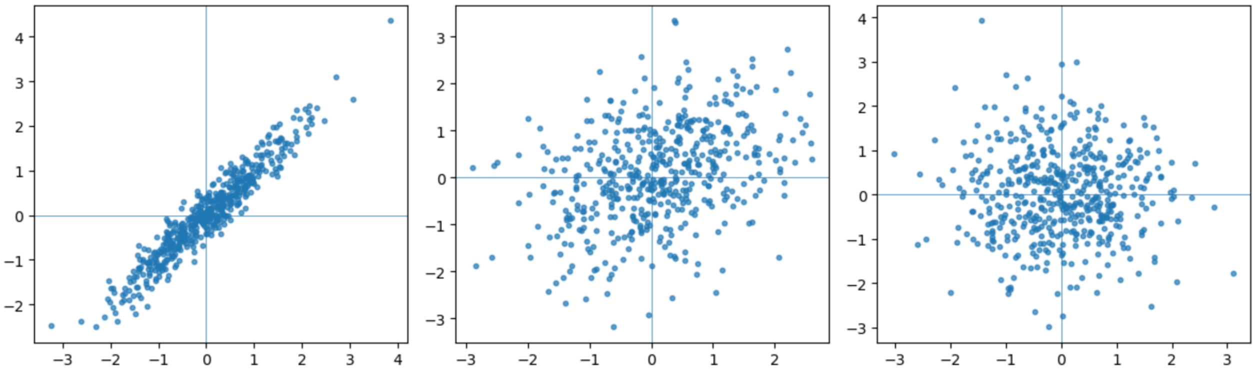
7. cvičení

Kovariance a korelace

- $\mathrm{cov}(X,Y) \overset{\text{def}}{=} \mathbb{E}\big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\big) \overset{\text{věta}}{=} \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$
- $\mathrm{cov}(X,X) = \mathrm{var}(X)$
- $\mathrm{cov}(X,Y)$ je lineární v obou složkách, tj. např. $\mathrm{cov}(X, a_1Y_1 + a_2Y_2) = a_1\mathrm{cov}(X,Y_1) + a_2\mathrm{cov}(X,Y_2)$
- $\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}$.

Úloha 1 (Coco taxi)
Nezávislé náhodné veličiny X, Y mají střední hodnotu 0 a rozptyl 1. Položíme $Z_0 = Y$, $Z_1 = X + 0.3Y$ a $Z_2 = 0.4X + Y$. Spočtěte $\mathrm{cov}(X, Z_i)$ a také korelaci $\rho(X, Z_i)$. Který obrázek odpovídá kterému vzorci?



Řešení
Máme $\mathbb{E}X = \mathbb{E}Y = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ a X, Y jsou nezávislé, tedy $\mathrm{cov}(X,Y) = 0$.

Pro $Z_0 = Y$: $\mathrm{cov}(X,Z_0) = \mathrm{cov}(X,Y) = 0, \qquad \rho(X,Z_0) = 0.$

Pro $Z_1 = X + 0.3Y$:

$$\mathrm{cov}(X,Z_1) = \mathrm{cov}(X,X) + 0.3\,\mathrm{cov}(X,Y) = 1.$$

Dále

$$\mathrm{Var}(Z_1) = \mathrm{Var}(X) + 0.3^2\mathrm{Var}(Y) = 1 + 0.09 = 1.09,$$

takže

$$\rho(X,Z_1) = \frac{\mathrm{cov}(X,Z_1)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Z_1)}} = \frac{1}{\sqrt{1.09}}.$$

Pro $Z_2 = 0.4X + Y$:

$$\mathrm{cov}(X,Z_2) = 0.4\,\mathrm{cov}(X,X) + \mathrm{cov}(X,Y) = 0.4.$$

Dále

$$\mathrm{Var}(Z_2) = 0.4^2\mathrm{Var}(X) + \mathrm{Var}(Y) = 0.16 + 1 = 1.16,$$

tedy

$$\rho(X,Z_2) = \frac{0.4}{\sqrt{1.16}}.$$

Nejsilnější závislost má Z_1 (levý obrázek), střední Z_2 (prostřední obrázek), nulovou Z_0 (pravý obrázek).

Distribuční funkce

Připomeňte si, že distribuční funkce F_X je definována vztahem

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

Úloha 2 (Distribuce)
Pro náhodnou veličinu X s distribuční funkcí F_X vyjádřete

a) $P(X \in (0, 1])$

b) $P(X > 0)$

c) * $P(X < 0)$

d) * $P(X \in [0, 1])$

Řešení
Nechť $F_X(x) = \mathbb{P}(X \leq x)$.

(a) $\mathbb{P}(X \in (0, 1]) = \mathbb{P}(X \leq 1) - \mathbb{P}(X \leq 0) = F_X(1) - F_X(0)$.

(b) $\mathbb{P}(X > 0) = 1 - \mathbb{P}(X \leq 0) = 1 - F_X(0)$.

(c) $\mathbb{P}(X < 0) = \mathbb{P}(X \leq 0^-) = F_X(0^-) := \lim_{x \uparrow 0} F_X(x)$.

(d) $\mathbb{P}(X \in [0, 1]) = \mathbb{P}(X \leq 1) - \mathbb{P}(X < 0) = F_X(1) - F_X(0^-)$.

Úloha 3
Nechť X splňuje $P(X = x) = 0$ pro každé x. Vyjádřete pomocí F_X distribuční funkci náhodných veličin

a) $-X$,

b) $X^+ = \max(0, X)$,

c) $|X|$.

Řešení
Předpokládáme, že $\mathbb{P}(X = x) = 0$ pro každé x.

(a) $-X$:

$$F_{-X}(x) = \mathbb{P}(-X \leq x) = \mathbb{P}(X \geq -x) = 1 - F_X(-x).$$

(b) $X^+ = \max(0, X)$:

$$F_{X^+}(x) = \begin{cases} 0, & x < 0, \\ F_X(x), & x \geq 0. \end{cases}$$

(c) $|X|$:

$$F_{|X|}(x) = \begin{cases} 0, & x < 0, \\ \mathbb{P}(-x \leq X \leq x) = F_X(x) - F_X(-x), & x \geq 0. \end{cases}$$

Exponenciální rozdělení

Říkáme, že X má exponenciální rozdělení, $X \sim Exp(\lambda)$, pokud

$$F_X(x) = 1 - e^{-\lambda x} \qquad \text{pro } x \geq 0, \text{ jinak } 0.$$

Z přednášky navíc máte vědět, že $\mathbb{E}(X) = 1/\lambda$.

Úloha 4
Střední doba života harddisku je 4 roky. Přepokládejme, že tato doba je popsána náhodnou veličinou s exponenciálním rozdělením[1].

a) Jaká je pravděpodobnost, že disk selže během prvních tří let?

b) Jaká je pravděpodobnost, že vydrží alespoň 10 let?

c) Po jaké době se rozbije 10 % disků?

Řešení
Pro $X \sim \text{Exp}(\lambda)$ platí $\mathbb{E}X = 1/\lambda$. Ze zadání tedy $\lambda = \frac{1}{4}$.

(a) $\mathbb{P}(X \leq 3) = 1 - e^{-3/4}$.

(b) $\mathbb{P}(X \geq 10) = e^{-10/4} = e^{-5/2}$.

(c) Hledáme t tak, aby $\mathbb{P}(X \leq t) = 0.1$. Tedy $1 - e^{-t/4} = 0.1 \implies t = -4\ln(0.9)$.

Úloha 5 (*Obliviate*)
Říkáme, že náhodná veličina X (resp. její rozdělení) *nemá paměť*, pokud

$$P(X > s + t | X \geq s) = P(X > t)$$

pro $s, t \geq 0$. Jinými slovy, doba, kterou jsme již čekali, nemá vliv na dobu, kterou budeme ještě čekat. Už jsme viděli, že geometrické rozdělení nemá paměť. Ukažte, že ani exponenciální rozdělení nemá paměť. Platí dokonce, že je to jediné spojité rozdělení na kladných čísel bez paměti (a geometrické je jediné diskrétní bez paměti), ale to dokazovat nemusíte.

Řešení
$P(X > s + t | X \geq s) = \frac{P[X \geq s+t \wedge X \geq s]}{P[X \geq s} = \frac{1-(1-e^{-\lambda(s+t)})}{1-(1-e^{-\lambda s})} = e^{-\lambda t} = 1 - (1 - e^{-\lambda t}) = P[X > t]$

Úloha 6
Sledujeme noční oblohu a vyhlížíme meteory. Pro dnešní noc víme, že pravděpodobnost spatření meteoru během časového úseku dt je (pro malá dt) rovna $c \cdot dt$, situace v disjunktních časových intervalech považujeme za nezávislé. Jaká je pravděpodobnost, že během doby t nic neuvidíme? [Klidně předpokládejte, že t/dt je celé číslo. Taky zkuste napsat $dt = 1/n$ a vzpomenout si na limity z analýzy.]

Řešení
V malém intervalu délky dt je pravděpodobnost spatření meteoru přibližně $c\,dt$, tedy pravděpodobnost, že nic neuvidíme, je

$$1 - c\,dt.$$

Rozdělíme-li čas t na $k = t/dt$ disjunktních intervalů, dostaneme

$$\mathbb{P}(\text{během doby } t \text{ nic neuvidíme}) = (1 - c\,dt)^{t/dt}.$$

Vzpomeneme si na známou limitu $\lim_{n\to\infty}(1 + x/n)^n = e^x$. Pak nám pro $dt = 1/n \to 0$ vyjde

$$\lim_{dt\to 0}(1 - c\,dt)^{t/dt} = \lim_{n\to\infty}(1 - c/n)^{tn} = e^{-ct}.$$

[1] To není realistický předpoklad, vizte např. `https://www.backblaze.com/blog/how-long-do-disk-drives-last/`.

Bonusové úlohy

Úloha 7 (Minimum z exponenciálních n.v.)
Nechť $X_i \sim \mathrm{Exp}(\lambda_i)$ pro $i = 1, \ldots, n$ jsou nezávislé náhodné veličiny. Označme $M = \min(X_1, \ldots, X_n)$. Ukažte, že $M \sim \mathrm{Exp}(\lambda_1 + \cdots + \lambda_n)$.

Řešení
$P[M > t] = \prod_{i=1}^{n} P[X_i > t] = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{t \cdot \sum_{i=1}^{n} \lambda_i}$

Úloha 8 (Algoritmus ze dvou algoritmů)
Pro jistý problém máme k dispozici dva algoritmy, A a B. Algoritmus C spočívá v tom, že si náhodně vybereme, který z algoritmů A, B spustíme – A bude mít pravděpodobnost p, B pravděpodobnost $1 - p$. Dobu běhu A, B, C chápeme jako náhodné veličiny, označíme je X, Y, Z.
(Čas na výběr algoritmu v Z počítat pro jednoduchost nemusíte.)

a) Vyjádřete $\mathbb{E}(Z)$ pomocí $\mathbb{E}(X)$ a $\mathbb{E}(Y)$.

b) Určete F_Z pomocí F_X, F_Y.

c) Pokud jsou X, Y spojité, určete f_Z pomocí f_X, f_Y.

Řešení
Střední hodnota je vážená
$F_Z(x) = pF_X(x) + (1 - p)F_Y(x)$
$f_Z = pf_X + (1 - p)f_Y$