Statistika i osnovna mjerenja

Procjenitelji i pogreške

M. Makek

2017/2018

Procjena parametara populacije

- Uzmimo da populacija ima karakteristične parametre poput μ, σ, σ^2, momenti, itd.
- Te parametre je nemoguće odrediti na cijeloj populaciji, stoga se uzima ***uzorak*** i na temelju njega se procjenjuju parametri populacije
- Uzmimo slučajnu varijablu Θ kao ***procjenitelj*** parametra θ
- Θ je nepristrani procjenitelj parametra θ ako vrijedi:

$$\boxed{E(\Theta) = \theta}$$

tj. mu je očekivanje stvarno jednako tom parametru

Procjena parametara populacije

- Ako je $\boldsymbol{X_1, \dots, X_n}$ slučajni uzorak populacije, onda je slučajna varijabla $\bar{X}$ (prosjek uzorka) nepristrani procjenitelj očekivanja populacije $\boldsymbol{\mu}$

  Dokaz: pokazali smo da je očekivanje $\bar{X}$ jednako očekivanju populacije

$$\boxed{E(\bar{X}) = \mu_{\bar{X}} = \mu}$$

- Ako je $\boldsymbol{X_1, \dots, X_n}$ slučajni uzorak populacije, onda je slučajna varijabla $\boldsymbol{S^2}$ nepristrani procjenitelj varijance populacije $\boldsymbol{\sigma^2}$

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

- Dokaz slijedi

Procjena parametara populacije

- Pokazali smo da za bilo koju slučajnu varijablu ***Y*** vrijedi:

$$V(Y) = E(Y^2) - [E(Y)]^2$$

- Stoga za slučajnu varijablu $\boldsymbol{X_i}$ možemo pisati:

$$E(X_i^2) = [E(X_i)]^2 + V(X_i)$$

- Raspišemo izraz za $\boldsymbol{S^2}$:

$$S^2 = \frac{1}{n-1}\left[\sum X_i^2 - \frac{1}{n}\left(\sum X_i\right)^2\right]$$

Procjena parametara populacije

$$E(S^2) = \frac{1}{n-1}\left\{\sum E\left(X_i^2\right) - \frac{1}{n}E\left[\left(\sum X_i\right)^2\right]\right\}$$
$$= \frac{1}{n-1}\left\{\sum (\mu^2 + \sigma^2)\right.$$
$$\left. - \frac{1}{n}\left\{V\left[\sum X_i\right] + \left[E\left(\sum X_i\right)\right]^2\right\}\right\}$$
$$= \frac{1}{n-1}\left\{\sum (\mu^2 + \sigma^2) - \frac{1}{n}\{V[T_0] + [E(T_0)]^2\}\right\}$$
$$= \frac{1}{n-1}\left\{\sum (\mu^2 + \sigma^2) - \frac{1}{n}\{n\sigma^2 + [n\mu]^2\}\right\} = \sigma^2$$

- Što je i trebalo pokazati

Pogreške mjerenja

- U fizici se rezultati mjerenja pišu kao srednja vrijednost nezavisnih mjerenja u intervalu od 68% pouzdanosti: $\boxed{\bar{x} \pm M}$
- Gdje je $\boxed{M = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}}$ nepouzdanost prosjeka, koju nazivamo i ***standardnom greškom*** (sad znamo od kuda dolazi ova formula)
- Pri tome je rasipanje odn. Standardna devijacija uzorka: $\boxed{m = s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}}$
- Obzirom da je rasipanje dolazi zbog nemogućnosti kontrole slučajnih pogrešaka, kažemo da je ***m*** mjera ***preciznosti aparature***
- ***m se ne mijenja značajno s povećanjem broja mjerenja, no M=m/√n se smanjuje te kažemo da je rezultat pouzdaniji ako napravimo više mjerenja***

Intervali pouzdanosti

- Procjenom parametara dobivamo njihove iznose, no ne znamo koliko su ti iznosi pouzdani!
- Npr. ako izračunamo prosjek uzorka on neće biti identičan očekivanju populacije, no zanima nas koliko je njihovo odstupanje
- Prema središnjem graničnom teoremu, srednja vrijednost uzorka ima normalnu raspodjelu $\bar{X} \sim N(\mu, \sigma^2/n)$ gdje su $\boldsymbol{\mu}$ i $\boldsymbol{\sigma}$ očekivanje i varijanca populacije

Pouzdanost procjene prosjeka populacije

- Obzirom da srednja vrijednost uzorka ima normalnu raspodjelu, vrijedi:

$$P\left(-1 \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le 1\right) = 68\%$$

$$P\left(-1{,}96 \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le 1{,}96\right) = 95\%$$

$$P\left(-2{,}575 \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le 2{,}575\right) = 99\%$$

- Dakle možemo reći da se prava vrijednost očekivanja populacije nalazu i intervalu: $\boxed{\bar{x} \pm \frac{\sigma}{\sqrt{n}}}$

  s vjerojatnošću od 68%

Pouzdanost procjene prosjeka populacije

- u prethodnoj relaciji smo pretpostavili da je varijanca populacije σ poznata.
- Ako nije poznata onda je procjenjujemo pomoću slučajne varijable:

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

- Tada interval od 68% pouzdanosti postaje:

$$\boxed{\bar{x} \pm \frac{s}{\sqrt{n}}}$$

Mjerenja različitih statističkih težina

- Mjerimo neku fizikalnu veličinu u nekoliko različitih mjerenja s različitim pouzdanostima → zanima nas procjena prave vrijednosti te veličine na temelju tih mjerenja
- Izvedeno je ***k*** nizova mjerenja te fizikalne veličine te je za svaki niz dobivena srednja vrijednost i standardna pogreška:

$$x_i = \bar{x}_i \pm M_i, \qquad i = 1 \ldots k$$

- Tražimo najvjerojatniju vrijednost $\boldsymbol{x_p}$ mjerene fizikalne veličine na temelju tih mjerenja

Opća srednja vrijednost

- Funkcija gustoće vjerojatnosti je na temelju ***k*** serija mjerenja:

$$f\left(x_p; \bar{x}_1, \dots, \bar{x}_k\right) = \frac{1}{M_1 \dots M_k \left(\sqrt{2\pi}\right)^k} e^{-\sum_{i=1}^{k} \frac{\left(\bar{x}_i - x_p\right)^2}{2M_i^2}}$$

- Da bi našli za koji $\boldsymbol{x_p}$ je $\boldsymbol{f}$ maksimalan logaritmiramo i tražimo minimum sume:

$$\sum_{i=1}^{k} \frac{\left(\bar{x}_i - x_p\right)^2}{2M_i^2} = min.$$

Opća srednja vrijednost

- Iz uvjeta minimuma slijedi:

$$\frac{d}{dx_p}\sum_{i=1}^{k}\frac{\left(\bar{x}_i - x_p\right)^2}{2M_i^2} = -\sum_{i=1}^{k}\frac{\left(\bar{x}_i - x_p\right)}{M_i^2} = 0$$

$$x_p = \frac{\sum_{i=1}^{k}\frac{\bar{x}_i}{M_i^2}}{\sum_{i=1}^{k}\frac{1}{M_i^2}}$$

Pri tome definiramo statističke težine pojedine serije kao:

$$w_i = \frac{\frac{1}{M_i^2}}{\sum_{j=1}^{k}\frac{1}{M_j^2}}$$

- Slijedi opća srednja vrijednost:

$$\Rightarrow \quad \boxed{x_p = \sum_{i=1}^{k} w_i \bar{x}_i}$$

Nepouzdanost opće srednje vrijednosti

- Da bi odredili pouzdanost sjetimo se da je procjenitelj prave vrijednosti veličine, linearna kombinacija:

$$X_p = \sum_{i=1}^{k} w_i \bar{X}_i$$

- Prema relaciji za varijancu linearne kombinacije dobivamo:

$$\boxed{M^2 = V\left(X_p\right) = \sum_{i=1}^{k} w_i^2 V(\bar{X}_i) = \sum_{i=1}^{k} w_i^2 M_i^2}$$

- Primjetimo još da se statističke težine onose kao:

$$w_1 : w_2 : \ldots : w_n = \frac{1}{M_1^2} : \frac{1}{M_2^2} : \ldots : \frac{1}{M_n^2}$$

- I da vrijedi:

$$\sum_{i=1}^{k} w_i = 1$$

Propagacija pogreške

- Želimo odrediti neku fizikalnu veličinu preko njene funkcijske ovisnosti u neposredno mjerenim veličinama
- Primjeri:
  a) Mjerimo masu i volumen predmeta i želimo odrediti njegovu gustoću
  b) Mjerimo put i vrijeme koje auto prijeđe i želimo odrediti njegovu brzinu
- Pitanje je kako pogreška posredno mjerene veličine ovisi o pogreškama neposredno mjerenih veličina?
- Pri sljedećim razmatranjima pretpostavljamo da su pogreške mjerenja slučajne (normalno raspodijeljene) i malene tj. da vrijedi $\boldsymbol{M_i \ll \mu_i}$
- Razmotrimo prvo dva specifična primjera koja ćemo poopćiti

Propagacija pogreške

1. Posredna veličina ***f*** je linearna kombinacija izravno mjerenih veličina ***X*** i ***Y***: ***f(X,Y)=aX+bY***
   -> Za slučaj linearne kombinacije *nezavisnih* varijabli znamo izračunati očekivanje i varijancu, dakle:
   $$E[f(X,Y)] = aE(X) + bE(Y) \qquad M_f = a^2 M_X^2 + b^2 M_Y^2$$
2. Posredna veličina ***f*** je nelinearna funkcija veličine ***X***: f(X)
   -> za očekivanje vrijedi (uz pretpostavku malene pogreške):
   $$E[f(X)] \simeq f(\mu_X) = f(\bar{X})$$
   -> za varijancu vrijedi:
   $$M_f^2 = V[f(X)] = E\left[\left(f(X) - f(\mu_X)\right)^2\right]$$

Propagacija pogreške

- dalje razvijemo f(X) u red: $f(X) = f(\mu_X) + (X - \mu_X)\left.\frac{df}{dX}\right|_{\mu_X} + \cdots$
- slijedi da je:

$$E\left[\left(f(X) - f(\mu_X)\right)^2\right] \simeq E\left[\left((X - \mu_X)\left.\frac{df}{dX}\right|_{\mu_X}\right)^2\right]$$

$$= \left(\left.\frac{df}{dX}\right|_{\mu_X}\right)^2 M_X^2$$

- odnosno: $f(X) - f(\mu_X) \simeq (X - \mu_X)\left.\frac{df}{dX}\right|_{\mu_X}$
- Dakle za male pogreške vrijedi: $M_f = \left|\left.\frac{df}{dX}\right|_{\mu_X}\right| M_X$

Propagacija pogreške

3. Općeniti slučaj – posredna veličina ***f*** je funkcija ***n*** izravno mjerenih neposrednih *nezavisnih* veličina $\boldsymbol{X_1, \ldots, X_n}$

   -> pokazuje se da tada vrijedi:

$$\boxed{\bar{f}(X_1, \ldots, X_n) = f(\bar{x}_1, \ldots, \bar{x}_n)}$$

$$\boxed{M_f = \sqrt{\sum_{i=1}^{n} \left(\frac{\partial f}{\partial X_i} \bigg|_{\bar{X}_1, \ldots \bar{X}_n} \right)^2 M_{X_i}^2}}$$

   To su relacije koje smo uveli na početku, a sada je objašnjeno odakle one dolaze

Propagacija pogrešaka s korelacijom

- Prethodne relacije za propagaciju pogreške vrijede kada su slučajne varijable $\boldsymbol{X}$ i $\boldsymbol{Y}$ nezavisne
- Sada promatramo općenit slučaja kada dopuštamo da mjerenja $\boldsymbol{X}$ i $\boldsymbol{Y}$ nisu međusobno nezavisna
- Uzmimo da smo izvršili $\boldsymbol{n}$ mjerenja $\boldsymbol{X}$ i $\boldsymbol{Y}$ i kao rezultat dobili $(\boldsymbol{x_i},\boldsymbol{y_i})$, gdje $i=1\ldots n$. Na temelju dobivenih vrijednosti izračunamo $\boldsymbol{n}$ vrijednosti veličine $\boldsymbol{q_i=q(x_i,y_i)}$
- Zanima nas srednja vrijednost i pogreška veličine $\boldsymbol{q}$

Propagacija pogrešaka s korelacijom

- Pretpostavljamo da su pogreške opservabli X i Y relativno male tj. da su odstupanja $x_i - \bar{x}$ i $y_i - \bar{y}$ mala tako da pri razvoju u red možemo uzeti samo linearne članove:

$$q_i \approx q(\bar{x},\bar{y}) + (x_i - \bar{x})\left.\frac{\partial q}{\partial x}\right|_{\bar{x},\bar{y}} + (y_i - \bar{y})\left.\frac{\partial q}{\partial y}\right|_{\bar{x},\bar{y}}$$

- Prema definiciji ***srednja vrijednost*** je:

$$\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i = \frac{1}{n}\sum_{i=1}^{n}\left[q(\bar{x},\bar{y}) + \underbrace{(x_i - \bar{x})\left.\frac{\partial q}{\partial x}\right|_{\bar{x},\bar{y}}}_{0} + \underbrace{(y_i - \bar{y})\left.\frac{\partial q}{\partial y}\right|_{\bar{x},\bar{y}}}_{0}\right]$$

- Slijedi: $\boxed{\bar{q} = q(\bar{x},\bar{y})}$

Propagacija pogrešaka s korelacijom

- Prema definiciji standardnu devijaciju možemo dobiti prema:

$$\sigma_q^2 = \frac{1}{n}\sum_{i=1}^{n}(q_i - \bar{q})^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - \bar{x})\frac{\partial q}{\partial x}\bigg|_{\bar{x},\bar{y}} + (y_i - \bar{y})\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}}\right]^2$$

$$= \left(\frac{\partial q}{\partial x}\right)^2 \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2 \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$\Rightarrow \boxed{\sigma_q^2 = \left(\frac{\partial q}{\partial x}\bigg|_{\bar{x},\bar{y}}\right)^2 \sigma_x^2 + \left(\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}}\right)^2 \sigma_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\bigg|_{\bar{x},\bar{y}} \sigma_{xy}}$$

- gdje je σ_{xy} kovarijanca
- Gornja relacija daje standardnu devijaciju veličine q bez obzira na to jesu li veličine X i Y povezane.
- Za nezavisne varijable X i Y kovarijanca je 0, a relacija za standardnu devijaciju se svodi na prva dva člana

Propagacija pogrešaka s korelacijom

- Ako kovarijanca ne iščezava kažemo da su pogreške opservabli ***X*** i ***Y*** korelirane, a standardna devijacija $\boldsymbol{\sigma_q}$ je očito različita nego u slučaju neovisnih, slučajnih pogrešaka.
- Može se pokazati da relacija za $\boldsymbol{\sigma_q}$ daje opću gornju granicu standardne devijacije veličine ***q***.
- Da bi to dokazali iskoristit ćemo ***Cauchy-Schwarzovu*** nejednakost: $\boxed{\sigma_{xy} \leq \sigma_x \sigma_y}$
- Uvrštavanjem te nejednakosti u izraz za standardnu devijaciju:

$$\sigma_q^2 \leq \left(\left. \frac{\partial q}{\partial x} \right|_{\bar{x},\bar{y}} \right)^2 \sigma_x^2 + \left(\left. \frac{\partial q}{\partial y} \right|_{\bar{x},\bar{y}} \right)^2 \sigma_y^2 + 2 \left. \frac{\partial q}{\partial x} \frac{\partial q}{\partial y} \right|_{\bar{x},\bar{y}} \sigma_x \sigma_y$$

Propagacija pogrešaka s korelacijom

- Dalje slijedi: $$\sigma_q^2 \leq \left[\left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y\right]^2$$

- Čime dobivamo gornju granicu za standardnu devijaciju bez obzira na korelaciju varijabli ***X*** i ***Y***

$$\boxed{\sigma_q \leq \left|\frac{\partial q}{\partial x}\right|\sigma_x + \left|\frac{\partial q}{\partial y}\right|\sigma_y}$$