7. cvičení z STP

1. dubna 2021

Příklad 7.1 *Do pojišťovny přijdou průměrně 2 hlášení škody denně.*

(a) Jaká je pravděpodobnost, že za den přijdou alespoň 4?

(b) Jaká je pravděpodobnost, že do pojišťovny přijde nejbližší hlášení škody nejdříve třetí den?

Řešení:

(a) Máme tedy veličinu

$$X = \text{"počet hlášení za 1 den"} \ ,$$

s Poissonovým rozdělením Poiss(λ), kde $\lambda = E(X) = 2$ (jde o bezrozměrné číslo).

$$P(X \geq 4) = 1 - P(X < 4) = 1 - e^{-\lambda}\Big(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\Big) = 1 - e^{-2}\Big(1 + 2 + \frac{2^2}{2} + \frac{2^3}{3!}\Big) =$$

$$= 1 - \tfrac{19}{3}e^{-2} \doteq 1 - 0.857 = 0.143 \ .$$

(b) Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení. Exponenciální rozdělení bude mít veličina

$$Y = \text{"doba čekání na příchod dalšího hlášení".}$$

- *Pomocí exponenciálního:* Podle zadání má veličina X Poissonovo rozdělení s parametrem $\lambda = 2$. Tudíž pro náhodnou veličinu Y s exponenciálním rozdělením a střední hodnotou $E(Y) = \tau$ platí, že $\tau = \frac{T}{\lambda} = \frac{1\ den}{2} = 0.5\ dne$, kde $T = 1\ den$. Hledaná pravděpodobnost je pak

$$P(Y > 2\ dny) = 1 - P(Y \leq 2\ dny) = 1 - F_Y(2) = 1 - \Big(1 - e^{-\frac{2}{0.5}}\Big) = e^{-4} \doteq 0.018.$$

- *Pomocí Poissonova: Pomocí Poissonova:* Uvažujme veličinu

  X'= *"počet hlášení během 2 dnů"*

  Hledaná pravděpodobnost je dána jako

$$P(X' = 0) = \frac{(\lambda')^0}{0!} \cdot e^{-\lambda'} = e^{-\lambda'}$$

  kde λ' je parametr veličiny X' s Poissonovým rozdělením. K určení parametru λ' použijeme vztah mezi veličinami Y a X' a už známou hodnotu parametru veličiny X, tj. platí $\lambda' = \frac{T'}{\tau} = \lambda \cdot \frac{T'}{T}$ neboli

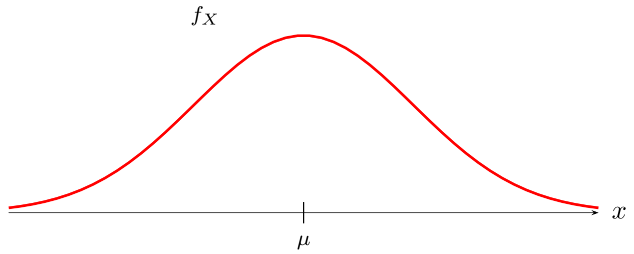
$$\frac{E(X')}{E(X)} = \frac{T'}{T}$$

  což odpovídá i výše zmíněnému požadavku, že průměrný počet událostí v časovém intervalu je úměrný jeho délce. Konkrétně tedy $\lambda' = 2 \cdot \frac{2\ dny}{1\ den} = 4$, kde $T' = 2\ dny$, a hledaná pravděpodobnost je pak

$$P(X' = 0) = e^{-4} \doteq 0.018.$$

Poznámky k normálnímu rozdělení:
Veličina X má *normální* (neboli Gaussovo) rozdělení $N(\mu, \sigma^2)$ (kde $\mu \in \mathbb{R}$ a $\sigma > 0$), jestliže má hustotu

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{pro} \quad x \in \mathbb{R} \ .$$



Je to tedy spojité rozdělení, $E(X) = \mu$, $\mathrm{var}(X) = \sigma^2$ a oborem hodnot veličiny X je celá reálná osa. Všimněme si ještě, že hustota f_X je symetrická vzhledem ke středu μ a proto platí $F_X(\mu) = \frac{1}{2}$.

Toto rozdělení je limitním rozdělením, které aproximuje součty nezávislých stejně (nebo podobně) rozdělených veličin (více později v Centrální limitní větě). Typicky se tedy objevuje u veličin, jejichž hodnoty jsou ovlivněny mnoha drobnými odchylkami (např. u chyb měření, výšky člověka apod.)

U zmíněné výšky člověka (která může být samozřejmě jen kladná) nebo u veličin s hodnotami omezenými na nějaký interval, je přesto použití normálního rozdělení (které může nabývat libovolných hodnot) přiměřené. Je to tím, že u dané veličiny Y předpokládáme aproximaci pomocí normálního rozdělení obvykle jen ve vhodném okolí kolem střední hodnoty $\mu := E(Y)$. Je to podobná situace, jako když aproximujeme funkci pomocí jejího Taylorova polynomu v okolí daného bodu.

Přesněji to vystihuje toto tvrzení:

Věta: Nechť Y je veličina s hustotou f_Y, střední hodnotou μ a rozptylem $\sigma^2 \neq 0$. Nechť $X \sim N(\mu, \sigma^2)$. Jestliže se hustoty f_X a f_Y rovnají na nějakém intervalu $(a, b) \subseteq \mathbb{R}$ takovém, že $\mu \in (a, b)$ a pokud $F_Y(\mu) = \frac{1}{2}$, pak

$$F_Y(t) = F_X(t) \quad \text{pro všechna} \quad t \in (a, b) \ .$$

Značení: Pro náhodnou veličinu X s konečnou střední hodnotou a konečným rozptylem, položme

$$\mathrm{norm}(X) := \frac{X - E(X)}{\sqrt{\mathrm{var}(X)}} \ .$$

Speciálně tedy vidíme, že $E\big(\mathrm{norm}(X)\big) = 0$ a $\mathrm{var}\big(\mathrm{norm}(X)\big) = 1$.

Platí: Pro takovouto veličinu X a konstanty $a > 0$ a $b \in \mathbb{R}$ je

$$\mathrm{norm}(aX + b) = \mathrm{norm}(X) \ .$$

Důležité vlastnosti normálního rozdělení:

- $X \sim N(\mu, \sigma^2) \ \Rightarrow \ \mathrm{norm}(X) = \frac{X-\mu}{\sigma} \sim N(0, 1)$ (je to tzv. normované normální rozdělení s hodnotami v tabulkách) dist. funkce pro $N(0, 1)$ se značí Φ.

  V tomto případě pak máme $F_X(t) = P(X \leq t) = P\Big(\underbrace{\frac{X-\mu}{\sigma}}_{=\mathrm{norm}(X)} \leq \frac{t-\mu}{\sigma} \Big) = \Phi\Big(\frac{t-\mu}{\sigma}\Big)$ pro všechna $t \in \mathbb{R}$.

- hustota $f_{N(0,1)}$ je sudá funkce $\Rightarrow$ $\Phi(t) + \Phi(-t) = 1$ pro všechna $t \in \mathbb{R}$.
- Nechť $X_i \sim N(\mu_i, \sigma_i^2)$, pro $i = 1, 2$, jsou nezávislé. Pak $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (tj. speciálně součet nezávislých normálních rozdělení je zase normální.)

Pro vybraná čísla $t \geq 0$ se dají hodnoty Φ najít ve statistických tabulkách. Pro záporná čísla si pak pomůžeme vztahem $\Phi(-t) = 1 - \Phi(t)$.

Příklad 7.2 *Výška dětí v* 1. *třídě je náhodná veličina* $X \sim N(130\,\mathrm{cm},\ 36\,\mathrm{cm}^2)$. *Jaká je pravděpodobnost, že náhodně vybrané dítě bude*

(a) větší než 136 *cm,*

(b) menší než 118 *cm,*

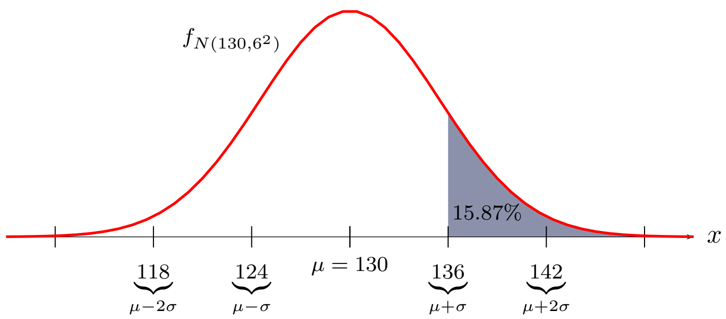
(c) mít výšku mezi 127 *a* 133 *cm?*

Řešení:
Nyní tedy máme $X \sim N(130,\ 36)$. Pro jednodušší zápis si ještě označme $Z := \mathrm{norm}(X)$.

(a)

$$P(X > 136) = P\left(\underbrace{\frac{X - 130}{\sqrt{36}}}_{Z} > \underbrace{\frac{136 - 130}{\sqrt{36}}}_{1}\right) = P(Z > 1) = 1 - P(Z \leq 1) =$$

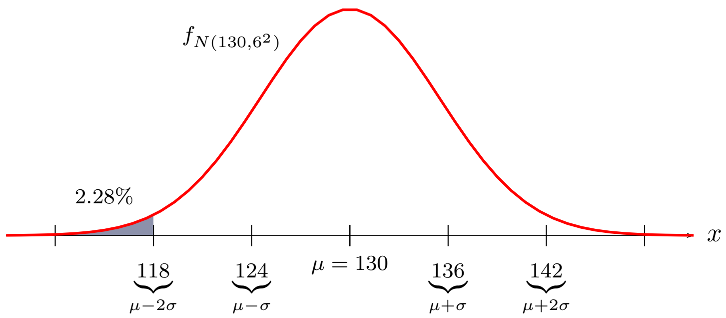
$$= 1 - \Phi(1) \doteq 1 - 0.8413 = 0.1587\ .$$



(b)

$$P(X < 118) = P\left(\frac{X - 130}{\sqrt{36}} < \frac{118 - 130}{\sqrt{36}}\right) = P(Z < -2) = \Phi(-2) =$$

$$= 1 - \Phi(2) \doteq 1 - 0.9772 = 0.0228\ .$$



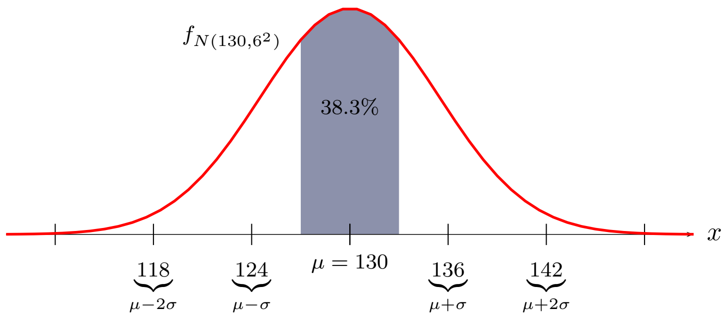
(c)

$$P(127 < X < 133) = P\left(\frac{127-130}{\sqrt{36}} < \frac{X-130}{\sqrt{36}} < \frac{133-130}{\sqrt{36}}\right) =$$

$$= P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z \leq -0.5) =$$

$$= \Phi(0.5) - \Phi(-0.5) = \Phi(0.5) - \Big(1 - \Phi(0.5)\Big) =$$

$$= 2 \cdot \Phi(0.5) - 1 \doteq 2 \cdot 0.6915 - 1 = 0.383\ .$$



Poznamenejme ještě, že hodnoty výšek, které nás zajímaly (tj. 136 cm, 118 cm atd.) se pohybují celkem blízko střední hodnoty $E(X) = 130$ cm, takže předpoklad o normálnosti rozdělení X byl přiměřený.

- -

Výpočty si můžeme i urychlit přímým vzorcem $F_X(t) = \Phi(\frac{t-\mu}{\sigma})$, kde $\mu = 130$ cm a $\sigma = 6$ cm:

(a) $P(X > 136) = 1 - F_X(136) = 1 - \Phi(\frac{136-130}{6}) = \ldots \doteq 0.1587$

(b) $P(X < 118) = F_X(118) = \Phi(\frac{118-130}{6}) = \ldots \doteq 0.0228$

(c) $P(127 < X < 133) = F_X(133) - F_X(127) = \Phi(\frac{133-130}{6}) - \Phi(\frac{127-130}{6}) = \ldots \doteq 0.383$

Příklad 7.3 *Oštěpařky Anna a Barbora mají střední hodnoty hodů po řadě* 67 *m a* 75 *m a směrodatné odchylky* 6 *m a* 3 *m. Předpokládejme nezávislá normální rozdělení. Odhadněte pravděpodobnost, že při jednom hodu hodí Anna dál.*

Řešení:
Náhodná veličina

$A =$"*délka hodu Anny*"

má rozdělení $N(67, 6^2)$ a veličina

$B =$"*délka hodu Barbory*"

má rozdělení $N(75, 3^2)$.

Zajímá nás $P(A > B) = P(A - B > 0)$. Protože veličiny A a B jsou nezávislé, tak veličina $Z := A - B$ má také normální rozdělení, a sice

$$Z \sim N(67 - 75, 6^2 + 3^2) = N(-8, 45) \ .$$

Takže

$$P(A > B) = P(Z > 0) = P\left(\underbrace{\frac{Z - (-8)}{\sqrt{45}}}_{\text{norm}(Z)} > \frac{0 - (-8)}{\sqrt{45}}\right) = 1 - P\left(\text{norm}(Z) \leq \frac{8}{\sqrt{45}}\right) =$$

$$= 1 - \Phi\left(\tfrac{8}{\sqrt{45}}\right) \doteq 1 - \Phi\left(1.1926\right) \doteq 1 - 0.883 = 0.117 \, .$$

POZOR! Zatímco střední hodnota je lineární zobrazení, tak rozptyl se chová jinak! Konkrétně je to takto:

Nechť X a Y jsou veličiny se střední hodnotou a konečným rozptylem. Pak

- $E(X \pm Y) = E(X) \pm E(Y)$
- $\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y) \pm 2 \cdot \text{cov}(X, Y) \ (\geq 0)$

Zde $\text{cov}(X, Y)$ je tzv. kovariance (viz poznámky níže). Speciálně, pokud X a Y jsou nezávislé, je $\text{cov}(X, Y) = 0$. Máme tedy:

- X a Y nezávislé $\Rightarrow$ $\text{var}(X \pm Y) = \text{var}(X) + \text{var}(Y)$

Tedy v tomto případě se rozptyly VŽDY sčítají!

Příklad 7.4 *Délka hrany krychle je náhodná veličina* $X \sim \text{Ro}(1, 2)$. *Určete distribuční funkci náhodné veličiny Y popisující plochu povrchu této krychle.*

Řešení:
Máme veličiny

$X =$ "*délka hrany krychle*"

$Y =$ "*plocha povrchu krychle*"

takže $Y = 6 \cdot X^2$ a pro distribuční funkci náhodné veličiny Y dostáváme

$$F_Y(y) = P(Y \leq y) = P(6X^2 \leq y) = P(X^2 \leq \tfrac{y}{6}) = \begin{cases} P\left(|X| \leq \sqrt{\tfrac{y}{6}}\,\right) = F_X\left(\sqrt{\tfrac{y}{6}}\,\right) & , \ y \geq 0 \\ P(\emptyset) = 0 & , \ y < 0 \ . \end{cases}$$

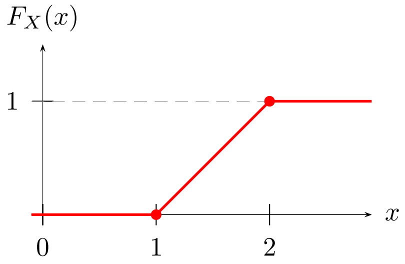
kde jsme využili toho, že obor hodnot pro X je $\langle 1, 2\rangle$, tedy $X \geq 0$ a speciálně tak platí, že $|X| = X$.

Teď si už si jen vyjádříme F_X a dosadíme:

Pro veličinu $X \sim \mathrm{Ro}(1,2)$ je její hustota $f_X(x) = \begin{cases} 1, & 1 \leq x \leq 2 \\ 0, & \text{jinak} \end{cases}$ a distribuční funkce

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \begin{cases} 0, & x < 1 \\ x - 1, & 1 \leq x \leq 2 \\ 1, & x > 2 \end{cases}$$

s grafem



Do F_X (správně!) dosadíme $x = \sqrt{\frac{y}{6}}$ (pro $y \geq 0$) a přepíšeme podmínky pro y:

$$1 \leq \sqrt{\frac{y}{6}} \leq 2 \quad \Leftrightarrow \quad 1 \leq \frac{y}{6} \leq 4 \quad \Leftrightarrow \quad 6 \leq y \leq 24$$

(zbylé podmínky jsou podobné) a dostaneme tak

$$F_Y(y) = \begin{cases} 0, & y < 6 \\ \sqrt{\frac{y}{6}} - 1, & 6 \leq y \leq 24 \\ 1, & y > 24 \end{cases}$$

s grafem