10 Poissonovo a exponenciální rozdělení pravděpodobnosti

Břetislav Fajmon, UMAT FEKT, VUT Brno

Této přednášce odpovídá část kapitoly 12 ze skript [1] a vše, co se nachází v kapitole 6 sbírky úloh [2] – tuto kapitolu 6 sbírky úloh je vhodné vzít jako cvičení tohoto tématu.

Pokračujeme v přehledu pravděpodobnostních popisů veličin, které jsou natolik důležité, že mají své jméno. Dvě rozdělení v této přednášce popisují dvě různé veličiny, které ovšem obě lze měřit v jedné konkrétní situaci. Jedna z nich je

- **diskrétní veličina (matematický popis: Poissonovo rozdělení)**
- **a druhá je spojitá (matematický popis: exponenciální rozdělení)**

10.1 Předpoklady tzv. „náhodné události“

Nejprve přistupme k popisu situace, ve které oba typy veličin, o kterých bude řeč, lze měřit. Jedná se o náhodný výskyt události, kde jsou splněny jisté předpoklady:

a) **Náhodná událost nenastává příliš často – přesněji řečeno, musí být možné rozdělit časovou osu na drobné intervaly stejné délky, že pravděpodobnost, že v jednom konkrétním časovém intervalku nastanou více než dva výskyty dané události, je rovna nule.**

   Například se jedná o – narození dítěte v jisté porodnici; příchod emailu na jistou adresu;

příchod zákazníka do fronty v supermarketu. U těchto událostí se může stát, že dojde ke dvěma výskytům této události blízko za sebou; ALE průměrně musí docházet k tomu, že v intervalu krátké délky nastane nanejvýš jeden (nebo spíše žádný) výskyt daného typu události. Jinak matematické popisy, o kterých budeme dnes mluvit, nebudou odpovídat reálným hodnotám dané veličiny.

b) Náhodná událost má vlastnost „zapomnětlivosti“ – výskyt této události nezávisí na minulých výskytech.

To znamená, že výskyt náhodné události je stejně

pravděpodobný (zjednodušeně řečeno) dnes, zítra i pozítří – zkrátka kdykoli začneme měřit, pravděpodobnost, že do padesáti minut od začátku měření dojde k výskytu náhodné události, je stejná, ať už začnu měřit dnes ráno v 8.00 hod, nebo zítra ve 12.00 hod.

(v reálu někdy se matematický popis dokáže vypořádat i s rozdíly měření veličin v průběhu dne, ale veličiny musí být dobře popsány, a také musíme pak „srovnávat srovnatelné“: například počet aut, která projedou jistou křižovatkou mezi půlnocí a jednou hodinou, se liší od počtu aut, která projedou stejnou křižovatkou mezi osmou

a devátou ráno – ovšem pokud upřesníme, že následující veličina popisuje situaci na dané křižovatce ve špičce pracovního dne, matematický popis této situace dopravní špičky může být docela přesný).

c) Náhodný výskyt událostí je generován na základě velkého počtu (v ideálním případě až nekonečně velkého počtu – zkrátka dostatečně velkého počtu) zdrojů.

Například příchod zákazníka do supermarketu v Brně lze dobře popsat pomocí veličin, které budou za chvíli uvedeny – potenciální zdroj lidí, kteří by totiž mohli každý den do supermarketu

přijít, je v řádu desetittisíců, možná statisíců. Na druhé straně, příchod zákazníka do samoobsluhy ve vesmírné orbitální stanici, kde je všehovšudy deset lidí, lze také matematicky popsat – ovšem jinými prostředky, než bude uvedeno dále. Příchod studenta do menzy, do které může potenciálně přijít i několik set studentů, zatímco teoreticky jich tam může během dne zamířit i několik tisíc, je ještě snesitelným příkladem události, které se budeme v dalším věnovat. Ovšem účast studenta na přednášce své skupiny je už spíše nepřijatelnou událostí pro model, který bude uveden dnes – jedná se o omezenou skupinu zpravidla méně než 250 studentů, a sice o jedinou

skupinu, která je potenciálním zdrojem přícházích studentů.

V následujícím tedy bude představen model popisující chování dvou veličin v situaci výskytu náhodné události, ke které dochází nepravidelně, a ne příliš často; potenciální zdroj výskytů události je dostatečně velký (minimálně několik set potenciálních výskytů, ovšem nikoli z jedné skupiny, ale z různých nezávislých zdrojů). A konečně, jednotlivé výskyty události jsou jeden na druhém nezávislé (a nebo lze vzájemné ovlivňování výskytů zanedbat vzhledem k jejich celkovému počtu).

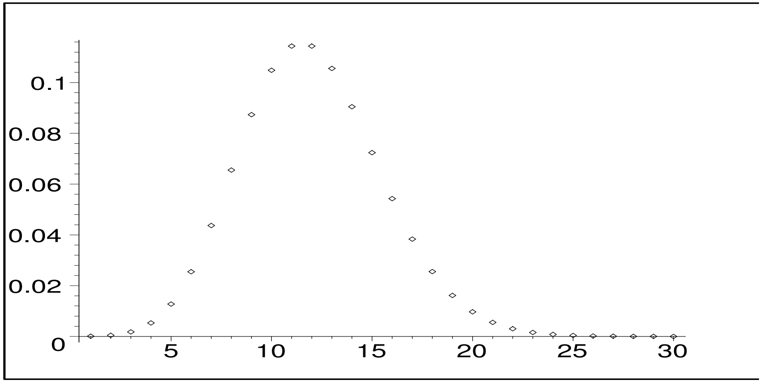
K popisu takto zhruba definované náhodné události potřebuje matematika jediný parametr (získaný např. z měření) – průměrný počet λ výskytů dané náhodné události za jednotku času.

10.2 Poissonovo rozdělení pravděpodobnosti

V situaci uvedené v předchozím oddílku nyní chceme matematicky popsat veličinu X = počet výskytů dané náhodné události za jednotku času.

Jakých hodnot může veličina X nabývat a s jakou pravděpodobností (= jaké je její rozdělení pravděpodobnosti)?

Nejprve je dobré si všimnout, že se jedná o <u>diskrétní veličinu</u>. Odtud lze dobře usoudit, že za určitý časový úsek může být výskytů jisté události $0, 1, 2, \ldots$**, a teoreticky to (kromě nuly) může být jakékoli PŘIROZENÉ číslo, záleží na povaze veličiny, na délce časové jednotky – při pevně stanovené časové jednotce bude asi pravděpodobnost výskytu vyššího a vyššího počtu událostí v dané době klesat, ale matematika tuto pravděpodobnost dokáže vyčíslit, i když bude velmi malá.**



Obrázek 1: Graf pravděpodobnostní funkce $p(k)$ rozdělení $Po(12)$.

Lze odvodit (skripta str. 187-191), že veličinu X **lze popsat pravděpodobnostní funkcí**

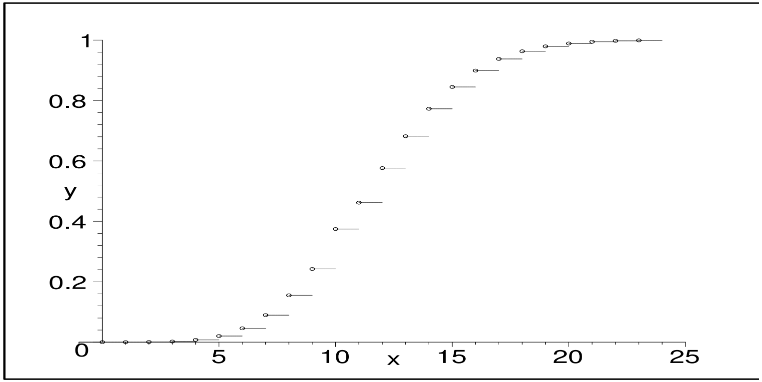
$$P(X=k)=\frac{\lambda^k}{k!}\cdot e^{-\lambda} \quad \textbf{pro } k=0,1,2,\dots \tag{1}$$

Dále lze odvodit dosazením těchto pravděpodobnostních hodnot $p(k)$ do příslušných vzorců, že

$$EX = \lambda, \quad DX = \lambda$$

(a to je na tomto Poissonově rozdělení pravděpodobnosti zajímavé – ze všech známých rozdělení je jediné, u kterého střední hodnota a rozptyl jsou stejné).

Pro distribuční funkcí $F(x)$ veličiny X platí $F(x) = \sum_{k<x} p(k)$.



Obrázek 2: Graf distribuční funkce $F(x)$ rozdělení $Po(12)$: funkce s nekonečně mnoha schody, která vyjadřuje kumulativní pravděpodobnosti $F(t) = P(Y < t)$.

10.3 Exponenciální rozdělení pravděpodobnosti

Ve stejné situaci jako v předchozím oddílku měřme nyní jinou veličinu, a sice Y = doba mezi dvěma následnými výskyty dané náhodné události.

Je asi rozumné si hned zkraje všimnout, že se jedná o spojitou veličinu. Odtud lze dobře usoudit (ještě za pomoci faktu, že se jedná o měření času počítané od nuly), že teoreticky hodnotou veličiny Y může být jakékoli KLADNÉ číslo. Možnost vyšší a vyšší naměřené hodnoty bude asi klesat, což se projeví na klesajícím charakteru hustoty $f(x)$.

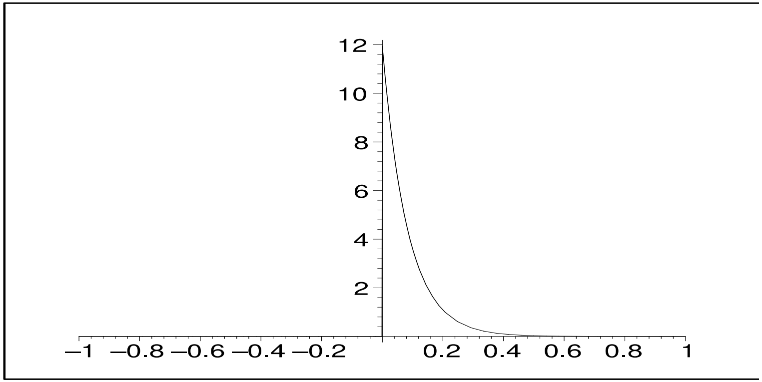
Lze odvodit (viz stejné – výše uvedené – strany ve skriptech), že hustota $f(x)$ veličiny Y má tvar

$$f(x) = \begin{cases} 0 & \textbf{pro } x \leq 0; \\ \lambda \cdot e^{-\lambda x} & \textbf{pro } x > 0. \end{cases} \tag{2}$$

Integrací od minus nekonečna po x z hustoty lze odvodit vztah pro distribuční funkcí $F(x)$ veličiny X:

$$F(x) = \begin{cases} 0 & \textbf{pro } x \leq 0; \\ 1 - e^{-\lambda x} & \textbf{pro } x > 0. \end{cases} \tag{3}$$

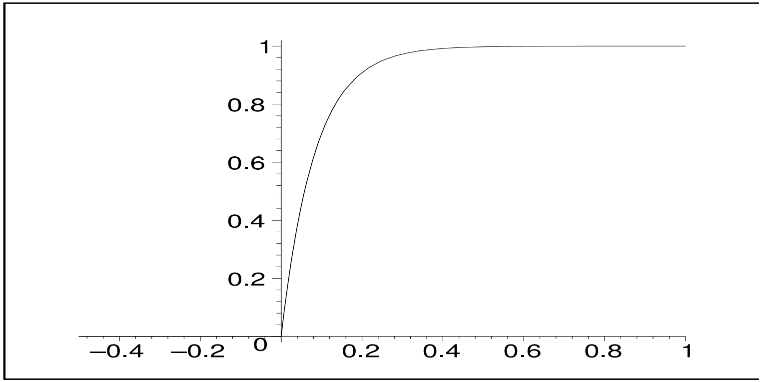
Použitím příslušných vzorců (pro spojitou veličinu) lze pomocí integrace per partes dospět k



Obrázek 3: Graf hustoty $f(x)$ rozdělení $Exp(12)$.

výpočtu střední hodnoty a rozptylu exponenciálně rozdělené veličiny:

$$EY = \frac{1}{\lambda}, \quad DX = \frac{1}{\lambda^2}$$



Obrázek 4: Graf distribuční funkce $F(x)$ rozdělení $Exp(12)$.

(znovu opakuji, že parametr λ má tentýž význam jako u Poissonova rozdělení: λ = průměrný počet výskytů náhodné události za jednotku času).

10.4 Konkrétní příklady

Příklad 10.1. **Zdravotnický úřad shromažďuje údaje o nově narozených dětech. Průměrně každé dvě hodiny se narodí další dítě. Určete**

a) **Průměrný počet narozených dětí za rok.**

b) **Pravděpodobnost, že v daném dnu se nenarodí žádné dítě.**

c) **Pravděpodobnost, že v jednom dnu se narodí** 20 **dětí.**

d) **Pravděpodobnost, že za** 4 **hodiny se narodí aspoň** 5 **dětí.**

Řešení:

ad a) *Z tohoto úkolu nebudeme dělat vědu. Průměrně jedno dítě za dvě hodiny dává dvanáct dětí za den a* $365 \cdot 12 = 4380$ *dětí za rok.*

ad b) *Základem dobrého využití exponenciálního nebo Poissonova popisu je zvolit si vhodnou časovou jednotku. Pokud hledáme určitý údaj za den, zvolme časovou jednotku jeden den. Druhým krokem po volbě časové jednotky je vypočtení parametru* λ*. V našem případě* $\lambda = 12$ *dětí za den (jedná se o průměrný údaj za časovou jednotku).*

V některých příkladech, máme možnost použít buď exponenciální, nebo Poissonovo rozdělení -

ukážeme si nyní obě možnosti.

Nejprve tedy označme Y *dobu mezi dvěma po sobě jdoucími výskyty narození dítěte. Podle podrobného odvození v předchozím oddílu má veličina* Y *exponenciální rozdělení s parametrem* $\lambda = 12$*. Pak pravděpodobnost, že daný den se nenarodí nikdo, je rovna*

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - F(1) = 1 - (1 - e^{-12 \cdot 1}) = e^{-12}.$$

(využili jsme raději distribuční funkce $F(x)$ než hustoty $f(x)$ exponenciálního rozdělení, protože ve funkci $F(x)$ máme vlastně hotovou integraci).

Druhá možná cesta je užít veličiny X, která udává počet narození za jeden den. X má Poissonovo rozdělení s parametrem $\lambda = 12$, čili hledaná pravděpodobnost je rovna

$$P(X=0) = \frac{12^0}{0!} \cdot e^{-12} = 0{,}00000614.$$

ad c) *Využijeme veličiny X zavedené v b) a dosadíme:*

$$P(X=20) = \frac{12^{20}}{20!} \cdot e^{-12} = 0{,}00968$$

Pro ilustraci - graf pravděpodobnostní funkce Poissonova rozdělení je uveden na obrázku 1, graf příslušné distribuční funkce na obr. 2.

ad d) *Poslední úkol tohoto příkladu je analogický, ovšem otázka je položena tak, že nás zajímá údaj dosažený za 4 hodiny. Musíme tedy změnit časovou jednotku na 4 hodiny. Tím pádem se mění průměrný počet narození za časovou jednotku na $\lambda = 2$. Označíme-li nyní $Y =$ počet dětí narozených za 4 hodiny, platí $Y \sim Po(\lambda = 2)$. A tedy*

$$\begin{aligned} P(Y \geq 5) &= p(5) + p(6) + p(7) + \cdots = \\ &= 1 - (p(0) + p(1) + p(2) + p(3) + p(4)) = \\ &= 1 - e^{-2} \cdot (\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!}) = 0{,}05265 \end{aligned}$$

(místo sečítání nekonečné řady jsme opět odečetli pravděpodobnost opačného jevu od jedničky). Jak je

uvedeno na posledním řádku výpočtu, pokud sečítáme několik pravděpodobností Poissonova rozdělení (zejména při písemce na kalkulačce), je vhodné člen $e^{-\lambda}$ vytknout, místo abychom jím násobili každý člen v závorce zvlášť - ušetříme si práci.

Příklad 10.2. **a poště mají být instalovány automaty na prodej známek, které po vhození mince vydají přesně za deset sekund žádanou známku. Předpokládáme, že průměrně bude chtít použít automatu šest osob za minutu. Kolik automatů bychom měli instalovat, aby s pravděpodobností** $0{,}95$ **byl i v době největší frekvence obsloužen každý zájemce bez čekání?**

Řešení: V dnešní hektické době jsou i ekonomické požadavky neúprosné: čekat deset sekund je nepřijatelné, na $95\,\%$ *musí být automat k dispozici okamžitě. Klíčem k tomuto příkladu je zjistit, s jakou pravděpodobností přijde jistý počet lidí za deset sekund - to je totiž doba, kdy automat eventuelně někoho obsluhuje a každý další příchozí musí čekat. Zvolme* *tedy v prvé řadě časovou jednotku rovnu deseti sekundám.*

Dále jestliže průměrně přijde šest za minutu, za deset sekund přijde jeden, čili $\lambda = 1$*. Označme* X = *počet příchozích zákazníků během deseti sekund. Bystrý čtenář již tuší, že na následujícím řádku prohlásím, že podle přechozího podrobného odvození má veličina* X *rozdělení Poissonovo s parametrem* $\lambda = 1$.

Položme si nyní následující otázku: Jaká je pravděpodobnost, že během deseti sekund nepřijde více než jeden zákazník (a tedy k okamžitému obsloužení stačí jeden automat)?

$$p = P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-1} \cdot \left(\frac{1^0}{0!} + \frac{1^1}{1!}\right) = e^{-1} \cdot (1+1) =$$

Tedy jediný automat je dostatečný v 73 % času. Ovšem v ostatních 27 % příchozí zákazník musí čekat, a to je nepřijatelné. Podívejme se, co říká teorie pro dva nainstalované automaty: Pravděpodobnost, že během deseti sekund přijdou maximálně dva zákazníci, je rovna

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 0{,}73 + P(X = 2) = 0$$

Tedy v 92 % času nový příchozí nemusí čekat. To je ovšem podle našeho zadání stále málo. Spočtěme dále pravděpodobnost, že během deseti sekund přijdou maximálně tři:

$$P(X \leq 3) = 0{,}92 + P(X = 3) > 0{,}95,$$

a tedy k naplnění požadavku ze zadání stačí tři automaty.

Příklad 10.3. Výrobní zařízení má poruchu v průměru jednou za 2000 hodin. Veličina Y představující dobu čekání na poruchu má exponenciální rozdělení. Určete dobu T_0 tak, aby pravděpodobnost, že přístroj bude pracovat delší dobu než T_0, byla $0{,}99$.

Řešení. Pravděpodobnost $0{,}99$ je dost vysoká - proto doba T_0 bezporuchového provozu s touto pravděpodobností bude mnohem nižší než 2000 hodin. Určeme nyní T_0 přesně.

V prvé řadě stanovíme časovou jednotku. Nabízí se jednotka 2000 hodin, tj. budeme teď počítat s čísly, kdy $1 = 2000$ hod. Za druhé stanovíme λ, tj. průměrný počet poruch za časovou jednotku: v našem případě

$\lambda = 1$. **A tak** $Y \sim Exp(\lambda = 1)$. **Hledejme teď takovou dobu** T_0**, aby** $P(Y \geq T_0) = 0{,}99$. **Využijeme opět distribuční funkce** $F(t)$**, protože její hodnoty jsou přímo rovny jistým kumulativním pravděpodobnostem - a jednu z nich můžeme do posledního vztahu dosadit:**

$$\begin{aligned} P(Y \geq T_0) &= 0{,}99 \\ 1 - P(Y < T_0) &= 0{,}99 \\ 1 - F(T_0) &= 0{,}99 \\ F(T_0) &= 0{,}01 \\ 1 - e^{-\lambda \cdot T_0} = 1 - e^{-T_0} &= 0{,}01 \\ T_0 &= 0{,}01005034 \end{aligned}$$

Musíme tento údaj prezentovat v rozumnějších jed-

notkách: Pokud $1 = 2000$ **hodin, tak**

$$T_0 = 0{,}01005034 = 2000 \cdot 0{,}01005034 \textbf{ hodin } = 20{,}1 \textbf{ hodin}.$$

Příklad 10.4. Tento příklad si klade za cíl dosvědčit, že Poissonovsky rozdělená veličina může mít i jiný charakter než jen počet náhodných událostí za jednotku času.

Další situace, které Poissonovský model může popsat, jsou ty, kde měříme počet náhodných událostí na jednotku délky!!! Například:

Počet znečišťujících částeček při výrobě optických disků má Poissonovo rozdělení pravděpodobnosti. Průměrný počet částeček na jeden centimetr čtvereční je $0{,}1$ (tj. průměrně je znečištěn každý desátý čtvereční centimetr). Plocha kontrolovaného disku je 100 cm^2. Určete pravděpodobnost, že více než 5 cm^2 tohoto disku bude znečištěno.

Řešení: v tomto příkladu nemáme zadánu jednotku času, ale jednotku plochy, a sice 1 cm^2.

Je také zadán průměrný počet výskytů znečištění na tuto jednotku, a sice $0{,}1$ znečištěných centimetrů. Tuto hodnotu si můžeme přepočítat na rozumnou: pokud jednotka plochy je 100 cm^2, tak průměrný počet znečištěných centimetrů čtverečních na tuto jednotku je $\lambda = 10$. Pak

$$\begin{aligned}
P(X > 5) &= p(6) + p(7) + p(8) + \cdots = \\
&= 1 - p(0) - p(1) - p(2) - p(3) - p(4) - p(5) = \\
&= 1 - e^{-10}\Big(\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!} + \\
&\quad + \frac{10^4}{4!} + \frac{10^5}{5!}\Big) \doteq 0{,}932914.
\end{aligned}$$

O dalších dvou význačných spojitých rozděleních pravděpodobnosti kromě exponenciálního rozdělení (o jednom triviálním a o jednom velmi často používaném pro pravděpodobnostní popis) bude informovat následující přednáška.

Otázky k opakování

Z dnešní přednášky vyvstávají dvě otázky k opakování pro studenty, jejichž odpovědi je záhodno mít s sebou na písemky a znát zpaměti:

otázka č. 17. **Poissonovo rozdělení – význačné diskrétní rozdělení číslo 5: V jaké situaci jej lze použít? Co měříme v proměnné** X**? Jakých hodnot veličina nabývá a s jakou pravděpodobností? Jaká je EX a DX?**

otázka č. 18. **Exponenciální rozdělení – význačné spojité rozdělení číslo 1: podotázky viz ot. 17.**

Literatura

[1] **Fajmon, B., Růžičková, I.: Matematika 3. Skriptum FEKT 2003 (identifikační číslo v informačním systému VUT: MAT103).** `http://www.rozhovor.cz/souvislosti/matematika3.pdf.`

[2] **Hlavičková, I., Hliněná, D.: Sbírka úloh z pravděpodobnosti. Skriptum FEKT 2008.**