Spojité náhodné veličiny a jejich charakteristiky, normální rozdělení

—

Supplementum ke cvičení 4ST210 Statistika pro finance

Lubomír Štěpánek[1, 2]

[1]Oddělení biomedicínské statistiky
Ústav biofyziky a informatiky
1. lékařská fakulta
Univerzita Karlova, Praha

[2]Katedra biomedicínské informatiky
Fakulta biomedicínského inženýrství
České vysoké učení technické v Praze

17. března 2020

(2020) Lubomír Štěpánek, CC BY-NC-ND 3.0 (CZ)

Dílo lze dále svobodně šířit, ovšem s uvedením původního autora a s uvedením původní licence. Dílo není možné šířit komerčně ani s ním jakkoliv jinak nakládat pro účely komerčního zisku. Dílo nesmí být jakkoliv upravováno. Autor neručí za správnost informací uvedených kdekoliv v předložené práci, přesto vynaložil nezanedbatelné úsilí, aby byla uvedená fakta správná a aktuální, a práci sepsal podle svého nejlepšího vědomí a svých „nejlepších“ znalostí problematiky.

Obsah

1. Opakování
2. Náhodná veličina
3. Spojitá náhodná veličina
4. Normální rozdělení
5. Limitní věty
6. Literatura

Příklad

- Na katederní schůzi, kde bylo původně šestnáct lidí, přišlo několik doktorandů pozdě. Jejich příchodem kleslo průměrné IQ v místnosti ze 140 na 137. Průměrné IQ ve skupince pozdě příchozích doktorandů bylo 125. Určete, kolik bylo doktorandů bylo ve skupince pozdě příchozích.

Příklad

- V souboru deseti čísel je aritmetický průměr roven 12 a rozptyl 4. Do souboru přidáme hodnotu 23. Jak se změní průměr a rozptyl souboru?

Příklad

- V souboru deseti čísel je aritmetický průměr roven 12 a rozptyl 4. Do souboru přidáme hodnotu 23. Jak se změní průměr a rozptyl souboru?
- *Řešení.* $\bar{x}_{\text{nový}} = 13$, $s^2_{x,\text{nový}} = 13\frac{7}{11}$ □

Příklad

- Loď pluje z říčního přístavu k ústí řeky rychlostí $v_1 > 0$, poté se vrací stejnou trasou zpět do přístavu rychlostí $v_2 > 0$. Určete průměrnou rychlost lodi na celkové trase z říčního přístavu k ústí řeky a zpět.

Příklad

- Průměrná cena letenky kanadských vnitrostátních letů vzrostla o $20\ \%$, zatímco rozptyl cen kanadských vnitrostátních letů vzrostl o $21\ \%$. Jak se změnil variační koeficient cen letenek kanadských vnitrostátních letů?

Příklad

- V urně je deset bílých a třicet černých koulí. S jakou pravděpodobností budou mezi šesti náhodně vytaženými koulemi právě čtyři černé, pokud
  - (i) každou kouli po vytažení vždy vrátíme zpět do urny?
  - (ii) koule po vytažení do urny zpět nevracíme?

Příklad

- V malé nemocnici zaznamenají vždy od pondělí do neděle průměrně 35 porodů.
  (i) Kolik porodů můžeme v dané nemocnici očekávat následující den?
  (ii) Kolik porodů můžeme v dané nemocnici očekávat během následujících tří dní?
  (iii) Jaký je očekávaný rozptyl počtu porodů v dané nemocnici během následujících tří dní?
  (iv) S jakou pravděpodobností dojde v dané nemocnici během následujících tří dní právě k 12 porodům?
  (v) S jakou pravděpodobností dojde v dané nemocnici během následujícího dne k nejvýše dvěma porodům?
  (vi) S jakou pravděpodobností dojde v dané nemocnici během následujícího dne k alespoň jednomu porodu?
  (vii) S jakou pravděpodobností dojde v dané nemocnici během následujícího dne právě ke čtyřem, pěti nebo šesti porodům?

Příklad

- Určitá léčba je úspěšná průměrně u $80\ \%$ pacientů. S jakou pravděpodobností bude v kohortě deseti pacientů léčba neúspěšná nejvýše u jednoho z pacientů?

Příklad

- Házíme desetkrát nespravedlivou (biasovanou) mincí, na které padne hlava s pravděpodobností $0{,}6$.
  (i) S jakou pravděpodobností padne ve všech deseti hodech desetkrát hlava?
  (ii) S jakou pravděpodobností padne mezi deseti hody právě čtyřikrát hlava?
  (iii) S jakou pravděpodobností padne mezi deseti hody alespoň osmkrát hlava?
  (iv) Jaký počet padnutí hlavy je mezi deseti hody nejpravděpodobnější?

Náhodná veličina

- je popsána číselným (nebo symbolickým) výsledkem náhodného pokusu
  - např. příjem náhodně vybrané domácnosti, doba do poruchy přístroje, počet dětí v rodině, počet porodů za noc v dané porodnici apod.
- náhodné veličiny se obvykle značí velkými písmeny z konce abecedy, tedy X, Y, Z, $\dots$, jejich konkrétní číselné realizace pak malými písmeny x, y, z, $\dots$
  - např. je-li X počet porodů za noc v dané porodnici, pak $x = 6$ značí, že zde došlo k $X = x = 6$ porodům za noc
  - X, Y, Z, $\dots$, mají tedy charakter proměnných, zatímco x, y, z, $\dots$, mají charakter konkrétních čísel (konstant)
- náhodné veličiny mohou být diskrétní, či spojité

Rozdělení náhodné veličiny

- dáno *zákonem rozdělení náhodné veličiny*
  - tj. že náhodná veličina nabude konkrétní hodnoty nebo hodnoty z nějakého intervalu s určitou pravděpodobností
- lze postupně popsat
  - řadou rozdělení
  - distribuční funkcí
  - pravděpodobnostní funkcí, resp. hustotou
  - momentovou funkcí, resp. charakteristickou funkcí
- vhodnými ukazateli jsou pak střední hodnota a variabilita náhodné veličiny

Spojitá náhodná veličina

- je popsána spojitým číselným výsledkem náhodného pokusu
  - např. příjem náhodně vybrané domácnosti, doba do poruchy přístroje, hmotnost náhodně vybraného balení kilogramové mouky apod.
- spojitá náhodná veličina X tedy může nabýt každé libovolné hodnoty z intervalu možných hodnot
- spojitá náhodná veličina X nabývá těchto hodnot s určitou „pravděpodobností", kterou však nelze smysluplně popsat pro konkrétní (jednu) hodnotu
  - např. pohybuje-li se doba životnosti přístroje mezi 1000 až 1200 hodinami, pak pravděpodobnost, že pro náhodně vybraný přístroj bude životnost $1158{,}5$ hodin je dle definice geometrické pravděpodobnosti rovna $\frac{S}{\Omega} = \frac{\text{bod odpovídající } 1158{,}5}{\text{úsečka mezi body } 1000 \text{ až } 1200} = \frac{0}{\infty} = 0$

Rozdělení spojité náhodné veličiny

- není (!) možné popsat tabulkou typu

x_1	x_2	$\dots$	x_k
$P(x_1)$	$P(x_2)$	$\dots$	$P(x_k)$

neboť

- hodnota k roste nade všechny meze
- hodnot $1, 2, \dots, k$ je nespočetně mnoho, tedy není smysluplně možné rozdělit interval možných hodnot spojité náhodné veličiny X na dílčí hodnoty $x_1, x_2, \dots, x_k$
- jednotlivé pravděpodobnosti $P(x_i)$ pro $\forall i \in \{1, 2, 3, \dots\}$ jsou nulové[1]

[1] pravděpodobnost typu „bod ku úsečce", neboli $\frac{0}{\infty}$, což vyplývá z geometrické definice pravděpodobnosti

Rozdělení spojité náhodné veličiny

- proto je vhodné pro spojitou náhodnou veličinu X popsat pravděpodobnost jejího výskytu v *intervalu* možných hodnot, nikoliv pouze pro bodovou hodnotu

Hustota pravděpodobnosti spojité náhodné veličiny

- je funkce typu $f_X : \mathbb{R} \to \langle 0, +\infty)$
- nemá (!) rozměr ani vlastnosti pravděpodobnosti[2]
- přesto však platí, že $\forall x \in \mathbb{R}$ je $f_X(x) \geq 0$ a dále že

$$f_X(x) = \lim_{\Delta x \to 0} \frac{P(x - \Delta x \leq X \leq x + \Delta x)}{\Delta x}$$

- méně přesně tak lze pro velmi malé kladné $\Delta x \approx 0$ psát, že

$$P(X = x) \approx \Delta x \cdot f_X(x)$$

- dále platí

$$\int_{-\infty}^{+\infty} f_X(x) \mathrm{d}x = 1$$

[2] běžně tedy může být např. $f_X(x) > 1$

Hustota pravděpodobnosti spojité náhodné veličiny

- pro spojitou náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1 \rangle$ tak, že každé z takových čísel je vráceno stejně „pravděpodobně“, platí

$$f_X(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}, & x \in \langle -1, 1 \rangle \\ 0, & x > 1 \end{cases}$$

Hustota pravděpodobnosti spojité náhodné veličiny

- stále uvažujme spojitou náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1\rangle$ tak, že každé z takových čísel je vráceno stejně „pravděpodobně“
- snadno ověříme, že platí $\int_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = 1$, neboť

$$\int\limits_{-\infty}^{+\infty} f_X(x)\mathrm{d}x = \int\limits_{-\infty}^{-1} f_X(x)\mathrm{d}x + \int\limits_{-1}^{+1} f_X(x)\mathrm{d}x + \int\limits_{+1}^{+\infty} f_X(x)\mathrm{d}x =$$

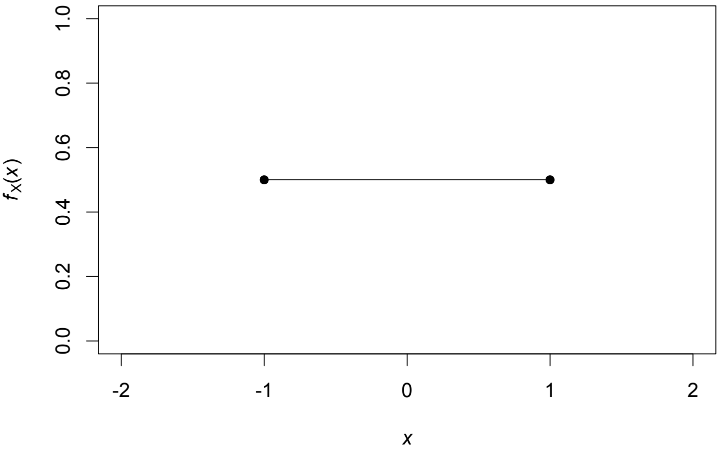
$$= \int\limits_{-\infty}^{-1} 0 \ \mathrm{d}x + \int\limits_{-1}^{+1} \frac{1}{2}\mathrm{d}x + \int\limits_{+1}^{+\infty} 0 \ \mathrm{d}x = 0 + \left[\frac{1}{2}x\right]_{-1}^{+1} + 0 =$$

$$= \frac{1}{2}(1-(-1)) = \frac{1}{2}\cdot 2 =$$

$$= \boxed{1}$$

Hustota pravděpodobnosti spojité náhodné veličiny

- pro náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1 \rangle$ takovou, že každé číslo je vráceno stejně „pravděpodobně“



Distribuční funkce (spojité) náhodné veličiny

- je funkce typu $F_X : \mathbb{R} \to \langle 0, 1 \rangle$, konkrétněji $F_X : x \to P(X \leq x)$
- vrací tedy pro libovolnou hodnotu $x \in \mathbb{R}$ pravděpodobnost $P(X \leq x)$, s jakou (spojitá) náhodná veličina X nepřevýší reálné číslo x
- píšeme tedy $F_X(x) = P(X \leq x)$
- dále je rovněž

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau$$

- a obdobně

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$$

Distribuční funkce (spojité) náhodné veličiny

- také platí

$$
\begin{aligned}
P(x_1 \leq X \leq x_2) &= P(X \in \langle x_1, x_2 \rangle) = \\
&= \underbrace{P(X \leq x_1) + P(X \in \langle x_1, x_2 \rangle)}_{P(X \leq x_2)} - P(X \leq x_1) = \\
&= F_X(x_2) - F_X(x_1) = \\
&= \int_{-\infty}^{x_2} f_X(\tau)\mathrm{d}\tau - \int_{-\infty}^{x_1} f_X(\tau)\mathrm{d}\tau = \\
&= \int_{x_1}^{x_2} f_X(\tau)\mathrm{d}\tau + \int_{-\infty}^{x_1} f_X(\tau)\mathrm{d}\tau - \int_{-\infty}^{x_1} f_X(\tau)\mathrm{d}\tau = \\
&= \int_{x_1}^{x_2} f_X(\tau)\mathrm{d}\tau
\end{aligned}
$$

Vlastnosti distribuční funkce (spojité) náhodné veličiny

$\forall x \in \mathbb{R}$ je $0 \leq F_X(x) \leq 1$

$\forall x \in \mathbb{R}$ je $F_X(x) = P(X \leq x)$ a nabývá tedy pouze hodnot z intervalu $\langle 0, 1 \rangle$ $\square$

$F_X(x)$ je neklesající funkce

$\forall a, b \in \mathbb{R} : a < b$ je $F_X(a) = P(X \leq a)$ a $F_X(b) = P(X \leq b) = P(X \leq a \cup X \in \langle a, b \rangle) = P(X \leq a) + P(X \in \langle a, b \rangle) \geq P(X \leq a) = F_X(a)$, tedy $F_X(x)$ je neklesající $\square$

Platí, že $P(a \leq X \leq b) = F_X(b) - F_X(a)$

$\forall a, b \in \mathbb{R} : a < b$ je $F_X(a) = P(X \leq a)$ a $F_X(b) = P(X \leq b) = P(X \leq a \cup X \in \langle a, b \rangle) = P(X \leq a) + P(X \in \langle a, b \rangle)$, tedy $F_X(b) - F_X(a) = P(X \leq a) + P(X \in \langle a, b \rangle) - P(X \leq a) = P(X \in \langle a, b \rangle)$, tudíž je $P(a \leq X \leq b) = F_X(b) - F_X(a)$ $\square$

Distribuční funkce spojité náhodné veličiny

- pro spojitou náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1 \rangle$ tak, že každé z takových čísel je vráceno stejně „pravděpodobně", platí

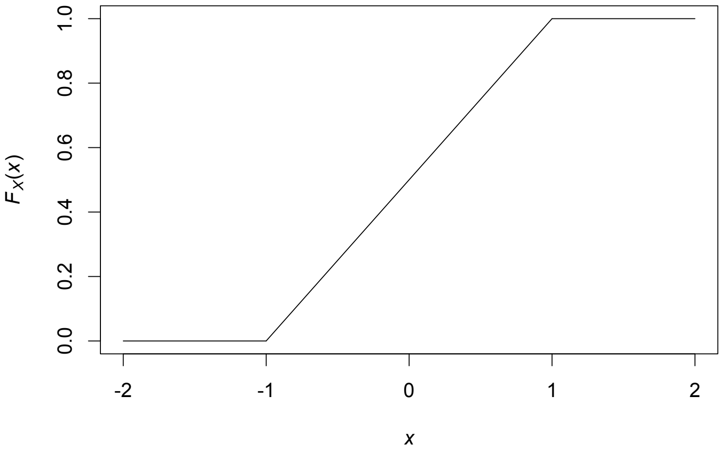
$$f_X(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}, & x \in \langle -1, 1 \rangle \\ 0, & x > 1 \end{cases}$$

- protože $\int \frac{1}{2} \mathrm{d}x = \frac{1}{2}x + c$ a $\int 0 \ \mathrm{d}x = d_{1,2}$ pro vhodná $c, d_{1,2} \in \mathbb{R}$, je

$$F_X(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}x + \frac{1}{2}, & x \in \langle -1, 1 \rangle \\ 1, & x > 1 \end{cases}$$

Distribuční funkce spojité náhodné veličiny

- pro náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1\rangle$ takovou, že každé číslo je vráceno stejně „pravděpodobně“



Střední hodnota spojité náhodné veličiny

- ať je spojitá náhodná veličina X popsána svou hustotou pravděpodobnosti $f_X(x)$
- pak střední hodnota $\mathbb{E}(X)$ spojité náhodné veličiny X je

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) \mathrm{d}x$$

- zřejmě dále platí
  - pro $a, b \in \mathbb{R}$ je $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$
  - pro libovolné náhodné veličiny $X_1, X_2, \ldots, X_p$ je $\mathbb{E}(X_1 + X_2 + \ldots + X_p) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \ldots + \mathbb{E}(X_p)$
  - pro nezávislé náhodné veličiny $X_1, X_2, \ldots, X_p$ je $\mathbb{E}(X_1 \cdot X_2 \cdot \ldots \cdot X_p) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdot \ldots \cdot \mathbb{E}(X_p)$

Střední hodnota spojité náhodné veličiny

- pro náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1\rangle$ takovou, že každé číslo je vráceno stejně „pravděpodobně“, platí

$$f_X(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}, & x \in \langle -1, 1\rangle \\ 0, & x > 1 \end{cases}$$

- a proto

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x =$$
$$= \int_{-\infty}^{-1} x f_X(x)\mathrm{d}x + \int_{-1}^{+1} x f_X(x)\mathrm{d}x + \int_{+1}^{+\infty} x f_X(x)\mathrm{d}x =$$

Střední hodnota spojité náhodné veličiny

$$\mathbb{E}(X) = \int_{-\infty}^{-1} x f_X(x)\ \mathrm{d}x + \int_{-1}^{+1} x f_X(x)\ \mathrm{d}x + \int_{+1}^{+\infty} x f_X(x)\ \mathrm{d}x =$$

$$= \int\limits_{-\infty}^{-1} x \cdot 0\ \mathrm{d}x + \int\limits_{-1}^{+1} x \cdot \frac{1}{2}\ \mathrm{d}x + \int\limits_{+1}^{+\infty} x \cdot 0\ \mathrm{d}x =$$

$$= 0 + \left[\frac{1}{2} \cdot \frac{x^2}{2}\right]_{-1}^{+1} + 0 =$$

$$= \left[\frac{1}{2} \cdot \frac{x^2}{2}\right]_{-1}^{+1} =$$

$$= \frac{1}{2}\left(\frac{1}{2} - \frac{1}{2}\right) =$$

$$= \frac{1}{2} \cdot 0 = \boxed{0}$$

Rozptyl spojité náhodné veličiny

- ať je spojitá náhodná veličina X popsána svou hustotou pravděpodobnosti $f_X(x)$
- pak rozptyl $\mathrm{var}(X)$ spojité náhodné veličiny X je

$$\mathrm{var}(X) = \int\limits_{-\infty}^{+\infty} (x - \mathbb{E}(X))^2 f_X(x) \mathrm{d}x =$$

$$= \int\limits_{-\infty}^{+\infty} x^2 f_X(x) \mathrm{d}x - \left(\int\limits_{-\infty}^{+\infty} x f_X(x) \mathrm{d}x \right)^2$$

- zřejmě dále platí
  - pro $a, b \in \mathbb{R}$ je $\mathrm{var}(aX + b) = a^2 \mathrm{var}(X)$
  - pro nezávislé náhodné veličiny $X_1, X_2, \ldots, X_p$ je $\mathrm{var}(X_1 + X_2 + \ldots + X_p) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \ldots + \mathrm{var}(X_p)$

Rozptyl spojité náhodné veličiny

- pro náhodnou veličinu X vracející náhodné reálné číslo z intervalu $\langle -1, 1\rangle$ takovou, že každé číslo je vráceno stejně „pravděpodobně", platí

$$f_X(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}, & x \in \langle -1, 1\rangle \\ 0, & x > 1 \end{cases}$$

- a proto

$$\operatorname{var}(X) = \int\limits_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{d}x - \left(\int\limits_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x\right)^2 =$$

$$= \int_{-\infty}^{+\infty} x^2 f_X(x)\mathrm{d}x - (\mathbb{E}(X))^2 =$$

Rozptyl spojité náhodné veličiny

$$\begin{aligned}
\mathrm{var}(X) &= \int_{-\infty}^{+\infty} x^2 f_X(x) \mathrm{d}x - (\mathbb{E}(X))^2 = \\
&= \int_{-\infty}^{-1} x^2 f_X(x)\ \mathrm{d}x + \int_{-1}^{+1} x^2 f_X(x)\ \mathrm{d}x + \int_{+1}^{+\infty} x^2 f_X(x)\ \mathrm{d}x \\
&- (\mathbb{E}(X))^2 = \\
&= \int\limits_{-\infty}^{-1} x^2 \cdot 0\ \mathrm{d}x + \int\limits_{-1}^{+1} x^2 \cdot \frac{1}{2}\ \mathrm{d}x + \int\limits_{+1}^{+\infty} x^2 \cdot 0\ \mathrm{d}x - (\mathbb{E}(X))^2 = \\
&= 0 + \left[\frac{1}{2} \cdot \frac{x^3}{3}\right]_{-1}^{+1} + 0 - (\mathbb{E}(X))^2 = \left[\frac{1}{2} \cdot \frac{x^3}{3}\right]_{-1}^{+1} - (\mathbb{E}(X))^2 = \\
&= \frac{1}{2}\left(\frac{1}{3} - \left(-\frac{1}{3}\right)\right) - 0^2 = \frac{1}{2} \cdot \frac{2}{3} - 0 = \boxed{\frac{1}{3}}
\end{aligned}$$

Kvantily spojité náhodné veličiny

- ať je spojitá náhodná veličina X popsána svou hustotou pravděpodobnosti $f_X(x)$ a distribuční funkcí $F_X(x)$
- pak $100p\%$ kvantil náhodné veličiny X je číslo x_p takové, že náhodná veličina X nabývá hodnoty nepřevyšující x_p s pravděpodobností $0 \leq p \leq 1$
- formálně tedy

$$P(X \leq x_p) = F_X(x_p) = \int\limits_{-\infty}^{x_p} f_X(\tau)\mathrm{d}\tau = p$$

Normální rozdělení

- náhodná veličina X sleduje normální rozdělení, tedy $X \sim \mathcal{N}(\mu, \sigma^2)$ o parametrech $\mu \in \mathbb{R}$ a $\sigma \in \mathbb{R} : \sigma \geq 0$, pokud pro její pravděpodobnostní hustotu $f_X(x)$ a pro libovolné $x \in \mathbb{R}$ platí

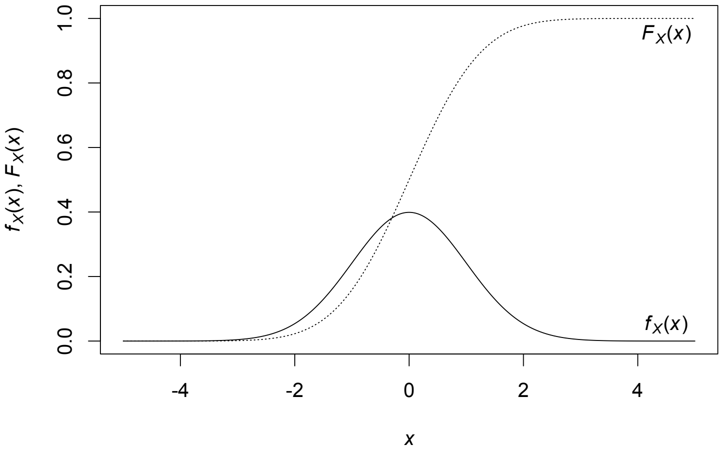
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- platí, že

$$\mathbb{E}(X) = \mu$$
$$\text{var}(X) = \sigma^2$$

- náhodná veličina X sledující normální rozdělení obvykle vhodně popisuje fenomény založené na vzájemném působení mnoha drobných nezávislých, eventuálně navzájem se rušících vlivů
  - např. veličina X může popsat distribuci náhodných chyb, tělesné výšky dané populace, inteligenčního kvocientu, sytost barvy vlasů (šedotónové pixely) dané populace a mnoho dalšího

Pravděpodobnostní hustota a distribuční funkce normálního rozdělení



Distribuční funkce normálního rozdělení

- ať náhodná veličina X sleduje normální rozdělení, tedy $X \sim \mathcal{N}(\mu, \sigma^2)$ o parametrech $\mu \in \mathbb{R}$ a $\sigma \in \mathbb{R} : \sigma \geq 0$
- její distribuční funkce $F_X(x)$ má tvar

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(\tau)\mathrm{d}\tau = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\tau-\mu)^2}{2\sigma^2}} \mathrm{d}\tau$$

- takový výraz ale nelze běžně integrovat, navíc numerické řešení integrálu by bylo nutné tabelizovat pro všechny kombinace (x, μ, σ), což není možné, proto zavádíme odvozenou náhodnou veličinu $U \equiv \frac{X-\mu}{\sigma}$, která sleduje standardní normální rozdělení $U \sim \mathcal{N}(0, 1^2)$

Standardní normální rozdělení

- ať náhodná veličina X sleduje normální rozdělení, tedy $X \sim \mathcal{N}(\mu, \sigma^2)$ o parametrech $\mu \in \mathbb{R}$ a $\sigma \in \mathbb{R} : \sigma \geq 0$
- náhodná veličinu $U \equiv \frac{X-\mu}{\sigma}$, sleduje standardní normální rozdělení $U \sim \mathcal{N}(0, 1^2)$ s pravděpodobnostní hustotou

$$f_U(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

distribuční funkcí

$$F_U(x) = \int_{-\infty}^{x} f_U(\tau) \mathrm{d}\tau = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}} \mathrm{d}\tau$$

a charakteristikami

$$\mathbb{E}(U) = 0$$

$$\mathrm{var}(U) = 1$$

Distribuční funkce standardního normálního rozdělení

- ať náhodná veličinu $U \equiv \frac{X-\mu}{\sigma}$, sleduje standardní normální rozdělení $U \sim \mathcal{N}(0, 1^2)$ a má distribuční funkci

$$F_U(x) = \int\limits_{-\infty}^{x} f_U(\tau)\mathrm{d}\tau = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{\tau^2}{2}}\mathrm{d}\tau$$

- takový výraz je stále nepraktický pro rutinní výpočty, ale díky standardizaci již lze smysluplně tabelizovat
- běžně píšeme

$$F_U(x) = P(U \leq x) = \Phi(x)$$

  a pro malé hodnoty $x \geq 0$ hodnoty $\Phi(x)$ tabelizujeme
- díky sudosti pravděpodobnostní hustoty $f_U(x)$ je $P(U \leq x) = 1 - P(U \leq -x)$, lze pro hodnoty $x < 0$ využít vztah

$$\Phi(-x) = P(X \leq -x) = 1 - P(X \leq x) = 1 - \Phi(x)$$

Příklad

- Pomocí úvahy nebo tabulek distribuční funkce standardního normálního rozdělení
  (i) najděte $\Phi(0)$.
  (ii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 0{,}5$.
  (iii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = \Phi(-x)$.
  (iv) najděte $\Phi(1{,}96)$.
  (v) najděte $\Phi(-1{,}96)$.
  (vi) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 0{,}025$.
  (vii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 1$.
  (viii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) \leq 2$.
  (ix) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = 0$.
  (x) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) = +\infty$.
  (xi) najděte $\Phi(+\infty)$.
  (xii) najděte $\Phi(-\infty)$.
  (xiii) najděte $x \in \mathbb{R}$ takové, aby $\Phi(x) - \Phi(-x) = 0{,}5$.
  (xiv) najděte $x \in \mathbb{R}$ takové, aby $P(|U| \leq x) = 0{,}95$.

Příklad

- Náhodná veličina X sleduje normální rozdělení $\mathcal{N}(20, 16)$. Jaká je pravděpodobnost, že nabude hodnoty
  - (i) menší než 16?
  - (ii) větší než 20?
  - (iii) v rozmezí mezi 12 a 28?
  - (iv) menší než 12 nebo větší než 28?

Příklad

- Náhodná veličina X sleduje normální rozdělení $\mathcal{N}(20, 16)$. Jaká je pravděpodobnost, že nabude hodnoty
  (i) menší než 16?
  (ii) větší než 20?
  (iii) v rozmezí mezi 12 a 28?
  (iv) menší než 12 nebo větší než 28?
- *Řešení.*
  (i) $0{,}158$.
  (ii) $0{,}500$.
  (iii) $0{,}955$.
  (iv) $0{,}046$. □

Příklad

- Náhodná veličina X vrací hodnotu chyby měření určitým přístrojem a sleduje normální rozdělení se střední hodnotu $\mu = 0{,}20$ a rozptylem $\sigma^2 = 0{,}64$.
  (i) Jaká je pravděpodobnost, že absolutní hodnota chyby měření bude menší než $1{,}0$?
  (ii) Jaká je horní hranice chyby měření, které se může přístroj dopustit s pravděpodobností $0{,}95$?

Příklad

- Náhodná veličina X vrací hodnotu chyby měření určitým přístrojem a sleduje normální rozdělení se střední hodnotu $\mu = 0{,}20$ a rozptylem $\sigma^2 = 0{,}64$.
  (i) Jaká je pravděpodobnost, že absolutní hodnota chyby měření bude menší než $1{,}0$?
  (ii) Jaká je horní hranice chyby měření, které se může přístroj dopustit s pravděpodobností $0{,}95$?
- *Řešení.*
  (i) $0{,}775$.
  (ii) $1{,}516$. □

Moivreova-Laplaceova centrální limitní věta

- buď X náhodná veličina daná součtem n vzájemně nezávislých náhodných veličin, z nichž každá sleduje alternativní rozdělení $\mathcal{A}(\pi)$
- pak jistě náhodná veličina X sleduje binomické rozdělení o parametrech n a π a střední hodnotě $n\pi$ a rozptylu $n\pi(1-\pi)$, tedy $X \sim \text{binom}(n, \pi)$
- dle Moivreovy-Laplaceovy centrální limitní věty platí

$$U \equiv \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \sim \mathcal{N}(0, 1^2)$$

- pro dostatečně velká n tedy binomické rozdělení konverguje k normálnímu

Příklad

- Pravděpodobnost narození chlapce je $0{,}515$. Jaká je pravděpodobnost, že mezi $10\,000$ narozenými novorozenci bude
  - (i) více děvčat než chlapců?
  - (ii) relativní četnost chlapců bude v mezích od $0{,}515$ do $0{,}517$?

Příklad

- Pravděpodobnost narození chlapce je $0{,}515$. Jaká je pravděpodobnost, že mezi $10\ 000$ narozenými novorozenci bude
  (i) více děvčat než chlapců?
  (ii) relativní četnost chlapců bude v mezích od $0{,}515$ do $0{,}517$?
- *Řešení.*
  (i) $0{,}00135$.
  (ii) $0{,}15500$. □

Lindebergova-Lévyho centrální limitní věta

- buď X náhodná veličina daná součtem n vzájemně nezávislých náhodných veličin X_i pro $\forall i \in \{1, 2, \ldots, n\}$, z nichž každá sleduje pro $\forall i \in \{1, 2, \ldots, n\}$ stejné rozdělení s konečnou střední hodnotou $\mathbb{E}(X_i) = \mu < \infty$ a konečným rozptylem $\mathrm{var}(X_i) = \sigma^2 < \infty$
- pak jistě náhodná veličina X má střední hodnotu $\mathbb{E}(X) = n\mu$ a rozptyl $n\sigma^2$
- dle Lindebergovy-Lévyho centrální limitní věty platí

$$U \equiv \frac{X - n\mu}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0, 1^2)$$

Příklad

- Zaměstnanec jezdí do zaměstnání pravidelně tam i zpět metrem. Doba čekání na příjezd soupravy metra se pohybuje rovnoměrně mezi 0 až 3 minutami. Jaká je pravděpodobnost, že celková doba čekání zaměstnance během 23 pracovních dní[3] bude kratší než 80 minut?

[3] zaměstnanec jede vždy dvakrát denně – do zaměstnání a zpět

Příklad

- Zaměstnanec jezdí do zaměstnání pravidelně tam i zpět metrem. Doba čekání na příjezd soupravy metra se pohybuje rovnoměrně mezi 0 až 3 minutami. Jaká je pravděpodobnost, že celková doba čekání zaměstnance během 23 pracovních dní[3] bude kratší než 80 minut?
- *Řešení.* $0{,}969$. □

[3] zaměstnanec jede vždy dvakrát denně – do zaměstnání a zpět

Ljapunova centrální limitní věta

- buď X náhodná veličina daná součtem n vzájemně nezávislých náhodných veličin X_i pro $\forall i \in \{1, 2, \ldots, n\}$, z nichž každá sleduje pro $\forall i \in \{1, 2, \ldots, n\}$ (ne nutně shodné) rozdělení s konečnou střední hodnotou $\mathbb{E}(X_i) < \infty$ a konečným rozptylem $\mathrm{var}(X_i) < \infty$
- dle Ljapunovy centrální limitní věty platí

$$U \equiv \frac{X - \sum_{i=1}^{n} \mathbb{E}(X_i)}{\sqrt{\sum_{i=1}^{n} \mathrm{var}(X_i)}} \sim \mathcal{N}(0, 1^2)$$

Literatura

- Hindls, Richard, Stanislava Hronová, Jan Seger a Jakub Fischer. *Statistika pro ekonomy*. Praha: Professional Publishing, 2007. ISBN: 978-80-86946-43-6.

- Marek, Luboš. *Statistika v příkladech*. Praha: Professional Publishing, 2015. ISBN: 978-80-7431-153-6.

Děkuji za pozornost!

lubomir.stepanek@vse.cz
lubomir.stepanek@lf1.cuni.cz
lubomir.stepanek@fbmi.cvut.cz

https://github.com/LStepanek/4ST210_Statistika_pro_finance