PŘEHLED VZORCŮ A HLAVNÍCH POJMŮ
na přednáškách KMA/PSA
5. a 6. týdne ZS 2015/16

(následující bude mj. promítnuto během těchto přednášek PSA - volná místa na stránkách jsou úmyslně : po vytištění lze používat k doplnění vlastních poznámek)

Vysvětlení, použití, grafy, příklady, etc. budou na přednášce.

---

7. KVANTILY SPOJITÝCH ROZDĚLENÍ

Nechť X je *spojitá* náhodná veličina, $p \in (0, 1)$.

Reálné číslo x_p se nazývá **100 p%-NÍ KVANTIL**, je-li

$$\boxed{P(X \leq x_p) = p}$$

$x_{0,5}$ ... **medián**
$x_{0,25}$ ... **dolní kvartil**
$x_{0,75}$ ... **horní kvartil**

Pomocí distribuční funkce $F(x)$ lze definici kvantilu x_p přepsat takto:

$$\boxed{F(x_p) = p}$$

Př.:

Určení kvantilů normálního rozdělení

Kvantily x_p rozdělení $N(\mu, \sigma^2)$ lze vyjádřit pomocí kvantilů u_p rozdělení $N(0,1)$, neboť pro $p \in (0,1)$ platí

$$\boxed{x_p = \mu + \sigma \cdot u_p}$$

K určení u_p používáme **tabulky**, pro $p < 0,5$ navíc rovnost

$$\boxed{u_p = -u_{1-p}}$$

[tedy např. $u_{0,05} = -u_{0,95}$]

která platí, neboť hustota $\varphi(u)$ je sudá funkce.

Př.:

8. TRANSFORMACE NÁHODNÉ VELIČINY

Dáno: X ... **spojitá** náh. veličina
$f(x)$ ... hustota X,
$F(x)$ ... distrib. fce X

Nechť $Y = h(X)$ (funkce náh. veličiny X).

Pak Y je náh. veličinou, označme:
$g(y)$ ... hustota Y,
$G(y)$ ... distrib. fce Y

$$g(y) = ?$$
$$G(y) = ?$$

Pozn.: Je-li fce h **konvexní** nebo **ryze monotónní**, je hledání $g(y)$, resp. $G(y)$ jednodušší - viz příklady. Některá takto vzniklá rozdělení ppsti mají své názvy a označení - např. normované normální rozdělení, logaritmicko-normální rozdělení, χ^2-rozdělení.

Pozn.: Y... spojitá náhodná veličina

• LOGARITMICKO-NORMÁLNÍ rozdělení (s parametry $\mu \in R$, $\sigma > 0$),
píšeme $Y \sim N(\mu, \sigma^2)$:
Y má hustotu

$$g(y) = \begin{cases} 0 & \text{pro } y \in (-\infty, 0] \\ \dfrac{1}{\sigma\, y} \cdot \varphi\left(\dfrac{\ln\, y - \mu}{\sigma}\right) & \text{pro } y \in (0, +\infty)\,, \end{cases}$$

kde φ je hustota pravděpodobnosti rozdělení $N(0, 1)$.

• χ^2-ROZDĚLENÍ [čteme:"chí-kvadrát"] o 1 stupni volnosti, píšeme $Y \sim \chi^2(1)$:
Y má hustotu

$$g(y) = \begin{cases} 0 & \text{pro } y \in (-\infty, 0] \\ \dfrac{1}{\sqrt{2\pi\, y}} \cdot e^{-\dfrac{y}{2}} & \text{pro } y \in (0, +\infty)\,, \end{cases}$$

[Později ve statistice bude $\chi^2(\nu)$, kde $\nu \in N$.]

NÁHODNÝ VEKTOR

$\vec{\mathbf{X}} = (\mathbf{X}, \mathbf{Y})$ ... (dvourozměrný) **náhodný vektor**

X, Y ... náh. veličiny (buď obě diskr. nebo obě spoj.)

$\boxed{F(x,y) = P(X \leq x), Y \leq y)}$ kde $x, y \in (-\infty, +\infty)$
... (sdružená) **distribuční funkce** náh. vektoru $\vec{X}$

$F_1(x) = P(X \leq x)$, resp. $F_2(y) = P(Y \leq y)$
... tzv. **marginální** distrib.funkce X, resp.Y

$\vec{X} = (X, Y)$ **diskrétní**:

$P(x, y)$ ... (sdružená) **pravděpodobnostní funkce**

$P_1(x)$, $P_2(y)$ ... **marginální** ppstní fce X, resp. Y

$\vec{X} = (X, Y)$ **spojitý**:

$f(x, y)$ ... (sdružená) **hustota** ppsti
$f_1(x)$, $f_2(y)$ ... **marginální** hustoty X, resp. Y

$E(\vec{X}) = (E(X), E(Y))$ ... vektor středních hodnot

Vztah mezi X a Y udává tzv. **KOVARIANCE**, píšeme $\mathrm{cov}(X,Y)$:

$$\boxed{\mathrm{cov}\ (X,Y) = E([X - E(X)] \cdot [Y - E(Y)])}$$

Roznásobením a úpravou dostaneme tzv. **výpočetní tvar kovariance**:

$$\mathrm{cov}\,(X,Y) \;=\; E(X \cdot Y) - E(X) \cdot E(Y)$$

Zřejmě: $\mathrm{cov}\,(X,Y) = \mathrm{cov}\,(Y,X)$

$$\mathrm{cov}\,(X,X) \;= D(X)$$

$$\mathrm{cov}\,(Y,Y) \;= D(Y)$$

Pro (nekonstantní) náh. veličiny X, Y definujeme tzv.

KORELAČNÍ KOEFICIENT, píšeme $\varrho(X, Y)$:

$$\boxed{\varrho = \varrho(X, Y) = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\,\sigma(Y)}\,,}$$

kde $\sigma(X)$, $\sigma(Y)$ jsou směrodatné odchylky veličin X, Y.

Platí :
- $-1 \leq \varrho \leq 1$
- $\varrho(X, Y) = \varrho(Y, X)$
- $\varrho(X, X) = \varrho(Y, Y) = 1$

ϱ je mírou statistické **lineární** závislosti veličin X, Y

$\varrho = +1 \Leftrightarrow Y = aX + b$, kde $a > 0$, $b \in R$

$\varrho = -1 \Leftrightarrow Y = aX + b$, kde $a < 0$, $b \in R$

Je-li $\varrho = 0$, říkáme, že X, Y jsou **nekorelované**.

X, Y nezávislé $\Rightarrow X, Y$ jsou nekorelované

(implikace "$\Leftarrow$"obecně neplatí)

Pro náh. vektor $\vec{X} = (X_1, X_2, \ldots, X_k)$ lze kovariance, resp. korelační koeficienty uspořádat do čtvercových matic $\Sigma = (\sigma_{ij})$, resp. $R = (\varrho_{ij})$, a to označením

$\sigma_{ij} = \mathrm{cov}\,(X_i,\, X_j)$, resp. $\quad \varrho_{ij} = \varrho\,(X_i,\, X_j)$

Pak

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix}$$

resp.

$$R = \begin{bmatrix} \varrho_{11} & \varrho_{12} & \cdots & \varrho_{1k} \\ \varrho_{21} & \varrho_{22} & \cdots & \varrho_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ \varrho_{k1} & \varrho_{k2} & \cdots & \varrho_{kk} \end{bmatrix}$$

Σ ... **kovarianční matice**, je symetrická, na hlavní diagonále jsou rozptyly $D(X_i)$

R ... **korelanční matice**, je symetrická, na hlavní diagonále jsou 1

Př.:

Dvourozměrné normální rozdělení

s vektorem středních hodnot $\vec{\mu} = (\mu_1, \mu_2)$ a kovarianční maticí Σ má hustotu

$$f(x,y) = \frac{1}{2\pi\sqrt{det\,\Sigma}} \cdot \mathrm{e}^{-\frac{1}{2}\,P(x,y)} \ ,$$

kde

$$P(x,y) = (x-\mu_1, y-\mu_2)\ \Sigma^{-1} \begin{pmatrix} x-\mu_1 \\ y-\mu_2 \end{pmatrix} .$$

Píšeme: $(X,Y) \sim N_2(\vec{\mu}, \Sigma)$.

Je-li $(X,Y) \sim N_2(\vec{\mu}, \Sigma)$, pak platí:

X, Y nezávislé $\Leftrightarrow$ X, Y nekorelované.
(tj. nelineární závislost není možná)