Matematika IV – 11. přednáška
Normální rozdělení, limitní vlastnosti, zákony velkých čísel

Michal Bulant

Masarykova univerzita
Fakulta informatiky

3. 5. 2010

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

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X,Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X,Y) = C(X,Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X,Y)| \leq 1$,

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X, Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X, Y) = C(X, Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X, Y)| \leq 1$,
- kvantily,

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X, Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X, Y) = C(X, Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X, Y)| \leq 1$,
- kvantily,
- další momenty (obecné, centrální) - momentová vytvořující funkce $M_X(t) = E(e^{tX})$

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X, Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X, Y) = C(X, Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X, Y)| \leq 1$,
- kvantily,
- další momenty (obecné, centrální) - momentová vytvořující funkce $M_X(t) = E(e^{tX})$

Věta

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X, Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X, Y) = C(X, Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X, Y)| \leq 1$,
- kvantily,
- další momenty (obecné, centrální) - momentová vytvořující funkce $M_X(t) = E(e^{tX})$

Věta

- *Pro nezávislé náhodné veličiny platí* $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X, Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X, Y) = C(X, Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X, Y)| \leq 1$,
- kvantily,
- další momenty (obecné, centrální) - momentová vytvořující funkce $M_X(t) = E(e^{tX})$

Věta

- *Pro nezávislé náhodné veličiny platí* $M_{X+Y}(t) = M_X(t)M_Y(t)$.
- *r-tý obecný moment μ'_r náhodné veličiny X je koeficient u $\frac{t^r}{r!}$ v rozvoji M_X do exponenciální mocninné řady (tedy např. $EX = \mu'_1, DX = \mu'_2 - (\mu'_1)^2$).*

Charakteristiky náhodných veličin – připomenutí

- střední hodnota $E(X)$,
- rozptyl $D(X) = E([X - E(X)]^2)$, směrodatná odchylka $\sqrt{D(X)}$
- kovariance $C(X, Y) = E([X - E(X)][Y - E(Y)])$, korelační koeficient $R(X, Y) = C(X, Y)/(\sqrt{D(X)}\sqrt{D(Y)})$, Cauchyova nerovnost $|R(X, Y)| \leq 1$,
- kvantily,
- další momenty (obecné, centrální) - momentová vytvořující funkce $M_X(t) = E(e^{tX})$

Věta

- *Pro nezávislé náhodné veličiny platí* $M_{X+Y}(t) = M_X(t)M_Y(t)$.
- *r-tý obecný moment* μ_r' *náhodné veličiny X je koeficient u* $\frac{t^r}{r!}$ *v rozvoji* M_X *do exponenciální mocninné řady (tedy např.* $EX = \mu_1'$, $DX = \mu_2' - (\mu_1')^2$*).*
- *Je-li* $Y = a + bX$, *pak* $M_Y(t) = e^{at} M_X(bt)$.

Plán přednášky

Momenty normálního rozdělení

Přímý výpočet střední hodnoty a rozptylu normovaného normálního rozdělení není triviální. S využitím momentové vytvořující funkce je ale poměrně jednoduchý.

Nechť $Z \sim N(0,1)$. Pak

$$
\begin{aligned}
M_Z(t) &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z = \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2 - 2tz + t^2 - t^2}{2}\right) \mathrm{d}z = \\
&= \exp\left(\frac{t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(z-t)^2}{2}\right) \mathrm{d}z = \exp\left(\frac{t^2}{2}\right).
\end{aligned}
$$

Poslední integrál je roven 1 díky tomu, že na místě integrované funkce je funkce s vlastnostmi hustoty.

Střední hodnota a rozptyl normálního rozdělení

S využitím předchozího výpočtu $M_Z(t) = \exp\left(\frac{t^2}{2}\right)$ snadno spočítáme, že

$$M'_Z(t) = t\exp\left(\frac{t^2}{2}\right),$$

$$M''_Z(t) = t^2\exp\left(\frac{t^2}{2}\right) + \exp\left(\frac{t^2}{2}\right).$$

Dosazením $t = 0$ pak dostaneme

$$E(Z) = 0, D(Z) = 1.$$

Střední hodnota a rozptyl normálního rozdělení

S využitím předchozího výpočtu $M_Z(t) = \exp\left(\frac{t^2}{2}\right)$ snadno spočítáme, že

$$M'_Z(t) = t\exp\left(\frac{t^2}{2}\right),$$

$$M''_Z(t) = t^2\exp\left(\frac{t^2}{2}\right) + \exp\left(\frac{t^2}{2}\right).$$

Dosazením $t = 0$ pak dostaneme

$$E(Z) = 0, D(Z) = 1.$$

Pro transformovanou náhodnou veličinu $Y = \mu + \sigma Z \sim N(\mu, \sigma^2)$ pak snadno odvodíme z vlastností střední hodnoty, resp. rozptylu, že $E(Y) = \mu, D(Y) = \sigma^2$ (což zpětně zdůvodňuje zápis $N(\mu, \sigma^2)$). Momentová vytvořující funkce má tvar

$$M_Y(t) = \exp\left(\mu t + \sigma^2\frac{t^2}{2}\right).$$

Příklad

Určete rozdělení součtu nezávislých náhodných veličin $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$.

Příklad

Určete rozdělení součtu nezávislých náhodných veličin $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$.

Řešení

Z vlastností momentové vytvořující funkce dostáváme

$$M_{X+Y}(t) = \exp\left(\mu_X t + \sigma_X^2 \frac{t^2}{2}\right) \exp\left(\mu_Y t + \sigma_Y^2 \frac{t^2}{2}\right) = \\ = \exp\left((\mu_X + \mu_Y)t + (\sigma_X^2 + \sigma_Y^2)\frac{t^2}{2}\right).$$

Proto $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Γ (gamma) rozdělení

Γ rozdělení se často používá u modelů čekání (např. v pojistné matematice je čas dožití často modelován pomocí gamma rozdělení).

Příklad

Určete konstantu c tak, aby funkce $cx^{a-1}e^{-bx}$ pro $x > 0$ a nulová jinde ($a, b > 0$ jsou parametry) byla hustotou náhodné veličiny.

Γ (gamma) rozdělení

Γ rozdělení se často používá u modelů čekání (např. v pojistné matematice je čas dožití často modelován pomocí gamma rozdělení).

Příklad

Určete konstantu c tak, aby funkce $cx^{a-1}e^{-bx}$ pro $x > 0$ a nulová jinde ($a, b > 0$ jsou parametry) byla hustotou náhodné veličiny.

Řešení

Hustota musí splňovat

$$
\begin{aligned}
1 &= \int_0^\infty cx^{a-1}e^{-bx}\,\mathrm{d}x = \\
&= \int_0^\infty c\left(\frac{t}{b}\right)^{a-1} e^{-t}\frac{1}{b}\,\mathrm{d}t = \\
&= \frac{c}{b^a}\int_0^\infty t^{a-1}e^{-t}\,\mathrm{d}t = \frac{c}{b^a}\Gamma(a).
\end{aligned}
$$

Poznámka

Funkce Γ je zobecnění faktoriálu ($\Gamma(n) = (n-1)!$ pro $n \in \mathbb{N}$), definované předpisem $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,\mathrm{d}x$. Často počítáme hodnoty této funkce s využitím vlastností
$\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(a+1) = a \cdot \Gamma(a)$.

Poznámka

Funkce Γ je zobecnění faktoriálu ($\Gamma(n) = (n-1)!$ pro $n \in \mathbb{N}$), definované předpisem $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,\mathrm{d}x$. Často počítáme hodnoty této funkce s využitím vlastností $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(a+1) = a \cdot \Gamma(a)$.

Definice

Rozdělení náhodné veličiny s hustotou

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

spočítanou v předchozím příkladu nazýváme **gamma rozdělení** s parametry a, b a značíme $\Gamma(a, b)$.

Poznámka

Funkce Γ je zobecnění faktoriálu ($\Gamma(n) = (n-1)!$ pro $n \in \mathbb{N}$), definované předpisem $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,\mathrm{d}x$. Často počítáme hodnoty této funkce s využitím vlastností $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(a+1) = a \cdot \Gamma(a)$.

Definice

Rozdělení náhodné veličiny s hustotou

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

spočítanou v předchozím příkladu nazýváme **gamma rozdělení** s parametry a, b a značíme $\Gamma(a, b)$. Momentová vytvořující funkce je pak $M(t) = (b/b-t)^a$, střední hodnota $E(X) = a/b$ a rozptyl $D(X) = a/b^2$.

Příklad (rozdělení χ^2 podruhé)

Nechť Z má normované normální rozdělení. Určete hustotu transformované náhodné veličiny $X = Z^2$.

Příklad (rozdělení χ^2 podruhé)

Nechť Z má normované normální rozdělení. Určete hustotu transformované náhodné veličiny $X = Z^2$.

Řešení

Již dříve jsme vypočetli přímým výpočtem přes distribuční funkci, že hustota

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

a řekli jsme si, že jde o (Pearsonovo) χ^2 rozdělení s jedním stupněm volnosti, které značíme $X \sim \chi^2(1)$.

Příklad (rozdělení χ^2 podruhé)

Nechť Z má normované normální rozdělení. Určete hustotu transformované náhodné veličiny $X = Z^2$.

Řešení

Již dříve jsme vypočetli přímým výpočtem přes distribuční funkci, že hustota

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

a řekli jsme si, že jde o (Pearsonovo) χ^2 rozdělení s jedním stupněm volnosti, které značíme $X \sim \chi^2(1)$. Nyní vidíme, že jde o speciální případ Γ-rozdělení, totiž $\Gamma(1/2, 1/2)$.

Obecně pro součet Y čtverců n nezávislých náhodných veličin s rozdělením $N(0, 1)$ obdobně odvodíme, že má rozdělení $\Gamma(n/2, 1/2)$ a říkáme, že Y má rozdělení $\chi^2(n)$ (*chí kvadrát s n stupni volnosti*). Toto rozdělení se ve statistice používá velmi často.

Další důležitá rozdělení

F-rozdělení

Jsou-li X, Y nezávislé náhodné veličiny s rozděleními $X \sim \chi^2(k)$, $Y \sim \chi^2(m)$, pak má transformovaná náhodná veličina

$$U = \frac{X/k}{Y/m}$$

takzvané Fisher-Snedecorovo F-rozdělení $F(k, m)$ s k a m stupni volnosti.

Další důležitá rozdělení

F-rozdělení

Jsou-li X, Y nezávislé náhodné veličiny s rozděleními $X \sim \chi^2(k)$, $Y \sim \chi^2(m)$, pak má transformovaná náhodná veličina

$$U = \frac{X/k}{Y/m}$$

takzvané Fisher-Snedecorovo F-rozdělení $F(k, m)$ s k a m stupni volnosti.

Studentovo **t-rozdělení**

Jsou-li $Z \sim N(0, 1)$ a $X \sim \chi^2(n)$ nezávislé náhodné veličiny, pak má veličina

$$T = \frac{Z}{\sqrt{X/n}}$$

tzv. Studentovo t-rozdělení $t(n)$ s n stupni volnosti.

Přehled rozdělení odvozených od normálního

$Z_1, \dots, Z_k \sim N(0,1)$ . . . . . **nezávislá** normovaná normální

$X_k^2 = \sum_{i=1}^k Z_i^2 \sim \chi^2(k)$ . . . . chí-kvadrát o k stupních volnosti

$F_{k,m} = \frac{X_k^2/k}{X_m^2/m} \sim F(k,m)$ . . F-rozdělení s k a m stupni volnosti

$T_k = \frac{Z}{\sqrt{X_k^2/k}} \sim t(k)$ . . . . . . t-rozdělení s k stupni volnosti

Přehled rozdělení odvozených od normálního

$Z_1, \ldots, Z_k \sim N(0,1)$ . . . . . **nezávislá** normovaná normální

$X_k^2 = \sum_{i=1}^k Z_i^2 \sim \chi^2(k)$ . . . . chí-kvadrát o k stupních volnosti

$F_{k,m} = \frac{X_k^2/k}{X_m^2/m} \sim F(k,m)$ . . F-rozdělení s k a m stupni volnosti

$T_k = \frac{Z}{\sqrt{X_k^2/k}} \sim t(k)$ . . . . . . t-rozdělení s k stupni volnosti

Zřejmě $Z^2 \sim \chi^2(1)$ a $T_k^2 \sim F(1,k)$.

Přehled rozdělení odvozených od normálního

$Z_1, \dots, Z_k \sim N(0,1)$ . . . . . **nezávislá** normovaná normální

$X_k^2 = \sum_{i=1}^k Z_i^2 \sim \chi^2(k)$ . . . . chí-kvadrát o k stupních volnosti

$F_{k,m} = \frac{X_k^2/k}{X_m^2/m} \sim F(k,m)$ . . F-rozdělení s k a m stupni volnosti

$T_k = \frac{Z}{\sqrt{X_k^2/k}} \sim t(k)$ . . . . . . t-rozdělení s k stupni volnosti

Zřejmě $Z^2 \sim \chi^2(1)$ a $T_k^2 \sim F(1,k)$.

rozdělení	střední hodnota	rozptyl
$N(\mu,\sigma^2)$	μ	σ^2
$\chi^2(k)$	k	$2k$
$t(k)$	0	$k/(k-2)$
$F(k,m)$	$m/(m-2)$	$2m^2(k+m-2)/k(m-2)^2(m-4)$

Plán přednášky

Motivace

S jedním případem limitní věty jsme se již setkali – de Moivre-Laplaceova věta říká, že binomické rozdělení $\mathrm{Bi}(n, p)$ lze za určitých podmínek aproximovat normovaným normálním rozdělením. Obvykle se k aproximaci přistupuje při splnění podmínky $np(1 - p) > 9$.

Motivace

S jedním případem limitní věty jsme se již setkali –
de Moivre-Laplaceova věta říká, že binomické rozdělení $\mathrm{Bi}(n, p)$ lze za určitých podmínek aproximovat normovaným normálním rozdělením. Obvykle se k aproximaci přistupuje při splnění podmínky $np(1-p) > 9$.
V této kapitole zformulujeme zobecnění této věty a rovněž další tvrzení umožňující odhadovat chování náhodných veličin při velkém počtu nezávislých opakování náhodného pokusu.

Čebyševova nerovnost

Věta

Pro libovolné $\epsilon > 0$ *platí*

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

Čebyševova nerovnost

Věta

Pro libovolné $\epsilon > 0$ *platí*

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

Důkaz.

Budeme odhadovat rozptyl DX ve spojitém případě (diskrétní analogicky):

$$DX = \int_{-\infty}^{\infty} (X - EX)^2 f(x)\,\mathrm{d}x \geq \int_{|x-EX|\geq\epsilon} (X - EX)^2 f(x)\,\mathrm{d}x \geq$$

$$\geq \int_{|x-EX|\geq\epsilon} \epsilon^2 f(x)\,\mathrm{d}x = \epsilon^2 P(|X - EX| \geq \epsilon).$$

□

Pomocí Čebyševovy nerovnosti můžeme odhadovat pravděpodobnost, s jakou se náhodná veličina s neznámým rozdělením odchýlí od své střední hodnoty o více než k-násobek směrodatné odchylky (zřejmě je totiž $P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$).

Pomocí Čebyševovy nerovnosti můžeme odhadovat pravděpodobnost, s jakou se náhodná veličina s neznámým rozdělením odchýlí od své střední hodnoty o více než k-násobek směrodatné odchylky (zřejmě je totiž $P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$).

Příklad

Nechť je $E(X) = \mu, D(X) = \sigma^2$.

1. Odhadněte $P(|X - \mu| \geq 3\sigma)$.

Pomocí Čebyševovy nerovnosti můžeme odhadovat pravděpodobnost, s jakou se náhodná veličina s neznámým rozdělením odchýlí od své střední hodnoty o více než k-násobek směrodatné odchylky (zřejmě je totiž $P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$).

Příklad

Nechť je $E(X) = \mu, D(X) = \sigma^2$.

1. Odhadněte $P(|X - \mu| \geq 3\sigma)$.
2. Vypočtěte $P(|X - \mu| \geq 3\sigma)$, jestliže navíc víte, že $X \sim N(0, 1)$.

Pomocí Čebyševovy nerovnosti můžeme odhadovat pravděpodobnost, s jakou se náhodná veličina s neznámým rozdělením odchýlí od své střední hodnoty o více než k-násobek směrodatné odchylky (zřejmě je totiž $P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$).

Příklad

Nechť je $E(X) = \mu, D(X) = \sigma^2$.

1. Odhadněte $P(|X - \mu| \geq 3\sigma)$.
2. Vypočtěte $P(|X - \mu| \geq 3\sigma)$, jestliže navíc víte, že $X \sim N(0, 1)$.

Pomocí Čebyševovy nerovnosti můžeme odhadovat pravděpodobnost, s jakou se náhodná veličina s neznámým rozdělením odchýlí od své střední hodnoty o více než k-násobek směrodatné odchylky (zřejmě je totiž $P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$).

Příklad

Nechť je $E(X) = \mu, D(X) = \sigma^2$.

1. Odhadněte $P(|X - \mu| \geq 3\sigma)$.
2. Vypočtěte $P(|X - \mu| \geq 3\sigma)$, jestliže navíc víte, že $X \sim N(0, 1)$.

Řešení

1. $1/9$,
2. 0,0027.

Zákon velkých čísel

Věta (Čebyševova)

Nechť jsou $X_1, X_2, \ldots$ po dvou nezávislé náhodné veličiny, které mají všechny stejnou střední hodnotu μ a stejný rozptyl σ^2. Pak pro libovolné $\epsilon > 0$ platí

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1.$$

Říkáme, že posloupnost aritmetických průměrů konverguje podle pravděpodobnosti ke střední hodnotě μ.

Zákon velkých čísel

Věta (Čebyševova)

Nechť jsou $X_1, X_2, \dots$ po dvou nezávislé náhodné veličiny, které mají všechny stejnou střední hodnotu μ a stejný rozptyl σ^2. Pak pro libovolné $\epsilon > 0$ platí

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1.$$

Říkáme, že posloupnost aritmetických průměrů konverguje podle pravděpodobnosti ke střední hodnotě μ.

Speciálním případem této věty je Bernoulliova věta, která říká, že je-li $Y_n \sim \mathrm{Bi}(n, p)$, pak posloupnost relativních četností Y_n/n konverguje podle pravděpodobnosti k p.

Věta (Bernoulliova)

Pro náhodnou veličinu s binomickým rozdělením $Y_n \sim \mathrm{Bi}(n, p)$ a pro libovolné $\epsilon > 0$ platí

$$P\left(\left|\frac{Y_n}{n} - p\right| > \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

Věta (Bernoulliova)

Pro náhodnou veličinu s binomickým rozdělením $Y_n \sim \mathrm{Bi}(n, p)$ a pro libovolné $\epsilon > 0$ platí

$$P\left(\left|\frac{Y_n}{n} - p\right| > \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

Důkaz.

Plyne snadno z Čebyševovy nerovnosti, neboť
$E(Y_n/n) = np/n = p$ a
$D(Y_n/n) = np(1-p)/n^2 = p(1-p)/n.$ □

Příklad

Při zkoušce bylo zjištěno, že mezi 600 kontrolovanými studenty je 5 studentů, kteří neumí ani malou násobilku. Odhadněte pravděpodobnost, že relativní četnost takových studentů se od jejich pravděpodobnosti výskytu liší o více než 0,01? (Můžete předpokládat, že pravděpodobnost výskytu studenta bez znalosti násobilky je menší než 0,02).

Centrální limitní věta

Centrální limitní věta dá odpověď na otázku, proč je normální rozdělení nejdůležitějším rozdělením. Ukazuje totiž, že rozdělení součtu dostatečně velkého počtu nezávislých a stejně rozdělených náhodných veličin lze aproximovat normálním rozdělením.

Věta

Nechť je $Y_1, Y_2, \ldots$ posloupnost **nezávislých stejně rozdělených** *náhodných veličin se střední hodnotou μ a rozptylem σ^2. Pak pro* **normované** *náhodné veličiny*

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - \mu}{\sigma}$$

platí

$$\lim_{n \to \infty} P(S_n < x) = \Phi(x),$$

kde Φ je distribuční funkce rozdělení $N(0,1)$.

Příklad

Mezi matematiky v ČR je jich 10% s příjmem přesahujícím celostátní průměr. Kolik matematiků je třeba pozvat na konferenci, aby s pravděpodobností aspoň 0,95 mezi nimi bylo 8 až 12 procent s nadprůměrným příjmem?

Příklad

Mezi matematiky v ČR je jich 10% s příjmem přesahujícím celostátní průměr. Kolik matematiků je třeba pozvat na konferenci, aby s pravděpodobností aspoň 0,95 mezi nimi bylo 8 až 12 procent s nadprůměrným příjmem?

Řešení

$Y_n \sim \mathrm{Bi}(n; 0,1)$, $E(Y_n) = 0,1 \cdot n$, $D(Y_n) = 0,1 \cdot 0,9 \cdot n$. Pak

$$
\begin{aligned}
0,95 \leq P(0,08n \leq Y_n \leq 0,12n) &= \\
&= P\left(\frac{0,08 - 0,1}{\sqrt{0,09n}} n \leq \frac{Y_n - 0,1n}{\sqrt{0,09n}} \leq \frac{0,12 - 0,1}{\sqrt{0,09n}} n\right) = \\
&= P\left(\frac{-\sqrt{n}}{15} \leq \frac{Y_n - 0,1n}{\sqrt{0,09n}} \leq \frac{\sqrt{n}}{15}\right) \approx \Phi\left(\frac{\sqrt{n}}{15}\right) - \Phi\left(-\frac{\sqrt{n}}{15}\right).
\end{aligned}
$$

Je tedy $\Phi\left(\frac{\sqrt{n}}{15}\right) \geq 0,975$, což je ekvivalentní $\sqrt{n}/15 \geq 1,96$, tj. $n \geq 865$.

Řešení (Pomocí Bernoulliovy nerovnosti)

Nyní využijme Bernoulliovu nerovnost – ta dává

$$P\left(\left|\frac{Y_n}{n} - 0,1\right| \leq 0,02\right) \geq 1 - \frac{0,1 \cdot 0,9}{n \cdot 0,02^2},$$

což má být alespoň 0,95. Odtud

$$n \geq \frac{0,09}{0,05 \cdot 0,02^2} = 4500.$$

Řešení (Pomocí Bernoulliovy nerovnosti)

Nyní využijme Bernoulliovu nerovnost – ta dává

$$P\left(\left|\frac{Y_n}{n} - 0,1\right| \leq 0,02\right) \geq 1 - \frac{0,1 \cdot 0,9}{n \cdot 0,02^2},$$

což má být alespoň 0,95. Odtud

$$n \geq \frac{0,09}{0,05 \cdot 0,02^2} = 4500.$$

Vidíme, že odhad prostřednictvím Bernoulliovy nerovnosti je podstatně slabší než odhad s využitím centrální limitní věty (resp. de Moivre-Laplaceovy věty).

Plán přednášky

Statistika zkoumá jevy na rozsáhlých **souborech** případů a zkoumá **statistické znaky** jednotlivých statistických **jednotek**. Obvykle nelze testovat všechny jednotky **základního souboru**, proto se omezujeme na prozkoumání některého **výběrového souboru** rozsahu n.

Statistika zkoumá jevy na rozsáhlých **souborech** případů a zkoumá **statistické znaky** jednotlivých statistických **jednotek**. Obvykle nelze testovat všechny jednotky **základního souboru**, proto se omezujeme na prozkoumání některého **výběrového souboru** rozsahu n.

Předpokládejme, že jsme na n statistických jednotkách naměřili **soubor hodnot**

$$x_1, \ldots, x_n$$

daného znaku. Znaky obvykle dělíme na *kvalitativní* (nominální, ordinální) a *kvantitativní* (intervalové, poměrové).

Počtu prvků souboru říkáme **rozsah**.

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram
- (výběrový) průměr, geometrický, harmonický průměr

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram
- (výběrový) průměr, geometrický, harmonický průměr
- medián, p-tý kvantil, percentil, kvartil

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram
- (výběrový) průměr, geometrický, harmonický průměr
- medián, p-tý kvantil, percentil, kvartil
- modus

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram
- (výběrový) průměr, geometrický, harmonický průměr
- medián, p-tý kvantil, percentil, kvartil
- modus
- rozptyl s_x^2, resp. $n/(n-1)s_x^2$

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram
- (výběrový) průměr, geometrický, harmonický průměr
- medián, p-tý kvantil, percentil, kvartil
- modus
- rozptyl s_x^2, resp. $n/(n-1)s_x^2$
- rozpětí, kvartilové rozpětí, průměrná odchylka (od mediánu)

Základní pojmy popisné statistiky

- absolutní (relativní) četnosti, četnostní tabulka
- histogram
- (výběrový) průměr, geometrický, harmonický průměr
- medián, p-tý kvantil, percentil, kvartil
- modus
- rozptyl s_x^2, resp. $n/(n-1)s_x^2$
- rozpětí, kvartilové rozpětí, průměrná odchylka (od mediánu)
- koeficient šikmosti, špičatosti

Diagramy

Krabicový diagram, box plot