15.5$, pak $\mathrm{var}(X) = 15.5 - 3.7^2 = 1.81$, dále $\sigma_X = \sqrt{1.81} \approx 1.345$.
(Přímý výpočet z definice by byl otravnější)

Úloha 5 (Je rozptyl taky lineární?)
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.

Řešení a) $\mathrm{var}(Y) = \mathrm{var}(aX + b) = \mathbb{E}[(aX + b)^2] - \mathbb{E}[aX + b]^2 = \mathbb{E}[a^2X^2 + 2abX + b^2] - (a\mathbb{E}[X] + b)^2 = a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - b^2 = a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2 \mathrm{var}(X)$

b) Okamžitě vidíme $\sigma_Y = a \cdot \sigma_X$, a pro $b = 0$ máme $CV_Y = CV_X$.

Úloha 6 (Můžeme rozptyl definovat jinak?)
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 $k = 3, 4, \ldots$?

Řešení a) $\mathbb{E}(X - \mathbb{E}(X)) = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$

b) $\mathbb{E}(|X - \mathbb{E}(X)|)$ – ano, ale bude to méně citlivé na velké odchylky

c) $\mathbb{E}((X - \mathbb{E}(X))^k)$ – ano, tyto k-té centrální momenty naopak budou citlivější na velké odchylky

Náhodné vektory

Úloha 7 (Opět házeme kostkami)
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 $XY = k$?

Řešení
$p_{Z_1}(1) = 1/16, p_{Z_1}(2) = 3/16, p_{Z_1}(3) = 5/16, p_{Z_1}(4) = 7/16$
$p_{Z_2}(1) = 1/16, p_{Z_2}(2) = 2/16, p_{Z_2}(3) = 2/16, p_{Z_2}(4) = 3/16, p_{Z_2}(6) = 2/16, p_{Z_2}(8) = 2/16, p_{Z_2}(9) = 1/16, p_{Z_2}(12) = 2/16, p_{Z_2}(16) = 1/16$

Úloha 8 (Čtení tabulky)

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

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 $\mathbb{E}(X)\mathbb{E}(Y)$.

Řešení

a) $p_X(0) = p_X(1) = 1/2$, $p_Y(0) = p_Y(1) = 5/12, p_Y(2) = 2/12$ $\mathbb{E}[X] = 1/2, \mathbb{E}[Y] = 9/12$

b) Ne, třeba $P[X = 0 \wedge Y = 0] = 1/4 \neq 1/2 \cdot 5/12 = P[X = 0] \cdot P[Y = 0]$

c) $p_{X+Y}(0) = 1/4, p_{X+Y}(1) = 2/6, p_{X+Y}(2) = 1/3, p_{X+Y}(3) = 1/12$ Pak $\mathbb{E}[X+Y] = 2/6+2\cdot 1/3+3\cdot 1/12 = 15/12 = 1/2 + 9/12$.

d) $p_{XY}(0) = 2/3, p_{XY}(1) = 1/4, p_{XY}(2) = 1/12$. Pak $E(XY) = 1/4 + 2 \cdot 1/12 = 5/12 \neq 1/2 \cdot 9/12 = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

Na procvičení

Úloha 9 (Poisson, střední hodnota a rozptyl)
Nechť $X \sim Poi(\lambda)$. Odvoďte vztahy $\mathbb{E}(X) = \lambda$ a $\text{var}(X) = \lambda$.

Řešení
TODO

Užitečné pojmy

- definice $\text{var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$
- věta $\text{var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
- odvozené veličiny:
  - směrodatná odchylka $\sigma_X = \sqrt{\text{var}(X)}$
  - variační koeficient $CV_X = \sigma_X/\mathbb{E}(X)$ (pokud $\mathbb{E}(X) > 0$)
- Sdružená pravděpodobnostní funkce je definována vztahem $p_{X,Y}(x, y) = P(X = x \wedge Y = y)$. „Jednorozměrné funkce“ p_X, p_Y se v tomto kontextu nazývají marginální pravděpodobnostní funkce. (Připomeňte si, jak je zjistit z $p_{X,Y}$.)