1 Výpočet číselných charakteristik náhodných veličin

Vyjádření k domácímu úkolu

Stará látka

Normální rozložení

- náhodná veličina
  - diskrétní
    - pstní fce
    - distr fce
  - spojitá
    - hustota
    - distr fce
- Náhodná veličina $X \sim N(\mu, \sigma^2)$
- Střední hodnota $E(X) = \mu$, rozptyl $D(X) = \sigma^2$.
- posunutí, rozptyl

standardizované normální rozložení

- $\mu = 0$, $\sigma^2 = 1$
- $U \sim N(0, 1)$
- distribuční fce:
$$\Phi(u) = \int_{-\infty}^{u} f(t)dt \tag{1}$$
- hodnoty distr. fce jsou tabelovány pro $u \geq 0$
- pro $u < 0$ používáme přepočtový vzorec $\Phi(-u) = 1 - \Phi(u)$.

1.1 Nová látka:

Vlastnosti normálního rozložení

- Pokud $X \sim N(\mu, \sigma^2)$, potom $\dfrac{X-\mu}{\sigma} \sim N(0,1)$
- Pokud $X \sim N(\mu, \sigma^2)$, potom $A = a + bX \sim N(a + b\mu, b^2\sigma^2)$
- $X_1 \dots X_n$; $X_i \sim N(\mu_i, \sigma_i^2)$, potom $Y = \sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$

Číselné charakteristiky náh.veličin

- F(x),p(x),f(x) …funkcionální charakteristiky
  - obsahují veškerou informaci o chování náh.veličiny
- někdy nás zajímají pouze rysy chování náh.veličiny
- k tomu slouží *číselné charakteristiky*
  - kvantily
  - střední hodnota
  - rozptyl / směrodatná odchylka
  - kovariance
  - korelace

α-kvantil

- α-kvantil náh.veličiny X
- $u_\alpha(X)$
- obdoba α-kvantilu v popisné statistice
- křivka hustoty:
  - plocha pod křivkou …pst … $= 1$
  - tuto plochu rozdělíme na 2 části
    - tmavá plocha α
    - světlá plocha $1 - \alpha$
  - pst, že náhodná veličina X se realizuje hodnotou menší nebo rovnou hodnotě $u_\alpha(X)$
  - číslo $u_\alpha(X)$ nazýváme $\alpha -$ *kvantilem* spojité náhodné veličiny.
  - $\alpha = P(X \leq u_\alpha(X))$
  - platí vztah
    $$u_\alpha = -u_{1-\alpha} \tag{2}$$
  - speciální kvantily

        * medián
        * 1.kvartil
        * 3.kvartil

  - Pearsonovo rozložení $X \sim \chi^2(n)$ ... nesymetrické
  - Studentovo rozložení $X \sim t(n)$ ... symetrické ... $t_\alpha = -t_{1-\alpha}$
  - Fisherovo-Snedecorovo rozložení $X \sim F(n_1, n_2)$

- R-příkazy

  - qnorm(α, mean=, sd=)
  - qchisq(α, n)
  - qt(α, n)
  - qf(α,n_1,n_2)

- Příklad:

  1. $U \sim N(0;1); \ldots u_{0.25} = -u_{0.75} = -0.674$
  2. $U \sim N(0;1); \ldots u_{0.5} = 0.5$
  3. $U \sim N(0;1); \ldots u_{0.75} = 0.674$
  4. $\chi^2_{0.025}(25) = 13.120$
  5. $t_{0.99}(30) = 2.457$
  6. $X \sim N(3,5) \ldots x_{0.25} =?$
     - Najdeme 0.25-kvantil pro $U \sim N(0;1) \rightarrow u_{0.25} = -0.674$
     - platí $U = \dfrac{X-\mu}{\sigma} \Rightarrow X = \sigma U + \mu \Rightarrow x_{0.25} = \sigma u_{0.25} + \mu \Rightarrow$

$$x_{0.25} = \sqrt{5}(-0.674) + 3 = 1.49 \tag{3}$$

střední hodnota μ

- nese informace o střední poloze
- střední hodnota diskrétní náhodné veličiny

$$E(X) = \sum_{-\infty}^{\infty} xp(x) \tag{4}$$

- střední hodnota spojité náhodné veličiny

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{5}$$

- může se stát, že náhodná veličina X nemá střední hodnotu
- jde o idealizovaný průměr; značí se μ a výběrový ekvivalent je aritm. průměr m
- střední hodnota diskrétní transformované náhodné veličiny

$$E(Y) = E(g(X)) = \sum_{\infty}^{\infty} g(x)p(x) \tag{6}$$

1.1.1 Vlastnosti střední hodnoty

1. $E(a) = a$
2. $E(a + bX) = a + bE(X)$
3. $E(X_1 + X_2) = EX_1 + EX_2$
4. $E(X_1 + \ldots, X_n) = EX_1 + \ldots EX_n$

rozptyl

- určuje, jak moc je náhodná veličina variabilní, jakou má tendenci realizovat se poblíž/daleko centrální polohy
- značíme $D(X)$, σ^2
- teoretický protějšek $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$; m je aritm.průměr, nebo medián
- čím více se hodnoty v souboru navzájem liší, tím je hodnota rozptylu vyšší, kdyby celý soubor byl složen z 1, pak by rozptyl byl 0
- $$D(X) = E([X - E(X)]^2) \tag{7}$$
- X porovnáme s její střední hodnotou, tuto odchylku umocníme na druhou a z kvadrátů všech X vypočítáme střední hodnotu
- rozptyl diskrétní náh.veličiny

$$D(X) = E([X - E(X)]^2) = \sum_{-\infty}^{\infty}[x - E(X)]^2 p(x) \tag{8}$$

- rozptyl spojité náh.veličiny

$$D(X) = E([X - E(X)]^2) = \int_{-\infty}^{\infty}[x - E(X)]^2 f(x)dx \tag{9}$$

- $\sqrt{D(X)}$ je směrodatná odchylka

1.1.2 Vlastnosti rozptylu

1. $D(X) = E(X^2) - [EX]^2$
2. $D(a) = 0$
3. $D(a + bX) = b^2 D(X)$
4. jsou-li náh.vel. stoch.nezáv, potom $D(X_1 + X_2) = D(X_1) + D(X_2)$
5. jsou-li náh.vel. stoch.nezáv, potom $D(X_1 + \ldots X_n) = D(X_1) + \ldots D(X_n)$

- Příklad: Náhodná veličina X udává počet ok při hodu kostkou. Vypočtěte střední hodnotu EX a rozptyl DX.
  - Náhodná veličina X je diskrétního charakteru
  - střední hodnota EX:
    * $E(X) = \sum_{-\infty}^{\infty} xp(x)$
    * k výpočtu střední hodnoty potřebujeme znát pstní fci pro všechny možné jevy:
    * $P(X = 1) = 1/6$, $P(X = 2) = 1/6$, $P(X = 3) = 1/6$, $P(X = 4) = 1/6$, $P(X = 5) = 1/6$, $P(X = 6) = 1/6$

$$\begin{aligned} E(X) &= 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} && (10) \\ &= 3.5 && (11) \end{aligned}$$

    * Střední hodnota počtu ok při hodu kostkou je 3.5.
  - rozptyl DX:
    * $D(X) = E([X - E(X)]^2) = \sum_{-\infty}^{\infty}[x - E(X)]^2 p(x)$
    * EX jsme vypočítali
    * $E(X^2)$ vypočteme podle vztahu $E(Y) = E(g(X)) = \sum_{\infty}^{\infty} g(x)p(x)$, kde $Y = g(X) = X^2$

$$\begin{aligned} E(X^2) &= 1^2\frac{1}{6} + 2^2\frac{1}{6} + 3^2\frac{1}{6} + 4^2\frac{1}{6} + 5^2\frac{1}{6} + 6^2\frac{1}{6} \\ &= 15.1 \end{aligned}$$

    * $DX = E(X^2) - (EX)^2 = 15.1 - 3.5^2 = 2.917$.
    * Rozptyl počtu ok při hodu kostkou je 2.917.

standardizovaná náh.veličina

- $$\frac{X - E(X)}{\sqrt{D(X)}} \qquad (12)$$

- realizace st.náh.vel. jsou bezrozměrná čísla, která nám říkají o kolikanásobek směrodatné odchylky $\sqrt{D(X)}$ je ralizace X posunuta doprava/doleva od střední hodnoty
- $E(X) = 0$
- $\sigma^2 = 1$
- centrování, škálování, standardizace

Kovariance $C(X,Y)$

- vždy mezi dvěmi náh.veličinami X,Y
- disrkétní náhodná veličina:
  $C(X,Y) = E([X - EX][Y - EY]) =$
  $$\sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}[x - EX][y - EY]p(x,y) \tag{13}$$
  kde $p(x,y)$ je simultánní pstní fce.
- spojitá náhodná veličina
  $C(X,Y) = E([X - EX][Y - EY]) =$
  $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[x - EX][y - EY]f(x,y)dx, \tag{14}$$
  kde $f(x,y)$ je simultánní hustota
- střední hodnota součinu centrovaných náh.veličin X,Y
- znaménko kovariance určuje, zda je vztah mezi náh.veličinami *přímý* nebo *nepřímý*

Korelace

- dvou náh.veličin X, Y
- střední hodnota součinu standardizovaných veličin X,Y
- charakterizuje těsnost LINEÁRNÍHO vztahu mezi X a Y
- $$R(X,Y) = E\left(\frac{X - EX}{\sqrt{DX}}\frac{Y - EY}{\sqrt{DY}}\right) = \frac{C(X,Y)}{\sqrt{DX}\sqrt{DY}} \tag{15}$$

- Příklad:
  Na pohřebišti se našlo několik koster, které pravděpodobně patřili obětem uctívacích rituálů. Kostrám chybí vždy na rukou buď jeden nebo dva prsty a na nohou tři nebo čtyři prsty. Máme dva znaky: Znak X - chybějící prsty na rukou má dvě varianty (X_1 - 1 prst; X_2 - 2 prsty). Znak Y - chybějící prsty na nohou má také dvě varianty (Y_1 - 3 prsty; Y_2 - 4 prsty). Pst kombinace R1+N3 je 0.1, pst kombinace R1+N4 je 0.3, pst kombinace R2+N3 je 0.35 a kombinace R2+N4 je 0.25. Určete kovarianci a korelaci znaků X a Y.
  Data můžeme uspořádat do přehledné tabulky: 0.1, 0.3, 0.35 a 0.25 jsou simultánní psti $p(x_i, y_j)$.

	N3	N4	p(x)
R1	0.1	0.3	0.4
R2	0.35	0.25	0.6
p(y)	0.45	0.55	1

$$C(X,Y) = \sum_{-\infty}^{\infty}\sum_{-\infty}^{\infty}[x - EX][y - EY]p(x,y)$$

- $EX = 1 * 0.4 + 2 * 0.6 = 1.6$
- $EY = 3 * 0.45 + 4 * 0.55 = 3.55$

$$\begin{aligned} C(X,Y) &= (1-1.6)(3-3.55)0.1 + (1-1.6)(4-3.55)0.3 + \\ &+ (2-1.6)(3-3.55)0.35 + (2-1.6)(4-3.55)0.25 \\ &= 0.33 * 0.1 - 0.27 * 0.3 - 0.22 * 0.35 + 0.18 * 0.25 = -0.08 \end{aligned}$$

$E(X^2) = 1^2 0.4 + 2^2 0.6 = 2.8$
$E(Y^2) = 3^2 0.45 + 4^2 0.55 = 12.85$
$DX = E(X^2) - (EX)^2 = 2.8 - 1.6^2 = 0.24$
$DY = E(Y^2) - (EY)^2 = 12.85 - 3.55^2 - 1.6^2 = 0.2475$
$R(X,Y) = \frac{C(X,Y)}{\sqrt{DX}\sqrt{DY}} = \dfrac{-0.08}{\sqrt{0.24}\sqrt{0.2475}} = -0.328.$