6. cvičení z PST

25. - 29. října 2021

Poznámky ke geometrickému rozdělení: Veličina

$$X = \text{“\textit{počet neúspěchů než nastane první úspěch}”}$$

má geometrické rozdělení, pokud můžeme opakovat libovolné množství nezávislých pokusů, které mají všechny stejnou pravděpodobnost úspěchu p.

Hodnoty veličiny X jsou $\{0, 1, 2, \dots\}$. Pro odvození rozdělení X si pro $i = 1, 2, 3 \dots$ označme jevy

$$A_i = \text{“\textit{i-tý pokus je úspěšný}”}$$

které budou nezávislé s budou mít pravděpodobnosti $P(A_i) = p$. Pak máme pravděpodobnostní funkci

$$p_X(k) = P(X = k) = P(\overline{A_1} \cap \dots \cap \overline{A_k} \cap A_{k+1}) = P(\overline{A_1}) \cdots P(\overline{A_k}) \cdot P(A_{k+1}) = (1-p)^k p$$

pro $k = 0, 1, 2, \dots$ Pro výpočet střední hodnoty a rozptylu potřebujeme pracovat s mocninnými řadami (např. použít vztah $\sum_{k=0}^{\infty}(1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$ a to, že derivace mocninné řady se dá dělat člen po členu):

$$E(X) = \sum_{k=0}^{\infty} k \cdot p_X(k) = \sum_{k=0}^{\infty} kp(1-p)^k = p(1-p) \underbrace{\sum_{k=0}^{\infty} k(1-p)^{k-1}}_{\frac{\mathrm{d}}{\mathrm{d}p}\left(-\sum_{k=0}^{\infty}(1-p)^k\right)} = p(1-p) \cdot \frac{\mathrm{d}}{\mathrm{d}p}\left(-\frac{1}{p}\right) = \frac{p(1-p)}{p^2} = \frac{1-p}{p} \; .$$

Podobně máme

$$E(X^2) = \sum_{k=0}^{\infty} \underbrace{k^2}_{k(k-1)+k} \cdot p_X(k) = \sum_{k=0}^{\infty} k(k-1)p(1-p)^k + \underbrace{\sum_{k=0}^{\infty} kp(1-p)^k}_{E(X)} =$$

$$= p(1-p)^2 \underbrace{\sum_{k=0}^{\infty} k(k-1)(1-p)^{k-2}}_{\frac{\mathrm{d}^2}{\mathrm{d}p^2}\left(\sum_{k=0}^{\infty}(1-p)^k\right)} + E(X) = p(1-p)^2 \cdot \frac{\mathrm{d}^2}{\mathrm{d}p^2}\left(\frac{1}{p}\right) + \frac{1-p}{p} = \frac{2p(1-p)^2}{p^3} + \frac{1-p}{p} = \frac{(1-p)(2-p)}{p^2}$$

takže

$$D(X) = E(X^2) - \Big(E(X)\Big)^2 = \frac{(1-p)(2-p)}{p^2} - \left(\frac{1-p}{p}\right)^2 = \frac{1-p}{p^2}\left(2 - p - (1-p)\right) = \frac{1-p}{p^2}$$

což si můžeme lépe zapamatovat jako

$$D(X) = \frac{1}{p} \cdot E(X) \; .$$

6.1 (geometrické rozdělení)

Alice a Bob hází na koš. Alice se trefí s pravděpodobností $p_1 = 0.05$ a Bob s pravděpodobností $p_2 = 0.08$. Začíná Alice a střídají se. S každým hodem Alice se začíná nové kolo. Náhodná veličina X je počet (celých) kol, které uběhnou před tím, než se někdo trefí.

(a) Určete rozdělení veličiny X.

(b) V kterém kole se průměrně někdo poprvé trefí? Jaká je směrodatná odchylka veličiny X?

(c) Jaký nejmenší počet kol musí začít, aby se během nich s pravděpodobností alespoň 90% někdo alespoň jednou trefil?

Řešení:

(a) Veličina X počítá počet neúspěšných kol. Podle poznámky výše má tedy geometrické rozdělení s oborem hodnot $\{0, 1, 2, \dots\}$ a pravděpodobnostní funkcí $p_X(k) = p(1-p)^k$ pro $k = 0, 1, 2, \dots$. Určíme hodnotu p:

V rámci daného kola označme jevy

$A =$ *"Alice se trefí"*

$B =$ *"Bob se trefí"*

$C =$ *"alespoň jeden se trefí"*

Pak je $C = A \cup B$.

Ze zadání neplyne, jakým způsobem celá hra probíhá při opakovaných pokusech: Konkrétně jde o to, že když se Alice trefí (a tím se určí hodnota X), jestli pak ještě Bob také hází, než se začne počítat zase nové kolo.

- Pokud si v takové situaci i tak může Bob hodit (čímž už pochopitelně hodnotu X neovlivní), můžeme se na opakování hry dívat jako na sled nezávislých hodů. V tom případě budou jevy A a B nezávislé a jejich pravděpodobností budou $P(A) = p_1$ a $P(B) = p_2$.

  Pak tedy
  $$P(C) = P(A \cup B) = P(A) + P(B) - P(A) \cdot P(B) = p_1 + p_2 - p_1 p_2$$

- Pokud ale Bobovi už neumožníme házet, jakmile se Alice trefí, a další kolo tak zase začne Alice, pak musíme jev B chápat jinak. V tomto případě bude $B \subseteq \overline{A}$ a pravděpodobnost toho, že se Bob trefí bude podmíněná, tedy $P(B|\overline{A}) = p_2$. Současně vidíme, že A a B jsou neslučitelné (tj. $A \cap B = \emptyset$). Z toho, že $P(A) = p_1$ pak dostaneme, že $P(B) = P(B \cap \overline{A}) = P(B|\overline{A}) \cdot P(\overline{A}) = p_2(1-p_1)$. To je prostě dáno tím, že jevů B dovolíme nastat až tehdy pokud určitě nenastane A. Pak máme
  $$P(C) = P(A \cup B) = P(A) + P(B) = p_1 + (1-p_1)p_2$$

Jak je ale vidět, ať tak či tak, pravděpodobnost $P(C)$ nakonec vychází (očekávatelně) stejně

$$p := P(C) = p_1 + (1-p_1)p_2 = 0.05 + 0.95 \cdot 0.08 = 0.126\ .$$

Do třetice ještě uvedeme odvození, které bude fungovat v obou přístupech (a my pak nemusíme rozlišovat, jak to přesně s jevem B je). Všimněme si, že v obou případech totiž určitě vždy máme $P(B|\overline{A}) = p_2$. Pak z věty o úplné pravděpodobnosti dostaneme

$$P(C) = P(A \cup B) = \underbrace{P(A \cup B|A)}_{1} \cdot \underbrace{P(A)}_{p_1} + \underbrace{P(A \cup B|\overline{A})}_{P(B|\overline{A})=p_2} \cdot \underbrace{P(\overline{A})}_{1-p_1} = p_1 + (1-p_1)p_2\ .$$

(b) Pro veličinu
$$Y = \text{“\textit{pořadí kola, ve kterém se někdo poprvé trefí}”}$$
je $Y = X + 1$. Tedy $E(Y) = E(X+1) = E(X) + 1 = \frac{1-p}{p} + 1 = \frac{1}{p} = \frac{1}{0.126} \doteq 7.94$. Pro směrodatnou odchylku platí $\sigma(X) := \sqrt{D(X)} = \sqrt{\frac{1-p}{p^2}} = \frac{\sqrt{1-p}}{p} = \frac{\sqrt{0.874}}{0.126} \doteq 7.42$.

Směrodatná odchylka je norma veličiny $X - E(X)$, kterou chápeme (při určitém ztotožnění) jako prvek vektorového prostoru se skalárním součinem daným předpisem $Y \bullet Z = E(Y \cdot Z)$.

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

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

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

$$\begin{aligned} 0.1 &\ge (1-p)^n = (1-0.126)^n = 0.874^n \\ \log 0.1 &\ge n \log 0.874 \\ n &\ge \tfrac{\log 0.1}{\log 0.874} \doteq 17.097 \end{aligned}$$

Pozor: logaritmus hodnoty 0.874 je záporný! Tedy musí začít alespoň $n = 18$ kol.

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

$$Z_n = \text{“\textit{počet tolika kol (z n možných), ve kterých se někdo trefil}”}$$

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

$$P(Z_n \ge 1) \ge 0.9 \ .$$

Máme tedy

$$0.9 \le P(Z_n \ge 1) = 1 - P(Z_n = 0) = 1-(1-p)^n$$

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

6.2 (geometrické rozdělení - střední hodnota, rozptyl)

Revizor najde v dané tramvaji alespoň jednoho černého pasažéra s pravděpodobností p. Náhodná veličina X je počet tramvají, které revizor projde před tím, než najde černého pasažéra.

(a) Určete rozdělení veličiny X.

(b) Kolik *dalších* tramvají musí revizor průměrně projít než opět nalezne dalšího černého pasažéra? Jaký je rozptyl veličiny X? Spočítejte obecně a pak pro $p = 26\%$.

(c) Jestliže je $p = 26\%$, kolik musí revizor zkontrolovat minimálně tramvají, aby s pravděpodobností alespoň 95% našel alespoň jednoho černého pasažéra?

Řešení:

(a) Veličina X počítá počet neúspěchů (nenalezení černého pasažéra) než nastane úspěch. Má tedy (viz výše) geometrického rozdělení s oborem hodnot $\{0, 1, 2, \dots\}$ a pravděpodobnostní funkci

$$p_X(k) = P(X = k) = \begin{cases} p(1-p)^k & , k \in \mathbb{N}_0 \\ 0 & , \text{jinak} \ . \end{cases}$$

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).

Distribuční funkce je po částech konstantní. Pro $t < 0$ je $F_X(t) = 0$ a pro $t > 0$ máme

$$F_X(t) = \sum_{k \le t} p_X(k) = \sum_{k=0}^{[t]} p(1-p)^k = p \cdot \frac{1-(1-p)^{[t]+1}}{1-(1-p)} = 1-(1-p)^{[t]+1}$$

kde $[t]$ je celá část t (tj. zaokrouhleni desetinného čísla t směrem dolů).

(b) Ptáme se na střední hodnotu veličiny X (tj. otázku zde chápeme jako střední počet tramvají *předtím* než se najde černý pasažér. Pokud bychom chtěli započítat i tu tramvaj, kde buče černý pasažér, vezmeme veličinu $X+1$).

Tedy

$$E(X) = \frac{1-p}{p} = \frac{0.74}{0.26} \doteq 2.85$$

$$D(X) = \frac{1-p}{p^2} = \frac{0.74}{0.26^2} \doteq 10.95$$

(c) Pro $n \in \mathbb{N}$ zřejmě máme, že

"revizor v prvních n tramvajích najde alespoň jednoho černého pasažéra"

právě když platí

$$X \le n-1 \ .$$

Chceme tedy znát nejmenší $n_0 \in \mathbb{N}$ takové, že $P\Big(X \le n_0 - 1\Big) \ge 0.95$. Takže

$$0.95 \le P\big(X \le n_0 - 1\big) = F_X(n_0-1) = 1-(1-p)^{n_0}$$

neboli

$$0.05 \ge (1-p)^{n_0} = (1-0.26)^{n_0} = 0.74^{n_0}$$

$$\log 0.05 \ge n_0 \cdot \log 0.74$$

$$9.95 \doteq \frac{\log 0.05}{\log 0.74} \le n_0$$

Pozor, logaritmus je záporný pro hodnoty menší než 1! Revizor tedy musí projít alespoň $n_0 = 10$ tramvají.

Uvědomte si rozdíl mezi tímto číslem $n_0 = 10$ a středním počtem tramvají $E(X) \doteq 3$.

Poznámky k Poissonovu rozdělení: Veličina

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

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

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

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

- počet událostí může nabývat libovolných (konečných) hodnot.
- jednotlivé události nenastávají současně (lze je časově oddělit) a z každého zdroje nastane událost nejvýše jednou,
- 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 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:

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

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 četnost" události) - viz níže.

Odvození Poissonova rozdělení: Za výše uvedených předpokladů dojdeme ke tvaru Poissonova rozdělení takto:
Časový interval si rozdělíme na n dílků tak malých, aby v každém byla maximálně jedna událost se stejnou pravděpodobností p_n. Dostaneme tak binomické rozdělení veličiny

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

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:

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

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

Z této limity je vidět, že Poissonovo rozdělení se dá použít jako aproximace binomického rozdělení X_n když:

$$1 - \frac{\lambda}{n} \doteq 1, \qquad 1 - \frac{i}{n} \doteq 1 \quad \text{pro} \quad 0 \le i \le k \quad \text{a} \quad \left(1 - \frac{\lambda}{n}\right)^n \doteq e^{-\lambda} \quad \left(\text{neboli} \quad n \cdot \ln\left(1 - \frac{\lambda}{n}\right) \doteq -\lambda\right)$$

tedy stručně řečeno, když

$$\lambda << n \quad \text{a} \quad k << n\ .$$

Neboli, když střední hodnota $E(X_n)$ je malá ve srovnání s velkým počtem n a zajímají nás hodnoty pravděpodobnosti veličiny X_n pro malá k.

Shrnutí: Mějme veličinu $W \sim \mathrm{Bi}(n,p)$.

- Jestliže p je blízké 0, můžeme W aproximovat Poissonovým rozdělením $\mathrm{Poiss}(np)$, ale jen pro hodnoty $k << n$ veličiny W.
- Jestliže p je blízké 1, přejdeme k veličině $W - n \sim \mathrm{Bi}(n, 1-p)$, kde budeme mít opět $1-p$ blízké 0. Veličinu W můžeme aproximovat veličinou $Z - n$, kde $Z \sim \mathrm{Poiss}(n(1-p))$, ale jen pro hodnoty k veličiny W takové, že $n - k << n$.
- Jestliže se p pohybuje kolem $\frac{1}{2}$, pak se (pro větší n) veličina W aproximuje pomocí normálního rozdělení $N\big(np, np(1-p)\big)$.

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

Souvislost mezi exponenciálním a Poissonovým rozdělením: Nechť

$Y =$ *"doba čekání na událost"*

je veličina s exponenciálním rozdělením $\mathrm{Exp}(\tau)$ se střední hodnotou $E(Y) = \tau$ (s jednotkou "čas").

Pak veličina

$X =$ *"počet událostí během doby T"*

má Poissonovo rozdělení $\mathrm{Poiss}(\lambda)$ se střední hodnotou $E(X) = \lambda$ (s bezrozměrnou jednotkou) a platí

$$\lambda = \frac{T}{\tau} \; .$$

6.3 (Poissonovo a exponenciální rozdělení)

Během hodiny přijme telefonní operátor průměrně 5 hovorů.

(a) Jaká je pravděpodobnost, že jich za hodinu přijme méně než 3?

(b) Jaká je pravděpodobnost, že když si operátor na 10 min odskočí, nikdo mu nezavolá?

(c) Operátorovi je přidělena skupina $n = 300$ lidí. Každý z nich mu bude během hodiny volat s pravděpodobností $p = 0.01$ (nezávisle na ostatních a nejvýše jednou). Jaká je pravděpodobnost toho, že během hodiny zavolají právě 4 lidé z této skupiny?

Řešení:

(a) Máme tedy veličinu

$$X = \text{"}\textit{počet hovorů během } 1 \textit{ hodiny}\text{"} \;\; ,$$

s Poissonovým rozdělením $\mathrm{Poiss}(\lambda)$, kde $\lambda = E(X) = 5$ (jde o bezrozměrné číslo).

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!}\right) =$$

$$= e^{-5}\left(1 + 5 + \frac{5^2}{2}\right) = \tfrac{37}{2}e^{-5} \doteq 0.12465 \; .$$

(b) Tuto úlohu můžeme řešit jak s využitím exponenciálního rozdělení, tak Poissonova rozdělení. Exponenciální rozdělení bude mít veličina

$$Y = \text{"}\textit{doba čekání na příchod dalšího hlášení}\text{"}.$$

- *Pomocí exponenciálního:* Podle zadání má veličina X Poissonovo rozdělení s parametrem $\lambda = 5$. Tudíž náhodná veličina Y s exponenciálním rozdělením má parametr $\tau = \frac{T}{\lambda} = \frac{60\ min}{5} = 12\ min$, kde $T = 60\ min$. Hledaná pravděpodobnost je pak

$$P(Y > 10\ min) = 1 - P(Y \le 10\ min) = 1 - F_Y(10) = 1 - \left(1 - e^{-\frac{10}{12}}\right) = e^{-\frac{5}{6}} \doteq 0.4346$$

- *Pomocí Poissonova:* Uvažujme veličinu

$X' =$ *"počet hovorů během 10 minut"*

Hledaná pravděpodobnost je dána jako

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

kde λ' je parametr veličiny X' s Poissonovým rozdělením. K určení parametru λ' použijeme vztah mezi veličinami Y a X' a už známou hodnotu parametru veličiny X, tj. platí $\lambda' = \frac{T'}{\tau} = \lambda \cdot \frac{T'}{T}$ neboli

$$\frac{E(X')}{E(X)} = \frac{T'}{T}$$

což odpovídá i výše zmíněnému požadavku, že průměrný počet událostí v časovém intervalu je úměrný jeho délce. Konkrétně tedy $\lambda' = \frac{10\ min}{60\ min} \cdot 5 = \frac{5}{6}$ a hledaná pravděpodobnost je pak

$$P(X' = 0) = e^{-\frac{5}{6}} \doteq 0.4346.$$

(c) Zde máme veličinu

Z_n= *"počet hovorů během jedné hodiny ze skupiny n lidí"*

Jde o binomické rozdělení $\mathrm{Bi}(n, p)$ a pravděpodobnost bude

$$P(Z_n = 4) = \binom{n}{4} \cdot p^4 \cdot (1-p)^{n-4} = \binom{300}{4} \cdot 0.01^4 \cdot 0.99^{296} \doteq 0.168877\ .$$

Protože zde násobíme velká čísla malými a počítáme vysoké mocniny, nabízí se pro výpočet použít aproximaci pomocí Poissonova rozdělení se stejnou střední hodnotou jako má veličina Z_n, tj. parametr Poissonova rozdělení bude $\lambda = E(Z_n) = n \cdot p = 300 \cdot 0.01 = 3$.

Hodnoty $\lambda = 3$ i $k = 4$ jsou malé ve srovnání s $n = 300$. To nás opravňuje použít tuto aproximaci. Dostáváme tak

$$P(Z_n = 4) \doteq \frac{\lambda^4}{4!} \cdot e^{-\lambda} = \frac{3^4}{4!} \cdot e^{-3} \doteq 0.168031\ .$$

6.4 (Poissonovo a exponenciální rozdělení)

Do pojišťovny přijdou průměrně 2 hlášení škody denně.

(a) Jaká je pravděpodobnost, že za den přijdou alespoň 4?

(b) Jaká je pravděpodobnost, že do pojišťovny přijde nejbližší hlášení škody nejdříve třetí den?

(c) Agent pojišťovny spravuje skupinu $n = 500$ klientů. Každý z nich mu během dne nahlásí událost s pravděpodobností $p = 0.001$ (nezávisle na ostatních a nejvýše jednou). Jaká je pravděpodobnost toho, že se mu během dne ozve právě 5 klientů z této skupiny?

Řešení:

Postupujeme stejně jako v příkladu **6.4**.

(a) $X =$ *"počet hlášení za den"* , $X \sim \mathrm{Poiss}(\lambda)$, $\lambda = 2$.

$$P(X \geq 4) = 1 - P(X < 4) = 1 - e^{-\lambda}\Big(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\Big) = 1 - e^{-2}\Big(1 + 2 + \frac{2^2}{2} + \frac{2^3}{3!}\Big) =$$

$$= 1 - \tfrac{19}{3}e^{-2} \doteq 1 - 0.857123 = 0.142877\ .$$

(b) Pomocí exponenciálního:
Y= *"doba čekání na příchod dalšího hlášení"*, $Y \sim \mathrm{Exp}(\tau)$, $\tau = \frac{T}{\lambda} = \frac{1\ den}{2} = 0.5\ dne$, kde $T = 1\ den$.

$$P(Y > 2\ dny) = 1 - P(Y \leq 2\ dny) = 1 - F_Y(2) = 1 - \left(1 - e^{-\frac{2}{0.5}}\right) = e^{-4} \doteq 0.01831.$$

Pomocí Poissonova:
X'= *"počet hlášení během 2 dnů"*, $X \sim \mathrm{Poiss}(\lambda')$, $\lambda' = \lambda \cdot \frac{T'}{T} = 2 \cdot \frac{2\ dny}{1\ den} = 4$, kde $T' = 2\ dny$.

$$P(X' = 0) = \frac{(\lambda')^0}{0!} \cdot e^{-\lambda'} = e^{-\lambda'} = e^{-4} \doteq 0.01831.$$

(c) Z_n= *"počet hlášení ze skupiny n lidí za den"*, $Z_n \sim \mathrm{Bi}(n, p)$,

$$P(Z_n = 5) = \binom{n}{5} \cdot p^5 \cdot (1-p)^{n-5} = \binom{500}{5} \cdot 0.001^5 \cdot 0.999^{495} \doteq 1.5555 \cdot 10^{-4}\ .$$

Pomocí Poissonova rozdělení s $\lambda = E(Z_n) = n \cdot p = 500 \cdot 0.001 = 0.5$:

$$P(Z_n = 5) \doteq \frac{\lambda^5}{5!} \cdot e^{-\lambda} = \frac{0.5^5}{5!} \cdot e^{-0.5} \doteq 1.5795 \cdot 10^{-4}\ .$$