6. cvičení z STP

25. - 29. březen 2019

Příklady 5.3, 5.4.

Poznámky k Poissonovu rozdělení: U veličiny

$$X = \text{„} \textit{ počet událostí během intervalu délky } T \text{"}$$

kde interval je obvykle časový (ale může být i délkový nebo měřený nějakou jinou jednotkou), můžeme předpokládat *Poissonovo* rozdělení 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. průměrný počet událostí za jednotku času se s průběhem doby nemění).

Pravděpodobnosti hodnot $k = 0, 1, 2, \dots$ jsou dány jako

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

kde $\lambda > 0$ je (bezrozměrný) parametr. Tento parametr představuje navíc střední hodnotu (tj. $E(X) = \lambda$) protože:

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

V praxi používáme Poissonovo rozdělení především jako aproximaci binomického rozdělení, pokud události pocházejí z velkého počtu nezávislých zdrojů a z každého zdroje nastane událost nejvýše jednou. Tento zdroj může být např. příchod zákazníka do fronty, ulovení ryby, průjezd auta atd.

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

Časový interval 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):

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

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

Poissonovo rozdělení je tak většinou spíše limitní případ pro binomické 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.

Použití: Hodnotu $P(Y = k)$ pro $Y \sim \mathrm{Bi}(n, p)$ aproximujeme pomocí Poissonova rozdělení jako

$$P(Y = k) \doteq \frac{\lambda^k}{k!} e^{-\lambda}$$

když je $k << n$ a $\lambda := E(Y) << n$.

Příklad 6.1 *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

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

kde $E(X) = \lambda$. Pro veličinu $\widetilde{X}$, označující počet kazů na 10 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

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

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 dobou 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

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

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, ze 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.

Ještě lepší analogií by ovšem bylo zvolit si tuto veličinu

$$Y' = \text{"}\textit{pořadové číslo pokusu, kdy nastal první úspěch}\text{"},$$

pro kterou máme jednoduchý vztah $Y' = \widetilde{Y} + 1$ a tedy oborem hodnot pro Y' je nyní $\{1, 2, \dots\}$. Proč je toto lepší analogie k exponenciálnímu rozdělení nežli veličina $\widetilde{Y}$, ukazují vztahy pro střední hodnotu a rozptyl:

$$E\big(Y'\big) = E\big(\widetilde{Y} + 1\big) = \frac{1-p}{p} + 1 = \frac{1}{p}$$

$$\mathrm{var}\big(Y'\big) = \mathrm{var}\big(\widetilde{Y} + 1\big) = \mathrm{var}\big(\widetilde{Y}\big) = \left(\frac{1}{p}\right)^2$$

které pěkně korespondují se vztahy pro exponenciální rozdělení veličiny Y.

Příklad 6.2 *Na zákaznickou linku přichází průměrně* 12 *hovorů za hodinu. 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.

Řešení:

(a) Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení. Použijeme následující vztah mezi oběma rozděleními:

Věta: 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

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

- *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

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

- *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

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

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)

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

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

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