

5. cvičení z STP

18. - 22. březen 2019

Připomenutí: Hustota pravděpodobnosti náhodné veličiny X je taková nezáporná integrabilní funkce $f_X : \mathbb{R} \rightarrow \mathbb{R}$, že

$$F_X(t) = \int_{-\infty}^t f_X(u) \, du$$

pro každé $t \in \mathbb{R}$.

Odsud snadno máme, že pokud $I \subseteq \mathbb{R}$ je interval (nebo i nějaká množina poskládaná z intervalů), pak

$$P(X \in I) = \int_I f_X(u) \, du$$

tedy pravděpodobnost, že hodnoty veličiny X padnou do I , zjistíme prostě zintegrováním hustoty přes I (podobně zjišťujeme např. hmotnost nějaké křivky, když zintegrujeme hustotu (hmotnosti) přes danou křivku).

Poznamenejme ještě důležitou věc a sice, že hustota f_X NENÍ zdaleka určena jednoznačně jako funkce (např. její změnou v konečně mnoha bodech se nezmění příslušné integrály, takže i změněná funkce bude také hustotou). Pokud veličina X hustotu má, pak platí, že derivace funkce F_X , tj. $(F_X)'$ je hustotou veličiny X (tato derivace sice nemusí všude existovat, ale to není na překážku - tam, kde neexistuje, si prostě zvolíme libovolné nezáporné hodnoty).

Příklad 5.1 Určete konstantu $c \in \mathbb{R}$ tak, aby funkce

$$f(x) = \begin{cases} c \cdot x e^{-2x} & , x \in (0, 1) \\ 0 & , \text{jinak.} \end{cases}$$

byla hustotou nějaké náhodné veličiny.

Řešení:

Máme charakterizační větu:

Nezáporná integrabilní funkce $f : \mathbb{R} \rightarrow \mathbb{R}$ je hustotou pravděpodobnosti nějaké náhodné veličiny X právě když $\int_{\mathbb{R}} f(x) \, dx = 1$.

Použitím integrace per partes dostaneme

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x) \, dx = \int_{-\infty}^0 0 \, dx + \int_0^1 c \cdot x e^{-2x} \, dx + \int_1^{\infty} 0 \, dx = \\ &= c \left[x \cdot \frac{e^{-2x}}{-2} \right]_0^1 + c \int_0^1 \frac{e^{-2x}}{2} \, dx = -\frac{c}{2} e^{-2} + c \left[\frac{e^{-2x}}{-4} \right]_0^1 = c \cdot \frac{1 - 3e^{-2}}{4}, \end{aligned}$$

a tudíž $c = \frac{4}{1 - 3e^{-2}}$. Protože $c > 0$, je splněna i nezápornost funkce f .

Příklad 5.2 Mějme funkci

$$f(x) = \begin{cases} 3e^{-3x} & x \in (0, \infty) \\ 0 & \text{jinak.} \end{cases}$$

- (a) Ověřte, že f je hustota pravděpodobnosti nějaké náhodné veličiny X .
- (b) Sestrojte distribuční funkci veličiny X příslušnou této hustotě.
- (c) Spočítejte pravděpodobnost $P(-1 \leq X \leq 1)$.
- (d) Spočítejte střední hodnotu $E(X)$ a rozptyl $\text{var}(X)$.

Řešení:

- (a) Funkce f je nezáporná a platí, že

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^0 0 dx + \int_0^{\infty} 3e^{-3x} dx = [e^{-3x}]_0^{\infty} = 1,$$

tudíž vlastnost $\int_{\mathbb{R}} f(x) dx = 1$ je splněna. Graf hustoty f viz níže.

- (b) Příslušná distribuční funkce F_X pro veličinu X je

$$\text{pro } t \in (0, \infty): F_X(t) = \int_{-\infty}^t f(x) dx = \int_{-\infty}^0 0 dx + \int_0^t 3e^{-3x} dx = [e^{-3x}]_0^t = 1 - e^{-3t}.$$

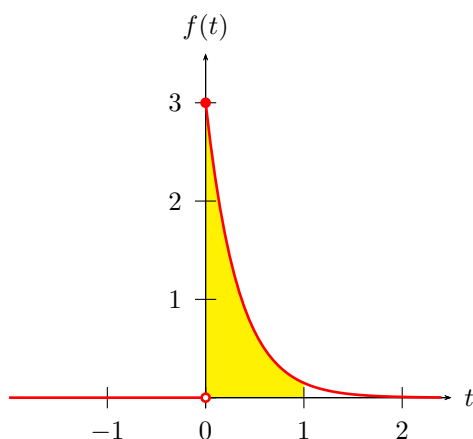
$$\text{pro } t \in (-\infty, 0): F_X(t) = \int_{-\infty}^x 0 dt = 0.$$

Graf distribuční funkce F_X viz níže.

- (c) Hledaná pravděpodobnost je

$$P(-1 \leq X \leq 1) = \int_{-1}^1 f(x) dx = \int_0^1 3e^{-3x} dx = 1 - e^{-3}.$$

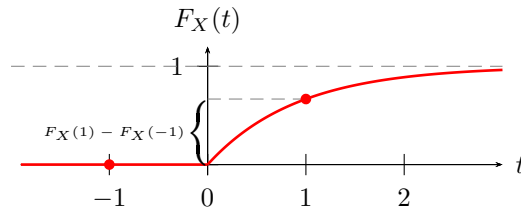
Geometrická interpretace této hodnoty je plocha pod grafem hustoty v intervalu $\langle -1, 1 \rangle$:



Lze využít také distribuční funkce:

$$P(-1 \leq X \leq 1) = \underbrace{P(X \leq 1)}_{F_X(1)} - \underbrace{P(X < -1)}_{\substack{\text{limita zleva } F_X(t) \\ \text{v bodě } -1}} \stackrel{(\text{spojitost } F_X)}{=} F_X(1) - F_X(-1) = 1 - e^{-3 \cdot 1} - 0 = 1 - e^{-3}.$$

Geometrická interpretace v tomto případě je rozdíl funkčních hodnot distribuční funkce:



(d) Použitím integrace per partes dostaneme

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} 3x e^{-3x} dx = \frac{1}{3},$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{\infty} 3x^2 e^{-3x} dx = \frac{2}{9}.$$

Střední hodnota je tudíž $E(X) = 1/3$ a rozptyl

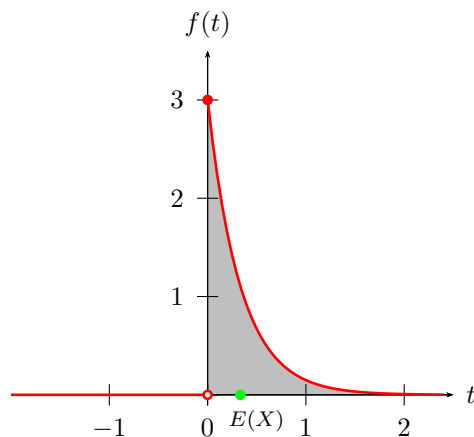
$$\text{var}(X) = E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2 = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$$

Poznamenejme ještě, že veličina X má tzv. *exponenciální* rozdělení, tj.

$$F_X(t) = \begin{cases} 0 & , t < 0 \\ 1 - e^{-\frac{t}{\tau}} & , t \geq 0 \end{cases}$$

kde $\tau > 0$ je parametr, jehož význam je, že $E(X) = \tau$. Dále ještě platí, že $\text{var}(X) = \tau^2$.

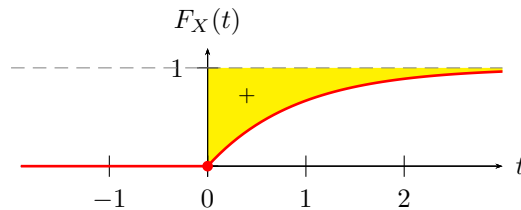
Geometrická interpretace hodnoty $E(X)$ - je to vodorovná souřadnice (zelený bod) těžiště plochy pod grafem hustoty (šedá plocha):



Ještě jedna geometrická interpretace hodnoty $E(X)$ - je to rozdíl velikosti plochy *nad* grafem distr. funkce F_X v intervalu $(0, +\infty)$ a velikosti plochy *pod* grafem distr. funkce F_X v intervalu $(-\infty, 0)$. Konkrétně:

$$E(X) = - \int_{-\infty}^0 F_X(t) dt + \int_0^{\infty} (1 - F_X(t)) dt$$

Tento vztah platí dokonce pro jakoukoliv náhodnou veličinu X (diskrétní, spojitou i smíšenou).



V tomto případě je velikost plochy pod grafem F_X na intervalu $(-\infty, 0)$ nulová, takže v obrázku nejde zvýraznit jako plocha se záporným znaménkem.

Připomeňme si **definici nezávislosti veličin**: Veličiny X_1, \dots, X_n jsou *nezávislé* \Leftrightarrow pro libovolné intervaly $I_1, \dots, I_n \subseteq \mathbb{R}$ platí, že

$$P(X_1 \in I_1, \dots, X_n \in I_n) = \prod_{i=1}^n P(X_i \in I_i)$$

konkrétně se lze (ekvivalentně) omezit jen na určité intervaly a pak máme:

Veličiny X_1, \dots, X_n jsou *nezávislé* \Leftrightarrow pro libovolné $a_1, \dots, a_n \in \mathbb{R}$ platí, že

$$P(X_1 \leq a_1, \dots, X_n \leq a_n) = \prod_{i=1}^n P(X_i \leq a_i)$$

Pro rozptyl jejich součtu pak platí:

Nechť X_1, \dots, X_n jsou nezávislé veličiny, pak

$$\text{var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{var}(X_i)$$

Příklad 5.3 Pravděpodobnost, že atlet v oddíle skočí do dálky přes 5 m, je 0.7. V oddíle je 6 atletů.

(a) Určete rozdělení náhodné veličiny

$$X = \begin{cases} 1, & \text{atlet skočil přes 5 m} \\ 0, & \text{atlet neskočil přes 5 m,} \end{cases}$$

její střední hodnotu $E(X)$ a rozptyl $\text{var}(X)$.

(b) Určete rozdělení náhodné veličiny

$Y =$ “ počet atletů v oddíle, kteří skočili přes 5 m ”

její střední hodnotu $E(Y)$ a rozptyl $\text{var}(Y)$.

(c) Jaká je pravděpodobnost, že přes 5 m skočí v oddíle alespoň 4 atleti?

Řešení:

(a) Veličina X má *alternativní* rozdělení, tj. $X \sim \text{Alt}(p)$, kde $p = P(X = 1) = 0.7$ je pravděpodobnost úspěšného pokusu a $P(X = 0) = 1 - p = 0.3$.

A dále máme

$$E(X) = \sum_{i \in \mathbb{R}} i \cdot P(X = i) = 0 \cdot (1 - p) + 1 \cdot p = p = 0.7$$

$$E(X^2) = \sum_{i \in \mathbb{R}} i^2 \cdot P(X = i) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p = 0.7$$

a

$$\text{var}(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p) = 0.7 \cdot 0.3 = 0.21 .$$

(b) Veličina Y představuje počet úspěchů při $n = 6$ nezávislých pokusech, s pravděpodobností úspěchu $p = 0.7$, takže Y má *binomické rozdělení* $\text{Binom}(n, p)$. Hodnoty veličiny Y jsou $k = 0, 1, \dots, n$ a jejich pravděpodobnosti jsou

$$P(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k} = \binom{6}{k} 0.7^k \cdot 0.3^{6-k} .$$

Pro další výpočty se hodí všimnout si, že $Y = \sum_{i=1}^n X_i$ kde

$$X_i = \begin{cases} 1, & i\text{-tý atlet skočil přes 5 m} \\ 0, & i\text{-tý atlet neskočil přes 5 m,} \end{cases}$$

jsou navzájem nezávislé náhodné veličiny a $X_i \sim \text{Alt}(p)$.

Pro střední hodnotu veličiny Y pak máme

$$E(Y) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \underbrace{E(X_i)}_p = n \cdot p = 6 \cdot 0.7 = 4.2$$

a z nezávislosti X_i pak pro rozptyl máme

$$\text{var}(Y) = \sum_{i=1}^n \underbrace{\text{var}(X_i)}_{p(1-p)} = np(1 - p) = 6 \cdot 0.7 \cdot 0.3 = 1.26 .$$

(c)

$$\begin{aligned} P(Y \geq 4) &= \sum_{k=4}^6 \binom{6}{k} 0.7^k \cdot 0.3^{6-k} = 15 \cdot 0.7^4 \cdot 0.3^2 + 6 \cdot 0.7^5 \cdot 0.3^1 + 1 \cdot 0.7^6 \cdot 1 = \\ &= 0.324135 + 0.302526 + 0.117649 = 0.74431 . \end{aligned}$$

Příklad 5.4 Pravděpodobnost narození chlapce je 0.51. Jaká je pravděpodobnost, že v dané porodnici dnes bylo nejpozději (v časovém pořadí) čtvrté narozené dítě holka?

Řešení:

Lze použít dva přístupy:

(a) Vezmeme si náhodnou veličinu

$X =$ “počet narozených chlapců před první narozenou holkou.”

Ta má *geometrické* rozdělení $\text{Geom}(p)$ s pravděpodobností $p = 1 - 0.51 = 0.49$ úspěšného pokusu (tj. narození holky). Hodnoty veličiny X jsou $k = 0, 1, 2, \dots$ s pravděpodobnostmi

$$P(X = k) = (1 - p)^k \cdot p = 0.51^k \cdot 0.49$$

Pro jednodušší výpočet si ještě označme $q = 1 - p = 0.51$. Hledaná pravděpodobnost je tedy

$$P(X < 4) = \sum_{k=0}^3 q^k \cdot (1 - q) = (1 - q) \cdot \underbrace{(1 + q + \dots + q^3)}_{\frac{1 - q^4}{1 - q}} = 1 - q^4 = 1 - 0.51^4 \doteq 0.9323.$$

(b) Vezmeme si náhodnou veličinu

$Y =$ “počet narozených holek mezi prvními 4 narozenými dětmi.”

Ta má binomické rozdělení $\text{Binom}(4, p)$, kde je opět $p = 0.49$. Hledaná pravděpodobnost je tedy

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{4}{0} p^0 \cdot (1 - p)^4 = 1 - q^4 = 1 - 0.51^4.$$