

Využití MATLABu při práci s exponenciálním rozložením

Základní poznatky o exponenciálním rozložení $Ex(\lambda)$

Náhodná veličina X udává dobu čekání na příchod nějaké události, která se může dostavit každým okamžikem se stejnou šancí bez ohledu na dosud pročekanou dobu. Přitom $1/\lambda$ vyjadřuje střední dobu čekání.

$$\text{Hustota: } \varphi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases}, \text{ distribuční funkce: } \Phi(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases},$$

$$\text{kvantilová funkce: } \Phi^{-1}(\alpha) = -\frac{1}{\lambda} \ln(1 - \alpha), \text{ kde } 0 < \alpha < 1.$$

$$\text{Střední hodnota: } E(X) = \frac{1}{\lambda}, \text{ rozptyl: } D(X) = \frac{1}{\lambda^2}.$$

Interval spolehlivosti pro střední hodnotu: Necht' X_1, \dots, X_n je náhodný výběr z rozložení $Ex(\lambda)$ a necht' m je realizace výběrového průměru. Pak meze $100(1-\alpha)\%$ přibližného empirického intervalu

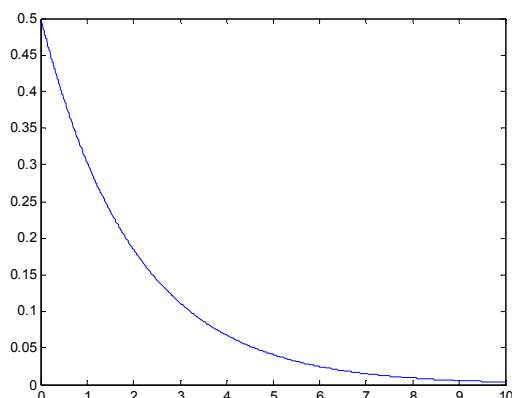
$$\text{spolehlivosti pro } E(X) = \frac{1}{\lambda} \text{ jsou: } d = \frac{2nm}{\chi^2_{1-\alpha/2}(2n)}, h = \frac{2nm}{\chi^2_{\alpha/2}(2n)}$$

Pozor, funkce v MATLABu pro práci s exponenciálním rozložením vyžadují zadávat převrácenou hodnotu parametru λ .

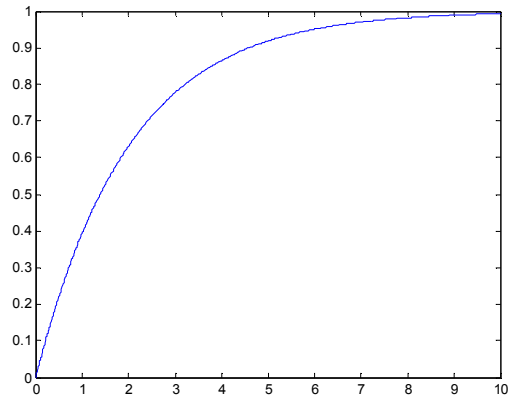
a) Kreslení grafu hustoty a distribuční funkce rozložení $Ex(1/2)$

```
x=[0:0.01:10]';  
f=expPDF(x,2);  
plot(x,f)  
df=expCDF(x,2);  
figure  
plot(x,df)
```

Graf hustoty

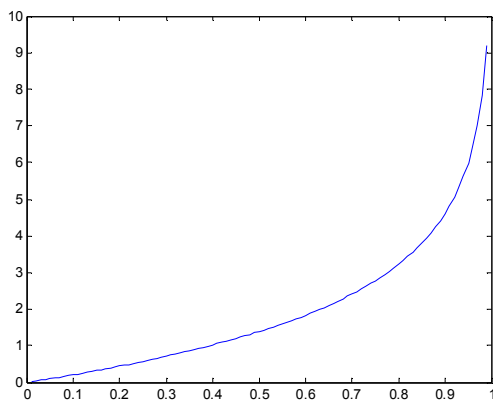


Graf distribuční funkce



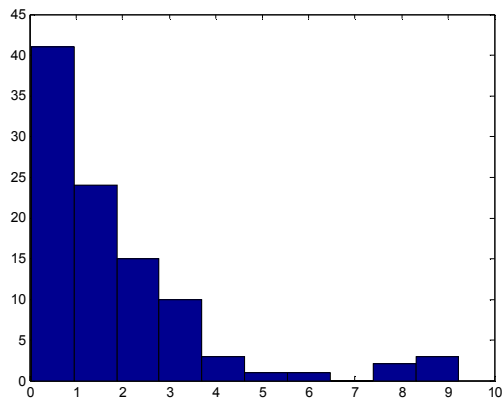
b) Kreslení grafu kvantilové funkce rozložení $Ex(1/2)$
`alfa=[0.01:0.01:0.99]';`
`kv=expinv(alfa,2);`
`plot(alfa,kv)`

Graf kvantilové funkce



c) Generování 100 realizací náhodné veličiny s rozložením $Ex(1/2)$ a kreslení histogramu s 10 třídícími intervaly
`r=exprnd(2,100,1);`
`hist(r)`

Histogram



d) Odhad střední hodnoty a meze intervalu spolehlivosti pro střední hodnotu na základě proměnné r

Hodnoty uložené v proměnné r považujeme za realizace náhodného výběru rozsahu 100 z rozložení $Ex(1/2)$

$[m, meze] = \text{expfit}(r)$

e) Výpočet střední hodnoty a rozptylu rozložení $Ex(1/2)$

$[m, v] = \text{expstat}(2)$

Využití exponenciálního rozložení při analýze příjmů

Úvod do problému: Je známo, že příjmy obyvatelstva ve společnosti jsou rozděleny nerovnoměrně. Jako první zkoumal toto rozdělení italský inženýr Vilfredo Pareto na konci 19. století. Zjistil, že příjmy lze modelovat mocninnou funkcí. V dalších letech se ukázalo, že tento tzv. Paretův zákon platí jen pro 5% nejbohatších lidí. Příjmy ostatních 95% obyvatel lze modelovat pomocí exponenciálního rozložení. (Proč to tak je? To je vysvětleno v článku F. Slaniny, Vesmír č. 9, rok 2001)

$$\text{Hustota: } \varphi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases}, \text{ distribuční funkce: } \Phi(x) = \begin{cases} 1 - e^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{pro } x \leq 0 \end{cases}$$

Nechť náhodná veličina X udává měsíční příjem náhodně vybraného zaměstnance.

Předpokládejme, že $X \sim \text{Ex}(\lambda)$. Podle údajů Českého statistického úřadu dosáhla průměrná hrubá mzda v ČR za 1. až 3. čtvrtletí roku 2007 hodnoty 21 119 Kč.

Úkol 1.: Zjistěte parametr λ pro náhodnou veličinu X .

$$E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \dots = \frac{1}{\lambda} = 21119 \Rightarrow \lambda = \frac{1}{21119} = 0,00004735$$

Řešení:

Úkol 2.: Odvoďte obecný vzorec pro výpočet α -kvantilu náhodné veličiny X a pak vyjádřete medián náhodné veličiny X . Co lze říci o vztahu střední hodnoty a mediánu?

$$\alpha = \Phi(K_{\alpha}(X)) = 1 - e^{-\lambda K_{\alpha}(X)} \Rightarrow K_{\alpha}(X) = -\frac{1}{\lambda} \ln(1 - \alpha)$$

Řešení:

$$K_{0,50}(X) = -\frac{1}{\lambda} \ln \frac{1}{2} = \frac{\ln 2}{\lambda} = 21119 \cdot \ln 2 = 14639$$

Výpočet mediánu:

Znamená to, že aspoň polovina osob má průměrnou hrubou mzdu nejvýše 14 639 Kč a aspoň polovina osob má průměrnou hrubou mzdu aspoň 14 639 Kč.

Protože exponenciální rozložení je rozložení s kladnou šikmostí (lze spočítat, že šikmost = 2), bude medián vždy menší než střední hodnota.

Úkol 3.: Kolik procent zaměstnanců má podprůměrnou hrubou mzdu?

$$P\left(X < \frac{1}{\lambda}\right) = \int_0^{\frac{1}{\lambda}} \lambda e^{-\lambda x} dx = 1 - e^{-1} = 0,6321$$

Řešení:

Znamená to, že téměř 2/3 zaměstnanců nedosáhnou na průměrnou mzdu. Průměr tedy není vhodnou charakteristikou střední úrovně mezd.

Práce se systémem MATLAB

Úkol 1.: Pomocí funkce `exprnd` náhodně vygenerujte příjmy $n = 1000$, 10 000 a 100 000 osob a vytvořte histogram vygenerovaných příjmů.

```
r = exprnd(20000,n,1);  
hist(r)
```

Úkol 2.: Vypočítejte průměrný příjem a vypočítejte medián příjmů.
 $m = \text{mean}(r)$; $x_{50} = \text{median}(r)$;
 Zjištěné hodnoty porovnejte s teoretickými hodnotami.

Úkol 3.: Zjistěte, kolik procent osob bude mít podprůměrné příjmy.
 $\text{pocet}=0$;
 $\text{pocet}=\text{sum}(r<m)$;
 $\text{procento}=100*\text{pocet}/n$
 Zjištěnou hodnotu porovnejte s teoretickou hodnotou 63,2%.

Příklady na využití exponenciálního rozložení

Příklad 1.: Doba do ukončení opravy v opravně obuvi je náhodná veličina, která se řídí exponenciálním rozložením se střední hodnotou 3 dny. Jaká je pravděpodobnost, že oprava bude ukončena do dvou dnů?

$$P(X \leq 2) = \int_0^2 \frac{1}{3} e^{-\frac{x}{3}} dx = \left[-e^{-\frac{x}{3}} \right]_0^2 = 1 - e^{-\frac{2}{3}} = 0,4866$$

Řešení: $X \sim \text{Ex}(1/3)$,
 V MATLABu: $p = \text{expcdf}(2,3)$

Příklad 2.: Životnost žárovky má exponenciální rozložení se střední hodnotou 600 h. Jaká je pravděpodobnost, že žárovka bude svítit dalších aspoň 200 h, jestliže již svítila aspoň 800 h?

Řešení: $X \sim \text{Ex}(1/600)$,

$$P(X \geq 800 + 200 / X \geq 800) = P(X \geq 200) = 1 - P(X \leq 200) + P(X = 200) = 1 - \int_0^{200} \frac{1}{600} e^{-\frac{x}{600}} dx =$$

$$= 1 - \left[-e^{-\frac{x}{600}} \right]_0^{200} = e^{-\frac{200}{600}} = e^{-\frac{1}{3}} = 0,7165$$

V MATLABu: $p = 1 - \text{expcdf}(200,600)$

Příklad 3.: Náhodné doby života dvou součástí jsou stochasticky nezávislé náhodné veličiny, přičemž $X_i \sim \text{Ex}(\lambda_i)$, $i = 1, 2$. Střední hodnota doby života první součástky je 2 roky, druhé součástky 3 roky. Jaká je pravděpodobnost, že druhá součástka přežije první?

$$P(X_2 > X_1) = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{3}} = \frac{\frac{1}{2}}{\frac{5}{6}} = 0,6$$

Řešení:

Příklad 4.: Doba (v hodinách), která uplyne mezi dvěma naléhavými příjmy v jisté nemocnici, se řídí exponenciálním rozložením se střední hodnotou 2 h. Jaká je pravděpodobnost, že uplyne více než 5 h bez naléhavého příjmu?

$$P(X > 5) = 1 - P(X \leq 5) = 1 - \int_0^5 \frac{1}{2} e^{-\frac{x}{2}} dx = 1 - \left[-e^{-\frac{x}{2}} \right]_0^5 = e^{-2,5} = 0,082$$

Řešení: $X \sim \text{Ex}(1/2)$,
 V MATLABu: $p = 1 - \text{expcdf}(5,2)$

Příklad 5.: Zkoumá se funkce dvou nezávisle na sobě pracujících přístrojů. Doba bezporuchové funkce i-tého přístroje je náhodná veličina $X_i \sim \text{Ex}(\lambda_i)$, $i = 1, 2$. Jaká je pravděpodobnost, že za dobu $t_0 > 0$ a) ani jeden přístroj neselže, b) selže aspoň jeden přístroj?

Řešení: ad a)

$$\begin{aligned} P(X_1 > t_0 \wedge X_2 > t_0) &= P(X_1 > t_0)P(X_2 > t_0) = [1 - P(X_1 \leq t_0)][1 - P(X_2 \leq t_0)] = \\ &= [1 - \Phi_1(t_0)][1 - \Phi_2(t_0)] = e^{-\lambda_1 t_0} e^{-\lambda_2 t_0} = e^{-t_0(\lambda_1 + \lambda_2)} \end{aligned}$$

$$\text{ad b) } P(X_1 \leq t_0 \vee X_2 \leq t_0) = 1 - e^{-t_0(\lambda_1 + \lambda_2)}$$

Příklad 6.: Najděte 5. percentil náhodné veličiny $X \sim \text{Ex}(0,1)$

Řešení: $0,05 = \Phi(K_{0,05}(X)) = 1 - \exp(-0,1K_{0,05}(X)) \Rightarrow K_{0,05}(X) = -10 \ln 0,95 = 0,5129$

V MATLABu: $K = \text{expinv}(0.05, 10)$