

7. cvičení z STP

30 března. - 3. dubna 2020

Příklad 7.1 Průměrný počet zákazníků během dne v první prodejně je 20, ve druhé prodejně 25. Předpokládáme, že oba počty se řídí Poissonovým rozdělením. Odvoďte rozdělení počtu zákazníků v obou prodejnách dohromady.

Řešení:

Označme si veličiny

$X =$ “počet zákazníků během dne v 1. prodejně”

$Y =$ “počet zákazníků během dne v 2. prodejně”

$Z =$ “počet zákazníků během dne v obou prodejnách dohromady”

kde X a Y budeme přirozeně pokládat za nezávislé. Máme

$$X \sim \text{Pois}(\lambda), \text{ kde } \lambda = E(X) = 20$$

$$Y \sim \text{Pois}(\mu), \text{ kde } \mu = E(Y) = 25$$

a

$$P(X = i) = \frac{\lambda^i}{i!} \cdot e^{-\lambda} \quad \text{a} \quad P(Y = j) = \frac{\mu^j}{j!} \cdot e^{-\mu}$$

Jelikož $Z = X + Y$ a případ $X + Y = k$ se rozloží na disjunktní možnosti $(X, Y) = (i, k - i)$ pro $i = 0, 1, \dots, k$, dostáváme

$$\begin{aligned} P(Z = k) &= P(X + Y = k) = \\ &= \sum_{i=0}^k P(X = i, Y = k - i) \stackrel{\text{(nezávislost)}}{=} \sum_{i=0}^k P(X = i) \cdot P(Y = k - i) = \\ &= \sum_{i=0}^k \frac{\lambda^i}{i!} e^{-\lambda} \cdot \frac{\mu^{k-i}}{(k-i)!} e^{-\mu} = e^{-(\lambda+\mu)} \sum_{i=0}^k \frac{1}{k!} \cdot \underbrace{\frac{k!}{i!(k-i)!}}_{\binom{k}{i}} \cdot \lambda^i \cdot \mu^{k-i} = \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^k \binom{k}{i} \lambda^i \cdot \mu^{k-i} \stackrel{\text{(binom. věta)}}{=} \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k \end{aligned}$$

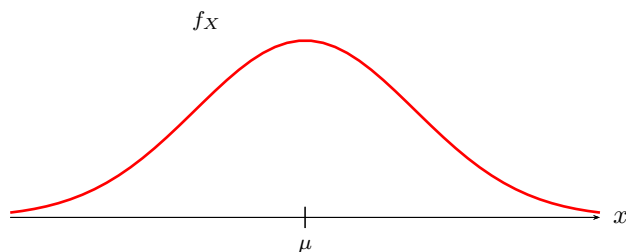
tj. $Z \sim \text{Pois}(\lambda + \mu)$ neboli $Z \sim \text{Pois}(45)$.

Poznámka: Představme si, že jednotlivým prodejnám přiřadíme (čistě účelově) nějaké velikosti a a b (např. velikost plochy prodejny v m^2) tak, aby průměrný počet zákazníků na jednotku plochy byl v obou prodejnách stejný, tj. $\frac{\lambda}{a} = \frac{\mu}{b}$. Veličiny X a Y pak můžeme chápat jako počty událostí v intervalech délky a a b , přičemž v obou intervalech je “hustota událostí” stejná. Při tomto přístupu bude veličina Z představovat počet událostí v intervalu délky $a + b$, takže Poissonovo rozdělení se pak dá skutečně očekávat.

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

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



Je to tedy spojité rozdělení, $E(X) = \mu$, $\text{var}(X) = \sigma^2$ 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

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

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

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

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

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

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

Důležité vlastnosti normálního rozdělení:

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

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

- hustota $f_{N(0,1)}$ je sudá funkce $\Rightarrow \Phi(t) + \Phi(-t) = 1$ pro všechna $t \in \mathbb{R}$.
- Necht' $X_i \sim N(\mu_i, \sigma_i^2)$, 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 $X \sim N(130 \text{ cm}, 36 \text{ cm}^2)$. Jaká je pravděpodobnost, že náhodně vybrané dítě bude

- větší než 136 cm,
- menší než 118 cm,
- mít výšku mezi 127 a 133 cm?

Řešení:

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

(a)

$$\begin{aligned} P(X > 136) &= P\left(\underbrace{\frac{X - 130}{\sqrt{36}}}_Z > \underbrace{\frac{136 - 130}{\sqrt{36}}}_1\right) = P(Z > 1) = 1 - P(Z \leq 1) = \\ &= 1 - \Phi(1) \doteq 1 - 0.8413 = 0.1587. \end{aligned}$$

(b)

$$\begin{aligned} P(X < 118) &= P\left(\frac{X - 130}{\sqrt{36}} < \frac{118 - 130}{\sqrt{36}}\right) = P(Z < -2) = \Phi(-2) = \\ &= 1 - \Phi(2) \doteq 1 - 0.9772 = 0.0228. \end{aligned}$$

(c)

$$\begin{aligned} P(127 < X < 133) &= P\left(\frac{127 - 130}{\sqrt{36}} < \frac{X - 130}{\sqrt{36}} < \frac{133 - 130}{\sqrt{36}}\right) = \\ &= P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z \leq -0.5) = \\ &= \Phi(0.5) - \Phi(-0.5) = \Phi(0.5) - (1 - \Phi(0.5)) = \\ &= 2 \cdot \Phi(0.5) - 1 \doteq 2 \cdot 0.6915 - 1 = 0.383. \end{aligned}$$

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

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

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

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

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

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

$A = \text{“délka hodu Anny”}$

má rozdělení $N(67, 6^2)$ a veličina

$B = \text{“délka hodu Barbory”}$

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

$$\begin{aligned} P(A > B) &= P(Z > 0) = P\left(\underbrace{\frac{Z - (-8)}{\sqrt{45}}}_{\text{norm}(Z)} > \frac{0 - (-8)}{\sqrt{45}}\right) = 1 - P\left(\text{norm}(Z) \leq \frac{8}{\sqrt{45}}\right) = \\ &= 1 - \Phi\left(\frac{8}{\sqrt{45}}\right) \doteq 1 - \Phi(1.1926) \doteq 1 - 0.883 = 0.117 . \end{aligned}$$

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

$X = \text{“délka hrany krychle”}$

$Y = \text{“plocha povrchu krychle”}$

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

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

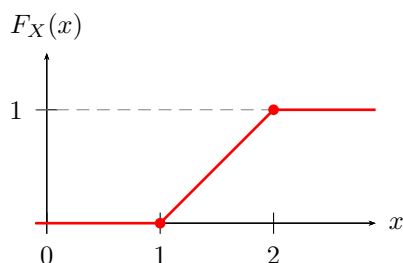
kde jsme využili toho, že obor hodnot pro X je $(1, 2)$, tedy $X \geq 0$ a speciálně tak platí, že $|X| = X$.

Tedy si už si jen vyjádříme F_X a dosadíme:

Pro veličinu $X \sim \text{Ro}(1, 2)$ 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



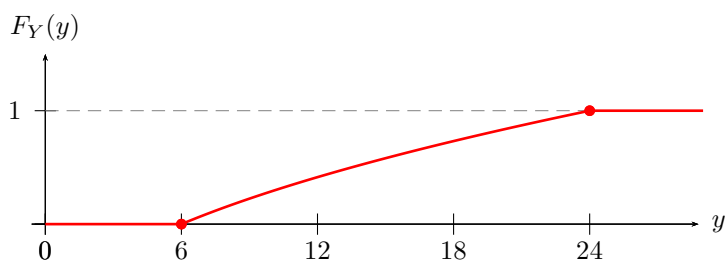
Do F_X (správně!) dosadíme $x = \sqrt{\frac{y}{6}}$ (pro $y \geq 0$) a přepíšeme podmínky pro y :

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

(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



Připomenutí: Kovariance pro X a Y je definována jako

$$\text{cov}(X, Y) := E\left((X - EX) \cdot (Y - EY)\right) = \dots = E(XY) - E(X) \cdot E(Y).$$

Speciálně je $\text{cov}(X, X) = \text{var}(X)$.

Kovariance $\text{cov}(\cdot, \cdot)$ má tyto vlastnosti (X, Y, Z jsou veličiny, $a, b, c, d \in \mathbb{R}$ jsou konstanty):

- je lineární v každé složce zvlášť (tj. je bilineární), tedy:

$$\text{cov}(aX + bY, Z) = a \cdot \text{cov}(X, Z) + b \cdot \text{cov}(Y, Z)$$

$$\text{cov}(Z, aX + bY) = a \cdot \text{cov}(Z, X) + b \cdot \text{cov}(Z, Y)$$

- symetrická, tj. $\text{cov}(X, Y) = \text{cov}(Y, X)$
- pozitivně semi-definitní, tj. $\text{cov}(X, X) \geq 0$, kde navíc platí, že:
 $\text{cov}(X, X) = 0 \Rightarrow \exists \alpha \in \mathbb{R}$, že $P(X = \alpha) = 1$ (neboli: X odpovídá konstantní veličině)
- $\text{cov}(X + c, Y + d) = \text{cov}(X, Y)$.

Platí: X a Y jsou nezávislé veličiny $\implies \text{cov}(X, Y) = 0$.

(POZOR: Opačná implikace obecně neplatí!!)

Příklad 7.5 *Nechť $X \sim \text{Ro}(0, 2)$ a $Y = X^2 + 1$.*

(a) *Sestrojte distribuční funkci náhodné veličiny Y .*

(b) *Spočtěte $\text{cov}(X, Y)$.*

(c) *Rozhodněte, zda jsou X a Y nezávislé a proč.*

Řešení:

- (a) Distribuční funkce F_Y se dá získat pomocí distribuční funkce F_X . Před výpočtem si ještě uvědomme, že $X \geq 0$ (protože její obor hodnot je $\langle 0, 2 \rangle$). Speciálně tedy $|X| = X$. Pro distribuční funkci náhodné veličiny Y máme

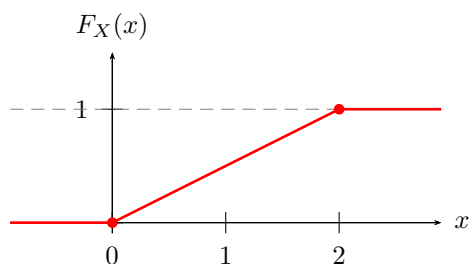
$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 + 1 \leq y) = \\ &= P(X^2 \leq y - 1) = \begin{cases} P(|X| \leq \sqrt{y-1}) = F_X(\sqrt{y-1}) & , y - 1 \geq 0 \\ P(\emptyset) = 0 & , y - 1 < 0 . \end{cases} \end{aligned}$$

Ted' už si stačí jen vyjádřit F_X a dosadit.

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

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

s grafem



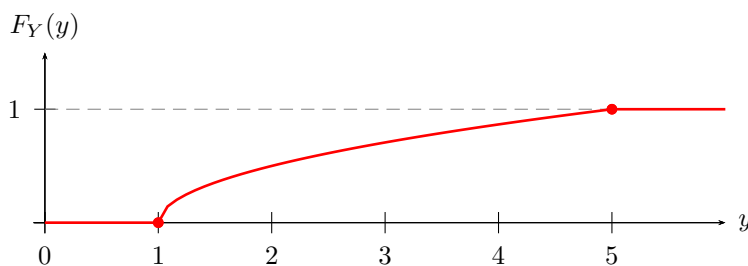
Do F_X (správně!) dosadíme $x = \sqrt{y-1}$ (pro $y-1 \geq 0$) a přepíšeme podmínky pro y :

$$0 \leq \sqrt{y-1} \leq 2 \Leftrightarrow 0 \leq y-1 \leq 4 \Leftrightarrow 1 \leq y \leq 5$$

(zbylé podmínky jsou podobné) a dostaneme tak

$$F_Y(y) = \begin{cases} 0, & y < 1 \\ \frac{\sqrt{y-1}}{2}, & 1 \leq y \leq 5 \\ 1, & y > 5 \end{cases}$$

s grafem



(b) Kovarianci vypočteme ze vztahu

$$\text{cov}(X, Y) = \text{cov}(X, X^2 + 1) = \text{cov}(X, X^2) = E(X \cdot X^2) - E(X) \cdot E(X^2)$$

Stačí si tedy pro $n \geq 1$ zjistit

$$E(X^n) = \int_{-\infty}^{\infty} x^n \cdot f_X(x) dx = \int_0^2 \frac{x^n}{2} dx = \left[\frac{x^{n+1}}{2(n+1)} \right]_0^2 = \frac{2^n}{n+1}$$

a dosazením dostaneme

$$\text{cov}(X, Y) = 2 - 1 \cdot \frac{4}{3} = \frac{2}{3}.$$

(c) Protože $\text{cov}(X, Y) \neq 0$, jsou veličiny X a Y závislé. Toto zjištění ovšem můžeme udělat i bez výpočtu kovariance:

Velichiny X a Y jsou funkčně propojené, takže stačí najít podmínky, které naráz nemůžou splnit, ale jednotlivě, s nenulovými pravděpodobnostmi, ano. Z předchozích úprav už víme, že pro $1 < y < 5$ platí

$$Y \leq y \Leftrightarrow X^2 + 1 \leq y \Leftrightarrow X \leq \sqrt{y-1}$$

a pravděpodobnosti těchto jevů jsou (z tvaru F_X a F_Y) ostře mezi 0 a 1. Takže např. z volby $y = 2$ dostaneme, že

$$Y \leq 2 \Leftrightarrow X \leq 1$$

takže

$$P(\underbrace{Y \leq 2, X > 1}_{\emptyset}) = 0 \neq \underbrace{P(Y \leq 2)}_{F_Y(2)=0.5} \cdot \underbrace{P(X > 1)}_{1-F_X(1)=0.5}$$

z čehož plyne, že veličiny X a Y jsou závislé.

Poznámka: Jak se dá očekávat, pokud jedna veličina závisí svými hodnotami na druhé, nejspíš nezávislé nebudou. Výjimkou je jen jeden případ a celá situaci se dá popsat takto:

Věta: Nechtě X a $h(X)$ jsou obě náhodné veličiny, kde $h : \mathbb{R} \rightarrow \mathbb{R}$ je borelovská funkce (např. spojitá). Pak X a $h(X)$ jsou nezávislé veličiny právě jen pokud

- $h(X)$ je konstantní veličina (přesněji: ex. $c \in \mathbb{R}$, že $P(h(X) = c) = 1$).

Příklad:

- (1) Pro představu, kdy může třeba nastat případ nezávislosti v předchozí větě, si vezmeme $X \sim \text{Ro}(1, 2)$ a funkci $h(x) = \max\{x, 3\}$. Vidíme, že ani veličina X ani funkce h nejsou konstantní, ale jejich složení $Y = h(X) = \max\{X, 3\} = 3$ už konstanta je. No a veličiny X a $Y = 3$ už samozřejmě nezávislé jsou.
- (2) Ukažme si příklad dvou veličin tvaru X a $Y = h(X)$, které budou závislé a přitom bude platit $\text{cov}(X, Y) = 0$.

Stačí zvolit $X \sim \text{Ro}(-1, 1)$ a $Y = X^2$. Podle věty výše X a Y budou závislé, protože Y není konstanta.

(Proč Y není (skoro všude) konstantní: protože X má spojitě rozdělení, tak je $P(X = a) = 0$ pro každé $a \in \mathbb{R}$. Odsud ihned máme, že pro $b \in \mathbb{R}$ je $P(Y = b) = P(X^2 = b) = 0$. Tedy Y nemůže splňovat $P(Y = c) = 1$ pro nějaké $c \in \mathbb{R}$ a tudíž nemůže být konstanta.)

Současně s tím vidíme, že $E(X) = \frac{-1+1}{2} = 0$ a dále, že

$$E(X \cdot Y) = E(X \cdot X^2) = \int_{-\infty}^{\infty} t^3 \cdot f_X(t) dt = \int_{-1}^1 \underbrace{t^3 \cdot \frac{1}{2}}_{\text{lichá funkce}} dt = 0$$

A dále, střední hodnota $E(Y) = E(X^2)$ existuje, i když už nulová nebude.

Tím tedy dostaneme

$$\text{cov}(X, Y) = \underbrace{E(XY)}_{=0} - \underbrace{E(X)}_{=0} \cdot E(Y) = 0.$$