Matematika IV – 11. přednáška
Náhodný vektor, náhodný výběr

Michal Bulant

Masarykova univerzita
Fakulta informatiky

5. 5. 2008

Obsah přednášky

Doporučené zdroje

- Martin Panák, Jan Slovák, **Drsná matematika**, *e*-text.
- Karel Zvára, Josef Štěpán, **Pravděpodobnost a matematická statistika**, Matfyzpress, 4. vydání, 2006, 230 stran, ISBN 80-867-3271-1.
- Marie Budíková, Štěpán Mikoláš, Pavel Osecký, **Popisná statistika**, Masarykova univerzita, 3. vydání, 2002, 48 stran, ISBN 80-210-1831-3.
- 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.

Plán přednášky

Náhodný vektor

Je-li $(\Omega, \mathcal{A}, P)$ pravděpodobnostní prostr a $X_1, \dots, X_n$ na něm definované náhodné veličiny s distribučními funkcemi $F_1, \dots, F_n$, pak **náhodným vektorem** je n-tice $X = (X_1, \dots, X_n)$ s distribuční funkcí definovanou vztahem

$$F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

V tomto kontextu nazýváme F *simultánní distribuční funkcí* náhodného vektoru X a F_i marginální distribuční funkcí náhodné veličiny X_i.

Podobně jako v případě diskrétní náhodné veličiny označuje $p(x_1, \ldots, x_n)$ pravděpodobnostní funkci **diskrétního náhodného vektoru** X, je-li

$$F(x_1, \ldots, x_n) = \sum_{t_1 \leq x_1} \cdots \sum_{t_n \leq x_n} p(t_1, \ldots, t_n).$$

[1]Obvykle zapisujeme ve statistice vektory do sloupců, proto bychom spíše měli psát $(X, Y)^T$.

Podobně jako v případě diskrétní náhodné veličiny označuje $p(x_1, \ldots, x_n)$ pravděpodobnostní funkci **diskrétního náhodného vektoru** X, je-li

$$F(x_1, \ldots, x_n) = \sum_{t_1 \leq x_1} \cdots \sum_{t_n \leq x_n} p(t_1, \ldots, t_n).$$

Funkci f_X nazveme **hustotou** normálního vektoru X, pokud pro libovolnou n-tici $(x_1, \ldots, x_n)$ platí

$$F_X(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \ldots, t_n)\, \mathrm{d}t_1 \ldots \mathrm{d}t_n.$$

[1] Obvykle zapisujeme ve statistice vektory do sloupců, proto bychom spíše měli psát $(X, Y)^T$.

Podobně jako v případě diskrétní náhodné veličiny označuje $p(x_1, \ldots, x_n)$ pravděpodobnostní funkci **diskrétního náhodného vektoru** X, je-li

$$F(x_1, \ldots, x_n) = \sum_{t_1 \leq x_1} \cdots \sum_{t_n \leq x_n} p(t_1, \ldots, t_n).$$

Funkci f_X nazveme **hustotou** normálního vektoru X, pokud pro libovolnou n-tici $(x_1, \ldots, x_n)$ platí

$$F_X(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \ldots, t_n)\, \mathrm{d}t_1 \ldots \mathrm{d}t_n.$$

Uvážíme-li diskrétní náhodný vektor (X, Y)[1], pak je vztah mezi sdruženým rozdělením vektoru (X, Y) a marginálním rozdělením promenné X určen rovností $P(X = x_i) = \sum_{j=1}^{\infty} P(X = x_i, Y = y_j)$, kde $y_1, \ldots$ tvoří úplný systém jevů.

[1]Obvykle zapisujeme ve statistice vektory do sloupců, proto bychom spíše měli psát $(X, Y)^T$.

Podobně jako v případě diskrétní náhodné veličiny označuje $p(x_1, \dots, x_n)$ pravděpodobnostní funkci **diskrétního náhodného vektoru** X, je-li

$$F(x_1, \dots, x_n) = \sum_{t_1 \le x_1} \cdots \sum_{t_n \le x_n} p(t_1, \dots, t_n).$$

Funkci f_X nazveme **hustotou** normálního vektoru X, pokud pro libovolnou n-tici $(x_1, \dots, x_n)$ platí

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n)\, \mathrm{d}t_1 \dots \mathrm{d}t_n.$$

Uvážíme-li diskrétní náhodný vektor (X, Y)[1], pak je vztah mezi sdruženým rozdělením vektoru (X, Y) a marginálním rozdělením promenné X určen rovností $P(X = x_i) = \sum_{j=1}^{\infty} P(X = x_i, Y = y_j)$, kde $y_1, \dots$ tvoří úplný systém jevů. Vztah pro spojitě rozdělený náhodný vektor je analogický.

[1] Obvykle zapisujeme ve statistice vektory do sloupců, proto bychom spíše měli psát $(X, Y)^T$.

(stochastická) Nezávislost náhodných veličin

Dříve uvedenou definici nezávislosti náhodných veličin $X_1, \ldots, X_n$ pomocí vztahu

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

pro libovolné $x_1, \ldots, x_n$, tak můžeme nyní přepsat pomocí vztahem mezi sdruženou distribuční funkcí náhodného vektoru $X = (X_1, \ldots, X_n)$ a marginálních distribučních funkcí náhodných veličin $X_1, \ldots, X_n$:

$$F_X(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

(stochastická) Nezávislost náhodných veličin

Dříve uvedenou definici nezávislosti náhodných veličin $X_1, \dots, X_n$ pomocí vztahu

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$$

pro libovolné $x_1, \dots, x_n$, tak můžeme nyní přepsat pomocí vztahem mezi sdruženou distribuční funkcí náhodného vektoru $X = (X_1, \dots, X_n)$ a marginálních distribučních funkcí náhodných veličin $X_1, \dots, X_n$:

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n).$$

Příklad

Házíme dvěma běžnými kostkami, jako náhodnou veličinu X označme součet bodů na obou kostkách, jako náhodnou veličinu Y absolutní hodnotu rozdílu. Určete sdružené rozdělení náhodného vektoru (X, Y), obě marginální rozdělení a odvoďte, jsou-li X a Y nezávislé.

Číselné charakteristiky náhodných vektorů

$E(X) = (E(X_1), \ldots, E(X_n))$ se nazývá vektor středních hodnot,

Číselné charakteristiky náhodných vektorů

$E(X) = (E(X_1), \ldots, E(X_n))$ se nazývá vektor středních hodnot,

$$\operatorname{var}(X) = \begin{pmatrix} D(X_1) & C(X_1, X_2) & \cdots & C(X_1, X_n) \\ \multicolumn{4}{c}{\dotfill} \\ C(X_n, X_1) & C(X_n, X_2) & \cdots & D(X_n) \end{pmatrix}$$

varianční (rozptylová) matice a

Číselné charakteristiky náhodných vektorů

$E(X) = (E(X_1), \dots, E(X_n))$ se nazývá vektor středních hodnot,

$$\operatorname{var}(X) = \begin{pmatrix} D(X_1) & C(X_1, X_2) & \cdots & C(X_1, X_n) \\ \multicolumn{4}{c}{\dotfill} \\ C(X_n, X_1) & C(X_n, X_2) & \cdots & D(X_n) \end{pmatrix}$$

varianční (rozptylová) matice a

$$\operatorname{cor} X = \begin{pmatrix} 1 & R(X_1, X_2) & \cdots & R(X_1, X_n) \\ \multicolumn{4}{c}{\dotfill} \\ R(X_n, X_1) & R(X_n, X_2) & \cdots & 1 \end{pmatrix}$$

je korelační matice.

Číselné charakteristiky náhodných vektorů

$E(X) = (E(X_1), \dots, E(X_n))$ se nazývá vektor středních hodnot,

$$\operatorname{var}(X) = \begin{pmatrix} D(X_1) & C(X_1, X_2) & \cdots & C(X_1, X_n) \\ \dots & \dots & \dots & \dots \\ C(X_n, X_1) & C(X_n, X_2) & \cdots & D(X_n) \end{pmatrix}$$

varianční (rozptylová) matice a

$$\operatorname{cor} X = \begin{pmatrix} 1 & R(X_1, X_2) & \cdots & R(X_1, X_n) \\ \dots & \dots & \dots & \dots \\ R(X_n, X_1) & R(X_n, X_2) & \cdots & 1 \end{pmatrix}$$

je korelační matice.

Snadno je po rozepsání po jednotlivých složkách vidět, že $\operatorname{var}(X) = E((X - E(X)) \cdot (X - E(X))^T)$

Ukážeme na příkladech, že pravděpodobnostní struktura náhodného vektoru (X, Y) není určena pouze marginálními rozděleními veličin X a Y. Podstatný je rovněž pravděpodobnostní vztah mezi X a Y, který je částečně popsán např. prostřednictvím korelačního koeficientu.

Ukážeme na příkladech, že pravděpodobnostní struktura náhodného vektoru (X, Y) není určena pouze marginálními rozděleními veličin X a Y. Podstatný je rovněž pravděpodobnostní vztah mezi X a Y, který je částečně popsán např. prostřednictvím korelačního koeficientu.

Příklad

Jsou-li X a Y náhodné veličiny, nabývající hodnot 0 a 1, pak

$$P(X = 1, Y = 1) - P(X = 1)P(Y = 1) = E(XY) - E(X)E(Y) =$$
$$= \operatorname{cov}(X, Y).$$

Ukážeme na příkladech, že pravděpodobnostní struktura náhodného vektoru (X, Y) není určena pouze marginálními rozděleními veličin X a Y. Podstatný je rovněž pravděpodobnostní vztah mezi X a Y, který je částečně popsán např. prostřednictvím korelačního koeficientu.

Příklad

Jsou-li X a Y náhodné veličiny, nabývající hodnot 0 a 1, pak

$$P(X = 1, Y = 1) - P(X = 1)P(Y = 1) = E(XY) - E(X)E(Y) =$$
$$= \operatorname{cov}(X, Y).$$

Odtud je snadno vidět, že pokud jsou X a Y nekorelované, jsou i nezávislé (což obecně neplatí).

Uveďme ještě příklad, ilustrující, že nekorelovanost nemusí implikovat nezávislost:

Uveďme ještě příklad, ilustrující, že nekorelovanost nemusí implikovat nezávislost:

Příklad

Buďte A a X nezávislé náhodné veličiny, splňující $X \sim N(0,1)$ a $P(A=1) = P(A=-1) = 1/2$. Položíme-li $Y = AX$, pak

$$P(Y < y) = \frac{1}{2}P(X < y) + \frac{1}{2}P(-X < y) = \Phi(y),$$

proto má rovněž Y rozdělení $N(0,1)$.

Uveďme ještě příklad, ilustrující, že nekorelovanost nemusí implikovat nezávislost:

Příklad

Buďte A a X nezávislé náhodné veličiny, splňující $X \sim N(0,1)$ a $P(A=1) = P(A=-1) = 1/2$. Položíme-li $Y = AX$, pak

$$P(Y < y) = \frac{1}{2}P(X < y) + \frac{1}{2}P(-X < y) = \Phi(y),$$

proto má rovněž Y rozdělení $N(0,1)$.

Dále $\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y) = E(AX^2) = E(A)E(X^2) = 0 \cdot 1 = 0$, přitom $P(X = Y) = P(X = -Y) = 1/2$ a X, Y zřejmě nejsou nezávislé.

Příklad

Nechť (X, Y) je náhodný vektor, který má rovnoměrné rozdělení na jednotkovém kruhu $K = \{(x, y) : x^2 + y^2 \leq 1\}$. Zřejmě je hustota tohoto rozdělení rovna $1/\pi$ pro $(x, y) \in K$ a 0 jinde a je rovněž vidět, že X, Y nejsou nezávislé.

Příklad

Nechť (X, Y) je náhodný vektor, který má rovnoměrné rozdělení na jednotkovém kruhu $K = \{(x, y) : x^2 + y^2 \leq 1\}$. Zřejmě je hustota tohoto rozdělení rovna $1/\pi$ pro $(x, y) \in K$ a 0 jinde a je rovněž vidět, že X, Y nejsou nezávislé. Označme $R = R(X, Y) a \Phi = \Phi(X, Y)$ polární souřadnice náhodného vektoru (X, Y) a určíme rozdělení vektoru (R, Φ).

Příklad

Nechť (X, Y) je náhodný vektor, který má rovnoměrné rozdělení na jednotkovém kruhu $K = \{(x, y) : x^2 + y^2 \leq 1\}$. Zřejmě je hustota tohoto rozdělení rovna $1/\pi$ pro $(x, y) \in K$ a 0 jinde a je rovněž vidět, že X, Y nejsou nezávislé. Označme
$R = R(X, Y) a \Phi = \Phi(X, Y)$ polární souřadnice náhodného vektoru (X, Y) a určíme rozdělení vektoru (R, Φ).
Pro $0 < r_1 \leq 1$ a $0 < \varphi_1 \leq 2\pi$ je

$$\begin{aligned} P(R < r_1, \Phi \leq \varphi_1) &= \frac{1}{\pi}\pi r_1^2 \frac{\varphi_1}{2\pi} = \\ &= \int_0^{r_1} \int_0^{\varphi_1} \frac{1}{2\pi} 2r \, \mathrm{d}\varphi \, \mathrm{d}r. \end{aligned}$$

Hustota je tedy rovna $f(r, \varphi) = \frac{r}{\pi}$ pro $0 < r \leq 1$, $0 < \varphi \leq 2\pi$ a rovna 0 všude jinde.

Příklad (pokr.)

Marginální hustoty $g(r)$ a $h(\varphi)$ veličin R a Φ se nyní snadno dopočtou:

$$g(r) = \int_{-\infty}^{\infty} f(r, \varphi)\,\mathrm{d}\varphi = \int_0^{2\pi} \frac{r}{\pi}\,\mathrm{d}\varphi = 2r$$

$$h(\varphi) = \int_{-\infty}^{\infty} f(r, \varphi)\,\mathrm{d}r = \int_0^{1} \frac{r}{\pi}\,\mathrm{d}r = \frac{1}{2\pi}.$$

Příklad (pokr.)

Marginální hustoty $g(r)$ a $h(\varphi)$ veličin R a Φ se nyní snadno dopočtou:

$$g(r) = \int_{-\infty}^{\infty} f(r,\varphi)\,\mathrm{d}\varphi = \int_0^{2\pi} \frac{r}{\pi}\,\mathrm{d}\varphi = 2r$$

$$h(\varphi) = \int_{-\infty}^{\infty} f(r,\varphi)\,\mathrm{d}r = \int_0^{1} \frac{r}{\pi}\,\mathrm{d}r = \frac{1}{2\pi}.$$

Veličina Φ má rovnoměrné rozdělení $(0, 2\pi\rangle$, odkud $E(\Phi) = \pi$ a $D(\Phi) = \pi^2/3$, snadno rovněž odvodíme $E(R) = 2/3$, $D(R) = 1/18$.

Příklad (pokr.)

Marginální hustoty $g(r)$ a $h(\varphi)$ veličin R a Φ se nyní snadno dopočtou:

$$g(r) = \int_{-\infty}^{\infty} f(r,\varphi)\,\mathrm{d}\varphi = \int_0^{2\pi} \frac{r}{\pi}\,\mathrm{d}\varphi = 2r$$

$$h(\varphi) = \int_{-\infty}^{\infty} f(r,\varphi)\,\mathrm{d}r = \int_0^{1} \frac{r}{\pi}\,\mathrm{d}r = \frac{1}{2\pi}.$$

Veličina Φ má rovnoměrné rozdělení $(0, 2\pi\rangle$, odkud $E(\Phi) = \pi$ a $D(\Phi) = \pi^2/3$, snadno rovněž odvodíme $E(R) = 2/3$, $D(R) = 1/18$.
Všimněme si ale zejména, že $f(r,\varphi) = g(r)h(\varphi)$, což znamená nezávislost veličin R a Φ.

Vlastnosti charakteristik náhodného vektoru

Věta

Pro náhodné vektory X, Y stejné dimenze, konstantní matici B a konstantní vektor a (odpovídajících dimenzí) platí

- $E(X + Y) = E(X) + E(Y)$,

Vlastnosti charakteristik náhodného vektoru

Věta

Pro náhodné vektory X, Y stejné dimenze, konstantní matici B a konstantní vektor a (odpovídajících dimenzí) platí

- $E(X+Y) = E(X) + E(Y)$,
- $E(a+BX) = a + B \cdot E(X)$,

Vlastnosti charakteristik náhodného vektoru

Věta

Pro náhodné vektory X, Y stejné dimenze, konstantní matici B a konstantní vektor a (odpovídajících dimenzí) platí

- $E(X+Y) = E(X) + E(Y)$,
- $E(a+BX) = a + B \cdot E(X)$,
- $\mathrm{var}(a + B \cdot X) = B\,\mathrm{var}(X) B^T$.

Vlastnosti charakteristik náhodného vektoru

Věta

Pro náhodné vektory X, Y stejné dimenze, konstantní matici B a konstantní vektor a (odpovídajících dimenzí) platí

- $E(X+Y) = E(X) + E(Y)$,
- $E(a+BX) = a + B \cdot E(X)$,
- $\operatorname{var}(a + B \cdot X) = B \operatorname{var}(X) B^T$.

Vlastnosti charakteristik náhodného vektoru

Věta

Pro náhodné vektory X, Y stejné dimenze, konstantní matici B a konstantní vektor a (odpovídajících dimenzí) platí

- $E(X+Y) = E(X) + E(Y)$,
- $E(a+BX) = a + B \cdot E(X)$,
- $\operatorname{var}(a + B \cdot X) = B \operatorname{var}(X) B^T$.

Důkaz.

Důkaz vyplývá z vlastností náhodných veličin a ze vztahu $\operatorname{var}(X) = E((X - E(X))(X - E(X))^T)$. $\square$

Mnohorozměrné normální rozdělení

Věta

Nechť jsou složky náhodného vektoru $Z = (Z_1, \dots, Z_n)$ *nezávislé a mají rozdělení* $Z_i \sim N(0,1)$, *dále nechť* Q *je ortonormální matice řádu* n. *Pak jsou rovněž složky náhodného vektoru* $U = Q^T Z$ *nezávislé a každá má rozdělení* $N(0,1)$.

Má tedy U (stejně jako Z) nulovou střední hodnotu a jednotkovou varianční matici a oba vektory jsou zobecněním normovaného normálního rozdělení. V následující definice zavedeme zobecnění normálního rozdělení s obecnými parametry:

Mnohorozměrné normální rozdělení

Věta

Nechť jsou složky náhodného vektoru $Z = (Z_1, \dots, Z_n)$ *nezávislé a mají rozdělení* $Z_i \sim N(0,1)$, *dále nechť* Q *je ortonormální matice řádu* n. *Pak jsou rovněž složky náhodného vektoru* $U = Q^T Z$ *nezávislé a každá má rozdělení* $N(0,1)$.

Má tedy U (stejně jako Z) nulovou střední hodnotu a jednotkovou varianční matici a oba vektory jsou zobecněním normovaného normálního rozdělení. V následující definice zavedeme zobecnění normálního rozdělení s obecnými parametry:

Definice

Nechť jsou složky náhodného vektoru $Z = (Z_1, \dots, Z_n)$ nezávislé a mají rozdělení $Z_i \sim N(0,1)$ a nechť $a \in \mathbb{R}^m$ je vektor konstant a B konstantní matice typu $m \times n$. Označme dále $V = V \cdot B^T$. Pak řekneme, že náhodný vektor $U = a + B \cdot Z$ má m-**rozměrné normální rozdělení** $N_m(a, V)$.

Pomocí vlastností charakteristik snadno spočítáme, že $E(U) = a, \mathrm{var}(U) = V = BB^T$. Pokud je matice V regulární, pak existuje hustota náhodného vektoru a je tvaru

$$f(u_1, \ldots, u_m) = (2\pi)^{-m/2} |V|^{-1/2} \exp\left(-\frac{1}{2}(u-a)^T V^{-1}(u-a)\right).$$

Pomocí vlastností charakteristik snadno spočítáme, že $E(U) = a, \mathrm{var}(U) = V = BB^T$. Pokud je matice V regulární, pak existuje hustota náhodného vektoru a je tvaru

$$f(u_1, \ldots, u_m) = (2\pi)^{-m/2} |V|^{-1/2} \exp\left(-\frac{1}{2}(u-a)^T V^{-1}(u-a) \right).$$

Pro úvahy ve statistice je důležitá následující věta.

Věta

Nechť má vektor U rozdělení $N_m(a, V)$, nechť $c \in \mathbb{R}^k$ a matice D typu $k \times m$ jsou konstanty. Pak má $c + D \cdot U$ k-rozměrné normální rozdělení $N_k(c + Da, DVD^T)$.

Pomocí vlastností charakteristik snadno spočítáme, že $E(U) = a, \mathrm{var}(U) = V = BB^T$. Pokud je matice V regulární, pak existuje hustota náhodného vektoru a je tvaru

$$f(u_1, \ldots, u_m) = (2\pi)^{-m/2}|V|^{-1/2} \exp\left(-\frac{1}{2}(u-a)^T V^{-1}(u-a)\right).$$

Pro úvahy ve statistice je důležitá následující věta.

Věta

Nechť má vektor U rozdělení $N_m(a, V)$, nechť $c \in \mathbb{R}^k$ a matice D typu $k \times m$ jsou konstanty. Pak má $c + D \cdot U$ k-rozměrné normální rozdělení $N_k(c + Da, DVD^T)$.

Důkaz.

Vyjádříme-li matici $V = BB^T$, dostáváme

$$c + DU = c + D(a + BZ) = (c + Da) + (DB)Z = \\ \sim N_k(c + Da, DBB^T D^T).$$

Speciálně je tedy marginální rozdělení podvektoru vektoru s mnohorozměrným normálním rozdělením opět mnohorozměrné normální a je-li navíc D jednořádková matice, dostáváme, že libovolná lineární funkce takového vektoru má normální rozdělení.

Speciálně je tedy marginální rozdělení podvektoru vektoru s mnohorozměrným normálním rozdělením opět mnohorozměrné normální a je-li navíc D jednořádková matice, dostáváme, že libovolná lineární funkce takového vektoru má normální rozdělení.

Připomeňme ještě jednou rozdělení odvozená od normálního:

rozdělení	transformace	střední hodnota	rozptyl
$N(\mu, \sigma^2)$	$\mu + \sigma Z$	μ	σ^2
$\chi^2(k)$	$X_k^2 = \sum_{j=1}^k Z_j^2$	k	$2k$
$t(k)$	$\frac{Z}{\sqrt{X_k^2/k}}$	0	$k/(k-2)$
$F(k, m)$	$\frac{x_k^2/k}{X_m^2/m}$	$m/(m-2)$	$\frac{2m^2(k+m-2)}{k(m-2)^2(m-4)}$

Plán přednášky

Definice

Náhodným výběrem rozsahu n rozumíme n-tici **nezávislých** a **stejně rozdělených** náhodných veličin $X_1, \ldots, X_n \sim F_X(x)$.

Definice

Náhodným výběrem rozsahu n rozumíme n-tici **nezávislých** a **stejně rozdělených** náhodných veličin $X_1, \ldots, X_n \sim F_X(x)$.
Náhodným výběrem rozsahu n **s** p**-rozměrného rozdělení** rozumíme n-tici **nezávislých** a **stejně rozdělených** p-rozměrných náhodných vektorů.

Definice

Náhodným výběrem rozsahu n rozumíme n-tici **nezávislých** a **stejně rozdělených** náhodných veličin $X_1, \ldots, X_n \sim F_X(x)$.
Náhodným výběrem rozsahu n **s** p**-rozměrného rozdělení** rozumíme n-tici **nezávislých** a **stejně rozdělených** p-rozměrných náhodných vektorů.

V matematické statistice často pracujeme s transformacemi náhodného výběru, takovým náhodným veličinám (příp. vektorům) říkáme **statistiky**. V následujícím zavedem několik důležitých statistik a ukážeme jejich souvislost s číselnými charakteristikami náhodných veličin.

Základní statistiky

Definice

Nechť $X_1, \ldots, X_n$ je náhodný výběr. Statistiku

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i$$

nazýváme **výběrový průměr**, statistiku

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - M)^2$$

výběrový rozptyl a statistiku $S = \sqrt{S^2}$ **výběrová směrodatná odchylka**. Analogicky se definují i výběrová kovariance, příp. výběrový korelační koeficient pro dvourozměrný náhodný výběr.

Vlastnosti statistik

Protože jsou uvedené statistiky náhodnými veličinami, lze se přirozeně ptát po jejich číselných charakteristikách.

Věta

Nechť $X_1, \dots, X_n$ je náhodný výběr rozsahu n z rozdělení se střední hodnotou μ a rozptylem σ^2 . Pak platí:

- $E(M) = \mu$,

Vlastnosti statistik

Protože jsou uvedené statistiky náhodnými veličinami, lze se přirozeně ptát po jejich číselných charakteristikách.

Věta

Nechť $X_1, \ldots, X_n$ je náhodný výběr rozsahu n z rozdělení se střední hodnotou μ a rozptylem σ^2 . Pak platí:

- $E(M) = \mu$,
- $D(M) = \mathrm{var}(M) = \sigma^2/n$,

Vlastnosti statistik

Protože jsou uvedené statistiky náhodnými veličinami, lze se přirozeně ptát po jejich číselných charakteristikách.

Věta

Nechť $X_1, \ldots, X_n$ je náhodný výběr rozsahu n z rozdělení se střední hodnotou μ a rozptylem σ^2 . Pak platí:

- $E(M) = \mu$,
- $D(M) = \mathrm{var}(M) = \sigma^2/n$,
- $E(S^2) = \sigma^2$.

Důkaz.

Ukážeme jen (nejsložitější) 3. tvrzení.
Snadno se odvodí, že platí

$$\sum (X_i - \mu)^2 = \sum (X_i - M)^2 + n(M - \mu)^2.$$

Důkaz.

Ukážeme jen (nejsložitější) 3. tvrzení.
Snadno se odvodí, že platí

$$\sum (X_i - \mu)^2 = \sum (X_i - M)^2 + n(M - \mu)^2.$$

Proto je

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} E(\sum (X_i - \mu)^2 - \frac{n}{n-1} E(M - \mu)^2 = \\ &= \frac{1}{n-1} \sum \operatorname{var}(X_i) - \frac{n}{n-1} \operatorname{var}(M) = \\ &= \frac{n}{n-1} \sigma^2 - \frac{1}{n-1} \sigma^2 = \sigma^2. \end{aligned}$$

$\square$

V předchozí větě jsme ukázali, že výběrový průměr M splňuje $E(M) = \mu$, jehostřední hodnota tedy rovna odhadovanému parametru μ. V takovém případě říkáme, že statistika M je **nestranným odhadem** parametru μ.

V předchozí větě jsme ukázali, že výběrový průměr M splňuje $E(M) = \mu$, jehostřední hodnota tedy rovna odhadovanému parametru μ. V takovém případě říkáme, že statistika M je **nestranným odhadem** parametru μ.
Podobně jsme viděli, že S^2 je nestranným odhadem parametru σ^2.

V předchozí větě jsme ukázali, že výběrový průměr M splňuje $E(M) = \mu$, jehostřední hodnota tedy rovna odhadovanému parametru μ. V takovém případě říkáme, že statistika M je **nestranným odhadem** parametru μ.
Podobně jsme viděli, že S^2 je nestranným odhadem parametru σ^2.
Všimněme si rovněž, že ,,přirozeněji'' definovaná statistika $\frac{1}{n}\sum(X_i - M)^2$ není nestranným odhadem σ^2, její střední hodnota je totiž $\frac{n-1}{n}\sigma^2$.

V předchozí větě jsme ukázali, že výběrový průměr M splňuje $E(M) = \mu$, jehostřední hodnota tedy rovna odhadovanému parametru μ. V takovém případě říkáme, že statistika M je **nestranným odhadem** parametru μ.
Podobně jsme viděli, že S^2 je nestranným odhadem parametru σ^2.
Všimněme si rovněž, že ,,přirozeněji" definovaná statistika $\frac{1}{n}\sum(X_i - M)^2$ není nestranným odhadem σ^2, její střední hodnota je totiž $\frac{n-1}{n}\sigma^2$. Rozmyslete si, je-li S nestranným odhadem směrodatné odchylky σ.

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \dots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \dots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- *$M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.*

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- *$M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.*
- *$K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.*

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \dots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- *$M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.*
- *$K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.*
- *$\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.*

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- $M \sim N(\mu, \sigma^2/n)$, *a tedy* $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.
- $K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.
- $\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.
- $T = (M - \mu)/(S/\sqrt{n}) \sim t(n-1)$.

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- *$M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.*
- *$K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.*
- *$\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.*
- *$T = (M - \mu)/(S/\sqrt{n}) \sim t(n-1)$.*

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- *$M \sim N(\mu, \sigma^2/n)$, a tedy $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.*
- *$K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.*
- *$\sum (X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.*
- *$T = (M - \mu)/(S/\sqrt{n}) \sim t(n-1)$.*

Poznámka

K odhadu μ, známe-li σ^2, slouží U, v opačném případě T.

Náhodný výběr z normálního rozdělení

Uvažme nyní speciální případ, kdy je $X_1, \ldots, X_n$ náhodný výběr z normálního rozdělení $N(\mu, \sigma^2)$.

Věta

- *M a S^2 jsou nezávislé náhodné veličiny.*
- $M \sim N(\mu, \sigma^2/n)$*, a tedy* $U = (M - \mu)/(\sigma/\sqrt{n}) \sim N(0, 1)$.
- $K = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.
- $\sum(X_i - \mu)^2/\sigma^2 \sim \chi^2(n)$.
- $T = (M - \mu)/(S/\sqrt{n}) \sim t(n-1)$.

Poznámka

K odhadu μ, známe-li σ^2, slouží U, v opačném případě T.
K odhadu σ^2, neznáme-li μ, slouží K, v opačném případě následující (bezejmenná?) statistika, která je vlastně statistikou K, v níž místo odhadu M použijeme přímo μ.

Důkaz.

Položme $Z_i = (X_i - \mu)/\sigma$, což jsou zřejmě nezávislé náhodné veličiny s normovaným normálním rozdělením. Zřejmě je $X = a + \sigma E_n Z$, kde a je vektor samých μ, a proto má podle předchozího X mnohorozměrné normální rozdělení a je-li dále d vektor ze samých $1/n$, pak má náhodná veličina $M = d^T X$ (jednorozměrné) normální rozdělení se střední hodnotou $d^T a = \mu$ a rozptylem $d^T \sigma^2 E_n d = \sigma^2/n$.

Důkaz.

Položme $Z_i = (X_i - \mu)/\sigma$, což jsou zřejmě nezávislé náhodné veličiny s normovaným normálním rozdělením. Zřejmě je $X = a + \sigma E_n Z$, kde a je vektor samých μ, a proto má podle předchozího X mnohorozměrné normální rozdělení a je-li dále d vektor ze samých $1/n$, pak má náhodná veličina $M = d^T X$ (jednorozměrné) normální rozdělení se střední hodnotou $d^T a = \mu$ a rozptylem $d^T \sigma^2 E_n d = \sigma^2/n$.
Ostatní tvrzení se dokážou obdobně. $\square$

Příklad

V roce 1951 bylo rozsáhlým statistickým průzkumem zjištěno, že střední hodnota výšky desetiletých chlapců je $136,1$ cm se směrodatnou odchylkou $\sigma = 6,4$ cm.

Příklad

V roce 1951 bylo rozsáhlým statistickým průzkumem zjištěno, že střední hodnota výšky desetiletých chlapců je $136,1$ cm se směrodatnou odchylkou $\sigma = 6,4$ cm.
V roce 1961 byla zjištěna výška pouze u 15 náhodně vybraných chlapců:

130	140	136	141	139	133	149	151
139	136	138	142	127	139	147	

Otázkou je, zda se v porovnání s rokem 1951 změnila střední výška chlapců, pokud předpokládáme, že variabilita výšek se v různých generacích příliš nemění.

Řešení

Vzhledem k tomu, že základní soubor všech desetiletých chlapců je rozsáhlý, lze zmíněná data považovat za náhodný výběr[a]. Zjistíme, že $M = 139,133$, $n = 15$ a s využitím statitstiky U dostáváme, že s 95% pravděpodobností leží hodnota μ v intervalu

$$(M - 1,96\sigma/\sqrt{n}; M + 1,96\sigma/\sqrt{n}) = (135,9; 142,4).$$

Protože i střední hodnota výšek z roku 1951 leží v tomto intervalu, nemá vážný důvod tvrdit, že se střední výška změnila.

Řešení

Vzhledem k tomu, že základní soubor všech desetiletých chlapců je rozsáhlý, lze zmíněná data považovat za náhodný výběr[a]. Zjistíme, že $M = 139,133$, $n = 15$ a s využitím statitstiky U dostáváme, že s 95% pravděpodobností leží hodnota μ v intervalu

$$(M - 1,96\sigma/\sqrt{n}; M + 1,96\sigma/\sqrt{n}) = (135,9; 142,4).$$

Protože i střední hodnota výšek z roku 1951 leží v tomto intervalu, nemá vážný důvod tvrdit, že se střední výška změnila. Pokud bychom ovšem připustili vyšší možnost omylu a stanovili interval se spolehlivostí pouze 90%, pak bychom na této hladině hypotézu, že střední výška se změnila, přijali – interval je nyní (136,41;141,85).

Řešení

Vzhledem k tomu, že základní soubor všech desetiletých chlapců je rozsáhlý, lze zmíněná data považovat za náhodný výběr[a]. Zjistíme, že $M = 139,133$, $n = 15$ a s využitím statitstiky U dostáváme, že s 95% pravděpodobností leží hodnota μ v intervalu

$$(M - 1,96\sigma/\sqrt{n}; M + 1,96\sigma/\sqrt{n}) = (135,9; 142,4).$$

Protože i střední hodnota výšek z roku 1951 leží v tomto intervalu, nemá vážný důvod tvrdit, že se střední výška změnila. Pokud bychom ovšem připustili vyšší možnost omylu a stanovili interval se spolehlivostí pouze 90%, pak bychom na této hladině hypotézu, že střední výška se změnila, přijali – interval je nyní (136,41;141,85). Podobně, pokud nás zajímá pouze **dolní odhad** střední hodnoty výšek chlapců (a vůbec tedy nepřipouštíme možnost, že by se střední výška snížila), pak s 95% pravděpodobností je střední výška větší než 136,41, a tedy nyní opět přijímáme hypotézu, že se střední výška zvýšila.