6. cvičení z STP

25. března 2021

Příklad 6.1 *Při hodu na koš se trefíme s pravděpodobností $p = 0.2$. Náhodná veličina X je počet hodů, než se trefíme.*

(a) Určete rozdělení veličiny X.

(b) V kterém kole se průměrně poprvé trefíme?

(c) Kolikrát musíme nejméně hodit, abychom se s pravděpodobností alespoň 90% alespoň jednou (v rámci těchto hodů) trefili?

Řešení:

(a) Veličina X počítá počet neúspěšných kol. Má tedy geometrické rozdělení Geom(0.2) s oborem hodnot $\{0, 1, 2, \dots\}$ a jejich pravděpodobnostmi

$$P(X = k) = p(1-p)^k = 0.2 \cdot 0.8^k$$

pro $k = 0, 1, 2, \dots$.

POZOR: Nepleťte si *geometrickou pravděpodobnost*, což je způsob počítání pravděpodobnosti jevu (název je odvozen od geometrických obrazců) a *geometrické rozdělení*, které zase přísluší náhodné veličině (název je odvozen od geometrické posloupnosti, kterou tvoří hodnoty pravděpodobnostní funkce dané veličiny).

(b) Hledáme střední hodnotu veličiny

$$Y = \text{\textit{"pořadí hodu, při kterém se poprvé trefíme"}} \,.$$

Zřejmě je $Y = X + 1$ a tedy

$$E(Y) = E(X+1) = E(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} = \frac{1}{0.2} = 5 \,.$$

(c) Abychom si lépe představili situaci, předpokládejme, že i po té, co se trefíme, pokračujeme v házení (a veličina Y zaznamená pouze ten první úspěšný pokus). Ptáme se tedy, jaký je nejmenší počet hodů $n \in \mathbb{N}$, abychom se v jejich průběhu alespoň jednou trefili. Chceme tudíž znát nejmenší $n \in \mathbb{N}$ takové, že $P\Big(Y \le n\Big) \ge 0.9$, neboli

$$0.9 \le P\Big(X + 1 \le n\Big) = P\Big(X \le n - 1\Big) = F_X(n-1) \,.$$

K tomu potřebujeme tudíž znát distribuční funkci X pro $k \in \mathbb{N}_0$:

$$F_X(k) = \sum_{i \le k} P(X = i) = \sum_{i=0}^{k} p(1-p)^i = p \cdot \frac{1 - (1-p)^{k+1}}{1 - (1-p)} = 1 - (1-p)^{k+1} \,.$$

Po dosazení tedy dostaneme podmínku

$$0.9 \le F_X(n-1) = 1 - (1-p)^n$$

neboli

$$\begin{aligned} 0.1 &\geq (1-p)^n = (1-0.2)^n = 0.8^n \\ \log 0.1 &\geq n \log 0.8 \\ n &\geq \tfrac{\log 0.1}{\log 0.8} \doteq 10.32 \end{aligned}$$

Pozor: logaritmus hodnoty 0.8 je záporný! Tedy musíme hodit alespoň $n = 11$ -krát.

- -

Úlohu (c) můžeme vyřešit i s pomocí binomického rozdělení. Pro $n \in \mathbb{N}$ uvažujme veličinu

$$Z_n = \text{"}\textit{počet tolika hodů (z n možných), ve kterých se trefíme}\text{"}$$

která zřejmě má binomické rozdělení $\mathrm{Bi}(n, p)$. Hledáme nyní nejmenší $n \in \mathbb{N}$ tak, aby

$$P(Z_n \geq 1) \geq 0.9 \; .$$

Máme tedy

$$0.9 \leq P(Z_n \geq 1) = 1 - P(Z_n = 0) = 1 - (1-p)^n$$

což je stejná nerovnost jako výše a tím dostaneme i stejné řešení.

Příklad 6.2 *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

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

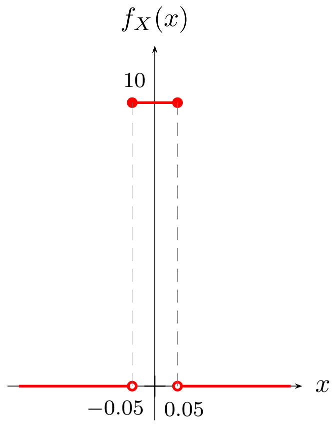
tj. pro $\omega \in \Omega = \langle -0.05,\ 0.05 \rangle$ to funguje jako

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

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 $\langle -0.05,\ 0.05 \rangle$. 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

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

s grafem



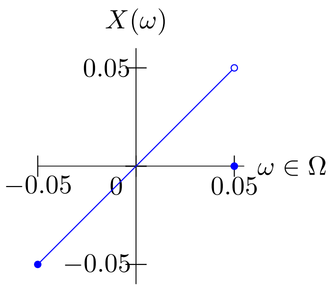
(Hodnoty f_X v krajních bodech intervalu nejsou podstatné.)

Hledaná pravděpodobnost je tudíž

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

- -

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



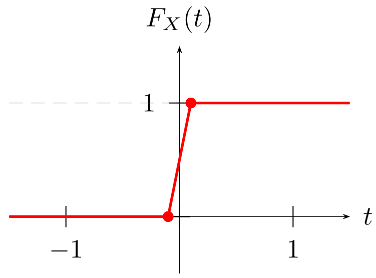
Odsud snadno vidíme, že pro $t \in \langle -0.05,\ 0.05)$ je

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

takže distribuční funkce je

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

s grafem



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

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

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

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):

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

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

Příklad 6.3 *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

$$X = \text{“ \textit{počet kazů na 50 m délky látky}”}$$

Pak $X \sim \text{Poiss}(\lambda)$ 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 m délky, předpokládáme také Poissonovo rozdělení se střední hodnotou $E(\widetilde{X}) = 1$. Protože střední hodnota má být úměrná délce intervalu dostaneme

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

Tudíž

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

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

(c) Označme si veličiny

$$X_1 = \text{“ \textit{počet kazů na prvních 20 m délky látky}”}$$
$$X_2 = \text{“ \textit{počet kazů na zbylých 30 m délky látky}”}\,.$$

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

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

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

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

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

$Y =$ *"doba čekání na událost"*

je veličina s exponenciálním rozdělením. Pak veličina

$X =$ *"počet událostí během doby T"*

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 6.4 *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:
Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení.

(a) • *Pomocí exponenciálního:* Podle věty má veličina

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

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

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

Hledaná pravděpodobnost pak je

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

Pravděpodobnost také můžeme spočítat pomocí hustoty f_Y veličiny Y:

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

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

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

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

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

tedy

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

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

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

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

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