VIS III. CIKLUS

UZORAK

Neka je X slučajna varijabla s razdiobom F. Za slučajne varijable $X_1, \ldots, X_n$ kažemo da su nezavisne kopije slučajne varijable X, ako one imaju svojstva:

1. međusobno su nezavisne
2. imaju razdiobu identičnu razdiobi slučajne varijable X.

Tako dobivenu n-torku slučajnih varijabli $(X_1, \ldots, X_n)$ nazivamo uzorak. Ako je x_1 realizacija varijable X_1, x_2 realizacija X_2 itd., tada se $(x_1, \ldots, x_n)$ naziva vrijednost ili realizacija uzorka $(X_1, \ldots, X_n)$. Broj n označava veličinu (dimenziju ili volumen) uzorka.

STATISTIKA, PROCJENITELJ i PROCJENA

→ sluč. varijabla

Slučajna varijabla $\Theta := g(X_1, \ldots, X_n)$ naziva se statistika. Statistikom nazivamo svaku funkciju koja ovisi o uzorku $X_1, \ldots, X_n$, a ne ovisi (eksplicitno) o nepoznatom parametru. Neka je ϑ nepoznati parametar u populaciji X. Za statistiku kažemo da je procjenitelj parametra ϑ. Vrijednost te statistike $\hat{\vartheta} := g(x_1, \ldots, x_n)$ nazivamo procjenom parametra ϑ.

PROCJENA OČEKIVANJA

Nepoznato očekivanje a populacije X procjenjujemo pomoću sredine uzorka:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

Za slučajnu varijablu vrijedi $E(\bar{X}) = a$, $D(\bar{X}) = \frac{\sigma^2}{n}$ gdje je σ^2 varijanca (disperzija) populacije.

NEPRISTRANI PROCJENITELJI

Za statistiku Θ kažemo da je nepristrani procjenitelj ili nepristrana statistika parametra ϑ, ukoliko vrijedi

$$E(\Theta) = \vartheta$$

USPOREDBA STATISTIKA

Neka je $(X_1, \dots, X_n)$ uzorak, ϑ nepoznati parametar te $\Theta_1(X_1, \dots, X_n)$, $\Theta_2(X_1, \dots, X_n)$ dvije nepristrane statistike za ϑ. Kažemo da je Θ_1 bolja (efikasnija) od Θ_2 ako je $D(\Theta_1) < D(\Theta_2)$.

VALJANE STATISTIKE - raste broj uzoraka mora biti bolja aproksim.

Statistiku $\Theta_n = \Theta(X_1, X_2, \dots, X_n)$ nazivamo valjanom procjenom parametra ϑ ako za svaki $\varepsilon > 0$ slučajna varijabla Θ_n konvergira prema ϑ po vjerojatnosti:

$$\lim_{n \to \infty} P(|\Theta_n - \vartheta| < \varepsilon) = 1$$

TEM. Da bi nepristrana statistika bila valjana, dovoljno je da joj disperzija teži u nulu (kad n teži u beskonačnost).

DOKAZ: slijedi iz Čebiševljeve nejednakosti:

$$P(|\Theta_n - \vartheta| < \varepsilon) \geq 1 - \frac{E[(\Theta_n - \vartheta)^2]}{\varepsilon^2} = 1 - \frac{D(\Theta_n)}{\varepsilon^2} \to 1$$

PROCJENE DISPERZIJE

Ako je očekivanje a populacije X poznato, nepristrana procjena nepoznate disperzije σ^2 računa se formulom

$$D^2 := \frac{1}{n} \sum_{i=1}^{n} (X_i - a)^2 .$$

Ako su očekivanje a i disperzija σ^2 populacije X nepoznati, onda se nepristrani procjenitelj za disperziju računa formulom

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

KRITERIJ NAJVEĆE IZGLEDNOSTI

Neka je $x_1, x_2, \ldots, x_n$ realizacija uzorka populacije X, čija funkcija gustoće $f(\vartheta, x)$ ovisi o nepoznatom parametru ϑ. Funkcija izglednosti definira se kao umnožak

$$L(\vartheta, x_1, \ldots, x_n) := f(\vartheta, x_1) f(\vartheta, x_2) \cdots f(\vartheta, x_n) .$$

Za procjenu parametra ϑ uzimamo onu vrijednost $\hat{\vartheta}$ za koju funkcija izglednosti poprima globalni maksimum.

KVANTILI

Realan broj x_p za koji vrijedi $F(x_p) = p$ to jest

$$\int_{-\infty}^{x_p} f(t)\,dt = p$$ naziva se kvantil reda p.

INTERVAL POVJERENJA SLUČAJNE VARIJABLE

Neka je $0 < p < 1$. Interval $[c_1, c_2]$ za koji vrijedi $P(c_1 < X < c_2) = p$ naziva se interval povjerenja reda p za slučajnu varijablu X.

$$c_2 - c_1 \longrightarrow \min$$

$$\int_{c_1}^{c_2} f(t)\,dt = p$$, f samo jednu točku lokalnog maksimuma

$\Rightarrow$ unimodalna funkcija gustoće

NIVO ZNAČAJNOSTI

Za zadani broj p, $0 < p < 1$ koji određuje interval povjerenja, veličina $\alpha = 1 - p$ naziva se nivo značajnosti (signifikantnosti). Pri tom za jednostrane kvantile vrijedi:

$$x_p = x_{1-\alpha} \qquad x_{1-p} = x_\alpha$$

, a za dvostrane:

$$x_{\frac{1}{2}(1-p)} = x_{\frac{\alpha}{2}} \qquad x_{\frac{1}{2}(1+p)} = x_{1-\frac{\alpha}{2}}$$

INTERVAL POVJERENJA ZA NEPOZNATI PARAMETAR

Pretpostavimo da postoje funkcije $\bar{\Theta}(x_1, \dots, x_n)$, $\underline{\Theta}(x_1, \dots, x_n)$ takve da za sve realizacije $x_1, \dots, x_n$ uzorka vrijedi:

$$P\left\{ \underline{\Theta}(x_1, \dots, x_n) < \vartheta < \bar{\Theta}(x_1, \dots, x_n) \right\} = p.$$

Interval $\langle \underline{\Theta}, \bar{\Theta} \rangle$ se zove interval povjerenja za ϑ reda p.

UZ POZNATU DISPERZIJU σ^2

RAZDIOBA SREDINE UZORKA

TRM. Ako populacija X ima normalnu $N(a, \sigma^2)$, onda za sredinu uzorka vrijedi:

$$\bar{X} = \frac{X_1, \dots, X_n}{n} \sim N\left(a, \frac{\sigma}{\sqrt{n}}^2\right)$$

INTERVALI POVJERENJA ZA OČEKIVANJE NORMALNE RAZDIOBE, UZ POZNATI σ^2

1. Zadaje se nivo pouzdanosti p i odredi $\alpha = 1-p$.
2. Iz tablice kvantila normalne razdiobe, odredi se kvantil $u_{1-\frac{\alpha}{2}}$.
3. Izračuna se sredina $\bar{x}$ uzorka $x_1, \dots, x_n$.
4. Izračuna se $u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.

Interval povjerenja je $P\left(\bar{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq a \leq \bar{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = p$

UZ POZNATO OČEKIVANJE

JEDNOSTRANI INTERVAL POVJERENJA ZA DISPERZIJU

1. Zadaje se nivo pouzdanosti.
2. Iz tablice kvantila hi kvadrat razdiobe s n stupnjeva slobode, odredi se odgovarajući kvantil $\chi^2_{n,1-p}$.
3. Izračuna se procjena disperzije ∂^2 iz uzorka $x_1, \dots, x_n$.
4. Izračuna se $\frac{n\partial^2}{\chi^2_{n,1-p}}$.

Jednostrani interval povjerenja je $P\left(0 \leq \sigma^2 \leq \frac{n\partial^2}{\chi^2_{n,1-p}}\right) = p$

DVOSTRANI INTERVAL POVJERENJA ZA DISPERZIJU

1. Zadaje se nivo pouzdanosti $p = 1-\alpha$
2. Iz tablice kvantila hi kvadrat razdiobe s n stupnjeva slobode, odrede se kvantili $c_1 = \chi^2_{n,\frac{\alpha}{2}}$ i $c_2 = \chi^2_{n,1-\frac{\alpha}{2}}$.
3. Izračuna se disperzija $\hat{d}^2$ uzorka $x_1, \dots, x_n$.
4. Izračunaju se $\Lambda_1 = \frac{n\hat{d}^2}{c_2}$, $\Lambda_2 = \frac{n\hat{d}^2}{c_1}$.

Dvostrani interval povjerenja je $P(\Lambda_1 \leq \sigma^2 \leq \Lambda_2) = p$.

TRM. Neka su $X_1, \dots, X_n$ nezavisne s $N(a, \sigma^2)$ razdiobom. Tada su $\bar{X}$ i S^2 nezavisne slučajne varijable! Pri tom $(n-1)\, S^2/\sigma^2$ ima χ^2_{n-1} razdiobu.

STUDENTOVA RAZDIOBA

$X, X_1, \dots, X_n$ nezavisne jedinične normalne varijable. Tada sluč. varijabla:

$$t := \frac{X}{\sqrt{(X_1^2 + \dots + X_n^2)/n}} = \frac{X}{\sqrt{X_n^2/n}}$$

ima Studentovu razdiobu (t - razdiobu)

gustoća t_n : $f(x) = C_n \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$

INTERVALI POVJERENJA ZA ~~DISPERZIJU~~ OČEKIVANJE NORMALNE RAZDIOBE, UZ NEPOZNATI σ^2

1. Zadaje se nivo pouzdanosti $p = 1-\alpha$
2. Iz tablica kvantila Studentove razdiobe s $n-1$ stupnjeva slobode, odredi se odgovarajući kvantil $t_{1-\frac{\alpha}{2}}$.
3. Izračuna se procjena sredine $\bar{X}$ uzorka
4. Izračuna se procjena disperzije $\hat{S}^2$ uzorka.
5. Izračuna se $t_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}$.

Interval povjerenja je:

$$P\left(\bar{X} - t_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}} \le a \le \bar{X} + t_{1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}\right) = p.$$

INTERVALI POVJERENJA ZA DISPERZIJU NORMALNE RAZDIOBE, UZ NEPOZNATO OČEKIVANJE a

1. Zadaje se nivo pouzdanosti p i odredi $\alpha = 1-p$.

2a. Za jednostrani interval, iz tablica kvantila hi kvadrat razdiobe s $n-1$ stupnjeva slobode, odredi se odgovarajući kvantil $c = \xi^2_{n-1,\alpha}$.

2b. Za dvostrani interval, iz tablica kvantila hi kvadrat razdiobe s $n-1$ stupnjeva slobode, odrede se kvantili $c_1 = \chi^2_{n-1,\frac{\alpha}{2}}$, $c_2 = \chi^2_{n-1,1-\frac{\alpha}{2}}$.

3. Izračuna se nepristrana procjena disperzije $\hat{S}^2$ uzorka.

4a. Izračuna se $\mu = \frac{(n-1)\hat{s}^2}{c}$.

4b. Izračunaju se $\mu_1 = \frac{(n-1)\hat{s}^2}{c_2}$, $\mu_2 = \frac{(n-1)\hat{s}^2}{c_1}$.

Jednostrani interval povjerenja je $P(0 \le \sigma^2 \le \mu) = p$.

Dvostrani interval povjerenja je $P(\mu_1 \le \sigma^2 \le \mu_2) = p$

Centralni granični teorem
Neka je (X_n) niz identički distribuiranih nezavisnih slučajnih varijabli s očekivanjem a i disperzijom σ^2. Označimo $Z_n = X_1 + \dots + X_n$. Onda vrijedi

$$\frac{Z_n - na}{\sigma\sqrt{n}} \xrightarrow{D} N(0,1)$$

INTERVALNA PROCJENA ZA VJEROJATNOST p DOGAĐAJA
Procjena vjerojatnosti p događaja A računa se pomoću relativne frekvencije pojavljivanja tog događaja: $\hat{p} = \frac{m}{n}$.

Interval povjerenja za vjerojatnost reda $1-\alpha$ jest

$$P(p_1 \le p \le p_2) = 1-\alpha$$

pri čemu se rubovi računaju formulom

$$p_{1,2} = \hat{p} \mp u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

SNAGA TESTA

Snaga testa je funkcija $S: \Lambda \to [0,1]$ definirana ovako:

$$S(\vartheta) = P(\{\text{prihvaćena je alternativa } H_1\})$$

POGREŠKE PRVE I DRUGE VRSTE

Definirajmo $\alpha = \sup_{\vartheta \in H_0} S(\vartheta)$.

Ovaj broj označava maksimalnu vjerojatnost da je prihvaćena alternativa H_1, iako je istinita H_0. Broj α naziva se pogreška prve vrste.

Na isti način: $\beta = \sup_{\vartheta \in H_1} [1 - S(\vartheta)]$ definira pogrešku druge vrste.

Ona daje maksimalnu vjerojatnost da je prihvaćena hipoteza H_0, iako je istinita H_1.

TEST

Izabere se kritična vrijednost x_a.
Ukoliko je realizacija statistike veća od x_a prihvaća se hipoteza H_0. (možda pogr. II. vrste)
Ukoliko je realizacija statistike manja od x_a odbacuje se hipoteza H_0 (možda pogr. I. vrste

TESTIRANJE PARAMETARSKIH HIPOTEZA - JEDNOSTRANI TESTOVI

Zadaje se nivo značajnosti α. Na temelju poznate razdiobe uz pretpostavljenu istinitost hipoteze H_0 odrede se
1. kvantil $x_{1-\alpha}$ (za $\vartheta > \vartheta_0$) ili
2. kvantil x_α (za $\vartheta < \vartheta_0$).

Hipoteza H_0 se odbacuje ukoliko vrijednost $\hat{x}$ varijable izračunate iz uzorka padne van intervala povjerenja reda $1-\alpha$: $\hat{x} > x_{1-\alpha}$ u prvom slučaju, odnosno $\hat{x} < x_\alpha$, u drugom slučaju. U protivnom se hipoteza ne može odbaciti.

TESTIRANJE PARAMETARSKIH HIPOTEZA - DVOSTRANI TEST

Zadaje se nivo značajnosti α. Na temelju poznate razdiobe uz pretpostavljenu istinitost hipoteze H_0 odrede se kvantili $x_{\frac{\alpha}{2}}$ i $x_{1-\frac{\alpha}{2}}$ za dvostranu alternativu $\vartheta \neq \vartheta_0$.

Hipoteza H_0 se odbacuje, ukoliko vrijednost $\hat{x}$ varijable izračunate iz uzorka padne van intervala povjerenja reda $1-\alpha$: $\hat{x} > x_{1-\frac{\alpha}{2}}$ ili $\hat{x} < \frac{\alpha}{2}$. U protivnom se hipoteza prihvaća.

u-test - nepoznato očekivanje i poznata disperzija

Zadaje se nivo značajnosti α. Na temelju njega odredi se kvantil
1. $u_{1-\alpha}$, u slučaju jednostranih alternativa, te
2. $u_{1-\frac{\alpha}{2}}$, u slučaju dvostranih alternativa.

Izračuna se vrijednost $\hat{u}$ statistike dobivena iz uzorka. Test glasi:
1. Ako je $\hat{u} > u_{1-\alpha}$ (za desnu alternativu) ili $\hat{u} < -u_{1-\alpha}$ (za lijevu) hipoteza H_0 se odbacuje.
2. Ako je $|\hat{u}| > u_{1-\frac{\alpha}{2}}$ (za dvostranu alternativu) hipoteza H_0 se odbacuje.

T-test - nepoznato očekivanje i poznata disperzija

$T = \frac{\overline{X} - x_0}{S/\sqrt{n}}$, H_0 istinita onda je Studentova razdioba

=> test isti kao za normalnu razdiobu

Hi kvadrat test

1. Uzorak $\{x_1, x_2, \ldots, x_n\}$ podjelimo u m razreda. Neka je n_k broj realizacija u pojedinom razredu, te p_k teorijska vjerojatnost pojedinog razreda. Minimalni volumen pojedinog razreda treba biti 5, razrede za koje je $n_j < 5$ spajamo s njima susjednim.

2. Statistika χ^2-testa dana je s $\chi_q^2 = \sum_{j=1}^{m} \frac{(n_j - np_j)^2}{np_j}$.

Slučajna varijabla χ_q^2 ima približno χ^2 razdiobu s $f = m - r - 1$ stupnjeva slobode. Tu je r broj parametara razdiobe izračunatih iz uzorka.

3. U tablicama pročitamo kvantil $\chi^2_{krit} = \chi^2_{f, 1-\alpha}$, za zadani nivo značajnosti α i broj stupnjeva slobode f.

4. Ako je $\chi_q^2 < \chi^2_{krit}$, prihvaća se hipoteza da se razdioba varijable X podvrgava določenom zakonu. U protivnom se hipoteza odbacuje.

$$F_n(x) = \frac{\mu_n(x)}{n}$$

TRM. Statistika $F_n(x)$ je nepristrana i valjana procjena za vrijednost slučajne varijable $F(x)$.

DOKAZ:

Označimo za svaki k iz skupa $\{1, \dots, n\}$:

$$p = P(X_k < x) = F(x)$$

$$q = P(X_k \geq x) = 1 - F(x).$$

Tada slučajna varijabla ima binomnu razdiobu $B(n, p)$.

Stoga vrijedi $E(F_n(x)) = \frac{1}{n} E(\mu_n(x)) = \frac{1}{n} \cdot np = p = F(x)$

Nadalje prema slabom zakonu velikih brojeva, vrijedi:

$$\left| \frac{B(n,p) - np}{n} \right| \xrightarrow{P} \emptyset$$

pa dobivamo $|F_n(x) - F(x)| \xrightarrow{P} \emptyset$

Dakle $F_n(x)$ je valjana procjena za $F(x)$.

TEOREM GLIVENKO-CANTELLIJA

Neka su (X_n) nezavisne kopije slučajne varijable X s razdiobom F. tada vrijedi

$$\lim_{n \to \infty} P\left(\sup_x \left| \frac{\mu_n(x)}{n} - F(x) \right| = \emptyset \right) = 1.$$

Dakle, $\frac{\mu_n(x)}{n}$ konvergira ka $F(x)$ skoro sigurno.