

Matematika III – 10. týden
Číselné charakteristiky – střední hodnota, rozptyl,
kovariance, korelace

Jan Slovák

Masarykova univerzita
Fakulta informatiky

28. 11 – 2. 12. 2016

Obsah přednášky

- 1 Literatura
- 2 Střední hodnota a rozptyl
- 3 Kovariance

Kde je dobré číst?

- Karel Zvára, Josef Štěpán, Pravděpodobnost a matematická pravděpodobnost statistika, Matfyzpress, 2006, 230pp.
- J. Slovák, M. Panák, M. Bulant, Matematika drsně a svižně, Muni Press, Brno 2013, v+773 s., elektronická edice www.math.muni.cz/Matematika_drsne_svizne
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, Teorie pravděpodobnosti a matematická statistika (sbírka příkladů), Masarykova univerzita, 3. vydání, 2004, 117 stran, ISBN 80-210-3313-4.
- Marie Budíková, Tomáš Lerch, Štěpán Mikoláš, Základní statistické metody, Masarykova univerzita, 2005, 170 stran, ISBN 80-210-3886-1.
- Riley, K.F., Hobson, M.P., Bence, S.J. Mathematical Methods for Physics and Engineering, second edition, Cambridge University Press, Cambridge 2004, ISBN 0 521 89067 5, xxiii + 1232 pp.

Střední hodnota

Nechť X je náhodná veličina s diskrétním rozdělením. Jestliže řada $\sum_{k=1}^{\infty} x_k P(X = x_k)$ konverguje absolutně (zejména tedy pro všechny X s konečně mnoha možnými hodnotami x_k), pak její součet EX nazýváme **střední hodnotou** X .

Je-li X náhodná veličina se spojitým rozdělením s hustotou $f(x)$ a nevlastní integrál $\int_{-\infty}^{\infty} xf(x)dx$ konverguje absolutně, pak jeho hodnota EX se nazývá **střední hodnota** X .

Je tedy $EX = np$, je-li $X \sim \text{Bi}(n, p)$, zatímco pro rovnoměrné rozdělení na intervalu (a, b) dostaneme dle očekávání

$$EX = \int_a^b \frac{x}{b-a} dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = \frac{1}{2}(a+b).$$

Vlastnosti střední hodnoty

Theorem

Uvažme náhodné veličiny X, Y , skaláry $a, b \in \mathbb{R}$, náhodný vektor $W = (X_1, \dots, X_n)$ a čtvercovou skalární matici B s n řádky.

- *Pro konstantní náhodnou veličinu $X = a \in \mathbb{R}$ je $E a = a$.*
- $E(a + bX) = a + b E X$.
- $E(X + Y) = E X + E Y$.
- $E(a + BX) = a + B(E X)$.

Theorem

Jsou-li veličiny X a Y nezávislé, pak $E(XY) = E X E Y$.

Rozptyl

Další charakteristika popisuje, jak moc se dá čekat, že se hodnoty náhodné veličiny „hemží“ kolem nějaké hodnoty.

Definition

Nechť X je náhodná veličina s konečnou střední hodnotou. Pak definujeme **rozptyl** veličiny X výrazem

$$\text{var } X = E(X - E X)^2,$$

pokud taková konečná hodnota existuje.

Odmocnina z rozptylu $\sqrt{\text{var } X}$ se nazývá **směrodatná odchylka** náhodné veličiny X .

Jde o zjevnou obdobu definice kvadrátu vzdálenosti vektorů nebo funkcí. Zachycujeme tak „očekávanou vzdálenost“ hodnot X od její střední hodnoty.

Theorem

Jestliže má náhodná veličina X konečný rozptyl, pro libovolné skaláry $a, b \in \mathbb{R}$ platí

- $\text{var } X = E X^2 - (E X)^2$
- $\text{var}(a + bX) = b^2 \text{var } X$
- $\sqrt{\text{var}(a + bX)} = |b| \sqrt{\text{var } X}$.

Občas přiřazujeme k X **normovanou** veličinu Z ,

$$Z = \frac{X - E X}{\sqrt{\text{var } X}},$$

kteřá má zjevně nulovou střední hodnotu a jednotkový rozptyl.

Normální rozdělení Z má hustotu $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ distribuční funkci $\Phi(z) = \int_{-\infty}^z \varphi(t) dt = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$.
Náhodná veličina $Y = \mu + \sigma Z$, $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ má distribuční funkci

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{\frac{y-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &\quad \{\text{substituce } x = \mu + \sigma z\} \\ &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \end{aligned}$$

Takové rozdělení je *normální*, píšeme $Y \sim N(\mu, \sigma^2)$.
Parametry odpovídají střední hodnotě a rozptylu.

Uvažme $Z \sim N(0, 1)$ a podívejme se na náhodnou veličinu $X = Z^2$.

$$\begin{aligned}F_X(x) &= P[Z^2 < x] \\&= \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\&= \int_0^x \frac{1}{\sqrt{2\pi}} t^{-1/2} e^{-t/2} dt\end{aligned}$$

s hustotou

$$f_X(x) = \frac{1}{\sqrt{2\pi}} t^{-1/2} e^{-t/2}.$$

Říkáme mu rozdělení χ^2 , píšeme $X \sim \chi^2(1)$.

kvantilová funkce

Je-li $F(x)$ distribuční funkce náhodné veličiny X , pak

$$F^{-1}(u) = \inf\{x \in \mathbb{R}; F(x) \geq u\}, \quad 0 < u < 1$$

je kvantilová funkce náhodné veličiny X .

Hodnota $F^{-1}(\alpha)$ se nazývá α -kvantil.

Tzv. kritické hodnoty pro veličinu X jsou pak $F^{-1}(1 - \alpha)$.

Čebyševova nerovnost

Theorem

Má-li X rozptyl a $\epsilon > 0$ je libovolné, pak platí

$$P(|X - EX| \geq \epsilon) \leq \frac{\text{var } X}{\epsilon^2}.$$

Kovariance veličin

Jsou-li X a Y dvě náhodné veličiny, pro které existují jejich konečné roptyly, pak definujeme jejich **kovarianci** vztahem

$$\text{cov}(X, Y) = E(X - E X)(Y - E Y).$$

Evidentně je $\text{cov}(X, X) = \text{var } X$ a $\text{cov}(X, Y) = \text{cov}(Y, X)$.

Theorem

Nechť existují konečné rozptyly veličin X a Y . Pak

- $\text{cov}(X, Y) = E(XY) - (E X)(E Y)$
- *pro jakékoliv skaláry a, b, c, d platí*
 $\text{cov}(a + bX, c + dY) = bd \text{cov}(X, Y)$
- $\text{var}(X + Y) = \text{var } X + \text{var } Y + 2 \text{cov}(X, Y)$.

Od kovariance snadno odvodíme tzv. **korelační koeficient** dvou náhodných veličin X a Y . Definujeme jej jako kovarianci příslušných normovaných veličin:

$$\rho_{X,Y} = \text{cov} \left(\frac{X - EX}{\sqrt{\text{var } X}}, \frac{Y - EY}{\sqrt{\text{var } Y}} \right) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var } X \text{ var } Y}}.$$

Theorem

- $\rho_{a+bX, c+dY} = \text{sign}(bd)\rho_{X,Y}$, pro $bd \neq 0$
- $\rho_{X,X} = 1$
- $\rho_{X,Y} = 0$, pokud jsou veličiny X a Y nezávislé.
- pokud je $\rho_{X,Y}$ definován, pak je roven jedné právě, když existují konstanty a, b, c tak, že $P(aX + bY = c) = 1$.

Varianční matice

Uvažme náhodný vektor $W = (X_1, \dots, X_n)$ takový, že pro všechny jeho komponenty existuje rozptyl. Pak **varianční matice** $\text{var } W$ je dána

$$\text{var } W = \begin{pmatrix} \text{var } X_1 & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{var } X_2 & \dots & \text{cov}(X_2, X_n) \\ & & \dots & \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \dots & \text{var } X_n \end{pmatrix}.$$

Theorem

Pro náhodný vektor X , skaláry a , matice skalárů B platí

$$\text{var}(a + BX) = B \text{var } XB^T.$$