INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odhady kvantilů a p-procentní životnosti

Kvantily jsou charakteristiky rozdělení pravděpodobností, které umožňují toto rozdělení pravděpodobností dobře charakterizovat. Jsou zvlášť užitečné při popisu asymetrických rozdělení a hojně se používají k popisu tzv. „chvostů“ rozdělení, tedy k popisu rozdělení v oblastech, kde hustota nebo pravděpodobnostní funkce nabývají hodnot blízkých k nule.

Pro dané číslo $\gamma \in (0,1)$ nazýváme γ-**kvantilem** rozdělení pravděpodobností náhodné veličiny X o distribuční funkci $F(x)$ číslo x_γ, když platí

$$x_\gamma = \inf\{x\colon F(x) \geq \gamma\}.$$

V uvedené definici označení „$\inf M$“ značí infimum číselné množiny M. Pro konečnou množinu je $\inf M$ rovno nejmenšímu číslu množiny M (tedy $\inf M$ je minimální prvek této množiny). Pro nekonečnou množinu M je $\inf M$ zobecnění pojmu minimální prvek pro tuto nekonečnou množinu. Např. když M je interval, $M = (5,6)$, je $\inf M = 5$. Je-li $\gamma = 0{,}5$, pak kvantil $x_{0,5}$ nazýváme **medián**, pro $\gamma = 0{,}25$, respektive $\gamma = 0{,}75$, kvantil $x_{0,25}$, respektive $x_{0,75}$, nazýváme **dolní**, respektive . Kvantily $x_{0,10}, x_{0,20}, \dots, x_{0,90}$ nazýváme **decily** a konečně kvantily $x_{0,01}, x_{0,02}, \dots, x_{0,98}, x_{0,99}$ nazýváme **percentily**.

Pro popis variability rozdělení pravděpodobností se užívají veličiny $Q_K = x_{0,75} - x_{0,25}$, $Q_D = x_{0,90} - x_{0,10}$ a $Q_P = x_{0,99} - x_{0,01}$, které se postupně nazývají **kvartilové**, **decilové** a **percentilové rozpětí**.

1 Kvantily – diskrétní náhodná veličina

Nechť X je náhodná veličina, $X \sim B(n,\theta),\ n = 3,\ \theta = 0{,}5$. Pak její pravděpodobnostní funkce je tvaru

$$p(x) = \binom{3}{x} 0{,}5^x \cdot 0{,}5^{3-x} = \frac{1}{8}\binom{3}{x}, \quad x \in \{0,1,2,3\},$$

její distribuční funkce je dána předpisem

$$F(x) = \begin{cases} 0 & \text{pro } x < 0, \\ \frac{1}{8} & \text{pro } 0 \leq x < 1, \\ \frac{4}{8} & \text{pro } 1 \leq x < 2, \\ \frac{7}{8} & \text{pro } 2 \leq x < 3, \\ 1 & \text{pro } 3 < x. \end{cases}$$

Pro $0{,}5$-kvantil (medián) tohoto rozdělení dostaneme

$$x_{0,5} = \inf\{x\colon F(x) \geq 0{,}5\} = 1.$$

Pro $0{,}75$-kvantil (horní kvartil) tohoto rozdělení dostaneme

$$x_{0,75} = \inf\{x\colon F(x) \geq 0{,}75\} = 2.$$

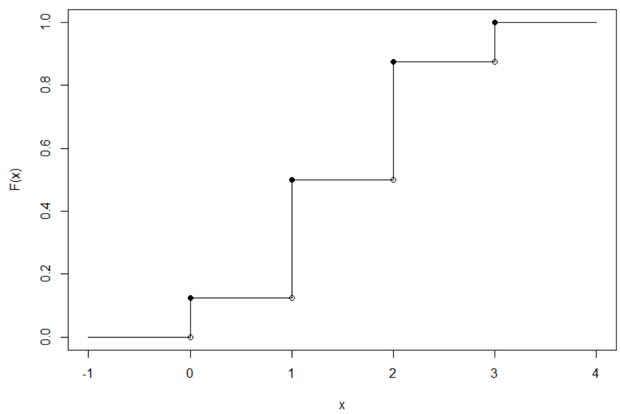
Z výsledků je zřejmé, že 50 % hodnot X je větších než medián $x_{0,5}$ a 25 % hodnot X je větších než horní kvartil.

Nechť $X \sim P_0(\lambda),\ \lambda = 0{,}3$. Pak pro pravděpodobnostní funkci náhodné veličiny X dostaneme

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!} = e^{-0,3}\frac{(0{,}3)^x}{x!}, \quad x = 0,1,2,\dots$$

Operační program Vzdělávání pro konkurenceschopnost
Název projektu: Inovace magisterského studijního programu Fakulty ekonomiky a managementu
Registrační číslo projektu: CZ.1.07/2.2.00/28.0326

PROJEKT JE SPOLUFINANCOVÁN EVROPSKÝM SOCIÁLNÍM FONDEM A STÁTNÍM ROZPOČTEM ČESKÉ REPUBLIKY.



Obr. 1: Distribuční funkce binomického rozdělení $B(3; 0.5)$

x	0	1	2	3	4	5	6
$p(x)$	0,740818	0,222245	0,033337	0,003334	0,000250	0,000015	0,000001
$F(x)$	0,740818	0,963063	0,996400	0,999734	0,999984	0,999999	1,000000

Užitím distribuční funkce snadno stanovíme kvantily $x_{0,5},\ x_{0,25},\ x_{0,75}$. Z třetího řádku tabulky hned vidíme, že $x_{0,5} = 0,\ x_{0,25} = 0,\ x_{0,75} = 1$.

2 Kvantily – spojitá náhodná veličina

Je-li distribuční funkce $F(x)$ náhodné veličiny X rostoucí na intervalu (a, b) a $F(a) = 0,\ F(b) = 1$, pak k funkci F existuje inverzní funkce F^{-1} a pomocí ní snadno vyjádříme příslušný γ-kvantil. Platí

$$x_\gamma = \inf\{x\colon F(x) \geq \gamma\} = \inf\{x\colon F(x) = \gamma\} = F^{-1}(\gamma).$$

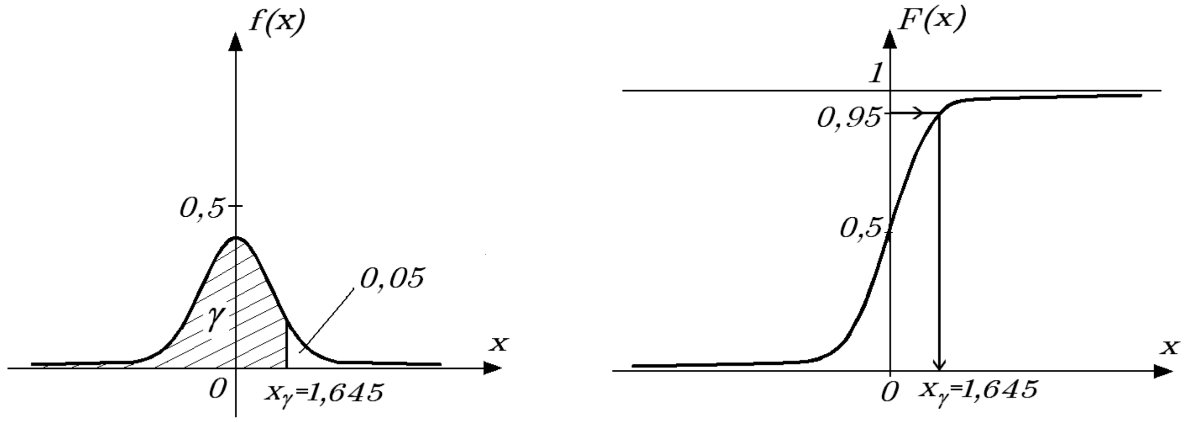
Tedy pro spojitou náhodnou veličinu X najdeme γ-kvantil ze vztahu $F(x_\gamma) = \gamma$. Tento vztah můžeme přepsat pomocí hustoty $f(x)$. Pro výpočet γ-kvantilu x_γ dostaneme rovnici

$$\int_a^{x_\gamma} f(x)dx = \gamma.$$

Nechť $X \sim N(\mu, \sigma^2),\ \mu = 0,\ \sigma^2 = 1$. Určete hodnoty kvantilů x_γ pro $\gamma = 0{,}5$ (tj. medián) a $\gamma = 0{,}95$. Příslušný γ-kvantil dostaneme řešením rovnice

$$\gamma = F(x_\gamma) = \int_{-\infty}^{x_\gamma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Nechť $X \sim N(\mu, \sigma^2),\ \mu = 5,\ \sigma^2 = 25$. Určete kvantil x_γ, nejprve pro obecnou hodnotu γ, poté speciálně pro $\gamma = 0{,}95$.



Obr. 2: Znázornění kvantilu $x_{0,95}$ normálního rozdělení $N(0,1)$

Standardizací $U = \frac{X-\mu}{\sigma} = \frac{X-5}{5}$ dostaneme náhodnou veličinu U, která má standardizované normální rozdělení pravděpodobností $N(0,1)$ a výpočet kvantilů náhodné veličiny X provedeme pomocí kvantilů náhodné veličiny U. Postupně dostaneme

$$\gamma = F(x_\gamma) = P(X \leq x_\gamma) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x_\gamma - \mu}{\sigma}\right) = P\left(U \leq \frac{x_\gamma - \mu}{\sigma}\right).$$

Položíme $u_\gamma = \frac{x_\gamma - \mu}{\sigma}$ a z výše uvedeného vztahu dostaneme

$$\gamma = P\left(U \leq \frac{x_\gamma - \mu}{\sigma}\right) = P(U \leq u_\gamma),$$

a tedy u_γ je γ-kvantil normálního rozdělení $N(0,1)$, jehož stanovení bylo diskutováno v předchozím případě. Pro kvantil x_γ dostáváme rovnici $x_\gamma = \mu + \sigma u_\gamma$. Speciálně pro $\gamma = 0{,}95$ je $u_\gamma = 1{,}645$ a

$$x_\gamma = x_{0,95} = 5 + 5 \cdot 1{,}645 = 13{,}225.$$

Nechť $X \sim Ex(\lambda)$. Stanovte kvantil x_γ, nejprve pro obecnou hodnotu γ, poté speciálně pro $\gamma = 0{,}95$. Distribuční funkce náhodné veličiny X je dána vztahem

$$\begin{aligned} F(x) &= 1 - e^{-\lambda x}, && x \geq 0, \\ F(x) &= 0, && x < 0. \end{aligned}$$

Pro γ-kvantil rozdělení $Ex(\lambda)$ dostaneme rovnici

$$\gamma = F(x_\gamma) = 1 - e^{-\lambda x_\gamma}$$

a jejím řešením obdržíme

$$x_\gamma = -\frac{1}{\lambda} \ln(1-\gamma).$$

Speciálně pro $\lambda = 0{,}2$ dostaneme pro 0,95-kvantil

$$x_{0,95} = -\frac{1}{0{,}2} \ln(1 - 0{,}95) = -5 \ln 0{,}05 = 14{,}9787.$$

Porovnáním výsledků z předchozích příkladů vidíme, že náhodné veličiny $X \sim N(5,25)$ a $Y \sim Ex(0{,}2)$ mají stejné střední hodnoty $E(X) = \mu = 5,\ E(Y) = \frac{1}{\lambda} = 5$, stejné rozptyly $D(X) = \sigma^2 = 25,\ D(Y) = = \frac{1}{\lambda^2} = 25$, ale 95 % kvantily x_γ se liší: 95 % kvantil rozdělení $Ex(0{,}2)$ je 14,9787, a tedy je větší než 95 % kvantil normálního rozdělení $N(5,25)$, který je 13,225. Znamená to, že hustota normálního rozdělení se rychleji blíží k nule než hustota rozdělení exponenciálního, i když střední hodnota i rozptyl obou rozdělení jsou shodné. Říkáme, že exponenciální rozdělení má **těžší chvosty** než rozdělení normální. Znamená to, že rychlost konvergence hustoty rozdělení s těžšími chvosty je malá ve srovnání s normálním rozdělením. Když náhodná veličina X popisuje dobu čekání na rizikový jev, dobu do poruchy, životnost zařízení nebo životnost jedince dané populace, potom její kvantil $x_{1-\gamma}$ představuje hodnotu, kterou tato náhodná veličina překročí s pravděpodobností γ. Tedy při opakovaných pozorováních náhodné veličiny X průměrně $100\gamma\%$ pozorování překročí hodnotu $x_{1-\gamma}$. Kvantil $x_{1-\gamma}$ značíme T_γ a veličinu T_γ nazýváme $100\gamma\%$-**ní životnost** nebo $100\gamma\%$ **zaručenou dobu bezporuchové činnosti**. Např. když $X \sim Ex(0{,}2)$, je 5%-ní životnost $T_{0{,}05}$ rovna $T_{0{,}05} = 14{,}9787$. Lze uzavřít, že v průměru 5% jedinců dané populace přežije čas $14{,}9787$.

Příklady k procvičení

1. Náhodná veličina X popisuje dobu do poruchy vybraného zařízení. Generujte náhodné výběry z rozdělení

   a) normálního $N(0,1)$
   b) normálního $N(10,16)$
   c) exponenciálního $Ex(5)$
   d) Weibullova $W(2,1)$
   e) gamma $\Gamma(2,1)$
   f) logaritmicko-normálního rozdělení $LN(2,4)$

   postupně rozsahu $n = 5, 10, 30, 50, 100$. Stanovte teoretické kvantily X_p a $100p\%$ teoretické životnosti pro $p \in \{0{,}01; 0{,}05; 0{,}; 0{,}25; 0{,}5; 0{,}75; 0{,}9; 0{,}95; 0{,}99\}$ a porovnejte je s kvantily a se $100p\%$ životnostmi, které získáte na simulovaných datech pomocí výběrové distribuční funkce. Popište, jak se mění kvalita odhadu s rostoucím rozsahem výběru n. Pro řešení využijte podle vlastní volby software STATISTICA, R nebo MATLAB.

2. Náhodná veličina X popisuje počet poruch vybraného zařízení za jednotku času. Generujte náhodné výběry z rozdělení

   a) binomického rozdělení $Bi(10; 0{,}1)$
   b) Poissonova rozdělení $Po(2)$
   c) negativně binomického rozdělení $NB(2; 0{,}5)$

   postupně rozsahu $n = 10, 30, 50, 100$. Stanovte teoretické kvantily X_p a $100p\%$ teoretické životnosti pro $p \in \{0{,}1; 0{,}25; 0{,}5; 0{,}75; 0{,}9\}$ a porovnejte je s kvantily a se $100p\%$ životnostmi, které získáte na simulovaných datech pomocí výběrové distribuční funkce. Popište, jak se mění kvalita odhadu s rostoucím rozsahem výběru n. Pro řešení využijte podle vlastní volby software STATISTICA, R nebo MATLAB.