

Základy matematickej štatistiky

1. Náhodný výber, výberové momenty a odhad parametrov

Aleš Kozubík

Katedra Matematických metód
Fakulta Riadenia a Informatiky
Žilinská Univerzita v Žiline

10. mája 2017

- 1 Náhodný výběr
- 2 Výberové momenty
- 3 Odhady parametrov
- 4 Intervalové odhady

Náhodný výber

V realite často potrebujeme analyzovať náhodné premenné, pričom ich rozdelenie nie je úplne známe.

Často napr. predpokladáme určitý typ rozdelenia, ale nepoznáme hodnoty parametrov.

Jediným spôsobom ako tieto informácie doplniť je vychádzať z výsledkov opakovania pokusov pri rovnakých podmienkach. Hľadané parametre sú potom funkciami týchto výsledkov.

Pre výsledky často používame termín *súbor hodnôt*. Na rozdiel od množiny hodnôt, v súbore sa môžu rovnaké hodnoty opakovať.

Náhodný výber

Teraz sformalizujeme proces získania súboru hodnôt, a to v podobe náhodného výberu.

Definícia (Náhodný výber)

Postupnosť nezávislých, rovnako rozdelených náhodných premenných X_1, X_2, \dots, X_n sa nazýva *náhodný výber*.
Číslo n nazývame *rozsah náhodného výberu*.

Niekedy sa pre pohodlnejšie vyjadrovanie hovorí o náhodnom výbere z určitého rozdelenia a náhodné premenné X_1, X_2, \dots, X_n sa potom nazývajú prvkami tohto náhodného výberu.

Štatistiky

Už sme spomenuli, že náhodný výber využívame na určenie neznámych parametrov rozdelenia. Tieto parametre sú obvykle nejakou funkciou tohto náhodného výberu. Preto zavádzame nasledujúci pojem.

Definícia (Štatistika)

Funkciu jednej alebo viac náhodných premenných ktorá nezávisí na hodnotách neznámych parametrov sa nazýva *štatistika*.

Ak X_1, X_2, \dots, X_n sú náhodné premenné, tak náhodná premenná $Y = \sum_{i=1}^n X_i$ je štatistikou v zmysle predchádzajúcej definície. Naproti tomu náhodná premenná $Y = \frac{X_1 - \mu}{\sigma}$ je štatistikou iba v prípade, že hodnoty μ a σ sú známe.

Usporiadany náhodný výber

Niekedy je nutné namerané hodnoty zoradiť podľa veľkosti. Usporiadajme teda náhodný výber X_1, \dots, X_n podľa veľkosti do novej postupnosti

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

Definícia (usporiadaný náhodný výber)

Postupnosť $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ nazývame *usporiadaný náhodný výber*.

Usporiadany náhodný výber

Veta (Rozdelenie usporiadaného náhodného výberu)

Nech $r \in \{1, 2, \dots, n\}$. Potom distribučná funkcia $G_r(x)$ náhodnej premennej $X_{(r)}$ je daná vzťahom

$$G_r(x) = \sum_{i=r}^n \binom{n}{i} F^i(x) [1 - F(x)]^{n-i},$$

kde $F(x)$ je distribučná funkcia náhodných premenných X_1, \dots, X_n .

Usporiadany náhodný výber

Dôkaz

Pravdepodobnosť, že medzi hodnotami X_1, \dots, X_n bude práve i takých, že ich hodnota je menšia než dané x je daná Bernoulliho vzorcom ako

$$\binom{n}{i} F^i(x) [1 - F(x)]^{n-i},$$

nakoľko počet takýchto hodnôt sa riadi binomickým rozdelením. Hodnota $X_{(r)}$ bude menšia než dané x vtedy, ak medzi X_1, \dots, X_n bude buďto r alebo $r + 1$ atď. alebo n takých, ktoré sú menšie než x . Tieto prípady predstavujú nezlučiteľné udalosti, preto výsledná pravdepodobnosť je súčtom ich pravdepodobností.

Usporiadany náhodný výber

Poznámka

Ako špeciálne prípad z vety vyplývajú vzorce pre distribučné funkcie najmenšieho a najväčšieho člena usporiadaného náhodného výberu. Dostávame tak

$$G_1(x) = 1 - [1 - F(x)]^n,$$

resp.

$$G_n(x) = F^n(x).$$

Všeobecný výberový moment

Definícia (Všeobecný výberový moment)

Nech X_1, \dots, X_n je náhodný výber zo základného súboru X . Pod *všeobecným výberovým momentom k -teho rádu v bode a* rozumieme hodnotu

$$M_k(a) = \frac{1}{n} \sum_{i=1}^n (x_i - a)^k.$$

Začiatkový výberový moment

Pre $a = 0$ získame definíciu začiatkových výberových momentov.

Definícia (Začiatkový výberový moment)

Nech X_1, \dots, X_n je náhodný výber zo základného súboru X . Pod *začiatkovým výberovým momentom k -teho rádu* rozumieme hodnotu

$$v_k = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Aritmetický priemer

Pre hodnotu $k = 1$ dostávame začiatkový moment prvého rádu, ktorý nazývame *aritmetický priemer* a píšeme

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Centrálny výberový moment

Pre $a = \bar{x}$ získame definíciu centrálnych výberových momentov.

Definícia (Centrálny výberový moment)

Nech X_1, \dots, X_n je náhodný výber zo základného súboru X . Pod *centrálnym výberovým momentom k -teho rádu* rozumieme hodnotu

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k.$$

Výberový rozptyl

Pre hodnotu $k = 2$ dostávame centrálny moment druhého rádu, ktorý nazývame *výberový rozptyl* a píšeme

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Charakteristiky polohy

Medzi charakteristiky polohy zaraďujeme

Aritmetický priemer

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Charakteristiky polohy

Medzi charakteristiky polohy zaraďujeme

Výberový medián

$X_{(1)}, \dots, X_{(n)}$ je usporiadaný náhodný výber. *Výberový medián* definujeme ako

$$\tilde{x} = \begin{cases} X_{(\frac{n+1}{2})} & n \text{ nepárne} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})}}{2} & n \text{ párne} \end{cases}$$

Charakteristiky polohy

Medzi charakteristiky polohy zaraďujeme

Výberový móduš

Nech X_1, \dots, X_n je náhodný výber. *Výberový móduš* definujeme ako hodnotu s najväčšou početnosťou a označujeme ho \hat{x} .

Charakteristiky polohy

Medzi charakteristiky polohy zaraďujeme

Geometrický priemer

Nech X_1, \dots, X_n je náhodný výber. Ak sú všetky hodnoty x_1, \dots, x_n kladné, tak definujeme *geometrický priemer* ako hodnotu:

$$\overline{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n}.$$

Charakteristiky polohy

Medzi charakteristiky polohy zaraďujeme

Harmonický priemer

Nech X_1, \dots, X_n je náhodný výber. Ak sú všetky hodnoty x_1, \dots, x_n kladné, tak definujeme *harmonický priemer* ako hodnotu:

$$\overline{x_H} = \frac{n}{x_1^{-1} + x_2^{-1} + \dots + x_n^{-1}},$$

Charakteristiky rozptýlenia

Medzi charakteristiky rozptýlenia zaraďujeme

Výberová disperzia

Nech X_1, \dots, X_n je náhodný výber. Výberovú disperziu (rozptyl) definujeme ako

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

a výberová smerodajná odchýlka $s = \sqrt{s^2}$.

Charakteristiky rozptýlenia

Medzi charakteristiky rozptýlenia zaradujeme

Výberová absolútna odchýlka

nech X_1, \dots, X_n je náhodný výber. *Výberovú absolútnu odchýlku* definujeme ako

$$A = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|.$$

Charakteristiky rozptýlenia

Medzi charakteristiky rozptýlenia zaradujeme

Výberový variačný koeficient

Nech X_1, \dots, X_n je náhodný výber. *Výberový variačný koeficient* definujeme ako podiel

$$V = \frac{s}{\bar{x}}.$$

Hodnotu

$$R = \max_i X_i - \min_i X_i,$$

nazývame *výberové variačné rozpätie*.

Dvozmerný náhodný výber

Definícia (Výberová kovariancia)

Nech $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ je náhodný výber z dvozmerného rozdelenia. *Výberovú kovarianciu* definujeme ako

$$s_{12} = \frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2).$$

a *výberový korelačný koeficient* ako

$$r_{12} = \frac{s_{12}}{s_1 s_2}.$$

Empirická distribučná funkcia

Definícia (Empirická distribučná funkcia)

Formálne je možné *empirickú distribučnú funkciu* definovať ako

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \chi_{(-\infty; x]}.$$

kde

$$\chi_A(x) = \begin{cases} 1 & \text{pre } x \in A \\ 0 & \text{pre } x \notin A \end{cases}$$

je charakteristická funkcia množiny A .

Vlastnosti výberových charakteristík

Veta

Nech X_1, \dots, X_n je náhodný výber z rozdelenia, ktoré má strednú hodnotu μ a konečný rozptyl σ^2 . Potom platí

a) $\mathbb{E}(\bar{x}) = \mu,$

b) $\mathbb{D}(\bar{x}) = \frac{\sigma^2}{n},$

c) $\mathbb{E}(s^2) = \frac{n}{n-1}\sigma^2,$ pre $n \geq 2.$

Vlastnosti výběrových charakteristik

Důkaz

a) Postupne dostávame

$$\mathbb{E}(\bar{x}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} n\mu = \mu.$$

Vlastnosti výberových charakteristík

Dôkaz

b) Postupne dostávame

$$\mathbb{D}(\bar{x}) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{D}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{1}{n}\sigma^2.$$

Vlastnosti výberových charakteristík

Dôkaz

c) Využijeme identitu

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

Z nej dostávame

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) &= \mathbb{E} \left(\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \right) \\ &= \sum_{i=1}^n \mathbb{D}(X_i) - n\mathbb{D}(\bar{X}) \\ &= n\sigma^2 - \frac{n\sigma^2}{n} = (n-1)\sigma^2, \end{aligned}$$

čo je ekvivalentné s tvrdením c)

Odhady parametrov

Členenie

Predpokladajme, že máme k dispozícii náhodný výber X_1, \dots, X_n z rozdelenia s hustotou $f(x, \theta)$, pričom $\theta = (\theta_1, \dots, \theta_m)$ sú neznáme parametre.

Na základe tohto výberu je potrebné určiť čo najpresnejší odhad parametra θ , o ktorom je známe len to, že patrí do nejakého parametrického priestoru \mathcal{P} .

Rozoznávame *bodové* a *intervalové* odhady.

Odhady parametrov

Bodový odhad

Definícia (Bodový odhad)

Nech náhodný výber $\mathbf{X} = (X_1, \dots, X_n)$ závisí od neznámeho parametra θ . *Odhadom* parametra rozumieme štatistiku $g(\mathbf{X})$, ktorá každej množine pozorovaných hodnôt náhodného výberu priradí hodnotu odhadu parametra, ktorú označíme $\hat{\theta}$.

Terminologická poznámka

Odhad teda predstavuje nejakú funkciu definovanú na \mathbb{R}^n , kde n je rozsah náhodného výberu, kým veličina $\hat{\theta}$ už je konkrétnou odhadnutou hodnotou parametra θ . V anglickej literatúre sa tieto dva pojmy rozlišujú ako *estimator*, čo je funkcia a *estimate*, ktorý predstavuje konkrétnu hodnotu odhadu.

Odhady parametrov

Intervalový odhad

Definícia (Intervalový odhad)

Intervalovým odhadom parametra θ rozumieme nejakú vhodnú množinu, ktorá s dostatočne veľkou pravdepodobnosťou pokrýva hodnotu parametra θ . Niekedy sa používa tiež termín *interval spol'ahlivosti* pre parameter θ a zmienená pravdepodobnosť, s ktorou hodnoty parametra θ patria do tejto množiny sa označuje ako *spol'ahlivosť* tohto odhadu.

Odhady parametrov

Vlastnosti odhadov

Jednou z častých požiadaviek, ktoré kladieme na odhady je ich nevychýlenosť v zmysle nasledujúcej definície.

Definícia (Nevychýlenosť odhadu)

Štatistiku $T = g(\mathbf{X})$ nazývame *nevychýleným* odhadom parametra θ , ak platí

$$\mathbb{E}(T) = \theta \quad \forall \theta \in \mathcal{P}. \quad (1)$$

Poznámka

Nevychýlenosť je síce dôležitou vlastnosťou odhadov, ale nie vždy je možné na tejto požiadavke trvať. Môže sa dokonca stať, že nevychýlený odhad parametra ani nemusí existovať.

Vlastnosti odhadov

Príklady

Už sme si ukázali, že pre aritmetický priemer platí $\mathbb{E}(\bar{x}) = \mu$, teda aritmetický priemer je nevychýleným odhadom strednej hodnoty základného súboru.

Pre výberový rozptyl naopak platilo $\mathbb{E}(s^2) = \frac{n}{n-1}\sigma^2$, teda výberový rozptyl nie je nevychýleným odhadom rozptylu základného súboru.

V ďalšom sa teda pokúsime skonštruovať nevychýlený odhad rozptylu.

Vlastnosti odhadov

Príklady

Už sme odvodili vzťah

$$\mathbb{E} \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) = (n-1)\sigma^2,$$

a teda

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) = \frac{n-1}{n} \sigma^2.$$

Aby sme na pravej strane rovnice dostali hodnotu σ^2 , je potrebné je vynásobiť zlomkom $\frac{n}{n-1}$. Pre nevychýlený odhad rozptylu tak dostávame

$$s_{n-1}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Vlastnosti odhadov

Poznámka k príkladu

Ak by sme porovnali strednú kvadratickú odchýlku nevychýleného odhadu rozptylu a výberového rozptylu od skutočnej hodnoty parametra σ^2 , zistili by sme, že paltí

$$\mathbb{E}((s^2 - \sigma^2)^2) < \mathbb{E}((s_{n-1}^2 - \sigma^2)^2).$$

To je dôvodom, prečo sa možno stretnúť s tým, že niektorí štatistici uprednostňujú výberový rozptyl pred nevychýleným odhadom rozptylu. Zároveň tento príklad ukazuje, že nevychýlený odhad nemusí byť najlepší, pokiaľ sa týka strednej kvadratickej odchýlky od skutočnej hodnoty.

Vlastnosti odhadov

Definícia (Konzistentný odhad)

Nech θ je jednorozmerný parameter a X_1, \dots, X_n je náhodný výber z rozdelenia, ktoré je závislé od tohto parametra. Nech pre každé n je definovaný odhad $T_n = g(X_1, \dots, X_n)$. Povieme, že odhad T_n je *konzistentný* ak pre $n \rightarrow \infty$ platí $T_n \xrightarrow{\mathbb{P}} \theta$.

Vlastnosti odhadov

Veta

Nech $\mathbb{E}(T_n^2) < \infty$ pre každé prirodzené n . Ak sú splnené predpoklady

- 1 $\mathbb{E}(T_n) \rightarrow \theta$,
- 2 $\mathbb{D}(T_n) \rightarrow 0$,

tak T_n je konzistentným odhadom parametra θ .

Dôkaz

Pre každé $\epsilon > 0$ platí

$$\mathbb{P}(|T_n - \theta| < \epsilon) \geq \mathbb{P}\left(|T_n - \mathbb{E}(T_n)| < \frac{\epsilon}{2}, |\mathbb{E}(T_n) - \theta| < \frac{\epsilon}{2}\right).$$

Podľa Čebyševovej nerovnosti je

$$\mathbb{P}\left(|T_n - \mathbb{E}(T_n)| < \frac{\epsilon}{2}\right) \geq 1 - \frac{\mathbb{D}(T_n)}{\left(\frac{\epsilon}{2}\right)^2}.$$

Predpoklad 1 zaručuje, že existuje číslo n_0 také, že pre každé $n \geq n_0$ platí $|\mathbb{E}(T_n) - \theta| < \frac{\epsilon}{2}$.

Dôkaz

Na pravej strane máme pravdepodobnosť prieniku dvoch udalostí.

Pravdepodobnosť prvej z nich konverguje k jednej, druhá z nich je pre $n \geq n_0$ istá udalosť.

Z toho vyplýva

$$\mathbb{P}(|T_n - \theta| < \epsilon) \rightarrow 1.$$

Momentová metóda

Nech X_1, \dots, X_n je náhodný výber z rozdelenia, ktoré závisí od parametra $\theta = (\theta_1, \dots, \theta_m)$. Predpokladajme, že pre každé $\theta \in \mathcal{P}$ existujú počiatkové momenty $\nu_k = \mathbb{E}(X_i^k)$, pre $k = 1, \dots, m$.

Výberové počiatkové momenty sú dané vzťahom $v_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ pre $k = 1, \dots, m$.

Momentová metóda odhadu parametra θ spočíva v tom, že za odhad berieme riešenie rovníc

$$\nu_k = v_k, \quad k = 1, \dots, m.$$

Metóda maximálnej vierohodnosti

Nech X_1, \dots, X_n je náhodný výber z rozdelenia, ktoré závisí od parametra $\theta = (\theta_1, \dots, \theta_m)$. Ak má vektor \mathbf{X} spojité rozdelenie, tak jeho hustotu označme symbolom $L(\mathbf{x}, \theta)$. Ak má \mathbf{X} diskrétné rozdelenie, tak $L(\mathbf{x}, \theta) = \mathbb{P}(\mathbf{X} = \mathbf{x})$.

Funkciu $L(\mathbf{x}, \cdot)$ premennej θ nazývame *vierohodnostná funkcia*¹.

Hovoríme, že sme určili odhad parametra θ *metódou maximálnej vierohodnosti*, ak za odhad berieme takú hodnotu $\hat{\theta}$, že pre každé $\theta \in \mathcal{P}$ platí $L(\mathbf{x}, \hat{\theta}) \geq L(\mathbf{x}, \theta)$.

¹Preto označenie $L(\mathbf{x}, \theta)$ z anglického *likelihood*

Metóda maximálnej vierohodnosti

Poznámka k výpočtom

Vierohodnostná funkcia je, v dôsledku nezávislosti prvkov náhodného výberu, súčinom hustôt príp. pravdepodobnostných funkcií. Určenie jej lokálneho maxima si preto vyžaduje pomerne náročné derivovanie rozsiahleho súčinu.

Pre uľahčenie výpočtu sa preto obvykle prechádza od hľadania extrémov samotnej vierohodnostnej funkcie ku hľadaniu extrémov logaritmu vierohodnostnej funkcie $\ln L(\mathbf{x}, \hat{\theta})$.²

Prejdeme tak od derivovania súčinu ku derivovania súčtu, čo je z výpočtového hľadiska podstatne pohodlnejšie.

²Rozmyslite si, prečo sa logaritmovaním extrémů funkcie nezmenia.

Intervalové odhady

Definícia (Interval spoľahlivosti)

Nech \mathbf{X} je náhodný výber, ktorého rozdelenie závisí od neznámeho parametra θ a nech sú dané dve štatistiky $A = g_1(\mathbf{X})$ a $B = g_2(\mathbf{X})$ také, že $A < B$. Ak a a b sú nejaké realizácie A a B a $\mathbb{P}(a < \theta < b) = \alpha$. Interval $(a; b)$ nazývame **100 α percentným intervalom spoľahlivosti** pre θ a hodnotu α nazývame **hladinou spoľahlivosti**. Ak je $A = -\infty$ alebo $B = \infty$, tak hovoríme o **jednostranných intervaloch spoľahlivosti**.

Príklad

Interval spoľahlivosti pre strednú hodnotu normálneho rozdelenia pri známom rozptyle

Nech $X_i \sim N(\mu, \sigma^2)$ je náhodný výber z normálneho rozdelenia so známym rozptylom σ^2 . Potom $\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ a preto

$$\mathbb{P}\left(\left|\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}\right| < c\right) = 2\Phi(c) - 1, \quad \forall c > 0,$$

čo je ekvivalentné s

$$\mathbb{P}\left(\bar{X}_n - c\frac{\sigma}{\sqrt{n}} < \mu < \bar{X}_n + c\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(c) - 1.$$

Príklad

Interval spoľahlivosti pre strednú hodnotu normálneho rozdelenia pri známom rozptyle

\bar{X}_n je štatistika určená náhodným výberom.

Ak použijeme konkrétne realizácie, dostaneme hodnotu aritmetického priemeru \bar{x}_n a interval spoľahlivosti v tvare

$$\left(\bar{x}_n - c \frac{\sigma}{\sqrt{n}}; \bar{x}_n + c \frac{\sigma}{\sqrt{n}} \right).$$

Toto je interval spoľahlivosti pre strednú hodnotu normálneho rozdelenia μ s hladinou spoľahlivosti $\alpha = 2\Phi(c) - 1$.

Príklad

Interval spoľahlivosti pre strednú hodnotu normálneho rozdelenia pri známom rozptyle

Ak si uvedomíme, že platí

$$c = \Phi^{-1} \left(\frac{\alpha + 1}{2} \right),$$

tak pre príslušný interval spoľahlivosti na hladine 100α percent môžeme písať

$$\left(\bar{x}_n - \Phi^{-1} \left(\frac{\alpha + 1}{2} \right) \frac{\sigma}{\sqrt{n}}; \bar{x}_n + \Phi^{-1} \left(\frac{\alpha + 1}{2} \right) \frac{\sigma}{\sqrt{n}} \right).$$

Ak konkrétne zvolíme spoľahlivosť 95%, tak pre $\alpha = 0,95$ určíme $\Phi^{-1}(0,975) \approx 1,96$ a interval spoľahlivosti dostávame v tvare

$$\left(\bar{x}_n - 1,96 \left(\frac{\sigma}{\sqrt{n}} \right); \bar{x}_n + 1,96 \left(\frac{\sigma}{\sqrt{n}} \right) \right).$$

Konštrukcia intervalu spoľahlivosti

Pre zadanú hladinu spoľahlivosti α môžeme nájsť viacero intervalov spoľahlivosti. Našou snahou je vybrať ten najkratší, ktorý vo všeobecnosti vyzerá takto:

$$\mathbb{P}(T - t_\alpha \cdot \sigma(T) < \Theta < T + t_\alpha \cdot \sigma(T)) = \alpha,$$

kde:

Θ je odhadovaný parameter,

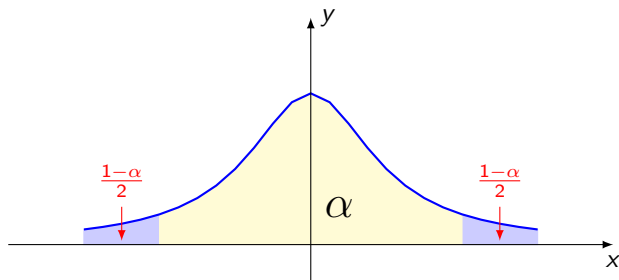
T je odhad na základe náhodného výberu,

t_α je tzv. *kritická hodnota* rozdelenia n.p. T ,

$\sigma(T)$ je smerodajná odchýlka odhadu T .

Normálne rozdelenie

Kritické hodnoty



$$\alpha = 0,95 \Rightarrow \frac{1-\alpha}{2} = 0.025 \Rightarrow t_{0,99} = \Phi^{-1}(0,975) = 1,96.$$

$$\alpha = 0,99 \Rightarrow \frac{1-\alpha}{2} = 0.005 \Rightarrow t_{0,99} = \Phi^{-1}(0,995) = 2,58.$$

Interval spoľahlivosti pre str. hodnotu normálneho rozdelenia

Rozptyl je známy

Náhodná premenná má rozdelenie $N(\mu, \sigma^2)$, kde μ je neznámy, odhadovaný parameter. Štandardizovaná náhodná premenná je potom

$$Z = \frac{X - \mu}{\sigma} \cdot \sqrt{n}$$

a $Z \sim N(0, 1)$.

Pre interval spoľahlivosti potom máme

$$\mathbb{P}\left(\mu \in \left(\bar{x} - t_\alpha \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + t_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right)\right) = \alpha,$$

čo sa zhoduje s výsledkom z riešeného príkladu.

Interval spoľahlivosti pre str. hodnotu normálneho rozdelenia

Rozptyl je neznámy

Namiesto známeho rozptylu použijeme jeho nevychýlený odhad $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ a $s = +\sqrt{s^2}$. Potom náhoná premenná

$$t = \frac{\bar{x} - \mu}{s} \cdot \sqrt{n}$$

má Studentovo rozdelenie $t \sim t(n-1)$.

Pre interval spoľahlivosti potom máme

$$\mathbb{P} \left(\mu \in \left(\bar{x} - t_\alpha \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_\alpha \cdot \frac{s}{\sqrt{n}} \right) \right) = \alpha,$$

kde t_α je kritická hodnota príslušného Studentovho rozdelenia. Pri veľkých hodnotách n ju aproximujeme kritickou hodnotou normálneho rozdelenia.

Interval spoľahlivosti pre rozptyl normálneho rozdelenia

Náhodná premenná má rozdelenie $N(\mu, \sigma^2)$, a určíme nevychýlený odhad rozptylu $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$, tak náhodná premenná

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^2$$

má rozdelenie $\chi^2 \sim \chi^2(n-1)$.

Pre interval spoľahlivosti potom máme

$$\begin{aligned} \mathbb{P}(c_1 < \chi^2 < c_2) &= \alpha \\ \mathbb{P}\left(c_1 < \frac{(n-1)s^2}{\sigma^2} < c_2\right) &= \alpha \\ \mathbb{P}\left(\sigma \in \left(\frac{(n-1)s^2}{c_2}; \frac{(n-1)s^2}{c_1}\right)\right) &= \alpha, \end{aligned}$$

kde c_1 a c_2 sú kritické hodnoty rozdelenie $\chi^2(n-1)$.

Interval spoľahlivosti pre pravdepodobnosť udalosti

Nevychýleným odhadom pravdepodobnosti p výskytu náhodnej udalosti je $\hat{p} = \frac{m}{n}$, kde m je počet výskytov sledovanej udalosti v sérii n náhodných pokusov.

Môžeme tiež interpretovať ako počet m prvkov so sledovanou vlastnosťou v náhodnom výbere o rozsahu n . Hovoríme potom o tzv. *ukazovateli štruktúry* základného súboru.

Je známe, že $\mathbb{E}(\hat{p}) = p$ a $\mathbb{D}(\hat{p}) = \frac{pq}{n}$, pričom pre veľké n platí $\frac{pq}{n} \approx \frac{\hat{p}\hat{q}}{n}$.

Potom náhodná premenná

$$Z = \frac{\frac{m}{n} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

má rozdelenie $Z \sim N(0, 1)$.

Interval spoľahlivosti pre pravdepodobnosť udalosti

Interval spoľahlivosti pre ukazovateľ štruktúry má potom tvar

$$\mathbb{P} \left(p \in \left(\hat{p} - t_\alpha \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}; \hat{p} + t_\alpha \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \right) \right) = \alpha,$$

kde t_α je kritická hodnota normálneho rozdelenia.

Potrebný rozsah výberu

Riešime obrátenú úlohu: pre danú spoľahlivosť α a danú maximálnu chybu ε určiť potrebný rozsah náhodného výberu.

Maximálna chyba ε predstavuje polovicu dĺžky intervalu spoľahlivosti, teda

$$\varepsilon = t_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad n = \frac{t_{\alpha}^2 \cdot \sigma^2}{\varepsilon^2},$$

pre odhad strednej hodnoty resp.

$$\varepsilon = t_{\alpha} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \Rightarrow \quad n = \frac{t_{\alpha}^2 \cdot \hat{p} \cdot \hat{q}}{\varepsilon^2},$$

pre odhad ukazovateľa štruktúry.