7. cvičení z STP

28. března - 1. dubna 2022

Poznámky k Poissonovu rozdělení: Pro veličinu

$$X = \text{“ } \textit{počet událostí během intervalu délky } T \text{”}$$

kde interval je obvykle časový (ale může být i délkový nebo měřený nějakou jinou jednotkou), můžeme předpokládat *Poissonovo* rozdělení

$$P(X = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \quad \text{kde} \quad k = 0, 1, 2, \dots$$

s (bezrozměrným) parametrem $\lambda > 0$, pokud jsou splněny následující podmínky

- počet událostí může nabývat libovolných (konečných) hodnot.
- jednotlivé události jsou nezávislé a nenastávají současně (lze je časově oddělit),
- *průměrný* počet událostí v libovolném časovém podintervalu je úměrný pouze časové délce tohoto podintervalu a ne jeho umístění v původním intervalu (tj. lze říct, že střední četnosti událostí za jednotku času se s průběhem doby nemění),

V praxi půjde např. o příchod zákazníka do fronty, chytání ryb, průjezd aut atd. a to během nějaké předem určené doby.

Parametr λ pak představuje střední hodnotu (tj. $E(X) = \lambda$) protože:

$$E(X) = \sum_{k=0}^{\infty} k \cdot P(X = k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \lambda e^{-\lambda} \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}}_{=e^{\lambda}} = \lambda$$

Poissonovo rozdělení je většinou spíše limitní případ a používá se jako aproximace binomického rozdělení $\mathrm{Bi}(n, p)$, u kterého sice neznáme n a p, ale víme, že n je (dostatečně) velké a známe střední hodnotu dané veličiny (viz níže).

V praxi tedy můžeme podmínky Poissonova rozdělení přibližně zajistit pokud budou události pocházet z velkého počtu nezávislých zdrojů (z každého jen jednou) a podmínky se během měření nebudou měnit (tj. nebude se náhle měnit “okamžitá střední četnost události”).

Poznámka: Ke tvaru Poissonova rozdělení se můžeme dostat pomocí binomického rozdělení (s využitím výše uvedených předpokladů) takto:

Časový interval délky T si rozdělíme na n dílků, a budeme předpokládat, že v každém se může stát maximálně jedna událost se stejnou pravděpodobností p_n. Dostaneme tak binomické rozdělení veličiny

$$X_n = \text{“}\textit{počet událostí v daném časovém úseku délky } T \textit{ rozděleném na } n \textit{ dílků}\text{”}$$

se střední hodnotou $\lambda = E(X_n) = n \cdot p_n$, kterou si vezmeme jako pevnou (neboli vlastně položíme $p_n := \frac{\lambda}{n}$). Tedy X_n má binomické rozdělení $\mathrm{Bi}(n, p_n)$. Spočítáme si teď limitu (pro pevně zvolené k):

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} = \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{\lambda^k}{k!} \lim_{n\to\infty} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to\, e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to\, 1} \underbrace{\prod_{i=0}^{k-1} \left(\frac{n-i}{n}\right)}_{\to\, 1} = \frac{\lambda^k}{k!} \cdot e^{-\lambda} \,.$$

Připomeňme si, že pojem nezávislosti pro jevy umožňuje přirozeně definovat nezávislost veličin :

Definice: Veličiny X a Y jsou *nezávislé* $\Leftrightarrow P(X \in I, Y \in J) = P(X \in I) \cdot P(Y \in J)$ pro libovolné intervaly $I, J \subseteq \mathbb{R}$.

Příklad 7.1 *Na látce (pevné šířky) je průměrně jeden kaz na* $10\,m$ *délky. Předpokládáme, že počet kazů se řídí Poissonovým rozdělením. Jaká je pravděpodobnost, že na* $50\,m$ *délky látky bude*

(a) přesně 10 *kazů,*

(b) maximálně 3 *kazy,*

(c) přesně 5 *kazů, z toho* 4 *na prvních* $20\,m$*?*

Řešení:
Označme si náhodnou veličinu

$$X = \text{„ počet kazů na } 50\,m \text{ délky látky"}$$

Pak $X \sim \mathrm{Poiss}(\lambda)$ a pravděpodobnosti hodnot jsou

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{pro} \quad k = 0, 1, \dots$$

kde $E(X) = \lambda$. Pro veličinu $\widetilde{X}$, označující počet kazů na $10\,\mathrm{m}$ délky, předpokládáme také Poissonovo rozdělení se střední hodnotou $E(\widetilde{X}) = 1$. Protože střední hodnota má být úměrná délce intervalu dostaneme

$$\frac{E(X)}{E(\widetilde{X})} = \frac{50\,\mathrm{m}}{10\,\mathrm{m}} = 5 \quad \Rightarrow \quad \lambda = E(X) = 5$$

Tudíž

(a) $P(X = 10) = \frac{5^{10}}{10!}\, e^{-5} \doteq 0.0181.$

(b) $P(X \leq 3) = e^{-5}\left(\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!}\right) = e^{-5}(1 + 5 + \frac{25}{2} + \frac{125}{6}) = \frac{118}{3} e^{-5} \doteq 0.265.$

(c) Označme si veličiny

$$X_1 = \text{„ počet kazů na prvních } 20\,m \text{ délky látky"}$$

$$X_2 = \text{„ počet kazů na zbylých } 30\,m \text{ délky látky"} .$$

Ty budou nezávislé. Analogicky k předešlému budeme mít

$$P(X_1 = k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1} \quad \text{a} \quad P(X_2 = k) = \frac{\lambda_2^k}{k!} e^{-\lambda_2},$$

kde $\lambda_1 = 2$ a $\lambda_2 = 3$. Hledaná pravděpodobnost je (díky nezávislosti X_1 a X_2) tedy

$$P(X_1 = 4,\ X_2 = 1) \overset{(\text{nezáv. } X_i)}{=} P(X_1 = 4) \cdot P(X_2 = 1) = \frac{2^4}{4!} e^{-2} \cdot \frac{3^1}{1!} e^{-3} = 2e^{-5} \doteq 0.01348.$$

Poznámky k exponenciálnímu rozdělení:
Exponenciální rozdělení popisuje pravděpodobnost veličiny

$$Y = \text{“}\textit{doba mezi dvěma následnými výskyty události}\text{”},$$

v systému, který nemá paměť na předchozí události. Tedy to, co se stane od určitého okamžiku, nezávisí na tom, co bylo předtím. V praxi jde např. o dobu, za kterou se porouchá zařízení, které se ”neopotřebovává” (např. polovodičové součástky), nebo o dobu radioaktivního rozpadu atd. Exponenciální rozdělení je jediné, které splňuje následující rovnici:

$$P(Y > s + t | Y > s) = P(Y > t)$$

pro všechna $s, t > 0$. Rovnice vyjadřuje to, že pravděpodobnost, že zařízení bude bez poruchy pracovat alespoň t hodin, je stejná v případě, že jsme jej právě zapnuli (pravá strana rovnice), jako za předpokladu, že předtím už bez poruchy pracovalo s hodin (levá strana rovnice).

Exponenciální rozdělení $\mathrm{Exp}(\frac{1}{\tau})$ je charakterizováno parametrem $\tau > 0$ (s fyzikálním rozměrem času), který představuje střední dobu čekání, tedy $E(Y) = \tau$ a dále ještě platí $\mathrm{var}(Y) = \tau^2$. Distribuční funkce pro $Y \sim \mathrm{Exp}(\frac{1}{\tau})$ je

$$F_Y(t) = \begin{cases} 0 & , \; t \leq 0 \\ 1 - e^{-\frac{t}{\tau}} & , \; t > 0 \; . \end{cases}$$

Doplnění: Diskrétní analogií exponenciálního rozdělení je *geometrické* rozdělení, které neměří čas spojitě ale pouze diskrétně. Jak už víme, je to rozdělení veličiny

$$\widetilde{Y} = \text{"}\textit{počet neúspěšných pokusů než nastane první úspěch}\text{"},$$

např. v situaci, že se chceme trefit míčem do koše atd. Hodnoty $\widetilde{Y}$ jsou $\{0, 1, 2, \dots\}$. Je to opět jediné takové diskrétní rozdělení splňující analogickou rovnici:

$$P(\widetilde{Y} > k + n | \widetilde{Y} > n) = P(\widetilde{Y} > k)$$

pro všechna $k, n \in \mathbb{N}_0$ s podobným významem jako u exponenciálního rozdělení.

Souvislost mezi exponenciálním a Poissonovým rozdělením:

Nechť

$$Y = \text{“}\textit{doba čekání na událost}\text{”}$$

je veličina s exponenciálním rozdělením. Pak veličina

$$X = \text{“}\textit{počet událostí během doby } T\text{”}$$

má Poissonovo rozdělení a platí

$$E(Y) = \frac{T}{E(X)}$$

kde doba T je vyjádřená ve stejných jednotkách, jaké má veličina Y. Neboli

$$\text{“}\textit{střední doba čekání}\text{”} = \frac{\text{“}\textit{délka intervalu}\text{”}}{\text{“}\textit{střední počet událostí v tomto intervalu}\text{”}} \; .$$

Příklad 7.2 *Na zákaznickou linku přichází průměrně 12 hovorů za hodinu. Ženy volají dvakrát častěji než muži. Doba čekání na hovor má exponenciální rozdělení.*

(a) Jaká je pravděpodobnost, že nejbližší hovor přijde nejdříve za 10 minut?

(b) Určete čas t takový, že nejbližší hovor přijde nejdříve za t minut s pravděpodobností 0.7.

(c) Jaká je pravděpodobnost, že v pracovní době mezi 10:00 a 10:40 zavolají maximálně 3 zákazníci a všechny budou ženy?

Řešení:

Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení.

(a)

- *Pomocí exponenciálního:* Podle věty má veličina

$$X = \text{“}\textit{počet hovorů během doby } 60 \ min\text{”}$$

Poissonovo rozdělení s parametrem $\lambda = E(X) = 12$. Tudíž náhodná veličina

$$Y = \text{“}\textit{doba čekání na hovor}\text{”} \ [\textit{v minutách}]$$

má exponenciální rozdělení se střední hodnotou $\tau = E(Y) = \frac{60 \ min}{12} = 5 \ min$. Její distribuční funkce je tedy

$$F_Y(t) = \begin{cases} 1 - e^{-t/5} & , \ t > 0 \\ 0 & , \ t \leq 0. \end{cases}$$

Hledaná pravděpodobnost pak je

$$P(Y > 10 \ min) = 1 - P(Y \leq 10 \ min) = 1 - F_Y(10) = 1 - \left(1 - e^{-\frac{10}{5}}\right) = e^{-2} \doteq 0.1353 \ .$$

Pravděpodobnost také můžeme spočítat pomocí hustoty f_Y veličiny Y:

$$f_Y(t) = \begin{cases} \frac{1}{5} e^{-t/5} & , \ t > 0 \\ 0 & , \ t \leq 0. \end{cases}$$

jako

$$P(Y > 10 \ min) = \int_{10}^{\infty} f_Y(t) \ dt = \int_{10}^{\infty} \frac{1}{5} e^{-t/5} \ dt = e^{-2} \ .$$

- *Pomocí Poissonova:* Podle věty opět víme, že veličina

$$X' = \text{“}\textit{počet hovorů během doby } 10 \ min\text{”}$$

má Poissonovo rozdělení s parametrem λ'. Hledaná pravděpodobnost je dána jako

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

K určení parametru λ' použijeme vztah mezi veličinami Y a X' a podobně mezi veličinami Y a X, tj.

$$\frac{E(X')}{10 \ min} = \frac{1}{E(Y)} = \frac{E(X)}{60 \ min}$$

což odpovídá i požadavku, že průměrný počet událostí v časovém intervalu je úměrný jeho délce. Konkrétně tedy $\lambda' = E(X') = \frac{10 \ min}{60 \ min} \cdot 12 = 2$ a hledaná pravděpodobnost je pak

$$P(X' = 0) = e^{-2} \ .$$

(b)

- *Pomocí exponenciálního:* Z předchozího víme, že

$$Y = \text{“}\textit{doba čekání na hovor}\text{”} \ [\textit{v minutách}]$$

má exponenciální rozdělení se střední hodnotou $\tau = E(Y) = 5 \ min$. Pak pro hledaný čas $t > 0$ máme

$$0.7 = P(Y > t) = 1 - P(Y \leq t) = 1 - F_Y(t) = 1 - (1 - e^{-t/5}) = e^{-t/5}$$

tedy

$$t = -5 \cdot \ln 0.7 \doteq 1.78 \ min \ .$$

- *Pomocí Poissonova:* Vezmeme veličinu (závislou na t - zde je to parametr)

$$\widetilde{X} = \text{“}\textit{počet hovorů během doby } t \textit{ minut}\text{”}$$

která bude mít Poissonovo rozdělení s parametrem $\widetilde{\lambda}$, pro který máme vztah

$$\widetilde{\lambda} = E(\widetilde{X}) = \frac{t}{\tau} = \frac{t}{5}$$

Hodnotu t pak dostaneme z rovnice

$$0.7 = P\left(\widetilde{X} = 0\right) = \frac{\left(\widetilde{\lambda}\right)^0}{0!} e^{-\widetilde{\lambda}} = e^{-t/5}$$

což vede pochopitelně na stejný výsledek

$$t = -5 \cdot \ln 0.7 \doteq 1.78 \; min \; .$$

(c) Označme si veličiny

$$X_0 = \text{“ }\textit{počet všech hovorů během } 40\,min\text{”} \sim Poiss(\lambda_0)$$

$$X_1 = \text{“ }\textit{počet hovorů od žen během } 40\,min\text{”} \sim Poiss(\lambda_1)$$

$$X_2 = \text{“ }\textit{počet hovorů od mužů během } 40\,min\text{”} \sim Poiss(\lambda_2) \; .$$

Budeme předpokládat nezávislost hovorů od žen a od mužů, tj nezávislost veličin X_1 a X_2. Analogicky k předešlému budeme mít $\lambda_0 = 12 \cdot \frac{40 \text{ min}}{60 \text{ min}} = 8$. Pro určení hodnot λ_1 a λ_2 teď využijeme toho, že $X_0 = X_1 + X_2$, tj. dostaneme $8 = E(X_0) = E(X_1) + E(X_2)$ a z poměru volání žen a mužů zase máme, že $E(X_1) : E(X_2) = 2 : 1$.

Takže $\lambda_1 = \frac{16}{3}$ a $\lambda_2 = \frac{8}{3}$. Hledaná pravděpodobnost je (díky nezávislosti X_1 a X_2) tedy

$$P(X_1 = 4,\ X_2 = 1) \overset{(\text{nezáv. } X_i)}{=} P(X_1 = 4) \cdot P(X_2 = 1) = \frac{\left(\frac{16}{3}\right)^4}{4!} e^{-\frac{16}{3}} \cdot \frac{\frac{8}{3}^1}{1!} e^{-\frac{8}{3}} = \frac{2^{16}}{3^6} e^{-8} \doteq 0.03.$$

Příklad 7.3 *Při numerickém výpočtu se reálná čísla zaokrouhlují na jedno desetinné místo. Jaká je pravděpodobnost, že vzdálenost skutečného čísla od zaokrouhleného bude větší než* 0.04*?*

Řešení:
Podobně jako v příkladu o házení mince na nekonečnou mřížku (**Příklad 2.2**) budeme předpokládat, že vstupní hodnoty (tj. čísla, která budeme zaokrouhlovat) pocházejí z nějakého referenčního intervalu $\Omega \subseteq \mathbb{R}$ délky 0.1 (zaokrouhlujeme na jedno desetinné místo), na kterém máme geometrickou pravděpodobnost, např. $\Omega = \langle -0.05,\ 0.05 \rangle$.

Označme si teď náhodnou veličinu $X : \Omega \to \mathbb{R}$ jako

$$X = \text{ “}\textit{skutečná hodnota}\text{” - “}\textit{zaokrouhlená hodnota}\text{”}$$

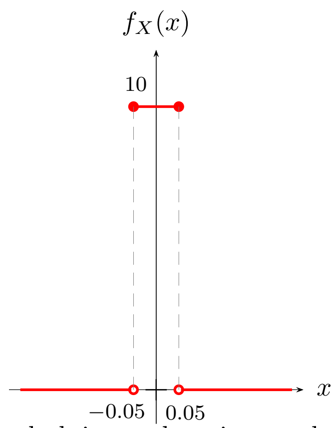
tj. pro $\omega \in \Omega = \langle -0.05,\ 0.05 \rangle$ to funguje jako

$$X(\omega) = \begin{cases} \omega - 0 = \omega & \text{pro} \quad -0.05 \leq \omega < 0.05, \\ 0.05 - 0.05 = 0 & \text{pro} \quad \omega = 0.05 \; . \end{cases}$$

A nás teď zajímá pravděpodobnost $P\big(|X| > 0.04\big)$. K tomu potřebujeme znát rozdělení veličiny X. Intuitivně tušíme, že X bude mít rovnoměrné rozdělení na své množině hodnot, tj. v intervalu $\langle -0.05,\ 0.05)$. Předpokládejme tedy, že $X \sim \mathrm{Ro}(-0.05,\ 0.05)$ (níže si to pak zdůvodníme). Veličina X má pak spojité rozdělení s hustotou pravděpodobnosti

$$f_X(x) = \begin{cases} \frac{1}{0.05-(-0.05)} = 10 & ,\ -0.05 \le x \le 0.05 \\ 0 & ,\ \text{jinak} \end{cases}$$

s grafem



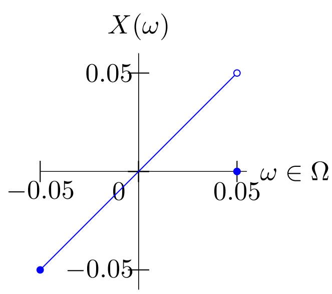
(Hodnoty f_X v krajních bodech intervalu nejsou podstatné.)

Hledaná pravděpodobnost je tudíž

$$P\Big(|X| > 0.04\Big) = 1 - P\Big(|X| \le 0.04\Big) = 1 - P(-0.04 \le X \le 0.04) =$$
$$= 1 - \int_{-0.04}^{0.04} 10\ dx = 1 - 10 \cdot 0.08 = 0.2\ .$$

- -

Z cvičných důvodu si teď výše zmíněné rovnoměrné rozdělení odvodíme. Veličinu X si můžeme znázornit tímto grafem:



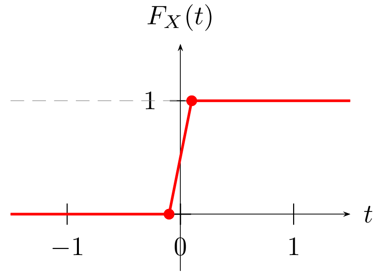
Odsud snadno vidíme, že pro $t \in \langle -0.05,\ 0.05)$ je

$$vol\Big(\{\omega \in \Omega \mid X(\omega) \le t\}\Big) = vol\Big(\langle -0.05,\ t\rangle\Big) = t + 0.05$$

takže distribuční funkce je

$$F_X(t) = P(X \le t) = \begin{cases} 0 & , t < -0.05 \\ \frac{vol\{\omega\in\Omega \ \mid\ X(\omega)\le t\}}{vol(\Omega)} = \frac{t+0.05}{0.1} = 10t + 0.5 & , t \in \langle -0.05,\ 0.05) \\ 1 & , t \ge 0.05 \end{cases}$$

s grafem



Distribuční funkce F_X je tedy spojitá a má výše uvedenou hustotu.