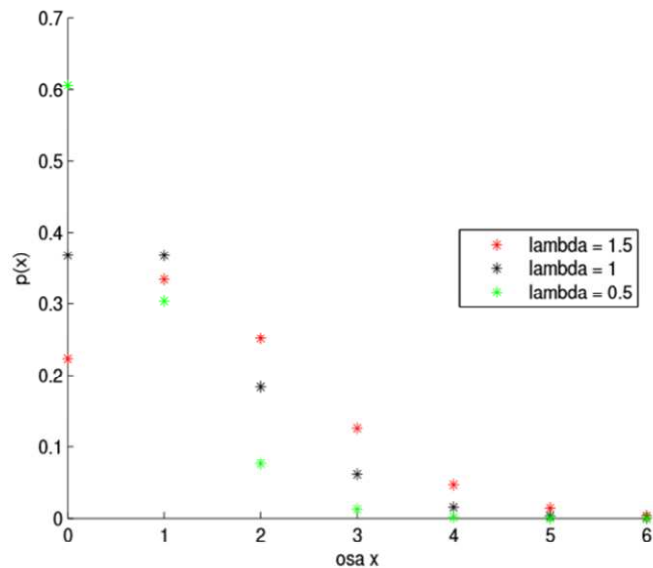


## 4. Poissonovo rozložení a jeho vlastnosti

**4.1. Definice:** Diskrétní náhodná veličina  $X$  má Poissonovo rozložení s parametrem  $\lambda > 0$ , jestliže pravděpodobnostní funkce  $\pi(x)$  má tvar:

$$\pi(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{jinak} \end{cases} . \text{ Zkráceně píšeme } X \sim \text{Po}(\lambda) .$$

Průběh pravděpodobnostní funkce Poissonova rozložení pro různé hodnoty parametru  $\lambda$ :



**4.2. Poznámka:** Lze odvodit, že náhodná veličina  $X \sim \text{Po}(\lambda)$  má tyto číselné charakteristiky:

a) střední hodnota  $E(X) = \lambda$

b) rozptyl  $D(X) = \lambda$

c) šikmost  $\alpha_3(X) = \frac{1}{\sqrt{\lambda}}$

d) špičatost  $\alpha_4(X) = \frac{1}{\lambda}$

**4.3. Poznámka:** Náhodná veličina  $X \sim \text{Po}(\lambda)$  udává počet událostí, které nastanou v jednotkovém časovém intervalu resp. jednotkové oblasti, přičemž tyto události nastávají náhodně, jednotlivě a nezávisle na sobě. Parametr  $\lambda > 0$  udává střední hodnotu (i rozptyl) počtu událostí.

Poissonovým rozložením se řídí např.

- počet výzev, které dojdou na TÚ za jednotkový časový interval
- počet mikroorganismů v jednotkové oblasti zorného pole mikroskopu
- počet požadavků v systému hromadné obsluhy za jednotkový časový interval
- atd.

**Upozornění:** Pokud  $X$  udává počet událostí, které nastanou v časovém intervalu délky  $t$  a střední hodnota počtu událostí v jednotkovém časovém intervalu je  $\lambda$ , pak  $X \sim \text{Po}(\lambda t)$ .

**4.4. Věta:** Necht'  $X \sim \text{Po}(\lambda)$  a  $Y \sim \text{Bi}(n, \vartheta_n)$ , přičemž  $\lim_{n \rightarrow \infty} \vartheta_n = 0$  a  $\lim_{n \rightarrow \infty} n\vartheta_n = \lambda$ . Pak pravděpodobnostní funkce náhodné veličiny  $Y$  konverguje pro  $n \rightarrow \infty$  k pravděpodobnostní funkci náhodné veličiny  $X$ .

**Důkaz:**

$$\begin{aligned} \lim_{n \rightarrow \infty} \binom{n}{y} \vartheta_n^y (1 - \vartheta_n)^{n-y} &= \lim_{n \rightarrow \infty} \frac{n!}{y!(n-y)!} \vartheta_n^y (1 - \vartheta_n)^{n-y} = \lim_{n \rightarrow \infty} \frac{n!}{y!(n-y)!} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y} = \\ &= \frac{\lambda^y}{y!} \lim_{n \rightarrow \infty} \frac{n \cdot (n-1) \dots (n-y+1)}{n^n} \left(1 - \frac{\lambda}{n}\right)^{n-y} = \frac{\lambda^y}{y!} \lim_{n \rightarrow \infty} \left[ 1 \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{y-1}{n}\right) \right] \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-y} = \\ &= \frac{\lambda^y}{y!} e^{-\lambda} \end{aligned}$$

**Upozornění:** Binomické rozložení můžeme dobře aproximovat Poissonovým rozložením, když pravděpodobnost výskytu jevu v jednom pokusu je velmi malá ( $\vartheta \leq 0,1$ ) a zároveň počet pokusů je dostatečně velký ( $n \geq 30$ ).

**4.5. Příklad:** Pravděpodobnost závady na 1 km telefonního kabelu je 0,01. Kabel je složen ze 100 vodičů a bude plnit svou funkci, když aspoň 99 vodičů bude v pořádku. Jaká je pravděpodobnost, že kabel bude plnit svou funkci?

**Řešení:** Označme  $Y$  náhodnou veličinu, která udává počet porouchaných vodičů. Z podmínek úlohy plyne, že  $Y \sim \text{Bi}(100;0,01)$ .

Přesný výpočet:

$$P(Y \leq 1) = \sum_{y=0}^1 \binom{100}{y} 0,01^y 0,99^{100-y} = \text{binocdf}(1,100,0.01) = 0,7358$$

Aproximace Poissonovým rozložením: podmínky  $n \geq 30$  a  $\vartheta \leq 0,1$  jsou splněny. Přitom  $\lambda = n \cdot \vartheta = 100 \cdot 0,01 = 1$ .

$$P(Y \leq 1) \approx \sum_{y=0}^1 \frac{1^y}{y!} e^{-1} = \text{poisscdf}(1,1) = 0,7358$$

Pravděpodobnost, že kabel bude plnit svou funkci, je 0,7358.

**4.6. Věta:** Necht'  $X \sim \text{Po}(\lambda)$ . Pak pro modus  $\hat{x}$  platí:  $\lambda - 1 \leq \hat{x} \leq \lambda$ . Je-li  $\lambda$  přirozené číslo, pak existují dvě modální hodnoty. Není-li  $\lambda$  přirozené číslo, je Poissonovo rozložení unimodální.

**Důkaz:** Protože modus je nejpravděpodobnější hodnota, musí současně vyhovovat dvěma nerovnostem:

$$\pi(\hat{x}) \geq \pi(\hat{x} - 1) \wedge \pi(\hat{x}) \geq \pi(\hat{x} + 1), \text{ tj. } \frac{\pi(\hat{x} - 1)}{\pi(\hat{x})} \leq 1 \wedge \frac{\pi(\hat{x} + 1)}{\pi(\hat{x})} \leq 1.$$

Dosadíme za pravděpodobnostní funkci a postupně upravujeme:

$$\frac{\frac{\lambda^{\hat{x}-1} e^{-\lambda}}{(\hat{x}-1)!}}{\frac{\lambda^{\hat{x}} e^{-\lambda}}{\hat{x}!}} \leq 1 \wedge \frac{\frac{\lambda^{\hat{x}+1} e^{-\lambda}}{(\hat{x}+1)!}}{\frac{\lambda^{\hat{x}} e^{-\lambda}}{\hat{x}!}} \leq 1$$

$$\frac{\hat{x}}{\lambda} \leq 1 \wedge \frac{\lambda}{\hat{x}+1} \leq 1 \Rightarrow \hat{x} \leq \lambda \wedge \hat{x} \geq \lambda - 1, \text{ tj. } \lambda - 1 \leq \hat{x} \leq \lambda.$$

**4.7. Příklad:** K holiči chodí průměrně 6 zákazníků za 1 h. Určete nejpravděpodobnější počet zákazníků u holiče během půl hodiny a určete pravděpodobnost tohoto počtu.

**Řešení:** Náhodná veličina  $X$  udává počet zákazníků u holiče během 1/2 h,  $X \sim \text{Po}(3)$ . Protože  $\lambda$  je přirozené číslo, existují dvě modální hodnoty, a to 2 a 3. Jejich pravděpodobnosti:

$$P(X = 2) = \frac{3^2}{2!} e^{-3} = 4,5e^{-3} = 0,224$$

$$P(X = 3) = \frac{3^3}{3!} e^{-3} = 4,5e^{-3} = 0,224$$

**4.8. Věta:** Necht'  $X_1, X_2$  jsou stochasticky nezávislé náhodné veličiny,  $X_i \sim \text{Po}(\lambda_i)$ ,  $i = 1, 2$ . Pak transformovaná náhodná veličina  $Y = X_1 + X_2 \sim \text{Po}(\lambda_1 + \lambda_2)$ .

**Důkaz:**

Podle věty o konvoluci dostáváme:

$$\begin{aligned} \pi_*(y) &= \sum_{x_1=-\infty}^{\infty} \pi_1(x_1)\pi_2(y-x_1) = |x_1 > 0, y-x_1 > 0 \Rightarrow 0 < x_1 < y| = \sum_{x_1=0}^y \frac{\lambda_1^{x_1}}{x_1!} e^{-\lambda_1} \frac{\lambda_2^{y-x_1}}{(y-x_1)!} e^{-\lambda_2} = \\ &= \frac{1}{y!} e^{-(\lambda_1+\lambda_2)} \sum_{x_1=0}^y \binom{y}{x_1} \lambda_1^{x_1} \lambda_2^{y-x_1} = \frac{(\lambda_1 + \lambda_2)^y}{y!} e^{-(\lambda_1+\lambda_2)} \text{ pro } y = 0, 1, 2, \dots, = 0 \text{ jinak} \end{aligned}$$

Vidíme, že  $Y \sim \text{Po}(\lambda_1 + \lambda_2)$ .

**4.9. Poznámka:** Tvrzení věty 4.8. lze zobecnit i na  $n$  stochasticky nezávislých veličin

$X_1, \dots, X_n$ ,  $X_i \sim \text{Po}(\lambda_i)$ ,  $i = 1, \dots, n$ . Pak transformovaná náhodná veličina

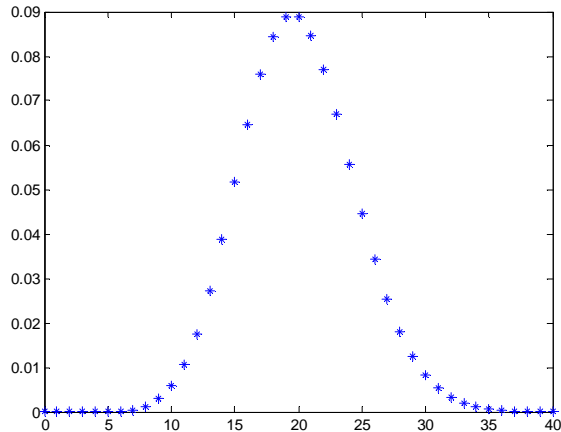
$Y = \sum_{i=1}^n X_i \sim \text{Po}\left(\sum_{i=1}^n \lambda_i\right)$ . Znamená to, že Poissonovo rozložení je uzavřené vzhledem k operaci sčítání.

**4.10. Věta:** Necht'  $X \sim \text{Po}(\lambda)$ , přičemž  $\lambda$  je přirozené číslo větší než 9. Pak rozložení náhodné veličiny  $X$  lze aproximovat rozložením  $N(\lambda, \lambda)$ .

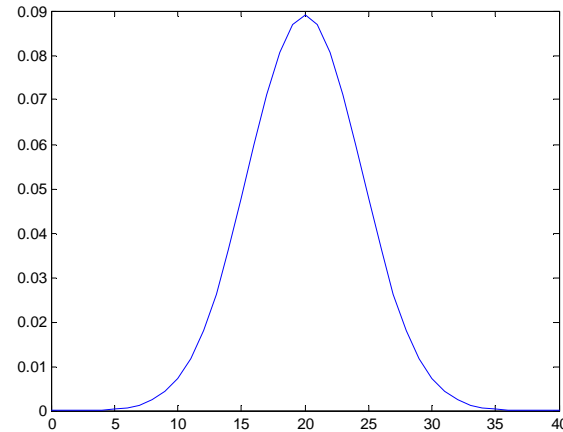
**Důkaz:** Podle poznámky 4.9. je  $X = \sum_{i=1}^{\lambda} X_i$ , přičemž stochasticky nezávislé náhodné veličiny  $X_1, \dots, X_{\lambda}$  se řídí rozložením  $\text{Po}(1)$ ,  $E(X_i) = 1$ ,  $D(X_i) = 1$ ,  $i = 1, \dots, \lambda$ . Podle CLV dostáváme, že

standardizovaná veličina 
$$U = \frac{\sum_{i=1}^{\lambda} X_i - \sum_{i=1}^{\lambda} E(X_i)}{\sqrt{\sum_{i=1}^{\lambda} D(X_i)}} = \frac{X - \lambda}{\sqrt{\lambda}} \approx N(0,1)$$
, tedy  $X = \sqrt{\lambda}U + \lambda \approx N(\lambda, \lambda)$ .

Graf pravděpodobnostní funkce  $\text{Po}(20)$



Graf hustoty  $N(20,20)$





**4.11. Věta:** Necht'  $X \sim \text{Po}(\lambda)$ , přičemž  $\lambda$  je přirozené číslo větší než 9. Pak pro nezáporná celá čísla  $a, b$ ,  $a < b$  platí:

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a-\lambda}{\sqrt{\lambda}}\right), \text{ kde } \Phi \text{ je distribuční funkce rozložení } N(0, 1).$$

**Důkaz:**

$$P(a \leq X \leq b) = P\left(\frac{a - E(X)}{\sqrt{D(X)}} \leq X \leq \frac{b - E(X)}{\sqrt{D(X)}}\right) = P\left(\frac{a - \lambda}{\sqrt{\lambda}} \leq X \leq \frac{b - \lambda}{\sqrt{\lambda}}\right) \approx \Phi\left(\frac{b - \lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \lambda}{\sqrt{\lambda}}\right)$$

**4.12. Poznámka:** Aproximace Poissonova rozložení normálním rozložením se zlepšuje, když použijeme tzv. opravu na nespojitost:

$$P(a \leq X \leq b) \approx \Phi\left(\frac{b - \lambda + \frac{1}{2}}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a - \lambda - \frac{1}{2}}{\sqrt{\lambda}}\right).$$

**4.13. Příklad:** Necht'  $X \sim \text{Po}(12)$ . Pomocí aproximace normálním rozložením stanovte  $P(8 \leq X \leq 20)$ .

**Řešení:**

$$\begin{aligned} P(8 \leq X \leq 20) &\approx \Phi\left(\frac{20 - 12 + \frac{1}{2}}{\sqrt{12}}\right) - \Phi\left(\frac{8 - 12 - \frac{1}{2}}{\sqrt{12}}\right) = \Phi(2,45) - \Phi(-1,3) = \\ &= \Phi(2,45) - 1 + \Phi(1,3) = 0,99286 - 1 + 0,9034 = 0,89606 \end{aligned}$$

Pro porovnání provedeme přesný výpočet:

$$P(8 \leq X \leq 20) = \sum_{x=8}^{20} \frac{12^x}{x!} e^{-12} = \text{poisscdf}(20,12) - \text{poisscdf}(7,12) = 0,8989$$

**4.14. Věta:** Necht'  $X_1, \dots, X_n$  je náhodný výběr z rozložení  $Po(\lambda)$ , přičemž  $n\lambda > 9$ . Označme

$M = \frac{1}{n} \sum_{i=1}^n X_i$  výběrový průměr. Pak meze  $100(1-\alpha)\%$  asymptotického intervalu spolehlivosti pro střední hodnotu  $\lambda$  jsou:

$$D = M - \sqrt{\frac{M}{n}} u_{1-\alpha/2}, \quad H = M + \sqrt{\frac{M}{n}} u_{1-\alpha/2}$$

**Důkaz:** Podle centrální limitní věty  $M \approx N(E(M), D(M))$ , kde  $E(M) = \lambda$ ,  $D(M) = \frac{\lambda}{n}$ .

Standardizací  $M$  dostaneme  $U = \frac{M - \lambda}{\sqrt{\frac{\lambda}{n}}} \approx N(0,1)$ . Konvergence k  $N(0,1)$  se neporuší, když  $\lambda$  ve

jmenovateli nahradíme  $M$ , tedy  $U = \frac{M - \lambda}{\sqrt{\frac{M}{n}}} \approx N(0,1)$ . Pak platí:

$$\forall \lambda > 0: 1 - \alpha \leq P \left( -u_{1-\alpha/2} < \frac{M - \lambda}{\sqrt{\frac{M}{n}}} < u_{1-\alpha/2} \right) = P \left( M - \sqrt{\frac{M}{n}} u_{1-\alpha/2} < \lambda < M + \sqrt{\frac{M}{n}} u_{1-\alpha/2} \right)$$

**4.15. Poznámka:** Meze  $100(1-\alpha)\%$  asymptotického intervalu spolehlivosti pro střední hodnotu  $\lambda$  se zpřesní, když použijeme tzv. opravu na nespojitost:

$$D = M - \frac{1}{2n} - \sqrt{\frac{M}{n}} u_{1-\alpha/2}, \quad H = M + \frac{1}{2n} + \sqrt{\frac{M}{n}} u_{1-\alpha/2}.$$

**4.16. Příklad:** Předpokládáme, že při výrobě určité tkaniny je počet kazů připadajících na 100 m této tkaniny náhodná veličina s rozložením  $Po(\lambda)$ . Při kontrole 25 balíků, z nichž každý obsahoval 100 m této tkaniny, bylo zjištěno, že celkový počet kazů je 30. Najděte 95% asymptotický interval spolehlivosti pro střední hodnotu počtu kazů připadajících na jeden balík.

**Řešení:**  $n = 25$ ,  $m = 30/25 = 1,2$ ,  $\alpha = 0,05$ .

$$d = m - \frac{1}{2n} - \sqrt{\frac{m}{n}} u_{1-\alpha/2} = 1,2 - \frac{1}{50} - \sqrt{\frac{1,2}{25}} 1,96 = 0,75$$

$$h = m + \frac{1}{2n} + \sqrt{\frac{m}{n}} u_{1-\alpha/2} = 1,2 + \frac{1}{50} + \sqrt{\frac{1,2}{25}} 1,96 = 1,65$$

S pravděpodobností aspoň 95 % lze očekávat, že střední hodnota počtu kazů připadajících na jeden balík se nachází v mezích od 0,75 do 1,65.

**4.17. Věta:** Necht'  $X_1, \dots, X_n$  je náhodný výběr z rozložení  $Po(\lambda)$ . Označme  $m = \frac{1}{n} \sum_{i=1}^n x_i$  realizaci výběrového průměru. Pak meze 100(1- $\alpha$ )% empirického intervalu spolehlivosti pro střední hodnotu  $\lambda$  jsou:  $d = \frac{1}{2n} \chi^2_{\alpha/2}(2nm)$ ,  $h = \frac{1}{2n} \chi^2_{1-\alpha/2}(2nm + 2)$ .

**Důkaz:** Viz HÁTLE JAROSLAV - LIKEŠ JIŘÍ. Základy počtu pravděpodobnosti a matematické statistiky, 2. vyd. Praha: SNTL – Nakladatelství technické literatury, 1974. 463 s.

**Upozornění:** Podle vztahů uvedených ve větě 4.17. počítá MATLAB pomocí funkce poissfit meze 100(1- $\alpha$ )% empirického intervalu spolehlivosti pro střední hodnotu  $\lambda$ .

**4.18. Příklad:** Pro údaje z příkladu 4.16. najděte 95% empirický interval spolehlivosti pro střední hodnotu počtu kazů připadajících na jeden balík.

**Řešení:**  $n = 25$ ,  $m = 30/25 = 1,2$ ,  $\alpha = 0,05$ .

$$d = \frac{1}{2n} \chi^2_{\alpha/2}(2nm) = \frac{1}{50} \chi^2_{0,025}(60) = \frac{1}{50} 40,48 = 0,81$$

$$h = \frac{1}{2n} \chi^2_{1-\alpha/2}(2nm + 2) = \frac{1}{50} \chi^2_{0,975}(62) = \frac{1}{50} 85,65 = 1,71$$

S pravděpodobností aspoň 95 % lze očekávat, že střední hodnota počtu kazů připadajících na jeden balík se nachází v mezích od 0,81 do 1,71.

**4.19. Poznámka:** Exponenciální a Poissonovo rozložení mají úzkou souvislost. Jestliže náhodná veličina  $X$ , která udává počet událostí za časovou jednotku, se řídí rozložením  $Po(\lambda)$ , pak náhodná veličina  $Y$ , která udává dobu mezi dvěma po sobě následujícími událostmi, se řídí rozložením  $Ex(\lambda)$ .

## Vzorce pro meze 100(1- $\alpha$ )% intervalu spolehlivosti pro střední hodnotu rozložení Po( $\lambda$ )

1. způsob: Využití Pearsonova rozložení chí-kvadrát:

$$d = \frac{1}{2n} \chi^2_{\alpha/2}(2nm), \quad h = \frac{1}{2n} \chi^2_{1-\alpha/2}(2nm + 2)$$

2. způsob: Využití standardizovaného normálního rozložení:

$$d = m - \sqrt{\frac{m}{n}} u_{1-\alpha/2}, \quad h = m + \sqrt{\frac{m}{n}} u_{1-\alpha/2}$$

