Procjenitelji u parametarskim statističkim modelima

Lončarević, Lorena

Sveučilište Josipa Jurja Strossmayera u Osijeku

Fakultet primijenjene matematike i informatike

Sveučilišni prijediplomski studij Matematika

Lorena Lončarević

Procjenitelji u parametarskim statističkim modelima

Završni rad

Osijek, 2023.

Sveučilište Josipa Jurja Strossmayera u Osijeku

Fakultet primijenjene matematike i informatike

Sveučilišni prijediplomski studij Matematika

Lorena Lončarević

Procjenitelji u parametarskim statističkim modelima

Završni rad

Mentor: doc.dr.sc. Ivan Papić

Osijek, 2023.

Sažetak:

Tema ovog rada su procjenitelji u parametarskim statističkim modelima. Za potrebe razumijevanja definirani su osnovni pojmovi iz teorije vjerojatnosti i statistike, iskazani su najbitniji rezultati te su navedeni primjeri parametarski zadanih distribucija. Pomoću uvedenih pojmova definiran je procjenitelj i neka njegova svojstva koja su poželjna te su definirani procjenitelji očekivanja, varijance i proporcije jednostavnog slučajnog uzorka. Iz navedenih primjera prametarski zadanih distribucija proizašli su konkretni parametarski statistički modeli. Za svaki navedeni parametarski statistički model napravljena je analiza, odnosno definiran je procjenitelj traženog parametra te su komentirana svojstva procjenitelja. Također, napravljene su simulacije koje ilustriraju spomenuta svojstva.
Ključne riječi: procjenitelj, statistički model, parametar, parametarski statistički model, očekivanje, varijanca, proporcija, χ^2 distribucija, normalna distribucija, binomna distribucija, Bernoullijeva distribucija, eksponencijalna distribucija

Estimators in parametric models

Abstract:

The topic of this bachelor's thesis are estimators of parameters in parametric models. Firstly, we defined some of the basic concepts and presented the most important results from the probability theory and statistics. Also, we presented examples of special probability distributions. The estimator and some of its desirable properties, and the estimators of expectation, variance and proportion of a random sample were defined using introduced terms. For each listed parametric distribution we defined parametric statistical model and an estimator of the required parameter. The properties of the estimator were analysed. Also, there are simulations which illustrate the mentioned properties.

Key words: estimator, statistical model, parameter, parametric statistical model, expectation, variance, proportion, χ^2 distribution, normal distribution, binomial distribution, Bernoulli distribution, exponential distribution

Sadržaj

Uvod

Svakodnevno na osnovu podataka koje prikupljamo iz okoline želimo doći do nekog zaključka što očekivati u budućnosti. Možemo reći da je statistika znanstvena disciplina koja se bavi razvojem metoda za manipuliranje podacima. Pod tim mislimo na njihovo prikupljanje, opisivanje i analiziranje te donošenje zaključka koji je utemeljen na prikupljenim podacima. Za statističko zaključivanje ključna je teorijska podloga teorije vjerojatnosti i njeno razumijevanje. U prvom poglavlju ovog rada nalaze se osnovni pojmovi iz teorije vjerojatnosti i statistike. Tu smo definirali parametarski statistički model i naveli primjere parametarski zadanih distribucija koje ćemo korisiti kod definiranja konkretnih parametarskih modela. Drugo poglavlje ovog rada bavi se procjenom parametara koji su učestali u problemima na koje nailazimo u statističkoj analizi podataka. Točnije, tu smo definirali pojam procjenitelja i poželjna svojstva te smo definirali procjenitelje očekivanja, varijance i proporcije. U trećem poglavlju se nalaze primjeri konkretnih parametarskih modela. Te modele smo analizirali te ilustrirali tražena svojstva procjenitelja.

1 Osnovni pojmovi

1.1 Osnovni pojmovi iz teorije vjerojatnosti

Kako bismo mogli uvesti pojmove vezane uz samu srž ovog rada prvo ćemo uvesti pojmove kao što su: slučajan pokus, prostor elementarnih događaja, σ-algebra skupova te familija događaja. To su sve pojmovi koji su nam potrebni kako bismo uveli aksiomatsku definiciju vjerojatnosti.

Definicija 1.1. *Za pokus kažemo da je slučajan ako njegov ishod nije jednoznačno određen.*

Definicija 1.2. *Elementarni događaj je svaki ishod jednog izvođenja slučajnog pokusa.*

Definicija 1.3. *Skup ili prostor elementarnih događaja je skup svih mogućih ishoda nekog slučajnog pokusa. Obično ga označavamo s* Ω.

Primjer 1. Kod bacanja jedne pravilne igraće kocke prostor elementarnih događaja je $\{1, 2, 3, 4, 5, 6\}$.

Kako bismo mogli definirati familiju događaja trebat će nam iduća definicija.

Definicija 1.4. *Neka je dan neprazan skup Ω. Familija $\mathcal{F}$ podskupova skupa Ω jest σ-algebra skupova na Ω ako vrijedi:*

i) $\emptyset \in \mathcal{F}$,

ii) ako je $A \in \mathcal{F}$, *onda je i* $A^c \in \mathcal{F}$,

iii) ako je dana prebrojiva familija skupova $(A_n, n \in \mathbb{N}) \subseteq \mathcal{F}$, *onda* $\mathcal{F}$ *sadrži i njihovu uniju tj.*

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}.$$

Familiju događaja $\mathcal{F}$ nekog pokusa možemo definirati kao bilo koju σ-algebru na prostoru elementarnih događaja Ω tog pokusa. Pojedine elemente familije $\mathcal{F}$ smatramo događajima. Sada možemo uvesti aksiomatsku definiciju vjerojatnosti te definiciju vjerojatnosnog prostora.

Definicija 1.5. *Neka je Ω neprazan prostor elementarnih događaja i $\mathcal{F}$ σ-algebra skupova na njemu. Funkciju $P : \mathcal{F} \to \mathbb{R}$ zovemo vjerojatnost na Ω ako zadovoljava sljedeće zahtjeve:*

A1. ***nenegativnost vjerojatnosti:*** $P(A) \geq 0$, *za sve* $A \in \mathcal{F}$,

A2. ***normiranost vjerojatnosti:*** $P(\Omega) = 1$,

A3. ***σ-aditivnost vjerojatnosti:*** *ako je dana prebrojiva familija međusobno disjunktnih skupova* $(A_i, i \in I) \subseteq \mathcal{F}, I \subseteq \mathbb{N}$, *tj.* $A_i \cap A_j = \emptyset$, $i \neq j$, *tada vrijedi*

$$P\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} P(A_i).$$

Zahtjeve A1. - A3. nazivamo ***aksiomima vjerojatnosti.***

Definicija 1.6. *Neka je Ω neprazan skup, $\mathcal{F}$ $\sigma-$algebra događaja na njemu, a P vjerojatnost na Ω. Uređenu trojku $(\Omega, \mathcal{F}, P)$ zovemo vjerojatnosni prostor.*

Definicija 1.7. *Neka je dan vjerojatnosni prostor $(\Omega, \mathcal{F}, P)$ i $\mathcal{B}(\mathbb{R})$ Borelova σ-algebra na $\mathbb{R}$. Slučajna varijabla X je funkcija $X : \Omega \to \mathbb{R}$ takva da je*

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} = \{X \in B\} \in \mathcal{F}$$

za sve $B \in \mathcal{B}(\mathbb{R})$.

Definicija 1.8. *Skup vrijednosti koje slučajna varijabla X može poprimiti zove se slika slučajne varijable i označava s $\mathcal{R}(X)$.*

Slučajne varijable dijelimo na diskretne i neprekidne obzirom na kardinalnost slike slučajne varijable.

Definicija 1.9. *Slučajna varijabla X na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ je diskretna ako postoji diskretan skup $D \subseteq \mathbb{R}$ takav da je $P(X \in D) = 1$, odnosno, ako je slika slučajne varijable prebrojiv skup.*

Diskretnu slučajnu varijablu X prikazujemo u obliku tablice:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix},$$

gdje je $x_i \in \mathcal{R}(X) = \{x_i : i \in I\}, I \subseteq \mathbb{N}$. Taj prikaz zovemo tablica distribucije ili distribucija slučajne varijable X.

Definicija 1.10. *Neka je dan vjerojatnosni prostor $(\Omega, \mathcal{F}, P)$ i funkcija $X : \Omega \to \mathbb{R}$ za koju vrijedi:*

- $\{\omega \in \Omega : X(\omega) \leq x\} = \{X \leq x\} \in \mathcal{F}$ *za svaki* $x \in \mathbb{R}$,
- *postoji nenegativna realna funkcija realne varijable, f, takva da vrijedi:*

$$P(\{\omega \in \Omega : X(\omega) \leq x\}) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)dt, \forall x \in \mathbb{R}.$$

Funkciju X zovemo apsolutno neprekidna slučajna varijabla na Ω ili neprekidna slučajna varijabla. Funkciju f tada zovemo funkcija gustoće vjerojatnosti slučajne varijable X ili, kraće, funkcija gustoće slučajne varijable X.

Definicija 1.11. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor i neka je X slučajna varijabla. Funkciju $F : \mathbb{R} \to [0, 1]$ koja realnom broju x pridružuje vjerojatnost da dana slučajna varijabla bude manja ili jednaka tom broju, tj. funkciju*

$$F(x) = P\{\omega \in \Omega : X(\omega) \leq x\} = P\{X \leq x\}$$

zovemo funkcija distribucije slučajne varijable X.

Napomena 1.1. *Funkcija distribucije diskretne slučajne varijable dana je s*

$$F(x) = P\{X \leq x\} = \sum_{x_i \leq x} p_i.$$

Funkcija distibucije neprekidne slučajne varijable dana je s

$$F(x) = P\{X \leq x\} = \int_{-\infty}^{x} f(t)dt.$$

Definicija 1.12. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor pridružen slučajnom pokusu. Funkciju $X = (X_1, \ldots, X_n)$ koja svakom ishodu pokusa pridružuje uređenu n-torku realnih brojeva $(x_1, \ldots, x_n)$ zovemo n-dimenzionalan slučajni vektor ako vrijedi*

$$\{X_1 \leq x_1\} \cap \cdots \cap \{X_n \leq x_n\} \in \mathcal{F}$$

za svaki $x_1 \in \mathbb{R}, \ldots, x_n \in \mathbb{R}$.

Definicija 1.13. *Neka je $(\Omega, \mathcal{F}, P)$ vjerojatnosni prostor pridružen slučajnom pokusu i $(X_1, \ldots, X_n)$ slučajni vektor. Funkciju*

$$F : \mathbb{R}^n \to [0, 1],$$

$$F(x_1, \ldots, x_n) = P\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$$

zovemo funkcija distribucije slučajnog vektora $(X_1, \ldots, X_n)$.

Definicija 1.14. *Neka su $X_1, \ldots, X_n$ slučajne varijable na nekom vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ s pripadnim funkcijama distribucije $F_1, \ldots, F_n$ i neka je F funkcija distribucije slučajnog vektora $(X_1, \ldots, X_n)$. Reći ćemo da su slučajne varijable $X_1, \ldots, X_n$ nezavisne ako vrijedi*

$$F(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n).$$

Numeričke karakteristike slučajne varijable

Definicija 1.15. *Neka je $(\Omega, \mathcal{F}, P)$ diskretan vjerojatnosni prostor i X slučajna varijabla na njemu. Ako red $\sum_{\omega \in \Omega} X(\omega)P(\omega)$ apsolutno konvergira (tj. ako konvergira red $\sum_{\omega \in \Omega} |X(\omega)||P(\omega)|$*

) onda kažemo da slučajna varijabla X ima matematičko očekivanje i broj

$$EX = \sum_{\omega\in\Omega} X(\omega)P(\omega)$$

zovemo matematičko očekivanje (očekivanje) slučajne varijable X.

Definicija 1.16. *Neka je X neprekidna slučajna varijabla s funkcijom gustoće f. Ako je konačan integral*

$$\int_{-\infty}^{\infty} |x|f(x)dx,$$

onda kažemo da slučajna varijabla X ima očekivanje i broj

$$\mu = EX = \int_{-\infty}^{\infty} |x|f(x)dx,$$

zovemo matematičko očekivanje neprekidne slučajne varijable X.

Napomena 1.2. *[2] Od brojnih svojstava matematičkog očekivanja izdvojit ćemo svojstvo linearnosti:*
Neka su a i b realni brojevi, a X slučajna varijabla koja ima očekivanje. Tada i slučajna varijabla $aX + b$ ima očekivanje i vrijedi: $E[aX + b] = aEX + b$.

Definicija 1.17. *Neka je X slučajna varijabla na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ i $r > 0$.*

- *Ako postoji $E(X^r)$, onda broj $\mu_r = E(X^r)$ zovemo moment r-tog reda ili r-ti moment od X.*
- *Ako postoji EX i $E(|X - EX|^r)$ onda broj $E(|X - EX|^r)$ zovemo r-ti centralni moment od X.*
- *Ako postoji $E(X - EX)^2$, onda taj nenegativan broj zovemo varijanca slučajne varijable X i označavamo s $VarX, \sigma_X^2$ ili σ^2.*

Napomena 1.3.

a) *Očekivanje slučajne varijable X je njezin moment prvog reda te ga se uobičajeno označava s μ.*
Varijanca je drugi centralni moment i predstavlja očekivano kvadratno odstupanje slučajne varijable od njezinog očekivanja.

b) [2] Jedno od svojstava varijance koje nam je bitno u ovom radu jest:

Neka je X slučajna varijabla koja ima varijancu te a i b proizvoljni realni brojevi. Tada vrijedi: $Var(aX + b) = a^2 Var X$.

Teorem 1.1. *[2] Neka su $X_1, \dots, X_n$ nezavisne slučajne varijable takve da postoji $VarX_k, k = 1, \dots, n$ i neka su $a_1, \dots, a_n$ realni brojevi. Tada vrijedi:*

$$Var\left(\sum_{i=1}^{n} a_k X_k\right) = \sum_{i=1}^{n} a_k^2\, VarX_k.$$

1.1.1 Osnovni granični rezultati

Teorem 1.2. ***(Slabi zakon velikih brojeva)**[2] Neka je $(X_n, n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranih slučajnih varijabli takav da je $VarX_1 < \infty$. Tada za svaki $\varepsilon > 0$*

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - EX_1\right| > \varepsilon\right) = 0,$$

gdje je $S_n = \sum_{i=1}^{n} X_i$.

Drugim riječima, slabi zakon velikih brojeva nam kaže kako vjerojatnost da aritmetička sredina niza slučajnih varijabli $(X_n, n \in \mathbb{N})$ odstupa od svoje očekivane vrijednosti za barem ε možemo učiniti proizvoljno malom birajući dovoljno veliki prirodni broj n.

Teorem 1.3. ***(Centralni granični teorem)** [2] Neka je $(X_n, n \in \mathbb{N})$ niz nezavisnih i jednako distribuiranoh slučajnih varijabli takav da je $VarX_1 < \infty$ i $S_n = \sum_{i=1}^{n} X_i$. Tada*

$$\frac{S_n - ES_n}{\sqrt{VarS_n}} \xrightarrow{D} Z \sim \mathcal{N}(0,1), \quad \text{kada } n \to \infty.$$

1.2 Osnovni pojmovi iz statistike

Prikupljene podatke koje obrađujemo obično organiziramo u tablicu u kojoj retci predstavljaju jedinke dok stupci predstavljaju zabilježena obilježja za te jedinke i njih nazivamo varijable te ih u statistici modeliramo korištenjem slučajnih varijabli. Uvedimo pojam statističkog modela.

Definicija 1.18. *Statistički model je familija funkcija distribucija slučajnog vektora $\mathbf{X}$ koja se uzima u obzir pri zaključivanju o promatranom problemu. Slučajni vektor $\mathbf{X} = (X_1, \dots, X_n)$*

nazivamo slučajan uzorak, a njegovu realizaciju $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ *uzorak.*

Definicija 1.19. *Jednostavni slučajni uzorak iz (funkcije) distribucije* F *je slučajni vektor* $\mathbf{X} = (X_1, \dots, X_n)$ *takav da*

i) $X_1, \dots, X_n$ *su nezavisne,*

ii) sve slučajne varijable $X_1, \dots, X_n$ *imaju funkciju distribucije* F.

Napomena 1.4. *Statistički model jednostavnog slučajnog uzorka je familija funkcija distribucija*

$$\mathbf{F}(t_1, \dots, t_n) = P(X_1 \leq t_1, \dots, X_n \leq t_n) = \prod_{i=1}^{n} P(X_i \leq t_i) = \prod_{i=1}^{n} F(t_i).$$

Tako se problem zaključivanja o distribuciji slučajnog vektora sveo na problem zaključivanja o distribuciji slučajne varijable.

Slučajna varijabla je u potpunosti zadana svojom distribucijom. Kada poznajemo distribuciju slučajne varijable, tada možemo izračunati vjerojatnosti vezane uz njezine realizacije. Distribucije nekih slučajnih varijabli mogu se jednostavno opisati koristeći jedan ili nekoliko realnih brojeva koje nazivamo parametrima (mogu se parametarski zadati). **Parametarski statistički model** $\mathcal{P} = \{\mathbf{F}_\theta : \theta \in \Theta\}$ je statistički model u kojem poznajemo ili pretpostavljamo analitički oblik od funkcije distribucije $\mathbf{F}_\theta$ do na neki k-dimenzionalni parametar $\theta \in \Theta$, gdje je $\Theta \subseteq \mathbb{R}^k$ prostor prametara odnosno skup svih dozvoljenih vrijednosti parametra.

Napomena 1.5. *Posebno, ako se radi o jednostavnom slučajnom uzorku* $(X_1, \dots, X_n)$*, gdje su slučajne varijable* $X_1, \dots, X_n$ *iz neke parametarski zadane distribucije s funkcijom distribucije* F_θ*, tada prema Napomeni 1.4 vrijedi:* $\mathbf{F}_\theta(t_1, \dots, t_n) = \prod_{i=1}^{n} F_\theta(t_i)$. *Parametarski statistički model jednostavnog slučajnog uzorka je tada* $\mathcal{P} = \{\mathbf{F}_\theta : \theta \in \Theta\} = \left\{\prod_{i=1}^{n} F_\theta : \theta \in \Theta\right\}$*, gdje je* Θ *prostor parametara. Primijetimo kako smo parametarski model jednostavnog slučajnog uzorka zapisali pomoću funkcije distribucije jedne komponente te umjesto da zaključujemo o funkciji distribucije slučajnog vektora* $(X_1, \dots, X_n)$ *možemo zaključivati o funkciji distribucije jedne komponente i promatrati model* $\mathcal{P} = \{F_\theta : \theta \in \Theta\}$.

Primjer 2. (primjer parametarskog statističkog modela)

Provode se izbori te birači imaju pravo birati između dva kandidata A i B. Mi želimo zaključivati o rezultatima izbora te anketiramo n slučajno odabranih birača o njihovom izboru. Odgovore možemo označiti kao

$$\text{0-kandidat A}$$
$$\text{1-kandidat B.}$$

Dobiveni uzorak $x_1, \ldots, x_n,\ x_i \in \{0, 1\}$ smatramo realizacijom slučajnog uzorka $X_1, \ldots, X_n$. Možemo pretpostaviti da su $X_1, \ldots, X_n$ nezavisne jer smo birače birali na slučajan način. Također pretpostavljamo da su jednako distribuirane. Distribuciju od X_i možemo zapisati kao

$$\begin{pmatrix} 0 & 1 \\ 1-\theta & \theta \end{pmatrix}.$$

Primjećujemo da se radi o jednostavnom slučajnom uzorku iz Bernoullijeve dstribucije (vidi poglavlje 1.3). Nepoznat parametar nam je θ, a prostor dozvoljenih prametara $\Theta = \langle 0, 1\rangle$. Radi se o parametarskom statističkom modelu $\mathcal{P} = \{\mathbf{F}_\theta : \theta \in \langle 0, 1\rangle\}$, gdje je $\mathbf{F}_\theta(t_1, \ldots, t_n) = \prod_{i=1}^{n} F_\theta(t_i)$, pri čemu je F_θ funkcija distribucije Bernoullijeve slučajne varijable.

1.3 Primjeri parametarski zadanih distribucija

Neke distribucije se tipično pojavljuju u određenim problemima pa ih imenujemo te ćemo u ovom poglavlju navesti neke primjere. Iz parametarske familije diskretnih distribucija upoznat ćemo se s Bernoullijevom i binomnom distribucijom. Iz parametarske familije neprekidnih distribucija navest ćemo normalnu, eksponencijalnu, χ^2 i studentovu distribuciju.

Bernoullijeva distribucija Za slučajnu varijablu X kažemo da ima Bernoullijevu distribuciju s parametrom p ako je $\mathcal{R}(X) = \{0, 1\}$ i $P(X = 1) = p$. Njome opisujemo slučajan pokus koji ima dva moguća ishoda (obično su te dvije moguće vrijednosti "uspjeh" i "neuspjeh" te je u tom slučaju parametar p vjerojatnost pojavljivanja "uspjeha"). Tablica distribucije Bernoullijeve distribucije s parametrom p je tada:

$$X = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}, \quad p \in \langle 0, 1\rangle, \quad q = 1 - p.$$

Očekivanje Bernoullijeve slučajne varijable X je $EX = p$, dok je varijanca $VarX = p(1-p)$.

Binomna distribucija

Definicija 1.20. *Neka je $n \in \mathbb{N}$ i $p \in \langle 0, 1\rangle$. Za slučajnu varijablu koja poprima vrijednosti iz skupa $\{0, 1, 2, \ldots, n\}$ s vjerojatnostima*

$$p_i = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

kažemo da ima binomnu distribuciju s parametrima n i p i pišemo $X \sim \mathcal{B}(n, p)$.

Binomna distribucija proizlazi iz nezavisnog ponavljanja istog Bernoullijevog pokusa n puta. Koristimo ju kad nas zanima broj uspjeha u n nezavisnih ponavljanja istog Bernoullijevog pokusa. Parametri n i p predstavljaju broj ponavljanja pokusa te vjerojatnost uspjeha.
Očekivanje binomne slučajne varijable X jest $EX = np$, dok je varijanca $VarX = np(1-p)$.

Normalna distribucija

Definicija 1.21. *Za neprekidnu slučajnu varijablu kažemo da ima Gaussovu ili normalnu distribuciju s parametrima μ i σ^2 ako je njezina funkcija gustoće dana izrazom*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

gdje su μ i σ^2 realni brojevi i $\sigma > 0$. Ako slučajna varijabla X ima normalnu distribuciju s parametrima μ i σ^2, koristimo oznaku $X \sim \mathcal{N}(\mu, \sigma^2)$.

Slika slučajne varijable iz normalne distribucije je skup realnih brojeva. Funkcija distribucije normalne slučajne varijable dana je izrazom

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \in \mathbb{R}.$$

Slika slučajne varijable s normalnom distribucijom je skup svih realnih brojeva. Normalnu distribuciju s parametrima $\mu = 0$ i $\sigma^2 = 1$ zovemo standardna ili jedinična normalna distribucija. Ako slučajna varijabla X ima normalnu distribuciju, $X \sim \mathcal{N}(\mu, \sigma^2)$, tada je očekivanje $EX = \mu$, dok je varijanca $VarX = \sigma^2$.

Teorem 1.4. *[1] Ako su $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad i = 1, \ldots, n$ normalne nezavisne slučajne vari-*

jable, tada

$$Y = \sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma^2\right).$$

Drugim riječima, linearna kombinacija normalnih nezavisnih slučajnih varijabli opet dolazi iz normalne distribucije.

Eksponencijalna distribucija

Definicija 1.22. *Neprekidna slučajna varijabla X ima eksponencijalnu distribuciju s parametrom $\lambda > 0$ ako je funkcija gustoće dana izrazom*

$$f(x) = \begin{cases} 0 & , \quad x < 0 \\ \lambda e^{-\lambda x} & , \quad x \geq 0. \end{cases}$$

Slika slučajne varijable iz eksponencijalne distribucije je skup nenegativnih realnih brojeva. Funkcija distribucije te slučajne varijable dana je izrazom

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ 1 - e^{-\lambda x} & , \quad x \geq 0. \end{cases}$$

Slučajna varijabla X s eksponencijalnom distribucijom može se koristiti za modeliranje vremena čekanja do pojave nekog događaja (npr. kvara uređaja). Očekivanje slučajne varijable X jest $EX = \frac{1}{\lambda}$, dok je varijanca $VarX = \frac{1}{\lambda^2}$

χ^2distribucija

Definicija 1.23. *Neprekidna slučajna varijabla ima χ^2 distribuciju s parametrom $n \in \mathbb{N}$ ako je funkcija gustoće dana izrazom:*

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{\langle 0,\infty \rangle}(x),$$

gdje $\mathbf{1_A}$ označava indikator funkciju

$$\mathbf{1}_A(x) = \begin{cases} 1 & , \quad x \in A, \\ 0 & , \quad x \notin A, \end{cases}$$

a $\Gamma(t)$ je Gama funkcija definirana s

$$\Gamma(t) = \int_0^\infty x^t e^{-x} dx.$$

Parametar n predstavlja broj stupnjeva slobode. Slika slučajne varijable iz χ^2 distribucije su pozitivni realni brojevi.

χ^2 distribucija je, među ostalima, važna zbog iduće tvrdnje.

Propozicija 1.1. *[1] Neka su $X, X_1, X_2, \ldots, X_n$ nezavisne slučajne varijable s distribucijom $\mathcal{N}(0,1)$. Tada vrijedi*

a) $X^2 \sim \chi^2(1)$(kvadrat standardne normalne ima χ^2 distribuciju s jednim stupnjem slobode)

b) $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$.

Očekivanje χ^2 distribucije s n stupnjeva slobode je $EX = n$, dok je varijanca $VarX = 2n$.

Studentova distribucija Studentova distribucija je vrlo slična standardnoj normalnoj, osim što su joj repovi puno teži nego u normalnoj distribuciji. To znači da vjerojatnosti ekstremnih vrijednosti sporije opadaju u nulu.

Definicija 1.24. *Za neprekidnu slučajnu varijablu X kažemo da ima studentovu distribuciju s parametrom $n \in \mathbb{N}$ ako je njena funkcija gustoće dana izrazom*

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

Oznaka je $X \sim \mathcal{T}_n$.

Parametar n predstavlja broj stupnjeva slobode. Što je on veći, to je veća sličnost s normalnom distribucijom. Slika slučajne varijable iz studetove distribucije je skup realih brojeva.

2 Procjenitelji

Kada distribucija slučajne varijable nije poznata javlja se problem jer ne možemo točno izračunati vjerojatnosti vezane za njezinu realizaciju niti neke od numeričkih karakteristika slučajne varijable.

Osnovu statističkog zaključivanja u parametarskom statističkom modelu čini zaključivanje o parametru. Kako želimo znati distribuciju slučajnog uzorka $(X_1, X_2, \ldots, X_n)$ iz parametarskog statističkog modela, procjenu parametra trebamo na neki način izračunati iz realizacije, odnosno, uzorka $(x_1, x_2, \ldots, x_n)$. Ponavljanjem istog istraživanja možemo dobiti nove podatke te će tada procjena biti drugačija. Zbog toga i procjenu parametra možemo smatrati kao realizaciju slučajne varijable ili slučajnog vektora. Definirajmo takvu slučajnu varijablu, tj. procjenitelj.

Definicija 2.1. *Neka je $(X_1, \ldots, X_n)$ slučajan uzorak iz parametarskog statističkog modela $\mathcal{P} = \{\mathbf{F}_\theta : \theta \in \Theta\}$ i $t : \mathbb{R}^n \to \Theta$ funkcija takva da je $t(X_1, \ldots, X_n)$ slučajna varijabla. Slučajnu varijablu $T = t(X_1, \ldots, X_n)$ nazivamo procjenitelj.*

Možemo reći da je procjenitelj određena statistika, tj. funkcija slučajnog uzorka.
U nastavku su definirana neka svojstva procjenitelja.

Definicija 2.2. *Procjenitelj T je nepristran ako za svaki $\theta \in \Theta$ vrijedi $E_\theta[T] = \theta$.*

Važno je da procjenitelj bude nepristran kako bi očekivana vrijednost procjenitelja parametra bila upravo stvarna vrijednost parametra.

Definicija 2.3. *Neka je $n \in \mathbb{N}$, $T_n = t_n(X_1, \ldots, X_n)$, $\quad t_n : \mathbb{R}^n \to \Theta$ procjenitelj u parametarskom statističkom modelu $\mathcal{P} = \{\mathbf{F}_\theta : \theta \in \Theta\}$. Kažemo da je niz procjenitelja $(T_n, n \in \mathbb{N})$ konzistentan ako za svaki $\theta \in \Theta$ i $\varepsilon > 0$ vrijedi*

$$\lim_{n\to\infty} P_\theta(|T_n - \theta| > \varepsilon) = 0.$$

Želimo da procjenitelj bude konzistentan jer očekujemo da se povećanjem broja podataka poveća točnost procjenitelja, odnosno da distribucija procjenitelja bude koncentriranija oko stvarne vrijednosti parametra kako veličina uzorka raste.

Definicija 2.4. *Neka je $n \in \mathbb{N}$, $T_n = t_n(X_1, \ldots, X_n)$, $\quad t_n : \mathbb{R}^n \to \Theta$ procjenitelj u parametarskom statističkom modelu $\mathcal{P} = \{\mathbf{F}_\theta : \theta \in \Theta\}$. Kažemo da je T_n asimptotski normalan procjenitelj ako niz slučajnih varijabli*

$$\frac{T_n - E[T_n]}{\sqrt{VarT_n}}$$

konvergira po razdiobi standardnoj normalnoj slučajnoj varijabli, tj. ako

$$\frac{T_n - E[T_n]}{\sqrt{VarT_n}} \xrightarrow{D} Z \sim \mathcal{N}(0,1), \text{ kada } n \to \infty.$$

Najefikasniji procjenitelji su nepristrani procjenitelji minimalne varijance.

U poglavlju 1.3 vidjeli smo koji su to parametri koji nam obično trebaju kako bismo znali egzaktnu distribuciju slučajne varijable. U nastavku ćemo spomenuti neke tipične probleme procjene.

2.1 Procjena očekivanja

Često iz uzorka želimo procijeniti očekivanu vrijednost nekog obilježja koje promatramo. Definirajmo procjenitelj očekivanja.

Neka je $X_1, X_2, \dots, X_n$ jednostavni slučajni uzorak iz distribucije F_μ slučajne varijable X, gdje je $\mu \in \mathbb{R}$ nepoznato očekivanje $EX = EX_i = \mu, i = 1, \dots, n$. Procjenitelj za očekivanje slučajne varijable X jest

$$\overline{X}_n = t_n(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

i nazivamo ga aritmetička sredina slučajnog uzorka.
Aritmetička sredina uzorka $x_1, \dots, x_n$ je realizacija procjenitelja $\overline{X}_n$ i procjena očekivanja slučajne varijable X.

Propozicija 2.1. *[2] Aritmetička sredina jednostavnog slučajnog uzorka je nepristran, konzistentan i asimptotski normalan procjenitelj očekivanja.*
Dokaz:

-nepristranost

$$E_\mu[\overline{X}_n] = E_\mu\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \overset{\text{linearnost očekivanja}}{=} \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

-konzistentnost

$\overline{X}_n$ *je konzistentan procjenitelj očekivanja* μ *jer prema Teoremu 1.2 za*
$\forall \varepsilon > 0 \lim\limits_{n\to\infty} P_\mu(|\overline{X}_n - \mu| > \epsilon) = 0.$

-asimptotska normalnost

Prema centralnom graničnom teoremu vrijedi:

$$\frac{S_n - ES_n}{\sqrt{VarS_n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right]}{\sqrt{\frac{1}{n^2}Var\sum_{i=1}^{n} X_i}} = \frac{\overline{X}_n - E\overline{X}_n}{\sqrt{Var\overline{X}_n}}$$
$$= \sqrt{n}\frac{\overline{X}_n - E\overline{X}_n}{\sigma} = Y_n \xrightarrow{D} Z \sim \mathcal{N}(0,1),\ \textit{kada } n \to \infty.$$

Drugim riječima, niz standardiziranih slučajnih varijabli konvergira prema standardnoj normalnoj slučajnoj varijabli. Y_n *se asimptotski ponaša kao standardna normalna slučajna varijabla, to jest* $\overline{X}_n$ *se asimptotski ponaša kao slučajna varijabla s normalnom distribucijom* $\mathcal{N}(\mu, \frac{\sigma^2}{n})$.

Izračunajmo varijancu procjenitelja:
$$\mathrm{Var}\overline{X}_n = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \overset{nezavisnost}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}X_i = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$
Jasno je da se povećanjem dimenzije uzorka varijanca smanjuje te konvergira u 0.

2.2 Procjena varijance

Neka je $X_1, \ldots, X_n$ jednostavni slučajni uzorak iz distribucije F_{σ^2} slučajne varijable X gdje je σ^2 nepoznata varijanca $VarX = VarX_i = \sigma^2, i = 1, \ldots, n$. Ako se vodimo istom idejom kao kod procjene očekivanja, tada varijancu slučajnog uzorka možemo definirati s

$$\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2. \tag{1}$$

Provjerimo je li varijanca slučajnog uzorka nepristran procjenitelj.

Označimo $\mu = EX$. Tada je

$$\overline{S}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu - (\overline{X}_n - \mu)\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(\overline{X}_n - \mu) + \frac{1}{n}\sum_{i=1}^{n}(\overline{X}_n - \mu)^2.$$

Uočimo:

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

Iz toga slijedi:

$$\overline{S}_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2-(\overline{X}_n-\mu)^2.$$

Sada možemo izračunati očekivanje:

$$\begin{aligned}E_{\sigma^2}[\overline{S}_n^2]&=\frac{1}{n}\sum_{i=1}^{n}E(X_i-\mu)^2-E(\overline{X}_n-\mu)^2\\&=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}X_i-\mathrm{Var}\overline{X}_n=\frac{1}{n}n\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2\neq\sigma^2.\end{aligned}\tag{2}$$

Vidimo da varijanca slučajnog uzorka nije nepristran procjenitelj, stoga za procjenu varijance koristimo modificiran procjenitelj:

Propozicija 2.2. *[2] Korigirana varijanca slučajnog uzorka*

$$S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\tag{3}$$

je nepristran procjenitelj varijance.

Dokaz:

$$E_{\sigma^2}[S_n^2]\overset{(1),(3)}{=}E_{\sigma^2}\left[\frac{n}{n-1}\overline{S}_n^2\right]\overset{(2)}{=}\frac{n}{n-1}\frac{n-1}{n}\sigma^2=\sigma^2$$

Napomena 2.1. *Može se pokazati da je korigirana varijanca slučajnog uzorka konzistentan procjenitelj pod određenim uvjetima. Pokazano u [1] na stranici 313. Example 9.4.2.*

2.3 Procjena proporcije

Propozicija 2.3. *[2] Neka je $X_1,\dots,X_n$ jednostavan slučajan uzorak iz slučajne varijable X koja ima Bernoullijevu distribuciju*

$$X\sim\begin{pmatrix}0&1\\1-\theta&\theta\end{pmatrix},$$

gdje je $\theta \in [0,1]$ nepoznati parametar (vjerojatnost "uspjeha" u pokusu s dva moguća ishoda). Tada je nepristran i konzistentan procjenitelj proporcije

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

Dokaz:

Očekivanje slučajne varijable X iz Bernoullijeve distribucije je upravo vjerojatnost uspjeha tj. $EX = \theta$. Problem procjene parametra θ se svodi na procjenu očekivanja. Svojstva nepristranosti i konzistentnosti su posljedica svojstava aritmetičke sredine slučajnog uzorka kao procjenitelja očekivanja (Vidi Propozicija 2.1.).

Napomena 2.2. *Obzirom da se radi o Bernoullijevoj slučajnoj varijabli, procjena proporcije je upravo relativna frekvencija jedinica u uzorku.*

3 Procjena parametara u konkretnim parametarskim modelima

U poglavlju 1.3 smo se upoznali s primjerima parametarski zadanih distribucija. Iz njih proizlaze parametarski statistički modeli te ćemo ih u nastavku analizirati.

3.1 χ^2 distribucija

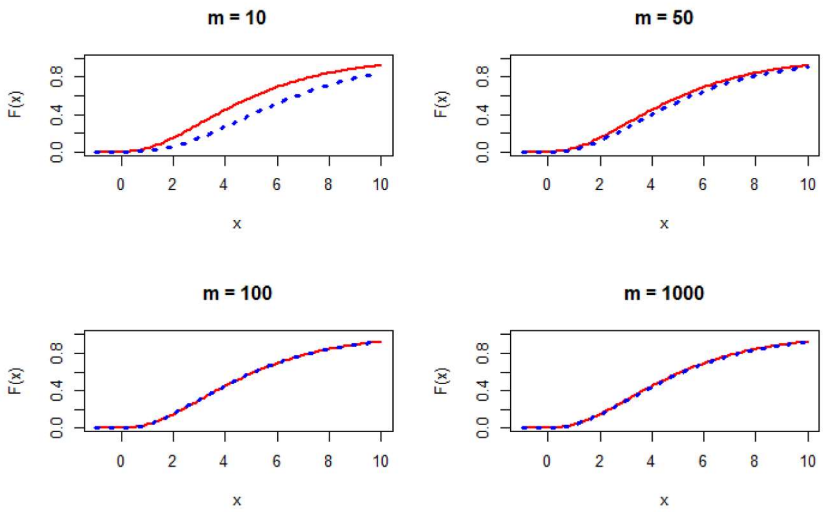
Neka je $(X_1, \dots, X_m)$ jednostavni slučajni uzorak iz χ^2 distribucije s $n \in \mathbb{N}$ stupnjeva slobode. Parametar n nam je nepoznat i trebamo ga procijeniti. Koristeći Napomenu 1.5 možemo zaključiti kako se tu radi o parametarskom statističkom modelu $\mathcal{P} = \{F_n : n \in \mathbb{N}\}$, gdje je F_n funkcija distrbucije slučajne varijable s χ^2 distribucijom.

U poglavlju 1.3 vidjeli smo kako je očekivanje slučajne varijable iz gore spomenute distribucije upravo broj stupnjeva slobode n, dok je varijanca $2n$. U poglavlju 2.1 definirali smo procjenitelj za očekivanje

$$\overline{X}_m = t_m(X_1, \dots, X_m) = \frac{1}{m}\sum_{i=1}^{m} X_i$$

te smo pokazali kako je taj procjenitelj nepristran i konzistentan. Upravo je $\overline{X_m}$ procjenitelj za broj stupnjeva slobode n.

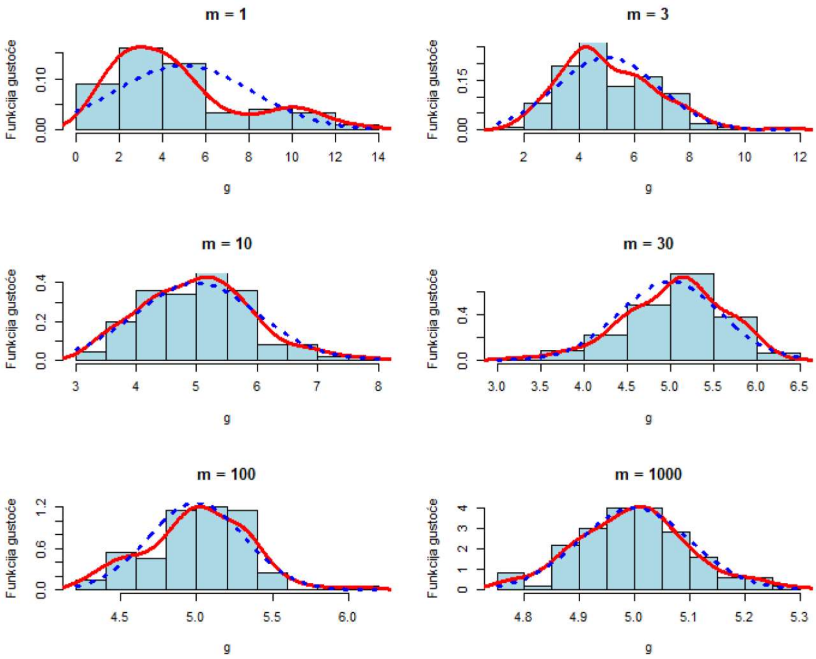
Ilustrirajmo procjenu i svojstva procjenitelja. Kako bismo ilustrirali nepristranost procjenitelja, za različit broj $m \in \mathbb{N}$, to jest za različit broj podataka u uzorku usporedit ćemo grafove funkcija distribucije. U tu svrhu za $m \in \{10, 50, 100, 1000\}$ generirat ćemo na slučajan način realizacije iz χ^2 distribucije s 5 stupnjeva slobode $(x_1, \ldots, x_m)$ te izračunati procijenjen stupanj slobode, odnosno aritmetičku sredinu uzorka $\overline{x}_m = \frac{1}{m}\sum_{i=1}^{m} x_i$. Na grafičkom prikazu 1, u ovisnosti o broju podataka, crvenom bojom smo prikazali χ^2 funkciju distribucije s 5 stupnjeva slobode, dok je s plavom bojom prikazana χ^2 funkcija distribucije kojoj je stupanj slobode aritmetička sredina generiranog uzorka te primijećujemo kako se s povećanjem broja podatka grafovi spomenutih distribucija sve više poklapaju.



Slika 1: Grafički prikaz funkcija distribucije s točnim i procijenjenim stupnjem slobode.

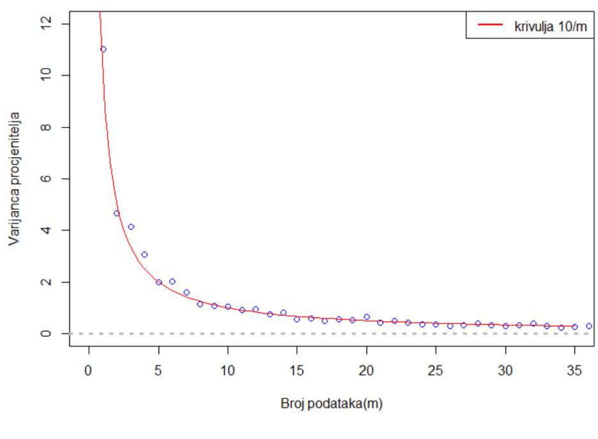
U poglavlju 2.1 dokazano je kako se aritmetička sredina jednostavnog slučajnog uzorka asimptotski ponaša kao slučajna varijabla s normalnom distribucijom $\mathcal{N}(\mu, \frac{\sigma^2}{n})$, gdje je n veličina slučajnog uzorka. Kod ovog parametarskog modela očekujemo da procjenitelj za broj stupnjeva slobode n asimptotski ima normalnu distribuciju $\mathcal{N}(n, \frac{2n}{m})$. Ilustrirajmo to svojstvo za konkretne m i n. Za svaki $m \in \{1, 3, 10, 30, 100, 1000\}$ generirat ćemo 100 slučajnih uzoraka iz χ^2 distribucije s $n = 5$ stupnjeva slobode te će aritmetička sredina svakog od uzoraka predstavljati jednu realizaciju procjenitelja stupnjeva slobode.

Na grafičkom prikazu 2 plavom bojom je prikazana funkcija gustoće slučajne varijable s normalnom distribucijom $\mathcal{N}(5, \frac{10}{m})$, dok je crvenom bojom prikazana procjena funkcije gustoće aritmetičke sredine slučajnog uzorka te ona prati histogram generiranog slučajnog uzorka našeg procjenitelja. Možemo primijetiti kako se povećanjem broja m distribucija prosjeka približava normalnoj distribuciji $\mathcal{N}(5, \frac{10}{m})$.



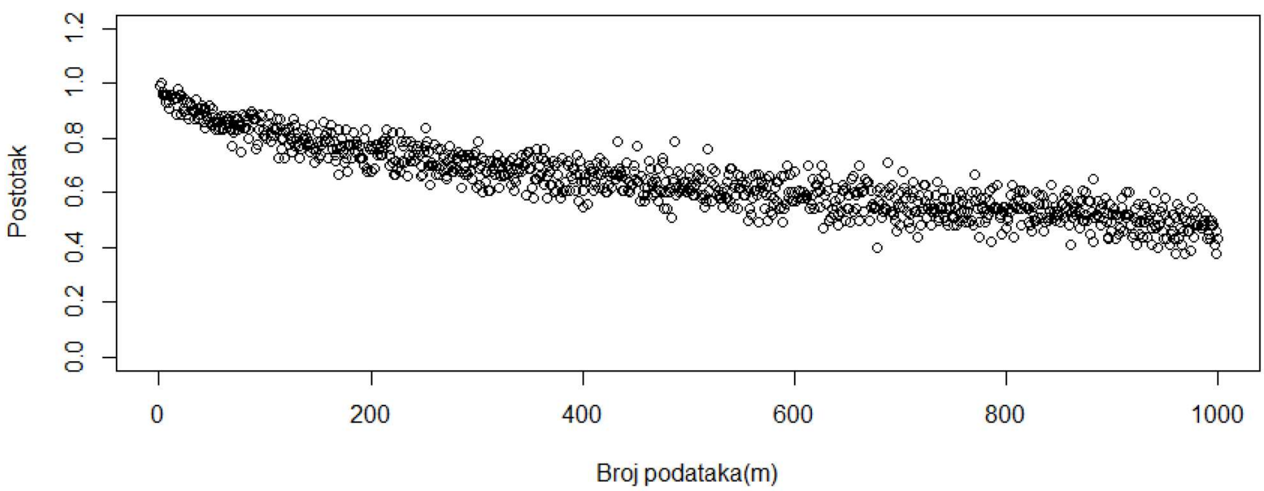
Slika 2: Grafički prikaz procjene funkcije gustoće procjenitelja stupnjeva slobode i funkcije gustoće normalne distribucije s pripadnim parametrima.

Na grafičkom prikazu 3 možemo vidjeti kako se ponaša varijanca procjenitelja stupnja slobode u konkretnom slučaju kada je stupanj slobode $n = 5$.



Slika 3: Grafički prikaz ovisnosti varijance procjenitelja o broju podataka.

Generirano je 100 uzoraka iz spomenute distribucije za broj podataka $m \in \mathbb{N}, 1 \leq m \leq 35$. Kada smo izračunali aritmetičku sredinu svakog uzorka dobili smo realizacije našeg procjenitelja stupnjeva slobode u ovisnosti o broju podataka. Računanjem njihovih varijanci dobili smo točke koje su na grafičkom prikazu 3 označene plavom bojom. Primijetimo kako smo u poglavlju 2.1 zaključili kako je varijanca aritmetičke sredine uzorka jednaka $\frac{\sigma^2}{n}$, gdje je n broj podataka, a σ^2 varijanca slučajne varijable. Također, znamo kako je varijanca slučajne varijable iz χ^2 distribucije s n stupnjeva slobode $2n$. Tada je funkcija koja opisuje ovisnost varijance procjenitelja o broju podataka $f(m) = \frac{2n}{m} = \frac{10}{m}$ te je ona na grafu 3 prikazana crvenom bojom. Iz grafičkog prikaza je vidljivo kako varijanca procjenitelja teži u 0.



Slika 4: Grafički prikaz postotka $P(|\overline{X}_m - 5| > 0.07)$ u ovisnosti o broju podataka.

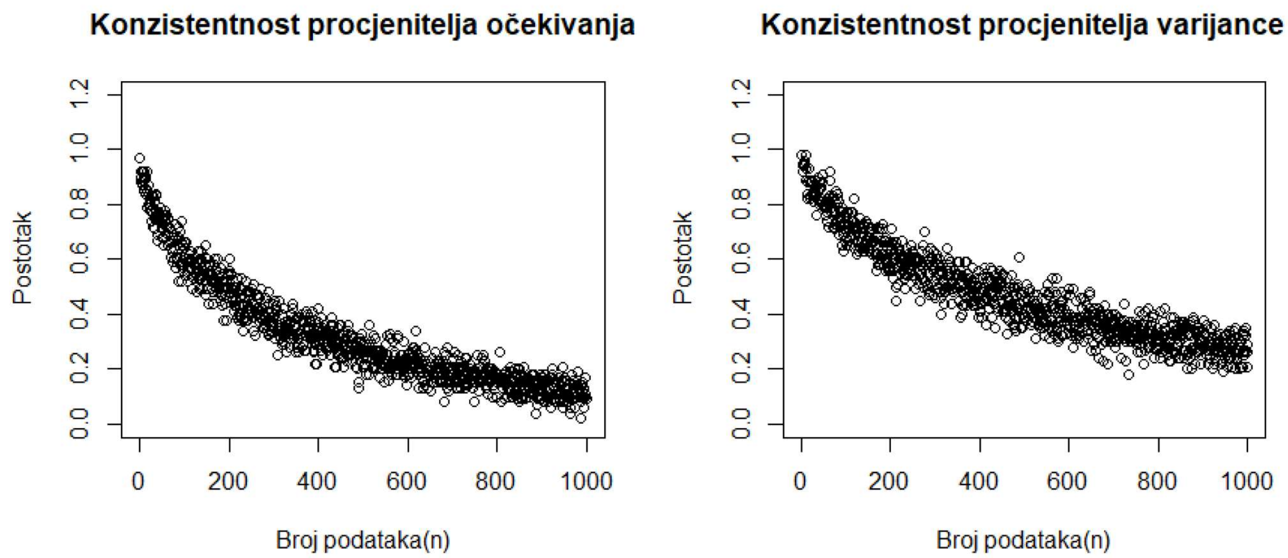
Kako bismo prikazali konzistentnost procjenitelja $\overline{X}_m$ generirali smo 100 uzoraka iz distribucije $\chi^2(5)$ te smo promatrali (u ovisnosti o broju podataka) postotak u kojem se procijenjena vrijednost parametra n razlikuje od stvarne vrijednosti parametra $n = 5$ za više od 0.07. Ovisnost postotka o broju podataka je vidljiva na grafičkom prikazu 4. Primijećujemo kako se traženi postotak smanjuje, ali relativno sporo što se može objasniti činjenicom da je varijanca $\chi^2(5)$ distribucije $2n = 10$.

3.2 Normalna distribucija

Neka je $(X_1, \ldots, X_n)$ jednostavni slučajni uzorak koji dolazi iz normalne distribucije s funkcijom distribucije F_{μ,σ^2} dok su nam parametri μ i σ^2 nepoznati. U pitanju je parametarski statistički model $\mathcal{P} = \{F_{\mu,\sigma^2} : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$. Iz poglavlja 1.3 znamo kako kod normalne distribucije parametar μ predstavlja očekivanje, a parametar σ^2 predstavlja varijancu slučajne varijable pa su procjenitelji za parametre μ i σ^2 procjenitelji očekivanja i varijance koje smo definirali u poglavljima 2.1 i 2.2.

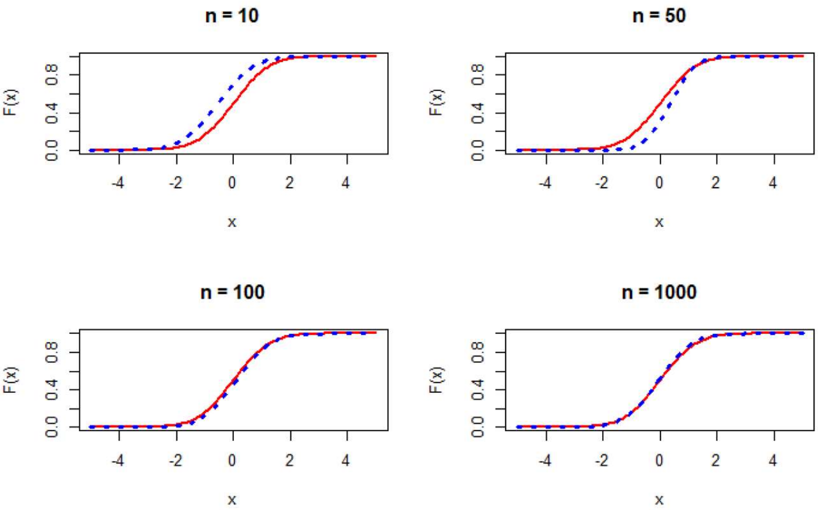
$$T_\mu = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad\qquad T_{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Svojstva nepristranosti i konzistentnosti nasljeđuju se iz nepristranosti i konzistentnosti procjenitelja očekivanja i varijance. Ilustraciju svojstva konzistentnosti za oba procjenitelja možemo vidjeti na grafičkom prikazu 5, gdje smo promatrali postotak (u ovisnosti o broju podataka n) u kojem se procijenjena vrijednost parametara razlikuje od stvarne vrijednosti za više od 0.05. Primijećujemo kako se postotak smanjuje povećanjem broja podataka n.



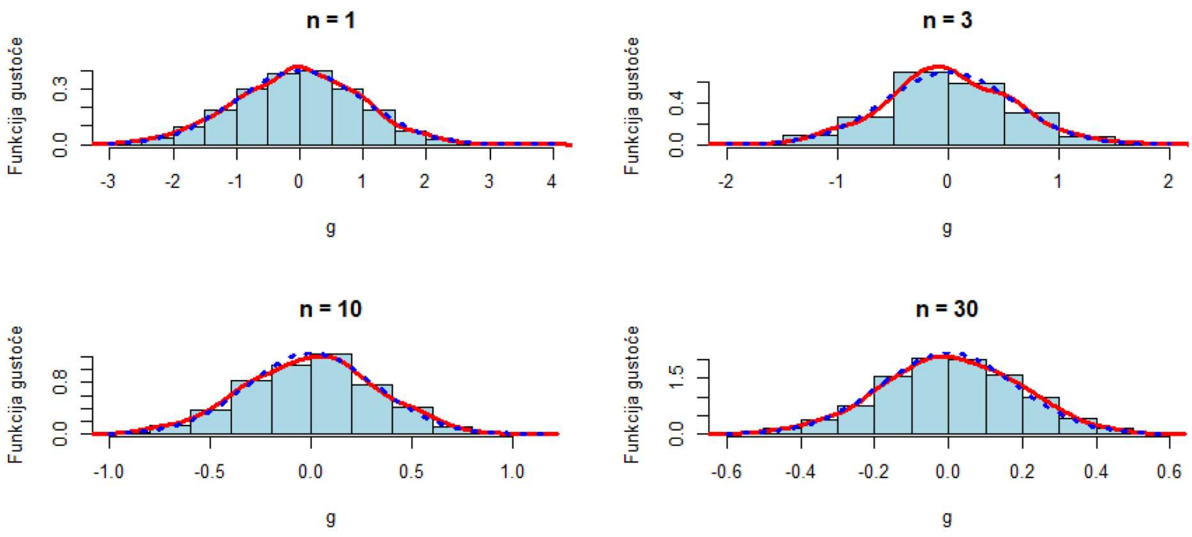
Slika 5: Ilustracija svojstva konzistentnosti procjenitelja T_μ i T_{σ^2}.

Svojstvo nepristranosti možemo ilustrirati na analogan način kao u prethodnom potpoglavlju. Ako generiramo uzorak iz standardne normalne distribucije $(x_1, \ldots, x_n)$ na slučajan način, tada za različit $n \in \mathbb{N}$ možemo usporediti funkciju standardne normalne distribucije i funkciju normalne distribucije s procijenjenim parametrima. Na grafičkom prikazu 6 crvenom bojom je označena funkcija standardne normalne distribucije, dok je plavom bojom označena funkcija normalne distribucije s procijenjenim parametrima. Primijećujemo kako se s povećenjem broja podataka u uzorku funkcije distribucije sve više poklapaju.

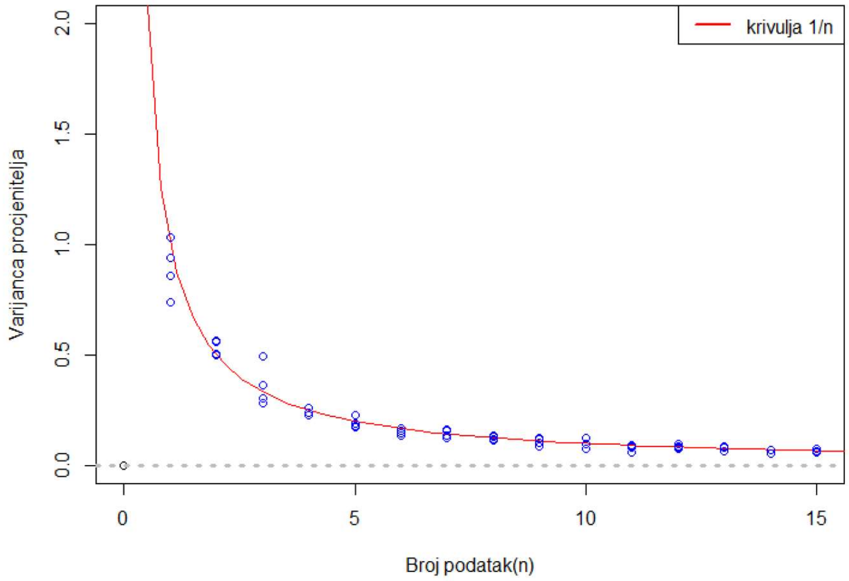


Slika 6: Grafički prikaz funkcija distribucije s točnim i procijenjenim parametrima standardne normalne distribucije.

Kako je procjenitelj očekivanja aritmetička sredina slučajnog uzorka, on je asimptotski normalan procjenitelj, ali možemo zaključiti i više od toga. Kako su podaci generirani iz normalne distribucije, tada je procjenitelj očekivanja linearna kombinacija normalnih nezavisnih slučajnih varijabli što prema Teoremu 1.4 znači kako procjenitelj ima egzaktnu normalnu distribuciju. Varijanca procjenitelja očekivanja konvergira u 0 s povećanjem broja podataka. Ilustracije navedenih svojstava su prikazane na grafičkim prikazima 7 i 8.



Slika 7: Grafički prikaz asimptotske normalnosti procjenitelja T_μ.



Slika 8: Grafički prikaz ovisnosti varijance procjenitelja T_μ o broju podataka.

3.3 Binomna distribucija

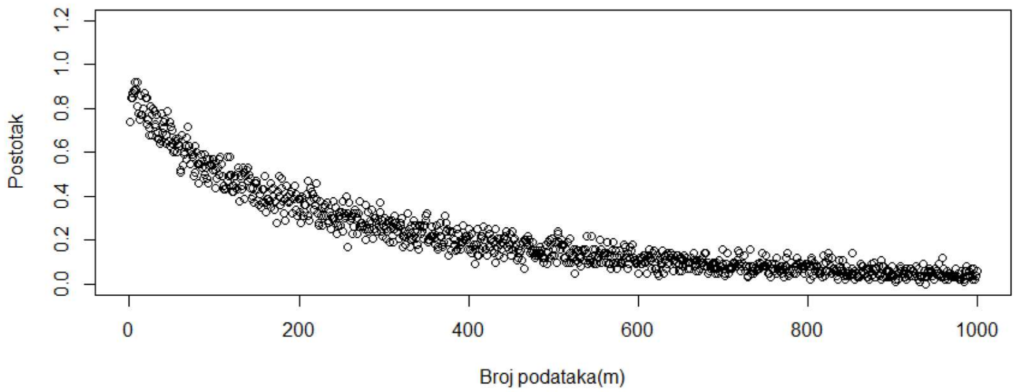
Neka je $(X_1, \ldots, X_m)$ slučajni uzorak iz binomne distribucije $\mathcal{B}(n, p)$, $n \in \mathbb{N}, 0 < p < 1$. U potpoglavlju 1.3 smo vidjeli da je očekivanje slučajne varijable s distribucijom $\mathcal{B}(n, p)$ jednako np, dok je varijanca $np(1-p)$. Izvest ćemo procjenitelj za parametar p kada nam je poznat parametar n, odnosno kada nam je poznat broj različitih vrijednosti koji slučajne varijable $X_1, \ldots, X_m$ poprimaju. Tada je $EX = np$ pa je smisleno pretpostaviti kako je procjenitelj parametra p definiran s

$$T_p = \frac{\overline{X}_m}{n} = \frac{1}{mn} \sum_{i=1}^{m} X_i.$$

Navedeni procjenitelj je nepristran jer:

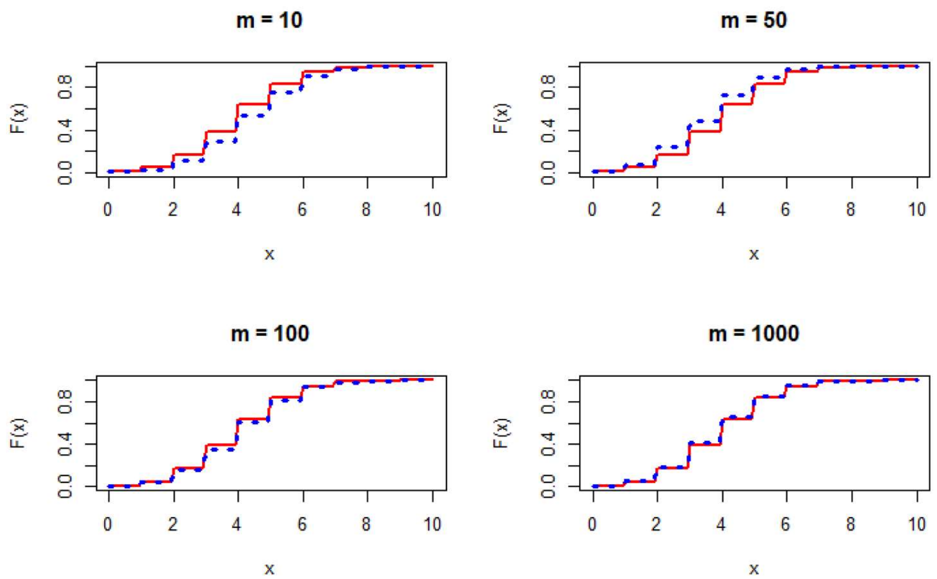
$$E_p[T_p] = E_p \left[\frac{1}{mn} \sum_{i=1}^{m} X_i \right] = \frac{1}{n} E_p \left[\overline{X}_m \right] = \frac{1}{n} \mu = \frac{1}{n} np = p.$$

Ilustracija konzistentnosti procjenitelja T_p može se vidjeti na grafičkom prikazu 9. Promatran je postotak u kojem se procjena parametra p razlikuje od stvarne vrijednosti parametra za više od 0.01. Ilustracija je simulirana za stvarne vrijednosti parametra $n = 10$, $p = 0.4$. Primijetimo kako se povećanjem broja podataka promatrani postotak smanjuje.



Slika 9: Grafički prikaz postotka $P(|T_p - 0.4| > 0.01)$ u ovisnosti o broju podataka.

Nepristranost ovog procjenitelja ilustrirat ćemo kao u prethodnom potpoglavlju. Ako generiramo uzorak $(x_1, \ldots, x_m)$ na slučajan način iz binomne distribucije s parametrima (10,0.4), tada možemo usporediti funkcije distribucija s točnim i procijenjenim parametrom. Na grafičkom prikazu 10 crvenom bojom je označena funkcija distribucije s parametrima $n = 10, \quad p = 0.4$ za različit broj podataka u uzorku, odnosno za $m \in \{10, 50, 100, 1000\}$. Plavo označen graf na istom prikazu jest graf funkcije distribucije s procijenjenim parametrom p. Primijećujemo kako se s povećanjem podataka grafovi sve više poklapaju.



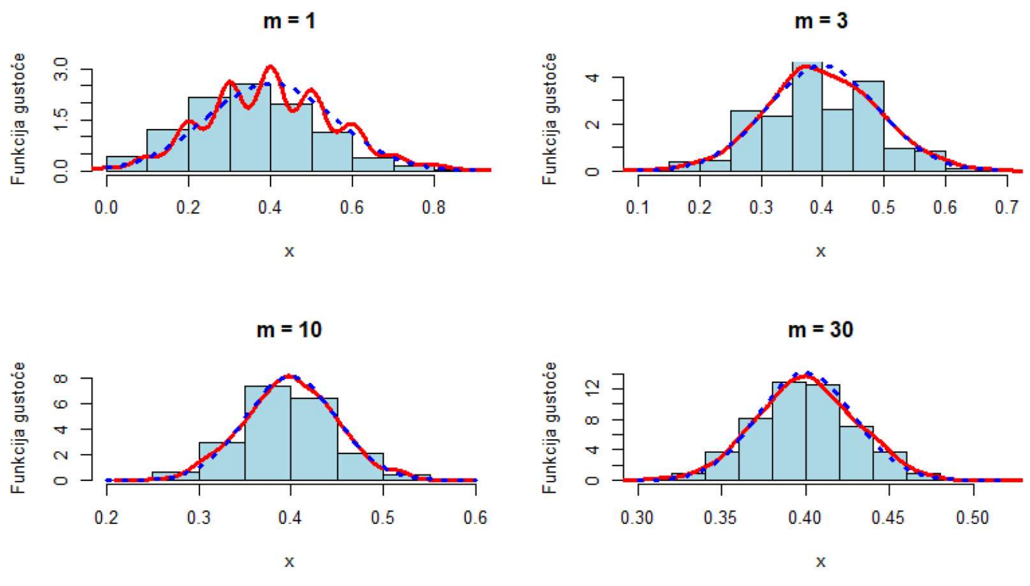
Slika 10: Grafički prikaz funkcija binomne distribucije s točnim i procijenjenim parametrom p, n poznat.

Pokažimo da je procjenitelj T_p asimptotski normalan. Znamo da je aritmetička sredina slučajnog uzorka asimptotski normalan procjenitelj. Stoga, prema centralnom graničnom teoremu

$$\frac{\overline{X}_m - \mu}{\sqrt{\frac{\sigma^2}{m}}} = \frac{\frac{1}{n}\overline{X}_m - \frac{1}{n}\mu}{\sqrt{\frac{\sigma^2}{n^2 m}}} = \frac{T_p - \frac{np}{n}}{\sqrt{\frac{np(1-p)}{n^2 m}}} = \frac{T_p - ET_p}{\sqrt{VarT_p}} \sim \mathcal{N}(0,1), n \to \infty,$$

$$VarT_p = \frac{\sigma^2}{n^2 m} = \frac{np(1-p)}{n^2 m} = \frac{p(1-p)}{nm}.$$

Primijetimo kako varijanca procjenitelja T_p s povećanjem broja podataka konvergira u nulu te da procjenitelj T_p asimptotski ima normalnu distribuciju $\mathcal{N}(p, \frac{p(1-p)}{nm})$. Na grafičkom prikazu 11 možemo vidjeti ilustraciju asimptotske normalnosti procjenitelja T_p za jednostavni slučajni uzorak iz binomne distribucije s parametrima $n = 10$ i $p = 0.4$. Za različit broj podataka u uzorku generirat ćemo 1000 uzoraka te izračunati procjenu parametra p te na taj način dobiti uzorak procjenitelja T_p. Kada usporedimo histogram i funkciju distribucije procjenitelja T_p (graf označen crvenom bojom) s funkcijom distribucije normalne slučajne varijable s parametrima $\mu = p = 0.04$ i $\sigma^2 = \frac{p(1-p)}{mn} = \frac{0.4 \cdot 0.6}{10m}$ (graf označen plavom bojom) primijećujemo kako se s povećanjem broja podataka u uzorku podudaraju.



Slika 11: Grafički prikaz asimptotske normalnosti procjenitelja parametra p binomne distribucije kada je $n = 10$

3.4 Bernoullijeva distribucija

Neka je $(X_1, \ldots, X_m)$ jednostavni slučajni uzorak iz Bernoullijeve distribucije s nepoznatim parametrom $p \in \langle 0, 1 \rangle$. Primijetimo da je Bernoullijeva distribucija specijalan slučaj binomne distirbucije gdje je $n = 1$. Tada je jednostavni slučajni uzorak $(X_1, \ldots, X_m)$ iz binomne distribucije s nepoznatim parametrom p, a $n = 1$, procjenitelj je definiran s:

$$T_p = \overline{X}_m.$$

Sva svojstva procjenitelja iz poglavlja 3.3 su naslijeđena.

3.5 Eksponencijalna distribucija

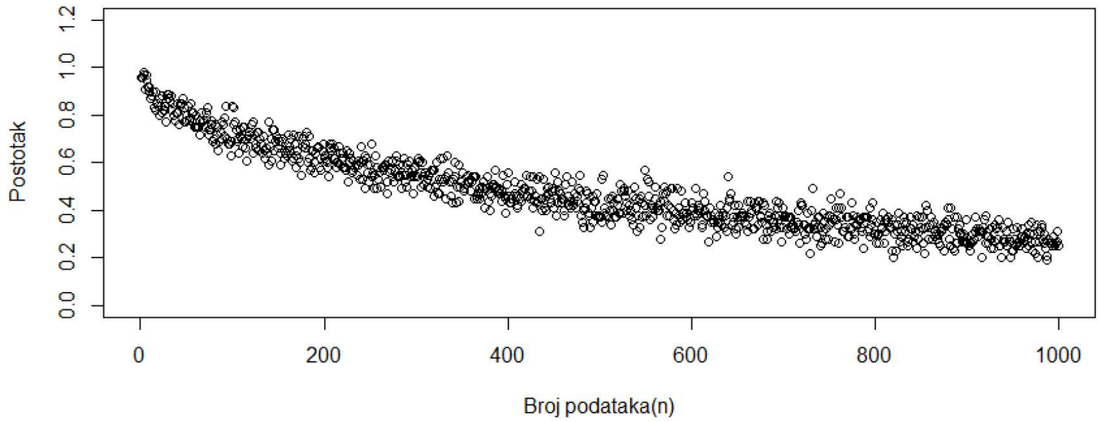
Neka je $(X_1, \ldots, X_n)$ slučajan uzorak iz eksponencijalne distribucije s parametrom $\lambda > 0$. Parametar λ nam je nepoznat i trebamo ga procijeniti. Tu se radi o parametarskom statističkom modelu $\mathcal{P} = \{F_\lambda : \lambda > 0\}$, gdje je F_λ funkcija distrbucije slučajne varijable s eksponencijalnom distribucijom.
U poglavlju 1.3 vidjeli smo kako je očekivanje slučajne varijable iz gore spomenute distribucije $\frac{1}{\lambda}$, dok je varijanca $\frac{1}{\lambda^2}$. U poglavljima 2.1 i 2.2 definirali smo procjenitelje za očekivanje i varijancu, $\overline{X}_n$ i S_n^2. U skladu s prijašnjim parametarskim statističkim modelima možemo promotriti dva procjenitelja parametra λ:

$$T_\lambda^1 = \frac{1}{\overline{X}_n} = \frac{n}{\sum_{i=1}^n X_i} \qquad \text{i} \qquad T_\lambda^2 = \sqrt{\frac{1}{S_n^2}}.$$

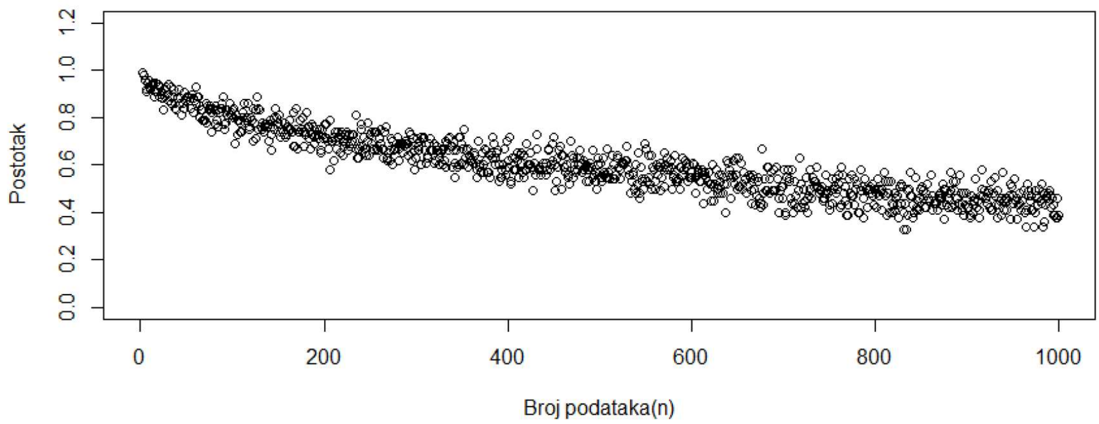
Napomena 3.1. *[4] Kriterij za vrednovanje različitih nepristranih procjenitelja istog nepoznatog parametra t svodi se na konstataciju da je najbolji onaj procjenitelj koji ima najmanju varijancu. Kažemo da je nepristran procjenitelj T_1 efikasniji od nepristranog procjenitelja T_2 ako vrijedi*

$$VarT_1 < VarT_2, \quad \forall t \in \Theta.$$

Konzistentnost procjenitelja prikazat ćemo kao u prethodnima poglavljima. Na grafičkim prikazima 12 i 13 promatran je postotak u kojem se procijenjena vrijednost parametra T_λ^1 i T_λ^2 razlikuje od stvarne vrijednosti $\lambda = 2$ za više od 0.07. Vidimo kako se promatrani postotak smanjuje povećanjem broja podataka.

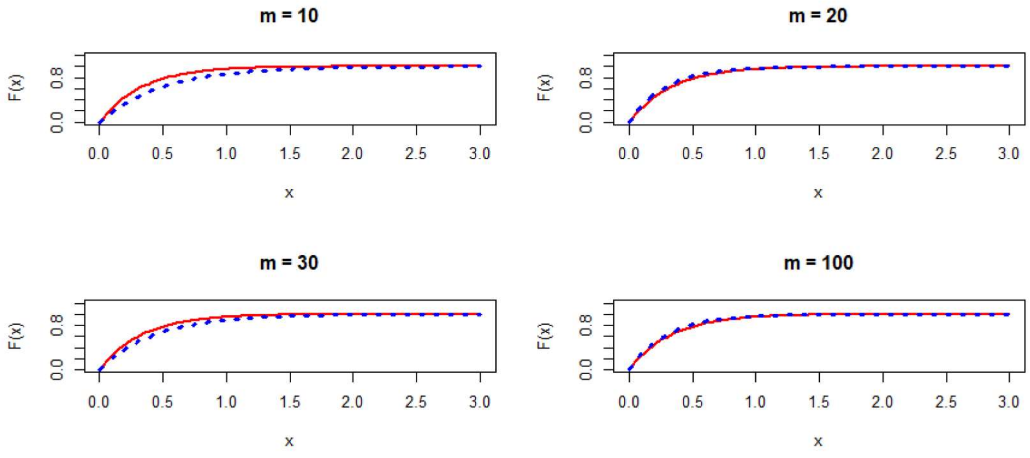


Slika 12: Grafički prikaz postotka $P(|T_\lambda^1 - 2| > 0.07)$ u ovisnosti o broju podataka.

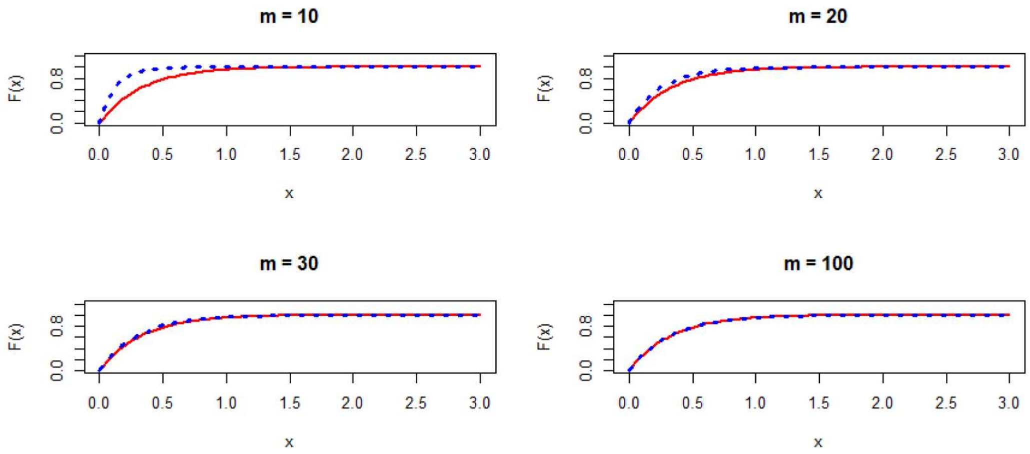


Slika 13: Grafički prikaz postotka $P(|T_\lambda^2 - 2| > 0.07)$ u ovisnosti o broju podataka.

Nepristranost ovih procjenitelja ilustrirat ćemo kao u prethodnim poglavljima. Na grafičkim prikazima 14 i 15 prikazane su usporedbe funkcije eksponencijalne distribucije s točnim parametrom $\lambda = 3$ i procijenjenim parametrom za procjenitelje T_λ^1 i T_λ^2.

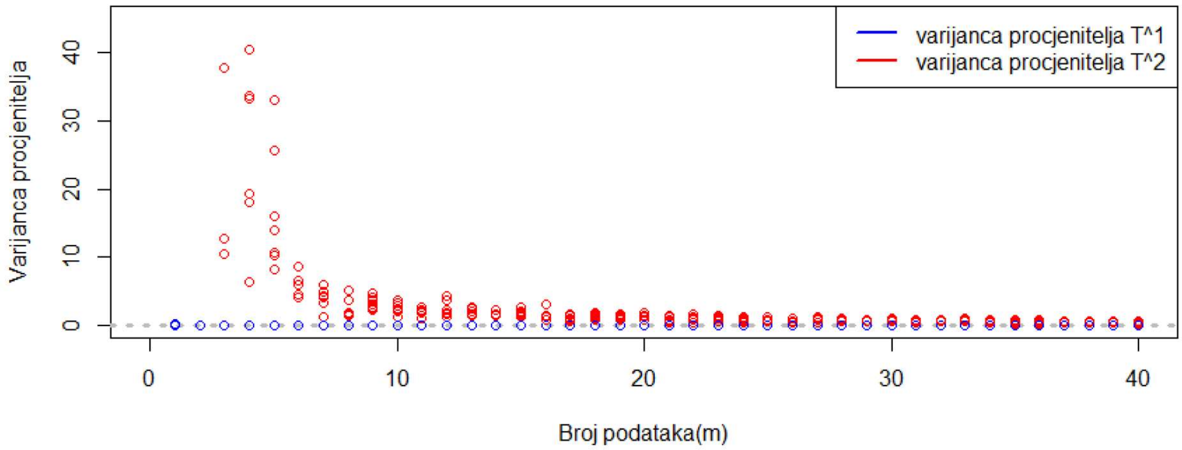


Slika 14: Grafički prikaz funkcija eksponencijalne distribucije s točnim i procijenjenim parametrom procjenitelja T_λ^1.



Slika 15: Grafički prikaz funkcija eksponencijalne distribucije s točnim i procijenjenim parametrom procjenitelja T_λ^2.

Na grafičkom prikazu 16 možemo vidjeti varijance procjenitelja T_λ^1 i T_λ^2. Osim što primijećujemo kako s povećanjem broja podataka varijanca konvergira u 0, vidimo kako je procjenitelj T_λ^1 efikasniji od procjenitelja T_λ^2 te je on onda bolji procjenitelj za nepoznati parametar λ.



Slika 16: Grafički prikaz varijance procjenitelja T^1_λ i T^2_λ.

Literatura

[1] L.J. Bain, M. Engelhardt - Introduction to Probability and Mathematical statistics(2nd edition), Duxbury-Thomson Learning, 1992.

[2] M. Benšić, N. Šuvak - Uvod u vjerojatnost i statistiku, Sveučilište J.J. Strossmayera, Odjel za matematiku, Osijek, 2014.

[3] M. Benšić, N. Šuvak - Primijenjena statistika, Sveučilište J.J. Strossmayera, Odjel za matematiku, Osijek, 2013.

[4] Ž. Pauše - Uvod u matematičku statistiku, Školska knjiga, Zagreb, 1993.