

7-8. OCENJIVANJE PARAMETARA RASPODELE (SI, MAST)

Profesor Milan Merkle
emerkle@etf.rs milanmerkle.etf.rs

Matematička Statistika-jesen 2019

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele ✓*
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)✓*
- *Testiranje hipoteza o parametrima raspodele ✓*
- *Testiranje neparametarskih hipoteza ✓*
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele ✓*
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)✓*
- *Testiranje hipoteza o parametrima raspodele ✓*
- *Testiranje neparametarskih hipoteza ✓*
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele ✓*
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)✓*
- *Testiranje hipoteza o parametrima raspodele ✓*
- *Testiranje neparametarskih hipoteza ✓*
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele ✓*
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)✓*
- *Testiranje hipoteza o parametrima raspodele ✓*
- *Testiranje neparametarskih hipoteza ✓*
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele ✓*
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)✓*
- *Testiranje hipoteza o parametrima raspodele ✓*
- *Testiranje neparametarskih hipoteza ✓*
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele ✓*
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)✓*
- *Testiranje hipoteza o parametrima raspodele ✓*
- *Testiranje neparametarskih hipoteza ✓*
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Uvod u statistiku

Polazi se od podataka (uzorka) iz jedne ili više raspodela (populacija)

- *Ocenjivanje parametara raspodele* ✓
- *Ocenjivanje raspodele (funkcija raspodele, zakon raspodele...)* ✓
- *Testiranje hipoteza o parametrima raspodele* ✓
- *Testiranje neparametarskih hipoteza* ✓
- *Regresija (fitovanje)*
- *Klasifikacija ...*

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo *uzorkom obima n* iz te raspodele.

- Parametri raspodele
- Mogu biti i višedimenzionalni, na primer (μ, σ^2) .
- U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra
- Parametri se ocenjuju pomoću uzorka
- $\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.
- Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se statistika.

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo *uzorkom obima n* iz te raspodele.

- *Parametri raspodele*
- *Mogu biti i višedimenzionalni, na primer (μ, σ^2) .*
- *U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra*
- *Parametri se ocenjuju pomoću uzorka*
- *$\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.*
- *Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se *statistika*.*

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo **uzorkom obima n** iz te raspodele.

- **Parametri raspodele**
 - Mogu biti i višedimenzionalni, na primer (μ, σ^2) .
 - U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra
 - Parametri se ocenjuju pomoću uzorka
 - $\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.
 - Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se **statistika**.

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo *uzorkom obima n* iz te raspodele.

- *Parametri raspodele*
- *Mogu biti i višedimenzionalni, na primer (μ, σ^2) .*
- *U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra*
- *Parametri se ocenjuju pomoću uzorka*
- *$\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.*
- *Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se *statistika*.*

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo **uzorkom obima n** iz te raspodele.

- *Parametri raspodele*
- *Mogu biti i višedimenzionalni, na primer (μ, σ^2) .*
- *U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra*
- *Parametri se ocenjuju pomoću uzorka*
- *$\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.*
- *Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se **statistika**.*

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo *uzorkom obima n* iz te raspodele.

- Parametri raspodele
- Mogu biti i višedimenzionalni, na primer (μ, σ^2) .
- U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra
- Parametri se ocenjuju pomoću uzorka
- $\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.
- Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se *statistika*.

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo *uzorkom obima n* iz te raspodele.

- Parametri raspodele
- Mogu biti i višedimenzionalni, na primer (μ, σ^2) .
- U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra
- Parametri se ocenjuju pomoću uzorka
- $\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.
- Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se *statistika*.

Osnovni pojmovi

Definicija 6.2. Skup n nezavisnih slučajnih promenljivih sa istom raspodelom nazivamo *uzorkom obima n* iz te raspodele.

- *Parametri raspodele*
- *Mogu biti i višedimenzionalni, na primer (μ, σ^2) .*
- *U opštem slučaju: θ - parametar, Θ - skup mogućih vrednosti parametra*
- *Parametri se ocenjuju pomoću uzorka*
- *$\hat{\theta}$ - ocena parametra θ , $\hat{\theta} = S(X_1, \dots, X_n)$.*
- *Funkcija koja zavisi samo od uzorka i ne sadrži nepoznate parametre, zove se *statistika*.*

Poželjne osobine ocena

- Ocena $\hat{\theta}_n$ (iz uzorka obima n) je *stabilna* ako $\hat{\theta}_n \rightarrow \theta$ za svako $\theta \in \Theta$ u verovatnoći. ✓
- Ocena $\hat{\theta}$ je *centrirana* ili *nepriistrasna* ako je $\mathbb{E}\hat{\theta} = \theta$ za svako $\theta \in \Theta$

Primer: Ocena matematičkog očekivanja $\hat{\mu}$.

Poželjne osobine ocena

- Ocena $\hat{\theta}_n$ (iz uzorka obima n) je *stabilna* ako $\hat{\theta}_n \rightarrow \theta$ za svako $\theta \in \Theta$ u verovatnoći. ✓
- Ocena $\hat{\theta}$ je *centrirana ili nepristrasna* ako je $\mathbb{E} \hat{\theta} = \theta$ za svako $\theta \in \Theta$

Primer: Ocena matematičkog očekivanja $\hat{\mu}$.

Poželjne osobine ocena

- Ocena $\hat{\theta}_n$ (iz uzorka obima n) je *stabilna* ako $\hat{\theta}_n \rightarrow \theta$ za svako $\theta \in \Theta$ u verovatnoći. ✓
- Ocena $\hat{\theta}$ je *centrirana* ili *nepriistrasna* ako je $E\hat{\theta} = \theta$ za svako $\theta \in \Theta$

Primer: Ocena matematičkog očekivanja $\hat{\mu}$.

Poređenje ocena

Kriterijum srednjeg kvadratnog odstupanja.

Od dve centrirane ocene parametra θ bolja je ona koja ima manju varijansu.

Primer. Imamo uzorak X_1, X_2, \dots, X_n iz raspodele sa matematičkim očekivanjem μ i varijansom σ^2 . Definišimo:

$$\hat{\mu} = \frac{X_1 + X_2 + \dots + X_n}{n}, \quad \mu^* = \frac{X_1 + 2X_2 + \dots + nX_n}{n(n+1)/2}.$$

Obe ocene su nepristrasne i stabilne, ali $\text{Var}(\hat{\mu}) < \text{Var}(\mu^)$, za svako n .*

Poređenje ocena

Kriterijum srednjeg kvadratnog odstupanja.

Od dve centrirane ocene parametra θ bolja je ona koja ima manju varijansu.

Primer. *Imamo uzorak X_1, X_2, \dots, X_n iz raspodele sa matematičkim očekivanjem μ i varijansom σ^2 . Definišimo:*

$$\hat{\mu} = \frac{X_1 + X_2 + \dots + X_n}{n}, \quad \mu^* = \frac{X_1 + 2X_2 + \dots + nX_n}{n(n+1)/2}.$$

Obe ocene su nepristrasne i stabilne, ali $\text{Var}(\hat{\mu}) < \text{Var}(\mu^)$, za svako n .*

Preciznost ocene

Za parametar θ i ocenu $\hat{\theta}$ definišemo maksimalnu grešku $\varepsilon > 0$ i verovatnoću $\alpha \in (0, 1)$ da greška bude veća od ε .

$$\exists n_0 \forall n \geq n_0 \quad |\hat{\theta} - \theta| \leq \varepsilon \quad \text{sa verovatnoćom } 1 - \alpha \text{ ili većom}$$

Potreban obim uzorka (n_0) može se naći:

- 1 Ako je $\hat{\theta}$ nepristrasna ocena i ako znamo njenu varijansu, preko nejednakosti Čebiševa.
- 2 Ako raspodela za $\hat{\theta} - \theta$ ne zavisi od θ i poznata je.
- 3 Ako znamo aproksimativnu raspodelu za $\hat{\theta} - \theta$

Šta ako ne možemo da biramo obim uzorka n ? (Primeri...)

Preciznost ocene

Za parametar θ i ocenu $\hat{\theta}$ definišemo maksimalnu grešku $\varepsilon > 0$ i verovatnoću $\alpha \in (0, 1)$ da greška bude veća od ε .

$\exists n_0 \forall n \geq n_0 \quad |\hat{\theta} - \theta| \leq \varepsilon$ sa verovatnoćom $1 - \alpha$ ili većom

Potreban obim uzorka (n_0) može se naći:

- 1 Ako je $\hat{\theta}$ nepristrasna ocena i ako znamo njenu varijansu, preko nejednakosti Čebiševa.
- 2 Ako raspodela za $\hat{\theta} - \theta$ ne zavisi od θ i poznata je.
- 3 Ako znamo aproksimativnu raspodelu za $\hat{\theta} - \theta$

Šta ako ne možemo da biramo obim uzorka n ? (Primeri...)

Preciznost ocene

Za parametar θ i ocenu $\hat{\theta}$ definišemo maksimalnu grešku $\varepsilon > 0$ i verovatnoću $\alpha \in (0, 1)$ da greška bude veća od ε .

$\exists n_0 \forall n \geq n_0 \quad |\hat{\theta} - \theta| \leq \varepsilon$ sa verovatnoćom $1 - \alpha$ ili većom

Potreban obim uzorka (n_0) može se naći:

- 1 Ako je $\hat{\theta}$ nepristrasna ocena i ako znamo njenu varijansu, preko nejednakosti Čebiševa.
- 2 Ako raspodela za $\hat{\theta} - \theta$ ne zavisi od θ i poznata je.
- 3 Ako znamo aproksimativnu raspodelu za $\hat{\theta} - \theta$

Šta ako ne možemo da biramo obim uzorka n ? (Primeri...)

Preciznost ocene

Za parametar θ i ocenu $\hat{\theta}$ definišemo maksimalnu grešku $\varepsilon > 0$ i verovatnoću $\alpha \in (0, 1)$ da greška bude veća od ε .

$\exists n_0 \forall n \geq n_0 \quad |\hat{\theta} - \theta| \leq \varepsilon$ sa verovatnoćom $1 - \alpha$ ili većom

Potreban obim uzorka (n_0) može se naći:

- 1 Ako je $\hat{\theta}$ nepristrasna ocena i ako znamo njenu varijansu, preko nejednakosti Čebiševa.
- 2 Ako raspodela za $\hat{\theta} - \theta$ ne zavisi od θ i poznata je.
- 3 Ako znamo aproksimativnu raspodelu za $\hat{\theta} - \theta$

Šta ako ne možemo da biramo obim uzorka n ? (Primeri...)

Preciznost ocene

Za parametar θ i ocenu $\hat{\theta}$ definišemo maksimalnu grešku $\varepsilon > 0$ i verovatnoću $\alpha \in (0, 1)$ da greška bude veća od ε .

$\exists n_0 \forall n \geq n_0 \quad |\hat{\theta} - \theta| \leq \varepsilon$ sa verovatnoćom $1 - \alpha$ ili većom

Potreban obim uzorka (n_0) može se naći:

- 1 Ako je $\hat{\theta}$ nepristrasna ocena i ako znamo njenu varijansu, preko nejednakosti Čebiševa.
- 2 Ako raspodela za $\hat{\theta} - \theta$ ne zavisi od θ i poznata je.
- 3 Ako znamo aproksimativnu raspodelu za $\hat{\theta} - \theta$

Šta ako ne možemo da biramo obim uzorka n ? (Primeri...)

Intervali poverenja

Definicija 8.5 Sa uzorkom iz raspodele sa nepoznatim parametrom θ , interval poverenja za θ sa nivoom poverenja $1 - \alpha$ je interval koji sa verovatnoćom $1 - \alpha$ sadrži θ .

Standardne vrednosti su $\alpha = 0.1, 0.05, 0.01$, odnosno 90%, 95%, 99% intervali poverenja.

Primer 138 Dat je uzorak X_1, \dots, X_n , $n = 100$ iz $\mathcal{N}(\mu, \sigma^2)$, $\sigma^2 = 1/4$, μ je nepoznato. Naći 90% i 95% interval poverenja za μ .

Polazimo od

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

R: $[\hat{\mu} - 0.08, \hat{\mu} + 0.08]$, $[\hat{\mu} - 0.1, \hat{\mu} + 0.1]$

Za $\hat{\mu} = 3$, dobija se interval $[2.92, 3.08]$. Da li je μ unutar *ovog intervala* sa verovatnoćom 0.9?

Intervali poverenja

Definicija 8.5 Sa uzorkom iz raspodele sa nepoznatim parametrom θ , interval poverenja za θ sa nivoom poverenja $1 - \alpha$ je interval koji sa verovatnoćom $1 - \alpha$ sadrži θ .

Standardne vrednosti su $\alpha = 0.1, 0.05, 0.01$, odnosno 90%, 95%, 99% intervali poverenja.

Primer 138 Dat je uzorak X_1, \dots, X_n , $n = 100$ iz $\mathcal{N}(\mu, \sigma^2)$, $\sigma^2 = 1/4$, μ je nepoznato. Naći 90% i 95% interval poverenja za μ .

Polazimo od

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

R: $[\hat{\mu} - 0.08, \hat{\mu} + 0.08]$, $[\hat{\mu} - 0.1, \hat{\mu} + 0.1]$

Za $\hat{\mu} = 3$, dobija se interval $[2.92, 3.08]$. Da li je μ unutar *ovog intervala* sa verovatnoćom 0.9?

Intervali poverenja

Definicija 8.5 Sa uzorkom iz raspodele sa nepoznatim parametrom θ , interval poverenja za θ sa nivoom poverenja $1 - \alpha$ je interval koji sa verovatnoćom $1 - \alpha$ sadrži θ .

Standardne vrednosti su $\alpha = 0.1, 0.05, 0.01$, odnosno 90%, 95%, 99% intervali poverenja.

Primer 138 Dat je uzorak X_1, \dots, X_n , $n = 100$ iz $\mathcal{N}(\mu, \sigma^2)$, $\sigma^2 = 1/4$, μ je nepoznato. Naći 90% i 95% interval poverenja za μ .

Polazimo od

$$Z = \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

R: $[\hat{\mu} - 0.08, \hat{\mu} + 0.08]$, $[\hat{\mu} - 0.1, \hat{\mu} + 0.1]$

Za $\hat{\mu} = 3$, dobija se interval $[2.92, 3.08]$. Da li je μ unutar **ovog intervala** sa verovatnoćom 0.9?

Interval poverenja za μ

Za uzorak obima n iz $\mathcal{N}(\mu, \sigma^2)$ sa poznatim parametrom σ ,

$$P\left(\mu \in \left[\hat{\mu} - \varepsilon_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \hat{\mu} + \varepsilon_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \alpha.$$

Dužina intervala poverenja: $2\varepsilon_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \rightarrow 0$ kad $n \rightarrow +\infty$ ✓

Interval poverenja je centriran u $\hat{\mu}$ jer takav interval ima najmanju dužinu za fiksirano n i nivo $1 - \alpha$. ✓

Interval poverenja za μ

Za uzorak obima n iz $\mathcal{N}(\mu, \sigma^2)$ sa poznatim parametrom σ ,

$$P\left(\mu \in \left[\hat{\mu} - \varepsilon_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \hat{\mu} + \varepsilon_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \alpha.$$

Dužina intervala poverenja: $2\varepsilon_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \rightarrow 0$ kad $n \rightarrow +\infty$ ✓

Interval poverenja je centriran u $\hat{\mu}$ jer takav interval ima najmanju dužinu za fiksirano n i nivo $1 - \alpha$. ✓

Tri oblika intervala poverenja

Interval oblika $[A, B]$ zove se dvostrani interval poverenja (A i B su slučajne promenljive).

Jednostrani intervali su oblika $[A, +\infty)$ ili $(-\infty, B)$ (umesto $\pm\infty$ mogu da budu i minimalna, odnosno maksimalna moguća vrednost parametra ako ima smisla).

Za jednostrane intervale parametra μ polazimo od činjenice da slučajne promenljive

$$\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \text{ ili } \frac{\mu - \hat{\mu}}{\sigma/\sqrt{n}}$$

imaju $\mathcal{N}(0, 1)$ raspodelu, i nalazimo kvantil $\varepsilon_{1-\alpha}$.

Ocene varijanse

Za poznato μ (redak slučaj) treba naći ocenu matematičkog očekivanja slučajne promenljive $Y = (X - \mu)^2$:

$$s_0^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2$$

Centrirana i stabilna ocena varijanse (u slučaju da je μ nepoznato)

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \hat{\mu})^2$$

Studentova t raspodela

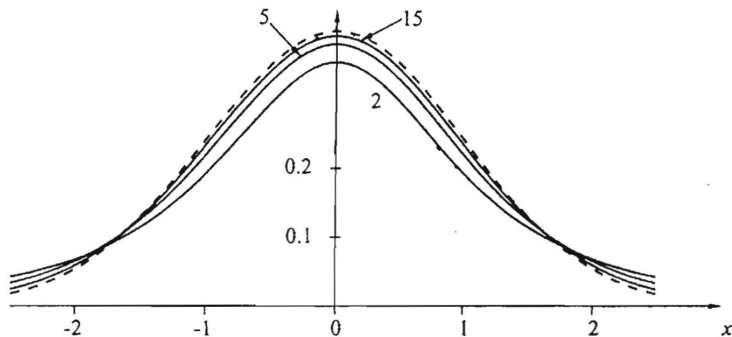
Neka su X_1, \dots, X_n nezavisne slučajne promenljive sa $\mathcal{N}(\mu, \sigma^2)$ raspodelom. U slučaju kad su oba parametra nepoznata, za nalaženje intervala poverenja koristi se statistika

$$t_n = \frac{\hat{\mu} - \mu}{s/\sqrt{n}}, \quad s = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (X_k - \hat{\mu})^2}$$

čija raspodela je poznata pod nazivom t raspodela, ili Studentova raspodela (William Gosset)

Studentova raspodela

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}, n > 0$$



Slika 34. Gustine $t(n)$ raspodele za $n = 2, 5, 15$ u poređenju sa normalnom $\mathcal{N}(0, 1)$ raspodelom (isprekidana linija).

Intervali poverenja za μ sa nepoznatim σ^2

Primenjuje se isti postupak kao i u slučaju poznate varijanse s tim da se umesto σ^2 uzima s^2 , a kvantili se nalaze iz tablica Studentove raspodele $t(n - 1)$

Za $n - 1 > 30$ uzimaju se kvantili iz standardne normalne raspodele

Intervali poverenja za μ sa nepoznatim σ^2

Primenjuje se isti postupak kao i u slučaju poznate varijanse s tim da se umesto σ^2 uzima s^2 , a kvantili se nalaze iz tablica Studentove raspodele $t(n - 1)$

Za $n - 1 > 30$ uzimaju se kvantili iz standardne normalne raspodele

Hi kvadrat raspodela

Raspodela zbira kvadrata n nezavisnih slučajnih promenljivih sa $\mathcal{N}(0, 1)$ raspedelom zove se **Hi kvadrat raspodela sa n stepeni slobode**

$$Z_1^2 + \dots + Z_n^2 \sim \chi^2(n)$$

Teorema 7.5 Za nezavisne $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$:

1 $\frac{n \cdot s_0^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \mu)^2 \sim \chi^2(n),$

2 $\frac{(n-1) \cdot s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \hat{\mu})^2 \sim \chi^2(n-1).$

Hi kvadrat raspodela

Raspodela zbira kvadrata n nezavisnih slučajnih promenljivih sa $\mathcal{N}(0, 1)$ raspedlom zove se **Hi kvadrat raspodela sa n stepeni slobode**

$$Z_1^2 + \dots + Z_n^2 \sim \chi^2(n)$$

Teorema 7.5 Za nezavisne $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$:

1
$$\frac{n \cdot s_0^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \mu)^2 \sim \chi^2(n),$$

2
$$\frac{(n-1) \cdot s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \hat{\mu})^2 \sim \chi^2(n-1).$$

Hi kvadrat raspodela

Raspodela zbira kvadrata n nezavisnih slučajnih promenljivih sa $\mathcal{N}(0, 1)$ raspedelom zove se **Hi kvadrat raspodela sa n stepeni slobode**

$$Z_1^2 + \cdots + Z_n^2 \sim \chi^2(n)$$

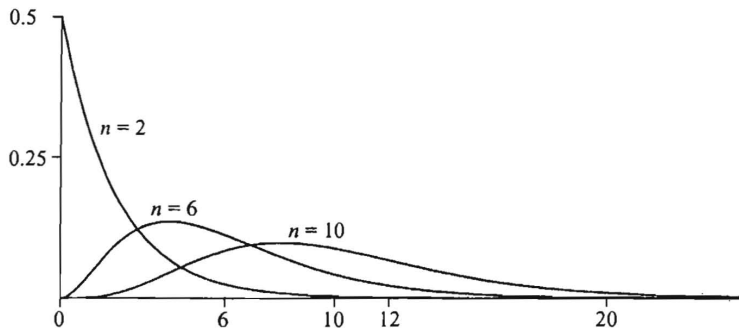
Teorema 7.5 Za nezavisne $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$:

$$\textcircled{1} \quad \frac{n \cdot s_0^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \mu)^2 \sim \chi^2(n),$$

$$\textcircled{2} \quad \frac{(n-1) \cdot s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n (X_k - \hat{\mu})^2 \sim \chi^2(n-1).$$

Hi kvadrat raspodela

$$f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad (x > 0)$$



Slika 33. Gustine hi kvadrat raspodele

Intervali poverenja za varijansu

X_1, \dots, X_n - uzorak iz $\mathcal{N}(\mu, \sigma^2)$ sa nepoznatim μ i σ^2 .

Dvostrani interval poverenja za σ^2 (Teorema 7.5.-2)

$$P\left(\sigma^2 \in \left[\frac{(n-1)s^2}{\varepsilon_{1-\alpha/2}}, \frac{(n-1)s^2}{\varepsilon_{\alpha/2}}\right]\right) = 1 - \alpha,$$

gde je ε kvantil iz $\chi^2(n-1)$.

Interval je simetričan u smislu da "reпови" imaju istu verovatnoću, $\varepsilon/2$.

Ako je μ poznato, onda je, prema teoremi 7.5.-1:

$$P\left(\sigma^2 \in \left[\frac{ns_0^2}{\varepsilon_{1-\alpha/2}}, \frac{ns_0^2}{\varepsilon_{\alpha/2}}\right]\right) = 1 - \alpha,$$

gde je ε kvantil iz $\chi^2(n)$.

Intervali poverenja za varijansu

X_1, \dots, X_n - uzorak iz $\mathcal{N}(\mu, \sigma^2)$ sa nepoznatim μ i σ^2 .

Dvostrani interval poverenja za σ^2 (Teorema 7.5.-2)

$$P\left(\sigma^2 \in \left[\frac{(n-1)s^2}{\varepsilon_{1-\alpha/2}}, \frac{(n-1)s^2}{\varepsilon_{\alpha/2}}\right]\right) = 1 - \alpha,$$

gde je ε kvantil iz $\chi^2(n-1)$.

Interval je simetričan u smislu da "reпови" imaju istu verovatnoću, $\varepsilon/2$.

Ako je μ poznato, onda je, prema teoremi 7.5.-1:

$$P\left(\sigma^2 \in \left[\frac{ns_0^2}{\varepsilon_{1-\alpha/2}}, \frac{ns_0^2}{\varepsilon_{\alpha/2}}\right]\right) = 1 - \alpha,$$

gde je ε kvantil iz $\chi^2(n)$.

Šta ako raspodela nije normalna?

Za uzorke obima većeg od 30 - aproksimativni intervali poverenja za μ i σ^2 koristeći iste formule kao za normalnu raspodelu (na osnovu CGT)

Još dva parametra:

- Verovatnoća događaja: Aproksimativni metod preko CGT ($n > 30$)- 8.3.1 ili egzaktni metod 8.3.2-neobavezno*
- Parametar λ Puasonove raspodele: Aproksimativni metod preko CGT ($n > 30, n\lambda > 10$) 8.4.2 ili egzaktni metod - neobavezno*

Šta ako raspodela nije normalna?

Za uzorke obima većeg od 30 - aproksimativni intervali poverenja za μ i σ^2 koristeći iste formule kao za normalnu raspodelu (na osnovu CGT)

Još dva parametra:

- 1 Verovatnoća događaja: Aproksimativni metod preko CGT ($n > 30$)-
8.3.1 ili egzaktni metod 8.3.2-neobavezno*
- 2 Parametar λ Puasonove raspodele: Aproksimativni metod preko CGT
($n > 30, n\lambda > 10$) 8.4.2 ili egzaktni metod - neobavezno*

Šta ako raspodela nije normalna?

Za uzorke obima većeg od 30 - aproksimativni intervali poverenja za μ i σ^2 koristeći iste formule kao za normalnu raspodelu (na osnovu CGT)

Još dva parametra:

- 1** *Verovatnoća događaja: Aproksimativni metod preko CGT ($n > 30$)-
8.3.1 ili egzaktni metod **8.3.2**-neobavezno*
- 2** *Parametar λ Puasonove raspodele: Aproksimativni metod preko CGT
($n > 30, n\lambda > 10$) **8.4.2** ili egzaktni metod - neobavezno*

Šta ako raspodela nije normalna?

Za uzorke obima većeg od 30 - aproksimativni intervali poverenja za μ i σ^2 koristeći iste formule kao za normalnu raspodelu (na osnovu CGT)

Još dva parametra:

- 1 Verovatnoća događaja: Aproksimativni metod preko CGT ($n > 30$)-
8.3.1 ili egzaktni metod **8.3.2**-neobavezno*
- 2 Parametar λ Puasonove raspodele: Aproksimativni metod preko CGT
($n > 30, n\lambda > 10$) **8.4.2** ili egzaktni metod - neobavezno*

Kako se dobijaju ocene parametra?

Metod momenata:

- Uzorak obima n : X_1, \dots, X_n .
- Moment reda k : $\mu_k = \mathbb{E} X^k$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i^k$
- Ako se parametar θ može izraziti preko momenata kao $\theta = g(\mu_1, \dots, \mu_k)$, $k \ll n$, ocena parametra θ po metodi momenata je $\hat{\theta} = g(\hat{\mu}_1, \dots, \hat{\mu}_k)$

Primer: $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E} X)^2 = \mu_2 - \mu_1^2 \implies \hat{\sigma}^2 = \hat{\mu}_2 - (\hat{\mu}_1)^2 = s_0^2$

Kako se dobijaju ocene parametra?

Metod momenata:

- Uzorak obima n : X_1, \dots, X_n .
- Moment reda k : $\mu_k = \mathbb{E} X^k$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i^k$
- Ako se parametar θ može izraziti preko momenata kao $\theta = g(\mu_1, \dots, \mu_k)$, $k \ll n$, ocena parametra θ po metodi momenata je $\hat{\theta} = g(\hat{\mu}_1, \dots, \hat{\mu}_k)$

Primer: $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E} X)^2 = \mu_2 - \mu_1^2 \implies \hat{\sigma}^2 = \hat{\mu}_2 - (\hat{\mu}_1)^2 = s_0^2$

Kako se dobijaju ocene parametra?

Metod momenata:

- Uzorak obima n : X_1, \dots, X_n .
- Moment reda k : $\mu_k = \mathbb{E} X^k$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i^k$
- Ako se parametar θ može izraziti preko momenata kao $\theta = g(\mu_1, \dots, \mu_k)$, $k \ll n$, ocena parametra θ po metodi momenata je $\hat{\theta} = g(\hat{\mu}_1, \dots, \hat{\mu}_k)$

Primer: $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E} X)^2 = \mu_2 - \mu_1^2 \implies \hat{\sigma}^2 = \hat{\mu}_2 - (\hat{\mu}_1)^2 = s_0^2$

Kako se dobijaju ocene parametra?

Metod momenata:

- Uzorak obima n : X_1, \dots, X_n .
- Moment reda k : $\mu_k = \mathbb{E} X^k$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i^k$
- Ako se parametar θ može izraziti preko momenata kao $\theta = g(\mu_1, \dots, \mu_k)$, $k \ll n$, ocena parametra θ po metodi momenata je $\hat{\theta} = g(\hat{\mu}_1, \dots, \hat{\mu}_k)$

Primer: $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E} X)^2 = \mu_2 - \mu_1^2 \implies \hat{\sigma}^2 = \hat{\mu}_2 - (\hat{\mu}_1)^2 = s_0^2$

Kako se dobijaju ocene parametra?

Metod momenata:

- Uzorak obima n : X_1, \dots, X_n .
- Moment reda k : $\mu_k = \mathbb{E} X^k$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i^k$
- Ako se parametar θ može izraziti preko momenata kao $\theta = g(\mu_1, \dots, \mu_k)$, $k \ll n$, ocena parametra θ po metodi momenata je $\hat{\theta} = g(\hat{\mu}_1, \dots, \hat{\mu}_k)$

Primer: $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E} X)^2 = \mu_2 - \mu_1^2 \implies \hat{\sigma}^2 = \hat{\mu}_2 - (\hat{\mu}_1)^2 = s_0^2$

Metod maksimalne verodostojnosti

- *Raspodela P_θ , nepoznati parametar θ - skalar ili višedimenzionalni.*
- *Uzorak obima n : $X_1 = x_1, \dots, X_n = x_n$.*
- *Funkcija verodostojnosti:- diskretna raspodela*

$$L(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = P_\theta(x_1) \cdots P_\theta(x_n)$$

- *Funkcija verodostojnosti:- neprekidna raspodela*

$$L(\theta) = f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$$

- *Ocena po metodi maksimalne verodostojnosti: $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.*

Metod maksimalne verodostojnosti

- *Raspodela P_θ , nepoznati parametar θ - skalar ili višedimenzionalni.*
- *Uzorak obima n : $X_1 = x_1, \dots, X_n = x_n$.*
- *Funkcija verodostojnosti:- diskretna raspodela*

$$L(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = P_\theta(x_1) \cdots P_\theta(x_n)$$

- *Funkcija verodostojnosti:- neprekidna raspodela*

$$L(\theta) = f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$$

- *Ocena po metodi maksimalne verodostojnosti: $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.*

Metod maksimalne verodostojnosti

- *Raspodela P_θ , nepoznati parametar θ - skalar ili višedimenzionalni.*
- *Uzorak obima n : $X_1 = x_1, \dots, X_n = x_n$.*
- *Funkcija verodostojnosti:- diskretna raspodela*

$$L(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = P_\theta(x_1) \cdots P_\theta(x_n)$$

- *Funkcija verodostojnosti:- neprekidna raspodela*

$$L(\theta) = f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$$

- *Ocena po metodi maksimalne verodostojnosti: $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.*

Metod maksimalne verodostojnosti

- *Raspodela P_θ , nepoznati parametar θ - skalar ili višedimenzionalni.*
- *Uzorak obima n : $X_1 = x_1, \dots, X_n = x_n$.*
- *Funkcija verodostojnosti:- diskretna raspodela*

$$L(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = P_\theta(x_1) \cdots P_\theta(x_n)$$

- *Funkcija verodostojnosti:- neprekidna raspodela*

$$L(\theta) = f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$$

- *Ocena po metodi maksimalne verodostojnosti: $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.*

Metod maksimalne verodostojnosti

- *Raspodela P_θ , nepoznati parametar θ - skalar ili višedimenzionalni.*
- *Uzorak obima n : $X_1 = x_1, \dots, X_n = x_n$.*
- *Funkcija verodostojnosti:- diskretna raspodela*

$$L(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = P_\theta(x_1) \cdots P_\theta(x_n)$$

- *Funkcija verodostojnosti:- neprekidna raspodela*

$$L(\theta) = f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$$

- *Ocena po metodi maksimalne verodostojnosti: $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.*

Metod maksimalne verodostojnosti

- Raspodela P_θ , nepoznati parametar θ - skalar ili višedimenzionalni.
- Uzorak obima n : $X_1 = x_1, \dots, X_n = x_n$.
- *Funkcija verodostojnosti*: - diskretna raspodela

$$L(\theta) = P_\theta(X_1 = x_1, \dots, X_n = x_n) = P_\theta(x_1) \cdots P_\theta(x_n)$$

- *Funkcija verodostojnosti*: - neprekidna raspodela

$$L(\theta) = f_\theta(x_1, \dots, x_n) = f_\theta(x_1) \cdots f_\theta(x_n)$$

- Ocena po metodi maksimalne verodostojnosti: $L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.

Metod maksimalne verodostojnosti-primeri

Primer 149. *Ocena verovatnoće, parametar p u Bernulijevoj raspodeli:
Neka je $X_1 + \dots + X_n = k$.*

$$L(p) = p^k(1 - p)^{n-k}.$$

Maksimum za $p = k/n$

Primer 150.

Normalna raspodela, parametar $\theta = (\mu, \sigma^2)$. Funkcija verodostojnosti:

$$L(\mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{(\sigma^2)^{n/2}} \exp\left(-\frac{\sum_{k=1}^n (x_k - \mu)^2}{2\sigma^2}\right).$$

Logaritmujemo i odbacujemo sve sabirke koji ne zavise od parametara:

$$l(\mu, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{k=1}^n (x_k - \mu)^2, \quad \mu \in \mathbb{R}, \sigma^2 > 0,$$

Tražimo maksimum:

$$\frac{\partial l}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{k=1}^n (x_k - \mu) = 0,$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{k=1}^n (x_k - \mu)^2 = 0.$$

$$\mu = \frac{1}{n} \sum_{k=1}^n x_k, \quad \sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \mu)^2.$$

Primer 153. X_1, \dots, X_n - nezavisan uzorak iz $\text{Unif}[0, \theta]$.

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & x_1, \dots, x_n \leq \theta, \\ 0 & \text{inače.} \end{cases}$$

Za vežbu: Primeri 139, 140, 141 ✓ , Zadaci 144, 145, 146, 147, 149, 151-161