Statistika i osnovna mjerenja

Teorija uzoraka

M. Makek

2017/2018

Slučajni uzorak

- Iz neke populacije uzmemo uzorak od ***n*** – elemenata, koje označimo sa $x_1, \ldots, x_n$
- U mnogim statističkim problemima se elementi u uzorku mogu poistovjetiti sa opaženim vrijednostima ***n*** slučajnih varijabli $X_1, \ldots X_n$
- Definicija: ako su $X_1, \ldots X_n$ *slučajne varijable*, one čine ***slučajni uzorak*** veličine ***n*** ako su:
  - Nezavisne
  - Imaju iste raspodjele vjerojatnosti

  ($X_1, \ldots X_n$ su nezavisne i identično raspodijeljene)

Slučajni uzorak

- Svaku veličinu koju možemo izračunati iz uzorka nazivamo ***statistikom*** (npr. prosjek uzorka, standardna devijacija, momenti., itd.)
- Obzirom da prije uzimanja uzorka ne možemo znati vrijednost statistike ona je također **slučajna varijabla**
- Primjena:
  - Mjerenja fizikalnih veličina
  - Uzimanje uzorka iz velike populacije

Prosjek slučajnog uzorka

- Je definiran kao: $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- Prosjek je linearna kombinacija nezavisnih slučajnih varijabli ***X_1,... X_n*** te je sam **slučajna varijabla**
- **Očekivanje i varijancu prosjeka** možemo odrediti prema relacijama za linearnu kombinaciju slučajnih varijabli:

$$E(\bar{X}) = E\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)$$

$$= \frac{1}{n}\left(E(X_1) + \cdots + E(X_n)\right) = \frac{n\,\mu}{n} \quad \Rightarrow \quad \boxed{\mu_{\bar{X}} = \mu}$$

$$V(\bar{X}) = V\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)$$

$$= \frac{1}{n^2}\left(V(X_1) + \cdots + V(X_n)\right) \quad \Rightarrow \quad \boxed{\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}}$$

Total slučajnog uzorka

- Je definiran kao: $T_0 = \sum_{i=1}^{n} X_i$
- Total je linearna kombinacija nezavisnih slučajnih varijabli ***X_1,… X_n*** te je sam **slučajna varijabla**
- **Očekivanje i varijancu totala** možemo odrediti prema relacijama za linearnu kombinaciju slučajnih varijabli:

$$E(T_0) = \mu_{T_0} = n\mu$$

$$V(T_0) = \sigma_{T_0}^2 = n\sigma^2$$

Raspodjela prosjeka i totala

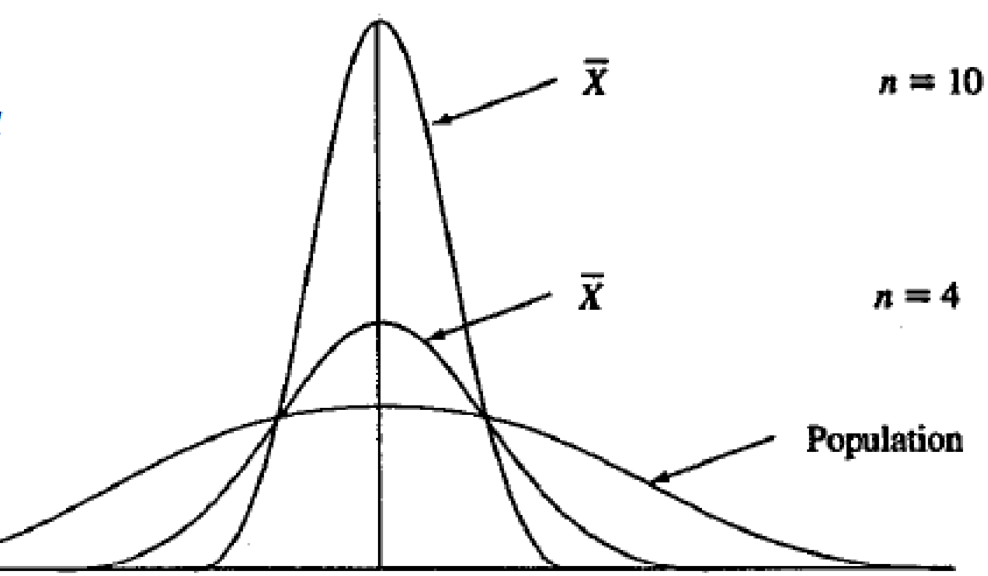
- Za slučajne varijable s **proizvoljnom raspodjelom** znamo ***očekivanje i varijancu*** prosjeka i totala no ne znamo točan oblik raspodjele
- Raspodjelu možemo dobiti iz *pravila vjerojatnosti* (za jedno-stavne raspodjele) ili pomoću *simulacijskog eksperimenta* (kad je nemoguće provesti analitički račun)

Raspodjela prosjeka normalne raspodjele

- <u>Specijalni slučaj:</u> normalna raspodjela
  - ako je uzorak sastavljen od ***n*** slučajnih varijabli s normalnom raspodjelom sa očekivanjem μ i varijancom σ^2, onda i prosjek ima normalnu raspodjelu s očekivanjem μ i varijancom σ^2/n (analogno vrijedi za total)
  - Što je ***n*** veći raspodjela prosjeka je uža (što imamo veći broj pokusa bol

$$E(\overline{X}) = \mu$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

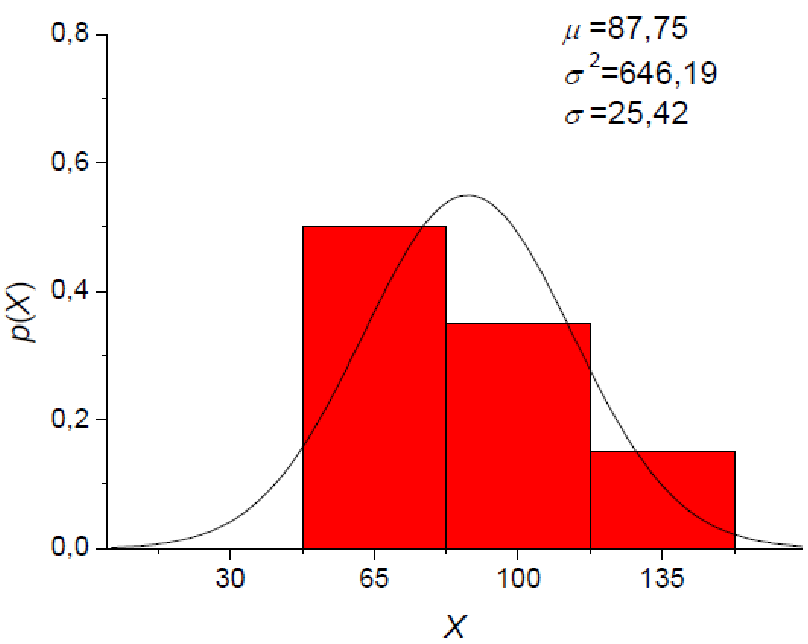


Primjer 1: diskretna raspodjela

Prodavač automobila prodaje 50% automobila niže klase po cijeni 65.000 kuna, 35% automobila srednje klase po 100.000 kuna i 15% automobila više klase po 135.000 kuna.
Definirajmo slučajnu varijablu

X=prihod od prodaje jednog automobila (u tisućama kuna)

x	$p(x)$
65	0,5
100	0,35
135	0,15



Primjer 1: diskretna raspodjela

Određenog dana najavila su se dva kupca. Neka su slučajne varijable:

X_1=prihod od prvog kupca

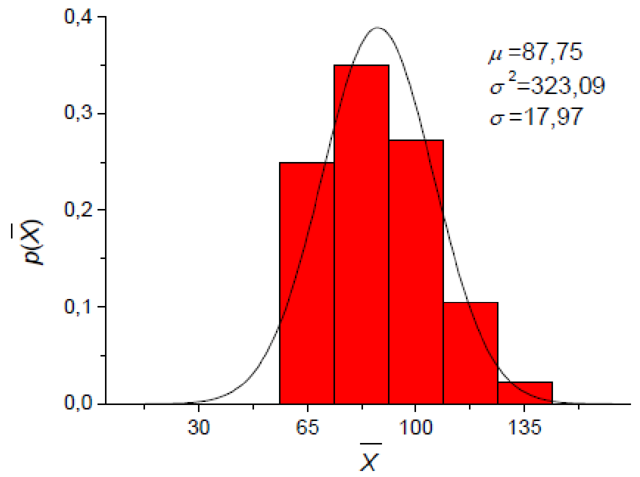
X_2=prihod od drugog kupca

Tablica mogućih prihoda:

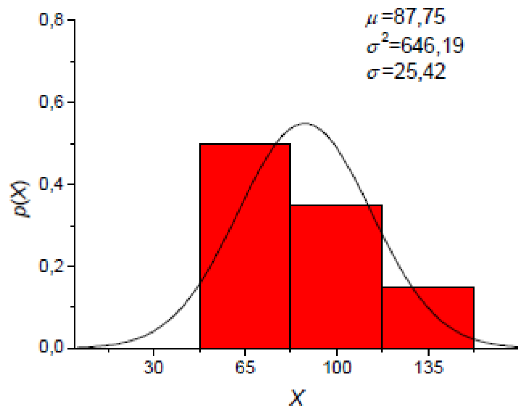
x_1	x_2	$p(x_1,x_2)$	x_1+x_2	$\bar{x}$
65	65	0,25	130	65
65	100	0,175	165	82,5
65	135	0,075	200	100
100	65	0,175	165	82,5
100	100	0,1225	200	100
100	135	0,0525	235	117,5
135	65	0,075	200	100
135	100	0,0525	235	117,5
135	135	0,0225	270	135

Slučajna varijabla $\overline{X}$ raspodijeljena je ovako:

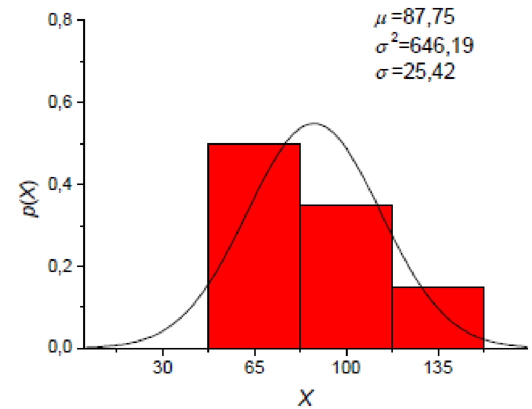
$\bar{x}$	$p(\bar{x})$
65	0,25
82,5	0,35
100	0,2725
117,5	0,105
135	0,0225

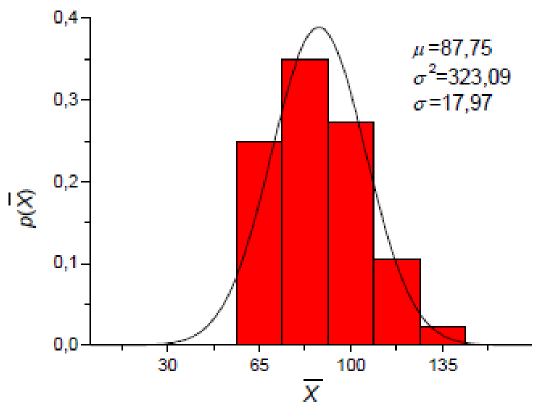


Primjer 1: diskretna raspodjela

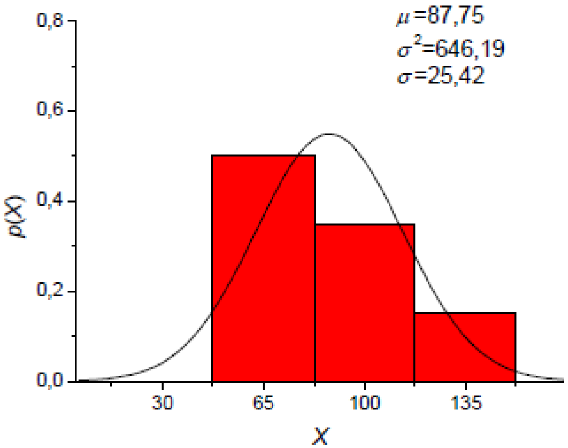


+

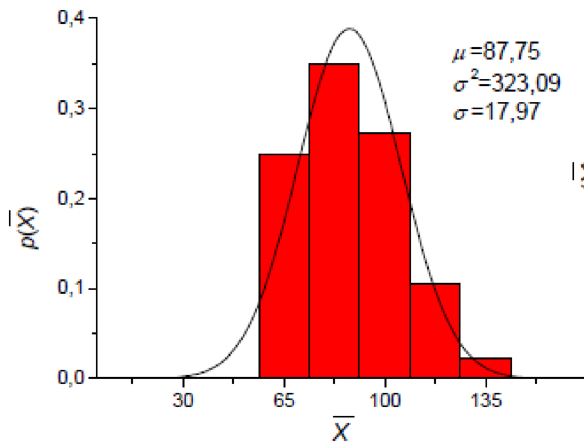




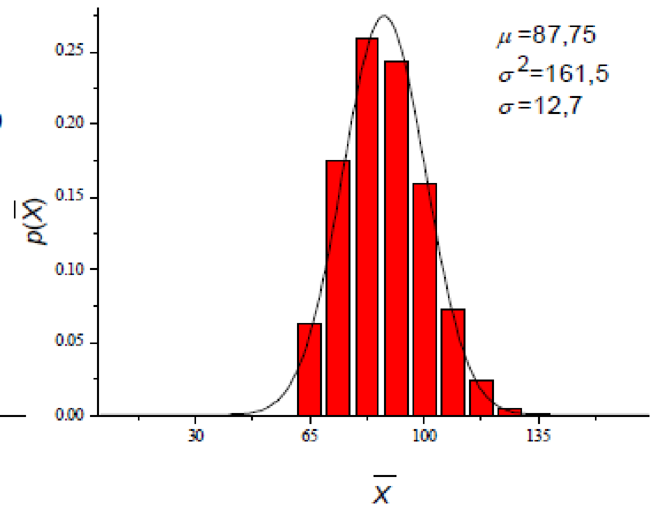
Primjer 1: diskretna raspodjela



$n = 1$



$n = 2$



$n = 4$

Primjer 2: kontinuirana raspodjela

Na putu do posla čekam autobus koji vozi svakih 10 minuta, a zatim tramvaj koji također vozi svakih 10 minuta.

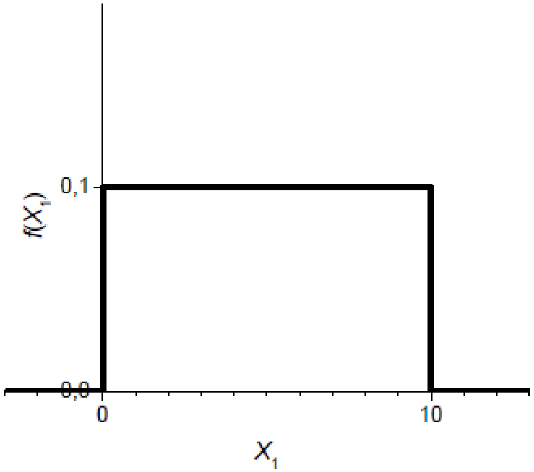
Neka su slučajne varijable:

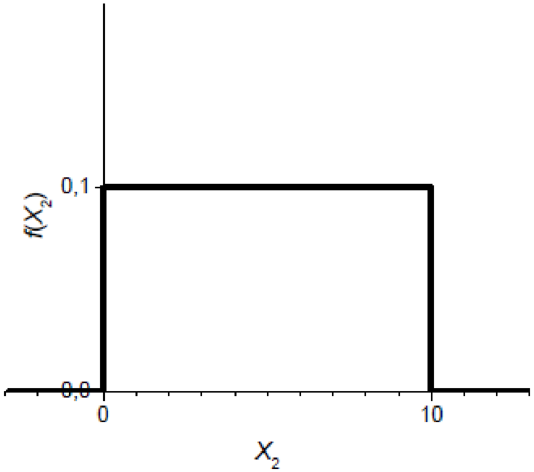
X_1 = vrijeme čekanja autobusa

X_2 = vrijeme čekanja tramvaja

$$f(x_1) = \begin{cases} 1/10 & , \quad 0 \le x_1 \le 10 \\ 0 & , \quad x_1 < 0 \quad ili \quad x_1 > 10 \end{cases}$$

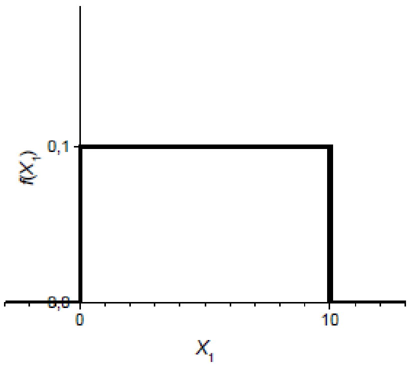
$$f(x_2) = \begin{cases} 1/10 & , \quad 0 \le x_2 \le 10 \\ 0 & , \quad x_2 < 0 \quad ili \quad x_2 > 10 \end{cases}$$



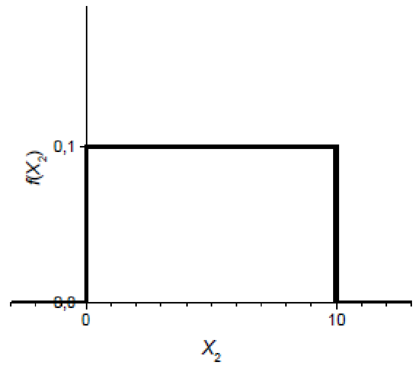


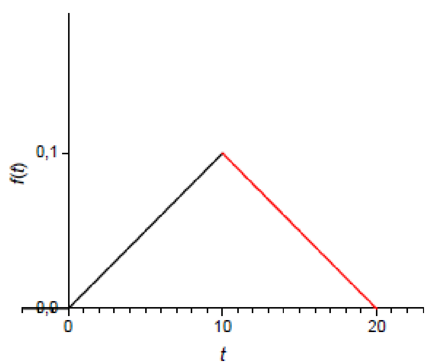
T_0 = ukupno vrijeme čekanja $= X_1 + X_2$

Primjer 2: kontinuirana raspodjela



+





$$f(t) = \begin{cases} \dfrac{t}{100}, 0 \leq t < 10 \\ \dfrac{20-t}{100}, 10 \leq t < 20 \\ 0, \quad inače \end{cases}$$

Uočiti

- Iz gornjih primjera vidi se da čak i kada populacija nije normalno raspodijeljena, ***raspodjela prosjeka i totala teži normalnoj raspodjeli*** kada je uzorak dovoljno velik

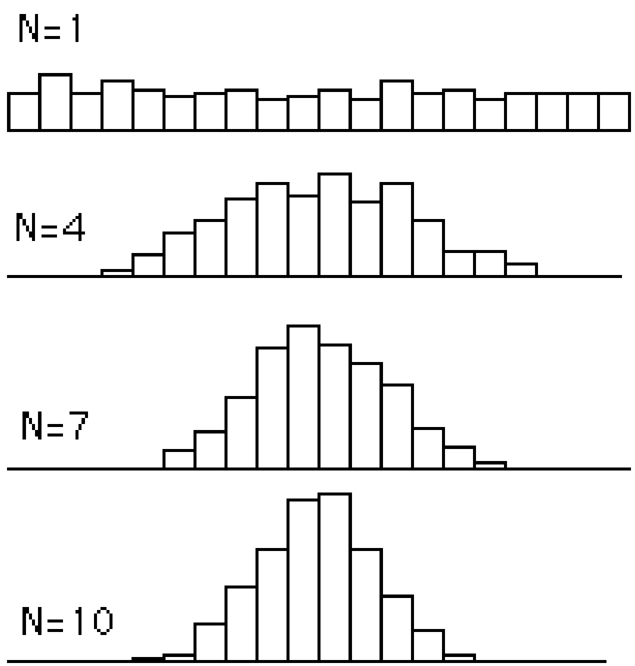
  → Centralni granični teorem

Centralni granični teorem

- Neka je $X_1, \ldots X_n$ slučajni uzorak ***proizvoljne*** raspodjele očekivanjem μ i standardnom devijacijom σ. Ako je ***n*** dovoljno velik $\bar{X}$ ima približno ***normalnu*** raspodjelu s očekivanjem $\mu_{\bar{X}} = \mu$ i standardnom devijacijom $\sigma_{\bar{X}}^2 = \sigma^2/n$
  T_0 ima približno ***normalnu*** raspodjelu s očekivanjem $\mu_{T_0} = n\mu$ i standardnom devijacijom $\sigma_{T_0}^2 = n\sigma^2$
- U praksi se teorem može primijeniti za n>30

Centralni granični teorem

- Primjer:



- Simulacija:

http://onlinestatbook.com/stat_sim/sampling_dist/