Intervalne ocene parametara raspodele

- Realizovana vrednost tačkaste ocene parametra može da odstupa od stvarne vrednosti parametra, pa se određuje interval koji, sa unapred zadatom verovatnoćom, sadrži nepoznati parametar.
- Neka je $X_1, X_2, ..., X_n$ prost slučajan uzorak obima n za obeležje X i θ je nepoznati parametar u raspodeli tog obeležja. Na osnovu posmatranog uzorka definišu se statistike $f(X_1, ..., X_n)$ i $g(X_1, ..., X_n)$ tako da važe uslovi:

$$P[f \leq g]=1$$

$$P[f \leq \theta \leq g]=\beta,\ \beta \in [0, 1]$$

- Tada kažemo da je $[f, g]$ ***interval poverenja*** za nepoznati parametar θ sa nivoom poveranja β.
- Za $\beta=0{,}9$ kaže se da je to 90%-tni interval poverenja.

Intervali poverenja za matematičko očekivanje

Interval poverenja za matematičko očekivanje m obeležja X, sa normalnom raspodelom i poznatom disperzijom

- Kako je gustina normalne raspodele simetrična u odnosu na pravu $x=m$, statistike f i g biramo simetrično u odnosu na $\overline{X}_n$

$$f = \overline{X}_n - \varepsilon \qquad g = \overline{X}_n + \varepsilon$$

Treba odrediti ε iz uslova

$$P[\overline{X}_n - \varepsilon \leq m \leq \overline{X}_n + \varepsilon] = \beta$$

$$P[\left|\overline{X}_n - m\right| \leq \varepsilon] = \beta$$

Intervali poverenja, nastavak

- Ako je σ^2 poznato, statistika $\bar{X}_n : N(m, \frac{\sigma^2}{n})$

$$\frac{\bar{X}_n - m}{\sigma}\sqrt{n} : N(0,1)$$

- z_β se može odrediti iz:

$$P\left[\left|\frac{\bar{X}_n - m}{\sigma}\sqrt{n}\right| \leq z_\beta\right] = \beta \qquad z_\beta = \frac{\varepsilon}{\sigma}\sqrt{n}$$

- Koristeći tablice za dato β nalazi se z_β, pa se dobija interval poverenja za matematičko očekivanje obeležja X

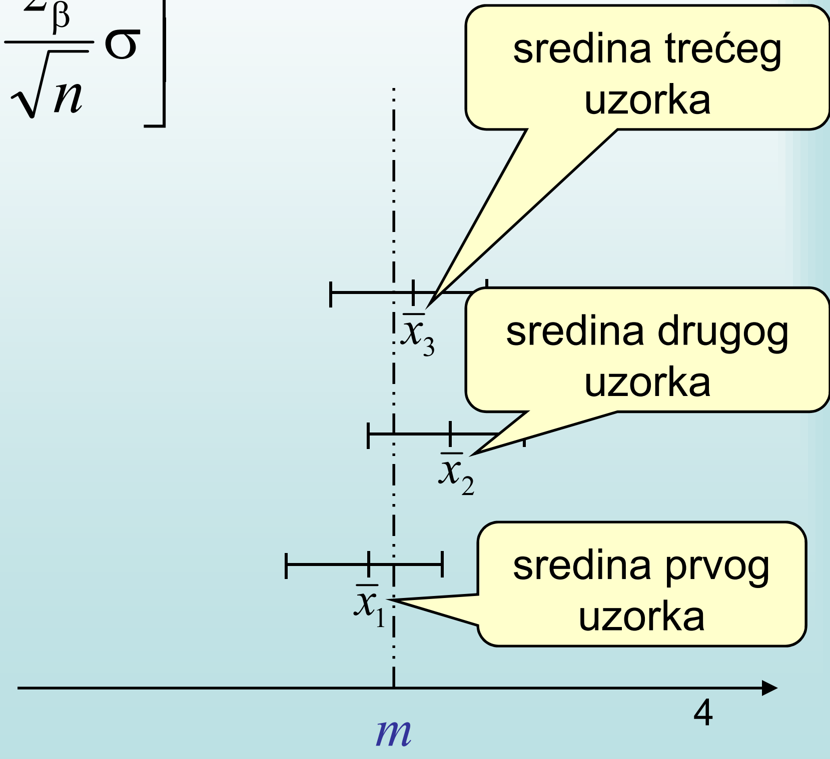
$$\left[\bar{X}_n - \frac{z_\beta}{\sqrt{n}}\sigma, \bar{X}_n + \frac{z_\beta}{\sqrt{n}}\sigma\right]$$

Intervali poverenja, nastavak

- Na osnovu vrednosti n i realizovane vrednosti $\overline{x}_n$ uzoračke sredine $\overline{X}_n$ dobija se realizovani interval poverenja

$$\left[\overline{x}_n - \frac{z_\beta}{\sqrt{n}}\sigma, \overline{x}_n + \frac{z_\beta}{\sqrt{n}}\sigma\right]$$

- Ako bismo za dato obeležje veliki broj puta uzeli uzorak obima n i izračunali intervale poverenja, tada bi približno $100\cdot\beta\%$ intervala sadržalo nepoznato matematičko očekivanje $E(X)=m$.



Primer

- Odrediti 90%-ni i 95%-ni interval poverenja za matematičko očekivanje obeležja X čija je raspodela $\mathcal{N}(m,2^2)$ na osnovu uzorka obima 20:

1,2	1,3	2,0	1,4	2,3	1,1	2,5	1,8	1,5	1,8
2,2	2,3	2,2	1,1	1,3	1,5	1,7	1,9	2,5	1,4

- Realizovana vrednost uzoračke sredine je $\bar{x}_{20} = \frac{1}{20}(1{,}2 + \ldots + 1{,}4) = 1{,}75$
- Neka je $Z = \frac{\bar{X}_n - m}{\sigma}\sqrt{n} = \frac{\bar{X}_{20} - m}{2}\sqrt{20}$
- Iz tablica za normalnu raspodelu se određuje z_β na osnovu $P\left[|Z| \le z_\beta\right] = 0{,}9$
- Dobija se z_β =1,64, pa je 90%-ni interval poverenja za mat. očekivanje

$$\left[\bar{x}_n - \frac{z_\beta}{\sqrt{n}}\sigma ; \bar{x}_n + \frac{z_\beta}{\sqrt{n}}\sigma\right] = \left[1{,}75 - \frac{1{,}64}{\sqrt{20}}2 ; 1{,}75 + \frac{1{,}64}{\sqrt{20}}2\right] = [1{,}02 ; 2{,}48]$$

- Za β=0,95, dobija se z_β =1,96, pa je 95%-ni interval poverenja za mat. očekivanje [0,87; 2,63].

Interval poverenja za MO m obeležja sa norm. rasp. i nepoznatom disperzijom

- Slično kao u slučaju sa poznatom disp., ali se koristi sp

$$\frac{\overline{X}_n - m}{\overline{S}_n}\sqrt{n-1}$$

koja ima Studentovu raspodelu sa n-1 stepeni slobode. Iz tablica za Studentovu raspodelu nalazi se $t_{n\text{-}1;\beta}$ tako da za dato β važi

$$P\left[\left|\frac{\overline{X}_n - m}{\overline{S}_n}\right|\sqrt{n-1} \le t_{n-1;\beta}\right] = \beta$$

dobija se interval poverenja za m na osnovu vrednosti iz uzorka za $n, \overline{X}_n, \overline{S}_n$

$$\left[\overline{X}_n - \frac{t_{n-1;\beta}}{\sqrt{n-1}}\overline{S}_n, \overline{X}_n + \frac{t_{n-1;\beta}}{\sqrt{n-1}}\overline{S}_n\right]$$

Intervali poverenja za disperziju

Interval poverenja za disperziju obeležja X sa normalnom raspodelom i poznatim matematičkim očekivanjem

- Kako je disperzija nenegativna veličina, može se odrediti jednostrani interval poverenja, tako što se za levi kraj IP uzima 0, a desni kraj treba odrediti iz uslova $P[0 \leq \sigma^2 \leq b] = \beta$.

$$P\left[0 \leq \sigma^2 \leq b\right] = P\left[\sigma^2 \leq b\right] = P\left[\frac{n\widetilde{S}_n^2}{\sigma^2} \geq \frac{n\widetilde{S}_n^2}{b}\right] = \beta$$

$X_1, X_2, \ldots, X_n$ je prost slučajan uzorak obima n za posmatrano obeležje. Statistika, kada je MO poznato

$$\widetilde{S}_n^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - m)^2.$$

IP za disperziju, poznato m

- Statistika $\dfrac{n\widetilde{S}_n^2}{\sigma^2}$ ima χ_n^2 raspodelu, iz tablica za χ^2 raspodelu nalazi se $\chi_{n;\beta}^2$

$$P\left[\frac{n\widetilde{S}_n^2}{\sigma^2} > \chi_{n;\beta}^2\right] = \beta$$

Na osnovu dobijene vrednosti iz tablica i vrednosti iz uzorka za n i $\widetilde{S}_n^2$ dobija se

$$b = \frac{n\widetilde{S}_n^2}{\chi_{n;\beta}^2}$$

Za $n>30$, koristi se aproksimacija hi-kvadrat raspodele, normalnom $\mathcal{N}(n,2n)$.

Dvostrani interval, poznato m

- Dvostrani interval poverenja $[b_1, b_2]$, gde je $0 < b_1 < b_2$, za nepoznatu disperziju obeležja X nalazimo iz:

$$P[b_1 \le \sigma^2 \le b_2] = \beta.$$

- Kako je:

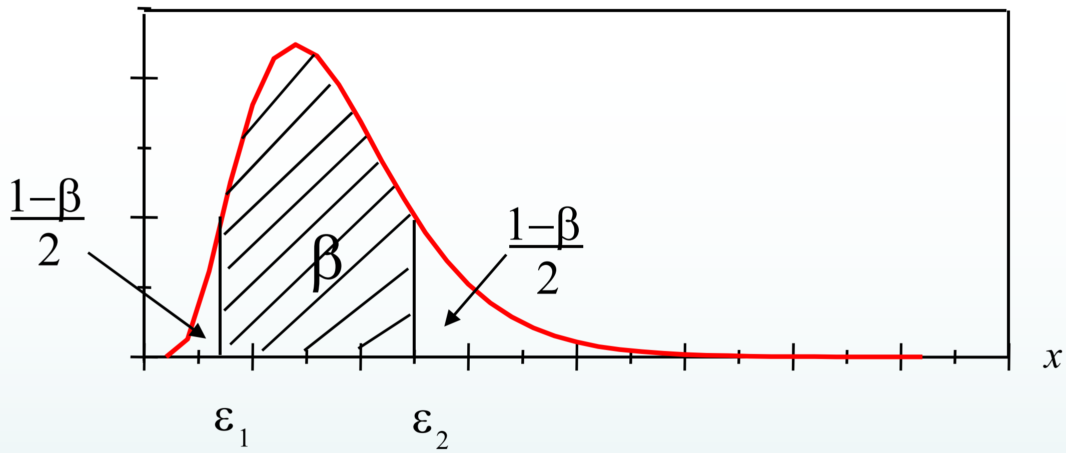
$$P\left[b_1 \le \sigma^2 \le b_2\right] = P\left[\frac{n\widetilde{S}_n^2}{b_2} \le \frac{n\widetilde{S}_n^2}{\sigma^2} \le \frac{n\widetilde{S}_n^2}{b_1}\right] = \beta$$

treba da važi

$$P\left[\varepsilon_1 \le \frac{n\widetilde{S}_n^2}{\sigma^2} \le \varepsilon_2\right] = \beta \qquad \varepsilon_1 = \frac{n\widetilde{S}_n^2}{b_2} \quad \varepsilon_2 = \frac{n\widetilde{S}_n^2}{b_1}$$

ima hi-kvadrat raspodelu sa n stepeni slobode

Hi-kvadrat raspodela



Postavljaju se uslovi, jer se ε_1 i ε_2 ne mogu jednoznačno odrediti

$$P\left[\chi_n^2 < \varepsilon_1\right] = \frac{1-\beta}{2} \qquad P\left[\chi_n^2 > \varepsilon_2\right] = \frac{1-\beta}{2}$$

Interval poverenja za nepoznatu disperziju će biti:

$$\left[\frac{n\widetilde{S}_n^2}{\varepsilon_2}, \frac{n\widetilde{S}_n^2}{\varepsilon_1}\right]$$

IP za disperziju, nepoznato m

Interval poverenja za disperziju obeležja X sa normalnom raspodelom i nepoznatim matematičkim očekivanjem

- Jednostrani interval poverenja $[0, b]$ za nepoznatu disperziju σ^2, dobijamo koristeći statistiku $\frac{n\bar{S}_n^2}{\sigma^2}$ čija je raspodela χ^2_{n-1}

  Iz tablica za χ^2 raspodelu se nalazi vrednost ε za koju je:

$$P\left[\chi^2_{n-1} > \varepsilon\right] = \beta$$

$$\bar{S}_n^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$$

$$b = \frac{n\bar{S}_n^2}{\varepsilon}$$

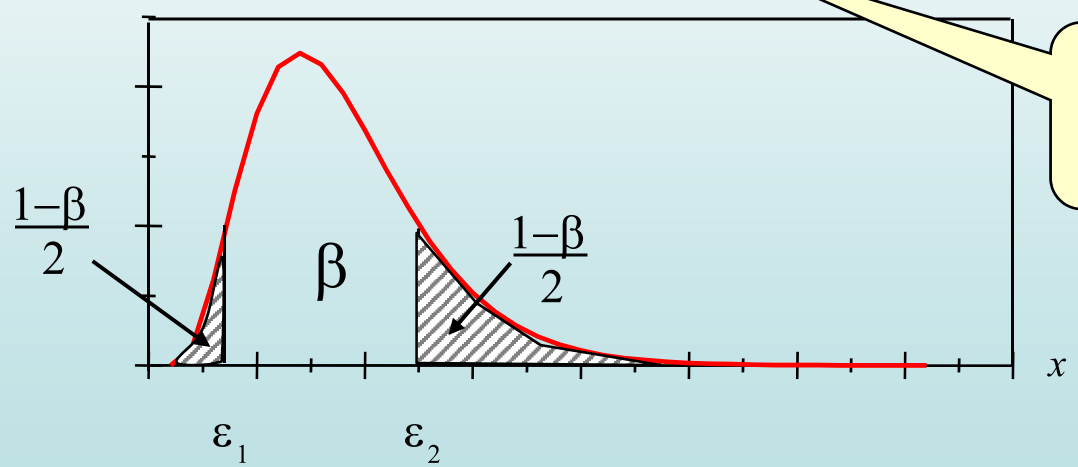
Dvostrani interval, nepoznato *m*

- Dvostrani interval poverenja $[b_1, b_2]$, gde je $0 < b_1 < b_2$ za nepoznatu disperziju σ^2 se dobija iz

$$P\left[b_1 \le \sigma^2 \le b_2\right] = \beta$$

$$P\left[b_1 \le \sigma^2 \le b_2\right] = P\left[\frac{n\overline{S}_n^2}{b_1} \le \frac{n\overline{S}_n^2}{\sigma^2} \le \frac{n\overline{S}_n^2}{b_2}\right] = \beta$$

ima hi-kvadrat raspodelu sa n-1 stepeni slobode



Interval poverenja za nepoznatu disperziju

$$\varepsilon_2 = \frac{n\overline{S}_n^2}{b_1} \qquad \varepsilon_1 = \frac{n\overline{S}_n^2}{b_2}$$

$$P\left[\chi_{n-1}^2 \geq \varepsilon_2\right] = \frac{1-\beta}{2} \qquad P\left[\chi_{n-1}^2 \leq \varepsilon_1\right] = \frac{1-\beta}{2}$$

Interval poverenja za nepoznatu disperziju će biti:

$$\left[\frac{n\overline{S}_n^2}{\varepsilon_2}, \frac{n\overline{S}_n^2}{\varepsilon_1}\right]$$

Primer

- Odrediti 90%-tni interval poverenja za disperziju obeležja X ako se smatra da je $E(X)=2$

1,2	1,3	2,0	1,4	2,3	1,1	2,5	1,8	1,5	1,8
2,2	2,3	2,2	1,1	1,3	1,5	1,7	1,9	2,5	1,4

- Realizovana vrednost statistike

$$\tilde{s}_{20}^2 = \frac{1}{20}((1{,}2-2)^2 + (1{,}3-2)^2 + \ldots + (1{,}4-2)^2) = 0{,}27$$

Iz tablica nalazimo $\chi^2_{20;0,9} = 12{,}443$

$$b = 0{,}4347$$

Interval poverenja za disperziju obeležja je [0; 0,4347].

Intervali poverenja za količnik disperzija

- Interval poverenja za količnik disperzija dva nezavisna obeležja sa normalnim raspodelama čija **su** MO **poznata**
- Neka su $X_1, \ldots, X_{n1}$ i $Y_1, \ldots, Y_{n2}$ prosti nezavisni slučajni uzorci

$$X : N(m_1, \sigma_1^2) \qquad Y : N(m_2, \sigma_2^2)$$

- Neka su m_1 i m_2 poznate vrednosti. Statistika $Z = \dfrac{\dfrac{\tilde{S}_{n_1}^2}{\sigma_1^2}}{\dfrac{\tilde{S}_{n_2}^2}{\sigma_2^2}}$

ima Fišerovu raspodelu $F_{n1,n2}$. Iz uslova $P[\varepsilon_1 \le \theta \le \varepsilon_2]=\beta$ postavljamo dva nova uslova

$$P(Z < \varepsilon_1) = \frac{1-\beta}{2} \qquad P(Z > \varepsilon_2) = \frac{1-\beta}{2}$$

Intervali poverenja za količnik disperzija, nastavak

- Iz tablica za Fišerovu raspodelu odrede se ε_1 i ε_2, iz uzorka se izračunaju uzoračke disperzije, pa se dobija interval poverenja

$$\varepsilon_1 \frac{\tilde{s}_{n_2}^2}{\tilde{s}_{n_1}^2} \leq \frac{\sigma_2^2}{\sigma_1^2} \leq \varepsilon_2 \frac{\tilde{s}_{n_2}^2}{\tilde{s}_{n_1}^2}$$

Interval poverenja za količnik disperzija

- Interval poverenja za količnik disperzija dva nezavisna obeležja sa normalnim raspodelama čija MO **nisu** poznata
- Neka su $X_1, \ldots, X_{n1}$ i $Y_1, \ldots, Y_{n2}$ prosti nezavisni slučajni uzorci

$$X : N(m_1, \sigma_1^2) \qquad Y : N(m_2, \sigma_2^2)$$

- Neka m_1 i m_2 nisu poznate vrednosti. Statistika $Z = \dfrac{\dfrac{\hat{S}_{n_1}^2}{\sigma_1^2}}{\dfrac{\hat{S}_{n_2}^2}{\sigma_2^2}}$

ima Fišerovu raspodelu $F_{n1-1,n2-1}$. Iz uslova $P[\varepsilon_1 \leq \theta \leq \varepsilon_2]=\beta$ postavljamo dva nova uslova

$$P(Z < \varepsilon_1) = \frac{1-\beta}{2} \qquad P(Z > \varepsilon_2) = \frac{1-\beta}{2}$$

- Interval poverenja je: $\varepsilon_1 \dfrac{\hat{s}_{n_2}^2}{\hat{s}_{n_1}^2} \leq \dfrac{\sigma_2^2}{\sigma_1^2} \leq \varepsilon_2 \dfrac{\hat{s}_{n_2}^2}{\hat{s}_{n_1}^2}$