Pravděpodobnost a statistika

Diskrétní rozdělení

Vilém Vychodil

KMI/PRAS, Přednáška 6

Vytvořeno v rámci projektu 2963/2011 FRVŠ

Přednáška 6: Přehled

1. Přehled základních rozdělení:
   - diskrétní empirické rozdělení,
   - diskrétní uniformní rozdělení,
   - alternativní (Bernoulliho) rozdělení.
2. Binomické a geometrické rozdělení:
   - posloupnosti nezávislých Bernoulliho pokusů,
   - binomické rozdělení, střední hodnota, rozptyl,
   - modifikace úlohy a počítání počtu pokusů,
   - geometrické rozdělení, střední hodnota, rozptyl.
3. Poissonovo rozdělení:
   - problém počtu změn ve spojitém prostoru,
   - přibližný Poissonův proces,
   - Poissonovo rozdělení a význam jeho parametru,
   - vztah binomického a Poissonova rozdělení.

Opakování: Diskrétní náhodné veličiny

Definice (Náhodná veličina)

Mějme pravděpodobnostní prostor $\langle \Omega, \mathcal{F}, P \rangle$. Zobrazení $X : \Omega \to \mathbb{R}$ nazýváme **náhodnou veličinou** v $\langle \Omega, \mathcal{F}, P \rangle$ (angl.: *random variable*) pokud

$$\{\omega \in \Omega \,|\, X(\omega) \leq a\} \in \mathcal{F}.$$

platí pro každé $a \in \mathbb{R}$. Množinu reálných čísel $\{X(\omega) \,|\, \omega \in \Omega\}$ nazveme **prostor** nebo **obor hodnot** náhodné veličiny X, angl.: *space*.

Definice (Diskrétní náhodná veličina)

Náhodná veličina X s rozdělením P_X se nazývá **diskrétní** (angl.: *discrete*) pokud existuje spočetná množina $C \subseteq \mathbb{R}$ taková, že $P_X(C) = 1$.

Věta

X je diskrétní právě tehdy, když P_X je diskrétní pravděpodobnostní míra.

Příklad (Motivace pro diskrétní empirické rozdělení)

Uvažujme výběrový soubor obsahující počty dětí ve 100 vybraných rodinách:

$$
\begin{array}{ccccccccccccccccccccccccc}
4 & 6 & 2 & 7 & 2 & 9 & 3 & 4 & 2 & 1 & 5 & 4 & 1 & 3 & 2 & 5 & 2 & 2 & 3 & 6 & 3 & 3 & 5 & 2 & 3 \\
1 & 5 & 2 & 2 & 3 & 4 & 0 & 4 & 2 & 0 & 4 & 2 & 4 & 3 & 5 & 0 & 3 & 4 & 5 & 1 & 3 & 7 & 4 & 2 & 2 \\
4 & 3 & 5 & 3 & 6 & 2 & 3 & 3 & 2 & 9 & 4 & 4 & 2 & 5 & 2 & 2 & 4 & 2 & 2 & 3 & 1 & 4 & 3 & 3 & 2 \\
5 & 6 & 3 & 2 & 2 & 3 & 3 & 3 & 2 & 2 & 4 & 2 & 4 & 8 & 2 & 2 & 5 & 2 & 4 & 3 & 6 & 2 & 3 & 1 & 5
\end{array}
$$

Pohled na data: Výběr byl získán pozorováním hodnot náhodné veličiny X.

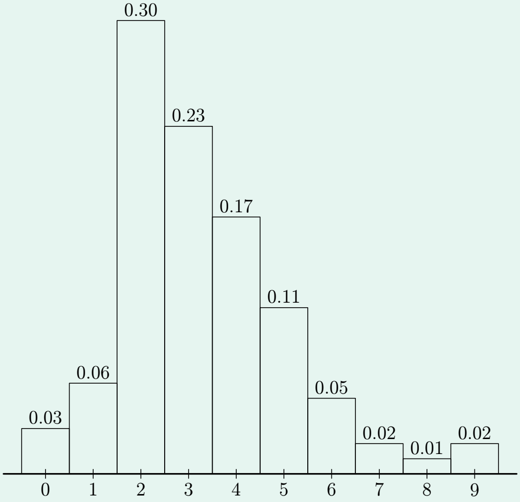
Otázka: *Jak na základě výběry získat (aproximaci) rozdělení X?*

Možnosti řešení:

- vyjádříme P_X jako rozdělení závislé na parametrech (ne vždy je tento postup možné provést, Přednáška 10);
- vyjádříme P_X jako (diskrétní) empirické rozdělení (relativní četnost hodnoty ve výběru interpretujeme jako pravděpodobnost).

Příklad (Tabulka a histogram relativních četností, Přednáška 1)

počet dětí	absolutní četnost	relativní četnost
0	3	0.03
1	6	0.06
2	30	0.30
3	23	0.23
4	17	0.17
5	11	0.11
6	5	0.05
7	2	0.02
8	1	0.01
9	2	0.02
$\sum$:	100	1.00



Diskrétní empirické rozdělení

Definice (Náhodná veličina s diskrétním empirickým rozdělením)

Uvažujme výběr $x_1, \ldots, x_m$ skládající se ze vzájemně různých hodnot $u_1, \ldots, u_n$, které mají relativní četnosti výskytu $f_1, \ldots, f_n$. Náhodná veličina X má **diskrétní empirické rozdělení** (angl.: *discrete empirical distribution*) P_X stanovené z výběru $x_1, \ldots, x_m$ pokud platí

$$P_X(A) = \sum\nolimits_{i=1}^{n} f_i \cdot \delta_{u_i}(A)$$

pro každou Borelovskou množinu $A \in \mathcal{B}$.

- P_X je diskrétní rozdělení, protože $P_X(C) = 1$ pro $C = \{u_1, \ldots, u_n\}$,
- $P(\{X = x\}) = P_X(x) = f_X(x) =$ relativní četnost hodnoty x ve výběru,
- *diskrétní rozdělení* P_X je *empirické* právě tehdy, když existuje *konečná* $C \subseteq \mathbb{R}$ tak, že $P_X(C) = 1$;
- úskalí (plyne z předchozího pozorování): *velmi obecný pojem.*

Příklad (Výpočet pravděpodobností)

Problém: Uvažujte data z předchozího příkladu a náhodnou veličinu X s příslušným diskrétním empirickým rozdělením.

Úkoly: Stanovte hodnoty následujících pravděpodobností:

1. *počet dětí je aspoň* 3,
2. *počet dětí je nejvýše* 4,
3. *počet dětí je nejvýše* 4, *ale vždy alespoň jedno,*
4. *buď žádné dítě, nebo aspoň* 5.

Řešení:

1. $P(\{X \geq 3\}) = 1 - P(\{X < 3\}) = 1 - (0.03 + 0.06 + 0.30) = 1 - 0.39 = 0.61$,
2. $P(\{X \leq 4\}) = 0.03 + 0.06 + 0.30 + 0.23 + 0.17 = 0.79$,
3. $P(\{0 < X \leq 4\}) = 0.06 + 0.30 + 0.23 + 0.17 = 0.76$,
4. $P(\{X = 0\} \cup \{X \geq 5\}) = 0.03 + 1 - P(\{X < 5\}) = 0.03 + (1 - 0.79) = 0.24$.

Diskrétní uniformní rozdělení

Definice (Náhodná veličina s diskrétním uniformním rozdělením)

Náhodná veličina X má **diskrétní uniformní rozdělení** (angl.: *discrete uniform distribution*) pokud existuje $m \in \mathbb{N}$ tak, že její pravděpodobnostní funkce f_X je

$$f_X(x) = \frac{1}{m} \qquad \text{pro } x \in 1, \ldots, m,$$

a $f_X(x) = 0$ jinak.

Poznámky:

- hodnota $m \in \mathbb{N}$ se nazývá *parametr* rozdělení;
- zřejmě $P_X(\{1, \ldots, m\}) = 1$, tedy X je zřejme diskrétní náhodná veličina;
- pravděpodobnost stejnoměrně koncentrovaná do bodů $1, \ldots, m$.

Věta (Druhý moment diskrétní uniformní veličiny)

Pokud má X diskrétní uniformní rozdělení s parametrem m, pak

$$E(X^2) = \frac{(m+1)(2m+1)}{6}.$$

Důkaz.

Za předpokladu, že platí pro X s m ověříme, že platí pro Y s parametrem $m+1$:

$$E(Y^2) = \sum_{y=1}^{m+1} \frac{y^2}{m+1} = \frac{1}{m+1} \cdot \left((m+1)^2 + \sum_{y=1}^{m} y^2\right) = \frac{(m+1)^2}{m+1} + \frac{m \cdot E(X^2)}{m+1}$$

$$= m + 1 + \frac{m(m+1)(2m+1)}{6(m+1)} = \frac{6(m+1)^2 + m(m+1)(2m+1)}{6(m+1)}$$

$$= \frac{6(m+1) + m(2m+1)}{6} = \frac{2m^2 + 7m + 6}{6} = \frac{(m+2)(2(m+1)+1)}{6}.$$

□

Věta (Střední hodnota a rozptyl diskrétní uniformní veličiny)

Pokud má X diskrétní uniformní rozdělení s parametrem m, pak

$$\mu_X = \frac{m+1}{2}, \qquad \sigma_X^2 = \frac{m^2-1}{12}.$$

Důkaz.

Využitím součtu prvků aritmetické posloupnosti dostáváme:

$$\mu_X = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m} \cdot \sum_{x=1}^{m} x = \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2}.$$

Vyjádřením rozptylu pomocí druhého momentu:

$$\sigma_X^2 = E(X^2) - \mu_X^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \frac{m^2-1}{12}.$$

□

Příklad (Házení s nefalšovanou kostkou)

Lze chápat jako výsledek diskrétní uniformní X s parametrem $m = 6$, tedy:

$$f_X(x) = \frac{1}{6}, \qquad \text{pro } x = 1, 2, 3, 4, 5, 6.$$

Pro distribuční funkci F_X tedy platí:

$$F_X(x) = \frac{x}{6}, \qquad \text{pro } x = 1, 2, 3, 4, 5, 6.$$

Střední hodnota a směrodatná odchylka:

$$\mu_X = \frac{m+1}{2} = \frac{6+1}{2} = \frac{7}{2} = 3.5, \qquad \sigma_X^2 = \frac{m^2-1}{12} = \frac{36-1}{12} = \frac{35}{12} \approx 2.917.$$

Pro $Y = 3X - 2$ máme (Přednáška 5):

$$\mu_Y = 3\mu_X - 2 = 10.5 - 2 = 8.5, \qquad \sigma_Y^2 = 3^2 \cdot \sigma_X^2 = 9 \cdot \frac{35}{12} = 26.25.$$

Alternativní (Bernoulliho) rozdělení

Definice (Náhodná veličina s alternativním rozdělením)

Náhodná veličina X má **alternativní rozdělení** (angl.: *Bernoulli distribution*) pokud existuje $p \in (0,1)$ tak, že její pravděpodobnostní funkce f_X je ve tvaru

$$f_X(x) = p^x \cdot (1-p)^{1-x} \qquad \text{pro } x = 0, 1,$$

a $f_X(x) = 0$ jinak.

Interpretace:

- náhodný pokus (Bernoulliho pokus) končící úspěchem 1 nebo neúspěchem 0;
- po provedení pokusu X nabude právě jedné z hodnot 0, 1;
- p je *parametr* interpretovaný jako **pravděpodobnost úspěchu**; platí totiž:

$$f_X(0) = p^0 \cdot (1-p)^1 = 1-p, \qquad f_X(1) = p^1 \cdot (1-p)^0 = p.$$

Věta (Střední hodnota a rozptyl veličiny s alternativním rozdělením)

Pokud má X alternativní rozdělení s parametrem p, pak $\mu_X = p$, $\sigma_X^2 = p \cdot (1-p)$.

Důkaz.

Vyjádřením μ_X přímo dostáváme:

$$\mu_X = \sum\nolimits_{x \in \{0,1\}} x \cdot p^x \cdot (1-p)^{1-x} = 0 + p^1 \cdot (1-p)^0 = p\,.$$

Rozptyl σ_X^2 lze analogicky vyjádřit jako

$$\begin{aligned}
\sigma_X^2 &= \sum\nolimits_{x \in \{0,1\}} (x - \mu_X)^2 \cdot f_X(x) = \sum\nolimits_{x \in \{0,1\}} (x-p)^2 \cdot p^x \cdot (1-p)^{1-x} \\
&= ((0-p)^2 \cdot p^0 \cdot (1-p)^{1-0}) + ((1-p)^2 \cdot p^1 \cdot (1-p)^{1-1}) \\
&= ((0-p)^2 \cdot (1-p)) + ((1-p)^2 \cdot p) = (p^2 \cdot (1-p)) + (p \cdot (1-p)^2) \\
&= p \cdot (p \cdot (1-p) + (1-p)^2) = p \cdot (p - p^2 + 1 - 2p + p^2) = p \cdot (1-p)\,.
\end{aligned}$$

□

Příklad (Posloupnost nezávislých Bernoulliho pokusů)

Pozorujeme výsledky několika Bernoulliho pokusů stejného typu (stejný parametr p).

Příklad: Provozovatel stírací loterie vydá celkem milion kusů stíracích losů, z toho jedna pětina je výherních.

Otázka: *Jaká je pravděpodobnost, že pokud si koupíme pět losů v řadě, tak právě čtvrtý z nich bude jediný výherní?*

Zjednodušení: Uvažujeme, že tažení jednoho výherního losu je Bernoulliho pokus s parametrem $p = 0.2$;

- podstata zjednodušení: nezáleží na tom, jaké losy už byly taženy;
- bez újmy, pokud jsou losy vybírány přibližně rovnoměrně $4 : 1$.

Za předpokladu nezávislosti jednotlivých výběrů, výsledná pravděpodobnost je:

$$0.8 \cdot 0.8 \cdot 0.8 \cdot 0.2 \cdot 0.8 = 0.2 \cdot 0.8^4 = 0.08192.$$

Příklad (Počet úspěchů v posloupnosti nezávislých Bernoulliho pokusů)

Pokračujeme v předchozím příkladu:

Otázka: *Pokud koupíme* 5 *losů, jaká je pravděpodobnost, že právě* 2 *budou výherní?*

Kolik je způsobů, jak dostat právě dva výherní losy mezi pěti koupenými?

- Tolik jako počet způsobů jak najít 2 pozice mezi 5 pozicemi $=$
- počet 2 prvkových kombinací z 5 (počet 2 prvkových podmnožin 5) $=$

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{120}{12} = 10\,.$$

Jaká je pravděpodobnost, že nastane jedna konkrétní z těchto 10 *možností?*

- $0.8^3 \cdot 0.2^2 = 0.512 \cdot 0.04 = 0.02048.$

Celkem:

$$\binom{5}{2} \cdot 0.8^3 \cdot 0.2^2 = 10 \cdot 0.02048 = 0.2048.$$

Binomické pokusy a binomické rozdělení

Definice (Náhodná veličina s binomickým rozdělením)

Náhodná veličina X má **binomické rozdělení** (angl.: *binomial distribution*) pokud existují $n \in \mathbb{N}$ a $p \in (0, 1)$ tak, že její pravděpodobnostní funkce f_X je

$$f_X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \qquad \text{pro } x = 0, 1, 2, \ldots, n,$$

a $f_X(x) = 0$ jinak; X se pak nazývá **binomická veličina** s rozdělením $b(n, p)$.

Binomický experiment (angl.: *binomial experiment*) je posloupnost pokusů:

- Bernoulliho pokus je proveden n krát (*parametr*);
- jednotlivé pokusy jsou nezávislé ve smyslu výskytu úspěchu či neúspěchu;
- pravděpodobost výskytu úspěchu každého z pokusů je rovna p (*parametr*);
- náhodná veličina $X =$ „počet úspěchů z celkových n pokusů“.

Příklad (f_X binomické proměnné je korektně definovaná)

Mějme binomickou náhodnou veličinu X s rozdělením $b(n, p)$.

Užitím binomické věty:

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$$

pro $a = p$ and $b = 1 - p$ dostáváme:

$$\sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x \cdot (1-p)^{n-x} = \big(p + (1-p)\big)^n = 1^n = 1.$$

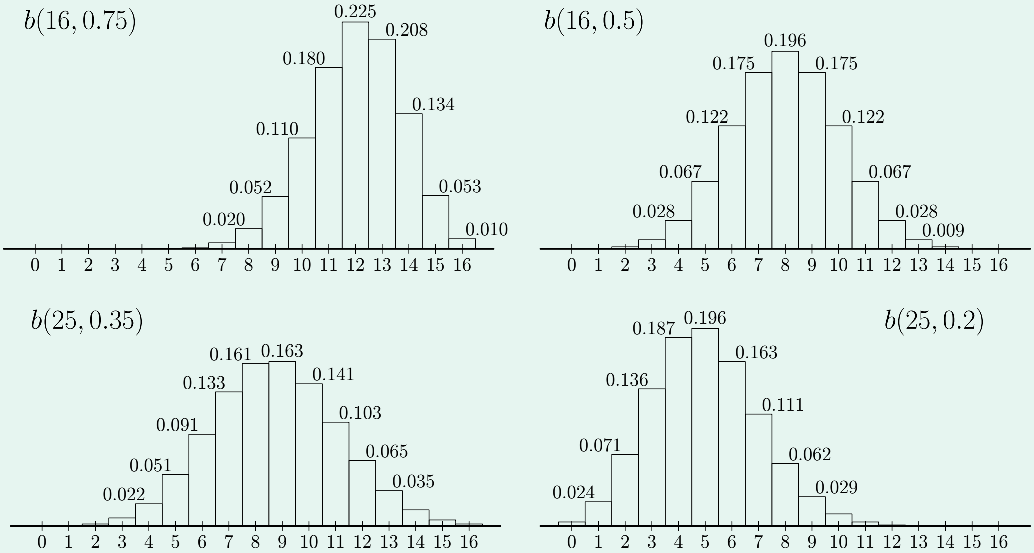
Zřejmě dále platí:

$$P_X(\{0, \ldots, n\}) = P(\{0 \leq X \leq n\}) = 1,$$
$$P_X(\{k\}) = P(\{X = k\}) = f_X(k).$$

Příklad (Speciální případy binomického rozdělení)

- $b(1, p)$ přechází v alternativní rozdělení s parametrem p.

Příklad (f_X pro veličiny s různým binomickým rozdělením)



Věta (Střední hodnota binomické veličiny)

Pokud má X rozdělení $b(n,p)$, pak $\mu_X = np$.

Důkaz (začátek).

S využitím krácení $\dfrac{x}{x!} = \dfrac{1}{(x-1)!}$ pro $x > 0$ dostáváme:

$$\mu_X = \sum\nolimits_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum\nolimits_{x=0}^{n} x \cdot \frac{n!}{x!(n-x)!} \cdot p^x (1-p)^{n-x}$$

$$= \sum\nolimits_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} \cdot p^x (1-p)^{n-x},$$

kde x jde od 1 do n. Předchozí můžeme ekvivalentně vyjádřit pro $k = x - 1$:

$$\mu_X = \sum\nolimits_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} \cdot p^{k+1} (1-p)^{n-k-1}.$$

Důkaz (dokončení).

S využitím předchozího vyjádření:

$$\mu_X = \sum_{k=0}^{n-1} \frac{n!}{k!(n-k-1)!} \cdot p^{k+1}(1-p)^{n-k-1}$$

můžeme dále vytknout np a zjednodušit:

$$\begin{aligned}\mu_X &= np \cdot \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} \cdot p^{k}(1-p)^{n-1-k} \\ &= np \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k}(1-p)^{n-1-k}.\end{aligned}$$

Použitím binomické věty dostáváme:

$$\mu_X = np \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k}(1-p)^{n-1-k} = np\big(p + (1-p)\big)^{n-1} = np \cdot 1 = np\,. \qquad \square$$

Věta (Druhý faktoriální moment binomické veličiny)

Pokud má X *rozdělení* $b(n, p)$, *pak* $E(X(X-1)) = n(n-1)p^2$.

Důkaz (začátek).

Použitím $\dfrac{x(x-1)}{x!} = \dfrac{1}{(x-2)!}$ pro $x > 1$ dostáváme

$$E(X(X-1)) = \sum\nolimits_{x=0}^{n} x(x-1)\binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum\nolimits_{x=0}^{n} \frac{x(x-1)n!}{x!(n-x)!} \cdot p^x (1-p)^{n-x} = \sum\nolimits_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} \cdot p^x (1-p)^{n-x}.$$

Dále pro $k = x - 2$ lze předchozí ekvivalentně zapsat jako:

$$E(X(X-1)) = \sum\nolimits_{k=0}^{n-2} \frac{n!}{k! \cdot (n-k-2)!} \cdot p^{k+2} \cdot (1-p)^{n-k-2}.$$

Důkaz (dokončení).

S využitím předchozího vyjádření:

$$E(X(X-1)) = \sum_{k=0}^{n-2} \frac{n!}{k!(n-k-2)!} \cdot p^{k+2}(1-p)^{n-k-2}$$

můžeme dále vytknout $n(n-1)p^2$ a zjednodušit:

$$\begin{aligned} E(X(X-1)) &= n(n-1)p^2 \cdot \sum_{k=0}^{n-2} \frac{(n-2)!}{k!(n-2-k)!} \cdot p^k(1-p)^{n-2-k} \\ &= n(n-1)p^2 \cdot \sum_{k=0}^{n-2} \binom{n-2}{k} p^k(1-p)^{n-2-k}. \end{aligned}$$

Analogicky jako v předchozím důkazu, použitím binomické věty dostáváme:

$$E(X(X-1)) = n(n-1)p^2 \cdot 1^{n-2} = n \cdot (n-1)p^2.$$

□

Věta (Rozptyl binomické veličiny)

Pokud má X rozdělení $b(n, p)$, pak $\sigma_X^2 = np(1-p)$.

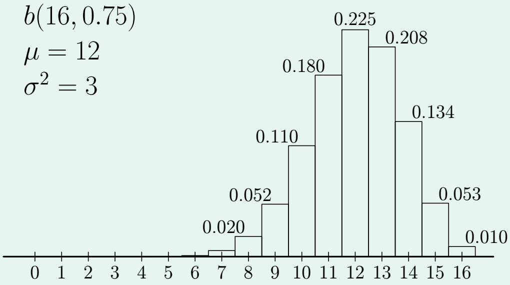
Důkaz.

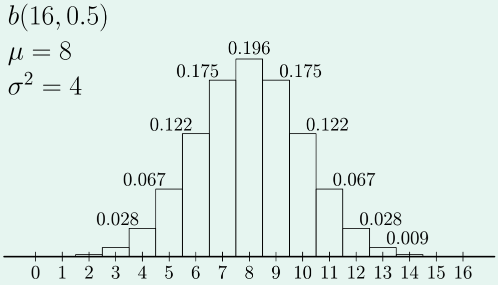
Použitím $\sigma_X^2 = E(X(X-1)) + \mu_X - \mu_X^2$ (Přednáška 5), to jest vyjádřením hodnoty rozptylu diskrétní náhodné veličiny X na základě druhého faktorálního momentu a střední hodnoty X dostáváme:

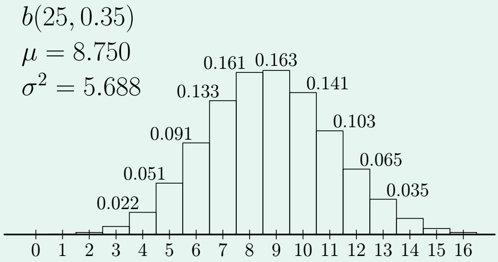
$$\begin{aligned}\sigma_X^2 &= E(X(X-1)) + \mu_X - \mu_X^2 = n(n-1)p^2 + np - (np)^2 \\ &= np((n-1)p + 1 - np) = np(np - p + 1 - np) = np(1-p)\,.\end{aligned}$$

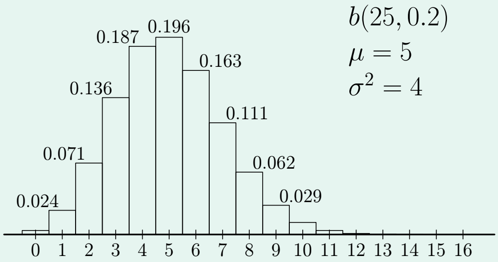
$\square$

Příklad (Hodnoty μ_X a σ^2_X pro různé binomické veličiny)









Příklad (Analýza pravděpodobnosti vzniku defektu při výrobě)

Problém: Dlouhodobým pozorováním jsme zjistili, že v průměru jeden výrobek z každých deseti vyrobených obsahuje defekt.

Otázka: *Jaká je pravděpodobnost, že pokud vezmeme náhodně pět výrobků, nejvýš jeden z nich bude obsahovat defekt?*

Analýza:

- X je náhodná veličina označující „počet defektních výrobků";
- pět nezávislých výběrů $n = 5$, pravděpodobnost nalezení defektu $p = 0.1$;
- X má binomické rozdělení $b(5, 0.1)$.

Řešení:

$$P(\{X \leq 1\}) = F_X(1) = \binom{5}{0}(0.1)^0(0.9)^5 + \binom{5}{1}(0.1)^1(0.9)^4 \approx 0.9185.$$

Pro X platí, že $\mu_X = 0.5$ a $\sigma_X^2 = 0.45$.

Příklad (Modifikace problému – počet nutných opakování)

Uvažujme posloupnost nezávislých Bernoulliho pokusů, ale modifikujeme otázku:

- původní otázka: *Kolik je úspěchů mezi n pokusy?*
- nová otázka: *Kolik je potřeba opakování pokusu, abychom viděli (první) úspěch?*

Rozdíl: Počet opakování Bernoulliho pokusu není předem dán (není omezen shora).

Otázka: *Jaká je pravděpodobnost, že počet potřebných opakování bude právě x?*

Vstupním parametrem je opět p (pravděpodobnost úspěchu jednoho pokusu).

Řešení:

- prvních $x - 1$ pokusů končí neúspěchem (pravděpodobnost $1 - p$);
- následující pokus končí úspěchem (pravděpodobnost p);
- s využitím nezávislosti jevů:

$$P(\{X = x\}) = (1 - p)^{x-1} \cdot p, \qquad \text{pro } x = 1, 2, \ldots$$

Geometrické rozdělení

Definice (Náhodná veličina s geometrickým rozdělením)

Náhodná veličina X má **geometrické rozdělení** (angl.: *geometric distribution*) pokud existuje $p \in (0, 1)$ tak, že její pravděpodobnostní funkce f_X je

$$f_X(x) = (1-p)^{x-1} \cdot p \qquad \text{pro } x = 0, 1, 2, 3, \ldots,$$

a $f_X(x) = 0$ jinak.

Pravděpodobnostní funkce f_X geometrické veličiny je dobře definovaná:

$$\sum_{x=1}^{\infty} f_X(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} \cdot p = \frac{p}{1-(1-p)} = 1,$$

což je speciální případ součtu prvků geometrické řady, protože pro $|r| < 1$ máme:

$$\sum_{k=1}^{\infty} a \cdot r^{k-1} = \sum_{k=0}^{\infty} a \cdot r^k = \frac{a}{1-r}.$$

Distribuční funkce geometrické náhodné veličiny

Použitím vztahu pro součet prvků geometrické řady můžeme vyjádřit:

$$P(\{X > k\}) = \sum_{x=k+1}^{\infty} (1-p)^{x-1} \cdot p = \frac{(1-p)^k \cdot p}{1-(1-p)} = (1-p)^k.$$

Odtud můžeme F_X zjednodušeně vyjádřit použitím faktu, že náhodné jevy $\{X \leq k\}$ a $\{X > k\}$ jsou komplementární:

Důsledek (Tvar distribuční funkce geometrické veličiny)

Pokud je X náhodná veličina s geometrickým rozdělením daným parametrem p, pak distribuční funkce F_X veličiny X je

$$F_X(k) = P(\{X \leq k\}) = 1 - P(\{X > k\}) = 1 - (1-p)^k.$$

Poznámka:

- hodnoty $P(\{X > k\})$ jsou v souladu s intuicí (alespoň k nezávislých neúspěchů)

Příklad (Pravděpodobnost nalezení klíčového slova)

Problém: Předpokládejme, že máme uloženo (mnoho) HTML souborů, ve kterých se zajímáme o pravděpodobnost výskytu klíčových slov. Dlouhodobým pozorováním jsme zjistili, že pravděpodobnost výskytu vybraného klíčového slova v náhodně zvoleném souboru je 0.25.

Úvaha: Pokud X označuje počet souborů, které je potřeba (náhodně) otevřít, než nelezneme soubor obsahující vybrané klíčové slovo, pak X je geometrická náhodná veličina s parametrem $p = 0.25$. Máme:

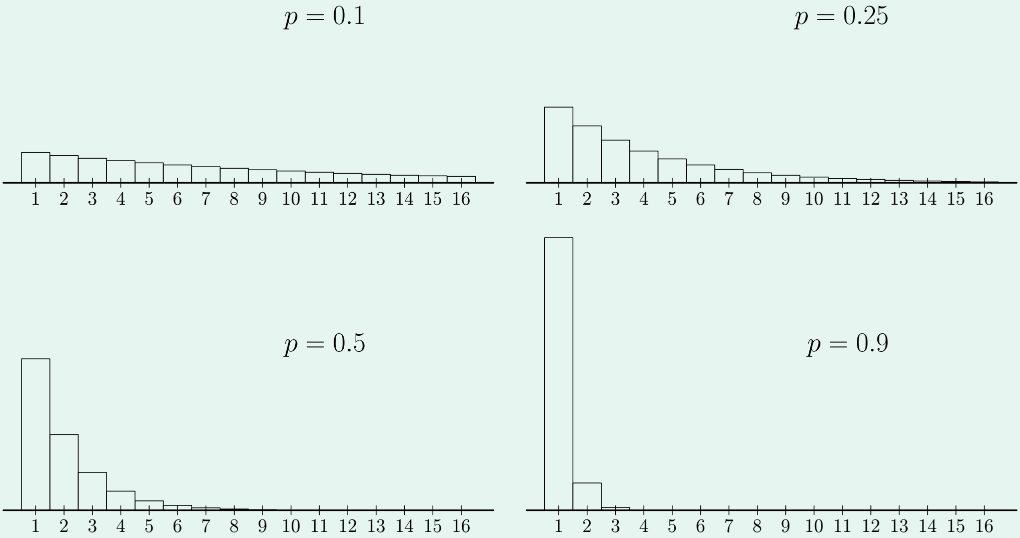
$$P(\{X \geq 4\}) = P(\{X > 3\}) = (1-p)^3 = 0.75^3 = \frac{27}{64} \approx 0.4219,$$

$$P(\{X \leq 4\}) = F_X(4) = 1 - P(\{X > 4\}) = 1 - 0.75^4 = \frac{175}{256} \approx 0.6836,$$

$$P(\{X = 4\}) = f_X(4) = (1-p)^{4-1} \cdot p = 0.75^3 \cdot 0.25 = \frac{27}{256} \approx 0.1055,$$

$$P(\{X = 4\}) = P(\{X \leq 4\}) - P(\{X \leq 3\}) = F_X(4) - F_X(3).$$

Příklad (f_X pro veličiny s různým geometrickým rozdělením)



Věta (Střední hodnota veličiny s geometrickým rozdělením)

Pokud má X geometrické rozdělení s parametrem p, pak $\mu_X = \frac{1}{p}$.

Důkaz.

Pro funkci g danou součtem prvků geometrické řady pro $|w| < 1$ platí:

$$g(w) = \sum_{k=0}^{\infty} a \cdot w^k = \frac{a}{1-w}.$$

Vyjádřením první derivace g v proměnné w dostáváme:

$$g'(w) = \sum_{k=1}^{\infty} a \cdot k \cdot w^{k-1} = \frac{a}{(1-w)^2}.$$

Speciálně pro $a = p$, $k = x$ a $w = 1 - p$ dostáváme:

$$\mu_X = \sum_{x=1}^{\infty} x \cdot (1-p)^{x-1} \cdot p = \frac{p}{(1-(1-p))^2} = \frac{p}{p^2} = \frac{1}{p}.$$

$\square$

Věta (Druhý faktoriální moment veličiny s geometrickým rozdělením)

Pokud má X geometrické rozdělení s parametrem p, pak

$$E(X(X-1)) = \frac{2(1-p)}{p^2}.$$

Důkaz.

Vyjádřením druhé derivace funkce g z předchozího důkazu dostáváme:

$$g''(w) = \sum\nolimits_{k=2}^{\infty} ak(k-1)w^{k-2} = \frac{2a}{(1-w)^3}.$$

Speciálně pro $a = p(1-p)$, $k = x$ a $w = 1-p$ dostáváme:

$$E(X(X-1)) = \sum\nolimits_{x=1}^{\infty} x(x-1)(1-p)^{x-1}p = \sum\nolimits_{x=2}^{\infty} p(1-p)x(x-1)(1-p)^{x-2}$$
$$= \frac{2p(1-p)}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}.$$

$\square$

Věta (Rozptyl náhodné veličiny s geometrickým rozdělením)

Pokud má X geometrické rozdělení s parametrem p, pak $\sigma_X^2 = \frac{1-p}{p^2}$.

Důkaz.

Vyjádřením σ_X^2 pomocí střední hodnoty a faktoriálního momentu dostáváme:

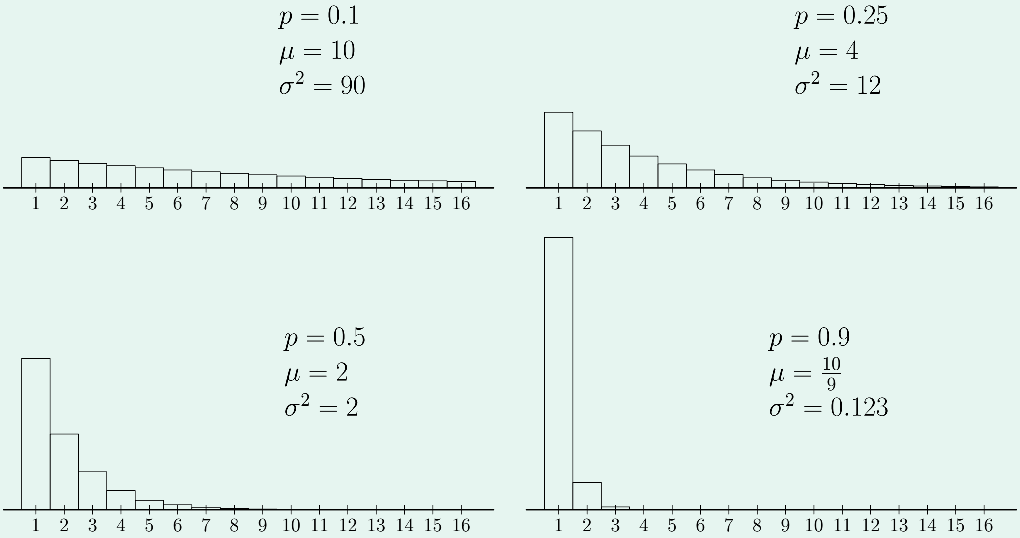
$$\sigma_X^2 = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2(1-p)+p-1}{p^2} = \frac{1-p}{p^2} = \frac{1-p}{p^2}.$$

□

Poznámky:

- vždy platí $f_X(1) = p$;
- neexistuje žádná konečná $C \in \mathcal{B}$ tak, že $P_X(C) = 1$;
- μ_X má dobrý intuitivní význam (pro $p = 0.25$ potřebujeme průměrně 4 pokusy);
- se vzrůstajícím p hodnota σ_X^2 rozptylu veličiny X klesá.

Příklad (Hodnoty μ_X a σ_X^2 pro X s geometrickým rozdělením)



Příklad (Nalezení osoby se stejným měsícem narození)

Problém: Mějme náhodnou veličinu X označující počet náhodně vybraných osob, které musíme oslovit, než nalezneme osobu narozenou v předem zvoleném měsíci.

Otázka: *Jaké má veličina X rozdělení?*

Analýza: X má geometrické rozdělení s parametrem $p = \frac{1}{12}$ (zjednodušení);

$$\mu_X = \frac{1}{\frac{1}{12}} = 12, \qquad \sigma_X^2 = \frac{1-p}{p^2} = \frac{\frac{11}{12}}{\left(\frac{1}{12}\right)^2} = 11 \cdot 12 = 132, \qquad \sigma_X = 11.489.$$

Dále máme například:

$$P(\{X > 23\}) = \left(\frac{11}{12}\right)^{23} \approx 0.1352,$$

$$P(\{X < 10\}) = 1 - P(\{X > 9\}) = 1 - \left(\frac{11}{12}\right)^{9} \approx 0.5430.$$

Motivace pro Poissonův proces

- počítání výskytu jevů (jako v případě binomických veličin),
- místo samostatných pokusů *počítáme ve spojitém prostoru*,
- spojitý prostor: časový úsek, vymezená část fyzického objektu, . . .

Příklad (Příklady počítání změn)

Může jít například o počítání

- počtu příchozích telefoních hovorů na jednu ústřednu mezi 9:00-15:00 hodinou;
- počtu fyzických poruch (poškození) na 100 metrech drátu;
- počtu zákazníků, kteří přijdou do obchodu poslední hodnu před zavírací dobou;
- počtu zrnek písku o průměru aspoň 1 mm v litru odebrané vody, . . .

Výskyt jevu se často interpretuje jako **změna** (oproti normálnímu stavu).

Přibližný Poissonův proces

Abychom mohli hovořit o (přibližném) Poissonově procesu, musí být splněny

Podmínky pro přibližný Poissonův proces

Předpokládejme, že počítáme počet změn ve spojitém prostoru reprezentovaném reálným intervalem. Řekneme, že výskyt změn ve spojitém intervalu odpovídá (**přibližnému**) **Poissonovu procesu** (angl.: *approximate Poisson process*) pokud existuje reálné číslo $\lambda > 0$ tak, že jsou splněny následující podmínky:

1. počty změn v disjunktních podintervalech jsou nezávislé;
2. pravděpodobnost, že nastane právě jedna změna v dostatečně malém podintervalu délky h je *přibližně* $\lambda \cdot h$;
3. pravděpodobnost, že nastanou dvě nebo více změn v dostatečně malém podintervalu je *v podstatě nulová*.

Pro naše účely je tato (zjednodušení) formulace Poissonova procesu dostatečná.

Příklad (Stanovení pravděpodobnosti počtu změn, *začátek*)

Problém: Předpokládáme přibližný Poissonův proces s parametrem λ.

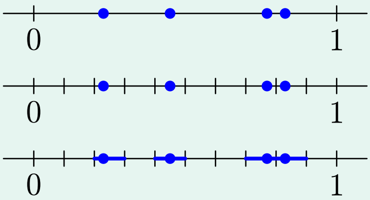
Jaká je pravděpodobnost, že počet změn (za danou jednotku času) je roven x?

Analýza: Uvažujme náhodnou veličini X označující počet změn za danou jednotku času. Pro jednoduchost budeme uvažovat, že spojitý prostor, ve kterém se pohybujeme je jednotkový interval $[0, 1] \subseteq \mathbb{R}$.

Úkolem je stanovit pravděpodobnost $P(\{X = x\})$ pro dané $x = 0, 1, 2, \ldots$

Jednotkový interval rozložíme na n podintervalů délek $\frac{1}{n}$ a uvažujeme:

Pokud je n *dost velké* (a tím pádem intervaly o délce $\frac{1}{n}$ dost malé), pak $P(\{X = x\})$ můžeme aproximovat jako pravděpodobnost, že *přesně v x podintervalech* (z celkového počtu n podintervalů) *nastane právě jedna změna.*



Příklad (Stanovení pravděpodobnosti počtu změn, *dokončení*)

Předchozí úvaha je korektní, protože:

1. pravděpodobnost, že *nastane právě jedna změna* v podintervalu délky $h = \frac{1}{n}$ (za předpokladu, že h je dost malé) je z podmínek přibližného Poissonova procesu rovna $\lambda \cdot h = \lambda \cdot \frac{1}{n}$;
2. výskyt změny v intervalu o délce $h = \frac{1}{n}$ lze chápat jako Bernoulliho pokus;
3. z podmínek přibližného Poissonova procesu plyne, že pozorování výskytů změn v různých intervalech lze chápat jako nezávislé pokusy.

Dohromady dostáváme, že $P(\{X = x\})$ může být aproximována jako pravděpodobnost, že náhodná veličina X_n s binomickým rozdělením $b(n, \frac{\lambda}{n})$ nabývá hodnoty x, to jest: $P(\{X = x\}) \approx P(\{X_n = x\})$. Čím větší n, tím lepší aproximace.

Zobecnění postupu z příkladu: $P(\{X = x\}) = \lim_{n \to \infty} P(\{X_n = x\})$.

Věta (Zákon vzácných jevů)

Pro náhodné veličiny X_n *s rozdělením* $b(n, \frac{\lambda}{n})$ *platí* $\lim\limits_{n\to\infty} P(\{X_n = x\}) = \dfrac{\lambda^x}{x!} \cdot e^{-\lambda}$.

Důkaz.

Pokud má X_n binomické rozdělení $b(n, \frac{\lambda}{n})$, pak platí

$$\lim_{n\to\infty} P(\{X_n = x\}) = \lim_{n\to\infty} \frac{n!}{x! \cdot (n-x)!} \cdot \left(\frac{\lambda}{n}\right)^x \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \frac{n \cdot (n-1) \cdots (n-x+1)}{n^x} \cdot \frac{\lambda^x}{x!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-x} = \frac{\lambda^x}{x!} \cdot e^{-\lambda},$$

což dostaneme jako přímý důsledek následujících rovností:

$$\lim_{n\to\infty} \frac{n \cdot (n-1) \cdots (n-x+1)}{n^x} = 1, \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}, \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1.$$

□

Poissonovo rozdělení

Definice (Náhodná veličina s Poissonovým rozdělením)

Náhodná veličina X má **Poissonovo rozdělení** (angl.: *Poisson distribution*) pokud existuje $\lambda > 0$ tak, že její pravděpodobnostní funkce f_X je ve tvaru

$$f_X(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!} \qquad \text{pro } x = 0, 1, \ldots$$

a $f_X(x) = 0$ jinak.

Poznámky:

- $f_X(x) > 0$ pro každé číslo $x = 0, 1, 2, \ldots,$
- $\lambda > 0$ je jediný parametr, který určuje rozdělení (nemusí být celočíselný),
- parametr λ: *průměrný počet změn v jednotce uvažovaného prostoru*,
- Poissonovo rozdělení vzniká limitním přechodem z binomických rozdělení.

Příklad (f_X je korektně definovaná)

Mějme náhodnou veličinu X s Poissonovým rozdělením s parametrem λ. Platí

$$\sum_{x=0}^{\infty} \frac{\lambda^x \cdot e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = e^{-\lambda+\lambda} = e^0 = 1,$$

což jsme získali vyjádřením e^{λ} prostřednictvím Maclaurinovy řady

$$g(x) = \sum_{n=0}^{\infty} \frac{g^{(n)}(0)}{x!} \cdot x^n = g(0) + g'(0) \cdot x + \frac{g''(0)}{2} \cdot x^2 + \frac{g'''(0)}{6} \cdot x^3 + \cdots$$

Z předchozího pro $g(x) = e^x$ speciálně dostáváme:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$$

To jest pravděpodobnostní funkce náhodné veličiny s Poissonovým rozdělením s parametrem $\lambda > 0$ je dobře definovaná, protože $\sum_{x=0}^{\infty} f_X(x) = 1$ a $f_X(x) \geq 0$ pro každé číslo $x \in \mathbb{R}$.

Rekurzivní vyjádření pravděpodobnostní funkce

Věta (Pravděpodobnostní funkce Poissonova rozdělení)

Pokud je f_X pravděpodobnostní funkce náhodné veličiny X s Poissonovým rozdělením daným parametrem λ, pak pro každé $x = 0, 1, 2, \ldots$ platí:

$$f_X(x) = \begin{cases} e^{-\lambda} & \textit{pokud } x = 0, \\ \dfrac{\lambda}{x} \cdot f_X(x-1) & \textit{jinak.} \end{cases}$$

Důkaz.

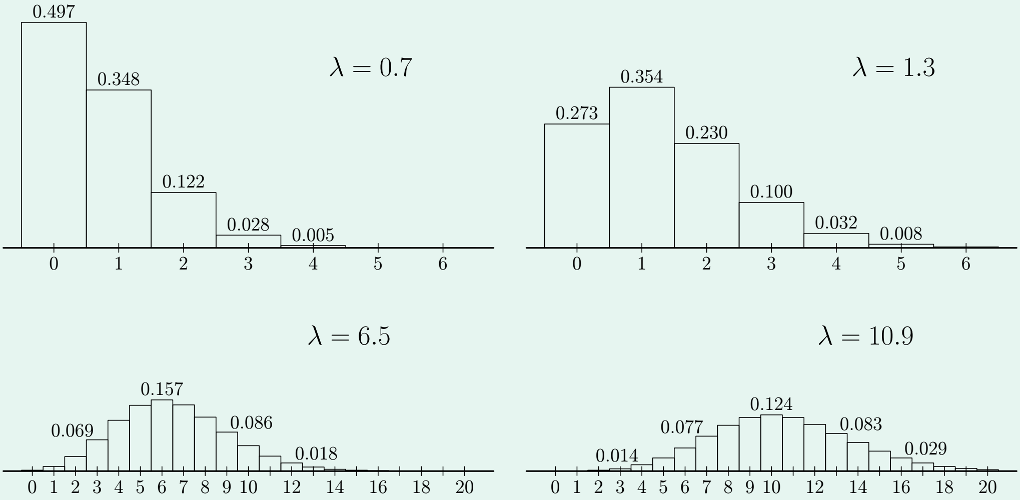
Přímo z předchozí definice dostáváme:

$$f_X(0) = \frac{\lambda^0 \cdot e^{-\lambda}}{0!} = \frac{1 \cdot e^{-\lambda}}{1} = e^{-\lambda},$$

$$f_X(x+1) = \frac{\lambda^{x+1} \cdot e^{-\lambda}}{(x+1)!} = \frac{\lambda \cdot \lambda^x \cdot e^{-\lambda}}{(x+1) \cdot x!} = \frac{\lambda}{x+1} \cdot \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{\lambda}{x+1} \cdot f_X(x).$$

□

Příklad (f_X pro veličiny s různým Poissonovým rozdělením)



Věta (Střední hodnota náhodné veličiny s Poissonovým rozdělením)

Pokud má X Poissonovo rozdělení s parametrem λ, pak $\mu_X = \lambda$.

Důkaz.

Využitím faktů, že $0 \cdot f_X(0) = 0$ a že pro $x > 0$ platí $\frac{x}{x!} = \frac{1}{(x-1)!}$, můžeme psát:

$$\mu_X = \sum_{x=0}^{\infty} x \cdot f_X(x) = \sum_{x=0}^{\infty} x \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum_{x=1}^{\infty} \frac{\lambda^x}{(x-1)!}.$$

Předchozí výraz můžeme vyjádřit pro $k = x - 1$ a zkrátit následujícím způsobem:

$$\mu_X = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} = \lambda \cdot e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda \cdot 1 = \lambda.$$

□

Poznámka: Naše interpretace λ jako „průměrného počtu změn“ je správná.

Věta (Druhý faktoriální moment veličiny s Poissonovým rozdělením)

Pokud má X Poissonovo rozdělení s parametrem λ, pak $E(X(X-1)) = \lambda^2$.

Důkaz.

Užitím faktu, že $\dfrac{x(x-1)}{x!} = \dfrac{1}{(x-2)!}$ pro $x > 1$ dostáváme:

$$E(X(X-1)) = \sum\nolimits_{x=0}^{\infty} x(x-1) \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum\nolimits_{x=2}^{\infty} \frac{\lambda^x}{(x-2)!},$$

Vyjádřením předchozího vztahu pro $k = x - 2$ můžeme provést zjednodušení:

$$E(X(X-1)) = e^{-\lambda} \cdot \sum\nolimits_{k=0}^{\infty} \frac{\lambda^{k+2}}{k!} = \lambda^2 e^{-\lambda} \cdot \sum\nolimits_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2 \cdot 1 = \lambda^2.$$

$\square$

Věta (Rozptyl náhodné veličiny s Poissonovým rozdělením)

Pokud má X Poissonovo rozdělení s parametrem λ, pak $\sigma_X^2 = \lambda$.

Důkaz.

Vyjádřením σ_X^2 pomocí střední hodnoty a faktorálního momentu dostáváme:

$$\sigma_X^2 = E(X(X-1)) + \mu_X - \mu_X^2 = \lambda^2 + \lambda - \lambda^2 = \lambda\,.$$

□

Poznámka (Počet změn v obecně velkém intervalu)

Pokud je průměrný počet změn v jednotkovém intervalu roven λ, pak v intervalu délky t (jednotek) lze očekávat průměrný počet změn roven $\lambda \cdot t$.

Pravděpodobnostní funkci f_X lze potom psát:

$$\boxed{f_X(x) = \frac{(\lambda t)^x \cdot e^{-\lambda t}}{x!}}\,, \quad \text{pro } x = 0, 1, 2, \ldots$$

Příklad (Počet chyb během datového přenosu)

Problém: Při bezdrátovém přenosu informace dojde v průměru k jedné chybě (bitová 0 se změní na bitovou 1 nebo obráceně) za $1\,200$ mikrosekund. *Jaká je pravděpodobnost, že během* $4\,800$ *mikrosekund*

1. *nedojde k žádné chybě,*
2. *dojde k nejvýše čtyřem chybám?*

Analýza: Za předpokladu, že počet chyb vyskytujících se v čase splňuje podmínky přibližného Poissonova procesu, má náhodná veličina X (počet chyb) Poissonovo rozdělení s parametrem $\lambda = 4$ (protože $4 \cdot 1\,200 = 4\,800$).

Řešení:

$$P(\{X = 0\}) = f_X(0) = \frac{4^0 \cdot e^{-4}}{0!} = e^{-4} = 0.018,$$

$$P(\{X \leq 4\}) = F_X(4) = \sum_{x=0}^{4} f_X(x) = \sum_{x=0}^{4} \frac{4^x \cdot e^{-4}}{x!} = 0.629.$$

Příklad (Vytížení webového serveru)

Problém: Webový server obslouží během tří hodin průměrně $2\,250$ klientů.

Otázka: *Jaká je pravděpodobnost, že webový server během sedmi minut obslouží více jak sto klientů?*

Řešení: Za předpokladu, že počet obsloužených klientů splňuje podmínky přibližného Poissonova procesu, má náhodná veličina X (počet obsloužených klientů) Poissonovo rozdělení s parametrem

$$\lambda = E(X) = \frac{2\,250}{3 \cdot 60} \cdot 7 = 87.5,$$

to jest $P(\{X > 100\})$ lze vyjádřit jako

$$1 - P(\{X \leq 100\}) = 1 - \sum_{x=0}^{100} \frac{87.5^x \cdot e^{-87.5}}{x!} = 1 - 0.915 = 0.085.$$

Aproximace binomických pravděpodobností

Pozorování: Pokud má X Poissonovo rozdělení s parametrem λ, pak

$$P(\{X = x\}) \approx \binom{n}{x} \cdot \left(\frac{\lambda}{n}\right)^{x} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-x},$$

pro *velká* n a *pravděpodobnost úspěchu* $p = \dfrac{\lambda}{n}$.

Důsledek:

Pokud má X binomické rozdělení $b(n, p)$, kde n *je dost velké* a p *je dost malé*, pak

$$f_X(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \approx \frac{(n \cdot p)^x \cdot e^{-n \cdot p}}{x!} = f_{X'}(x),$$

kde X' má Poissonovo rozdělení s parametrem $\lambda = n \cdot p$.

Poznámka: Dobrá aproximace pro $n \geq 20$ a $p \leq 0.05$, $n \geq 100$ a $p \leq 0.10, \ldots$

Příklad (Počty závadných čipů ve výrobě)

Problém: Výrobce počítačových čipů analýzou výrobního procesu zjistil, že $2\,\%$ vyrobených čipů je závadných.

Otázka: *Jaká je pravděpodobnost, že tovární balení obsahující* 100 *vyrobených čipů obsahuje nejvýše tři závadné?*

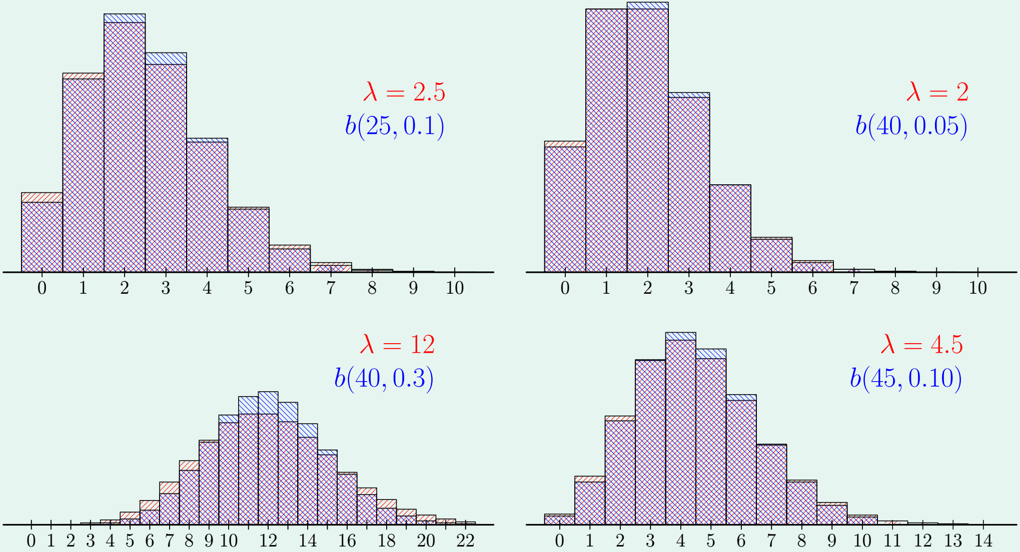
Řešení: Za předpokladu nezávislosti výskuty závad má náhodná veličina X (počet závadných čipů) Binomické rozdělení s parametry $p = 0.02$ (pravděpodobnost závady) a $n = 100$ (počet pozorování). Odtud

$$P(\{X \leq 3\}) = \sum_{x=0}^{3} f_X(x) = \sum_{x=0}^{3} \binom{100}{x} \cdot 0.02^x \cdot 0.98^{100-x} = 0.859.$$

Pokud má X Poissonovo rozdělení s parametrem $\lambda = 100 \cdot 0.02 = 2$, pak

$$P(\{X \leq 3\}) = \sum_{x=0}^{3} f_X(x) = \sum_{x=0}^{3} \frac{2^x \cdot e^{-2}}{x!} = 0.857.$$

Příklad (Poissonovo × binomické rozdělení)



Přednáška 6: Závěr

Pojmy:

- diskrétní uniformní rozdělení, alternativní (Bernoulliho) rozdělení
- posloupnost nezávislých Bernoulliho pokusů, binomické rozdělení
- geometrické rozdělení, přibližný Poissonův proces, Poissonovo rozdělení

Použité zdroje:

Capinski M., Zastawniak T. J.: *Probability Through Problems*
Springer 2001, ISBN 978–0–387–95063–1.

Devore J. L.: *Probability and Statistics for Engineering and the Sciences*
Duxbury Press, 7. vydání 2008, ISBN 978–0–495–55744–9.

Hogg R. V., Tanis E. A.: *Probability and Statistical Inference*
Prentice Hall; 7. vydání 2005, ISBN 978–0–13–146413–1.

Johnson J. L.: *Probability and Statistics for Computer Science*
Wiley-Interscience 2008, ISBN 978–0–470–38342–1.