Metoda Monte-Karlo

- Numerička metoda rešavanja matematičkih problema pomoću modeliranja slučajnih promenljivih i statističkog ocenjivanja karakteristika tih promenljivih.
- Primenljiva je na sve matematičke zadatke.
- Naziv potiče od članka „*The Monte-Carlo method*”, N. Metropolis, S.M. Ulam iz 1949. god.
- Ne može predstavljati dokaz da je neka teorija ispravna.
- Za osnovnu sp bira se sp sa uniformnom raspodelom
- Pseudoslučajni brojevi – brojevi iz intervala (0,1)

MATEMATIČKA STATISTIKA

- Statistika je prvobitno proučavala masovne pojave u društvu (lat. status - država). Oblast primene statistike se proširila.
- Statistika predstavlja analizu podataka metodama zasnovanim na teoriji verovatnoće.
- Uspešno se primenjuje u raznim granama nauke: društvenim naukama, biologiji (Pirson i Fišer), fizici, meteorologiji, hemiji, medicini, ...
- Početkom XIX veka značajan doprinos teoriji su dali Laplas i Gaus.
- Početkom XX veka Pirson - engleska škola.
- Savremena statistika vezana za ime Fišera.

Populacija, obeležje

- **Populacija** – skup objekata koji izučavamo u mnogim problemima (**osnovni skup** ili **generalna kolekcija**).
- Na tom skupu proučavamo određeno **obeležje** – varijabilna kvantitativna ili kvalitativna osobina.
- Obeležje može biti jednodimenzionalno, dvodimenzionalno ili višedimenzionalno.
- Ako na slučajan način izaberemo jedan element populacije, ne znamo unapred vrednost obeležja (**modalitet**) koju taj element ima. To znači da se vrednost obeležja može shvatiti kao vrednost slučajne promenljive.
- Raspodela verovatnoća te sp se naziva **raspodela obeležja**.

Primer obeležja i modaliteta

- Posmatra se stepen oblačnosti u oblasti aerodroma u 7h. Populaciju čine svi dani od početka rada aerodroma. Posmatrano obeležje može imati vrednosti od 1 (vedro) do 10 (potpuno oblačno).
- Posmatraju se brzina i pravac vetra u oblasti aerodroma u toku dana. Ako se merenja beleže svakog sata, tada imamo dvodimenzionalno obeležje (brzina i pravac vetra).
- Posmatra se zbir dobijen pri bacanju tri numerisane kocke. Vrednosti obeležja su 3, 4, ..., 18, a elementi populacije su prvo, drugo, ..., n-to bacanje ...

Čime se bavimo u statistici?

U statistici se bavimo:

- **Prikupljanjem podataka.**
- **Prikazivanjem podataka** (tabelarno ili grafički).
- **Analizom podataka.**
- **Zaključivanjem na osnovu podataka**.

Prikupljanje podataka

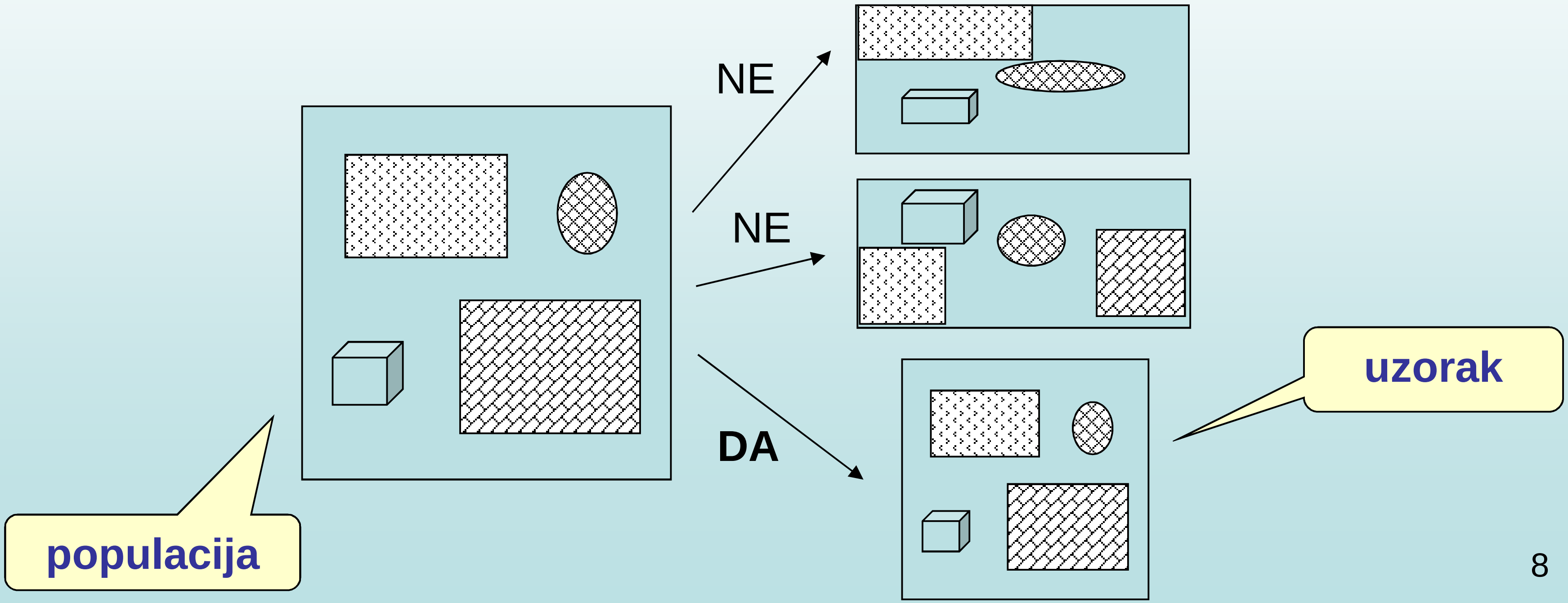
- Prema načinu dobijanja podataka imamo dve osnovne grupe:
- 1) Podaci dobijeni brojanjem
- 2) Podaci dobijeni merenjem.
- Podatke možemo prikupljati iz cele populacije, ili ako populacija sadrži veliki broj elemenata, iz cele populacije se na slučajan način uzima jedan deo koji se proučava.
- Taj deo se naziva ***uzorak***.
- Broj elemenata u uzorku je ***obim uzorka***.

Vrste uzorka

- Postoje razne metode za formiranje uzorka. Izbor metode zavisi od vrste, veličine i strukture populacije.
- Razlikuju se metode koje se primenjuju za male uzorke (obima $n \leq 30$) i za velike uzorke (obima $n > 30$). Za uzorke velikog obima tačne raspodele možemo aproksimirati drugim raspodelama (normalnom, na osnovu CGT).
- Uzorak se dobija:
- (a) izabrani element se posle beleženja osobina vraća u populaciju (***izbor sa vraćanjem***) ili
- (b) izabrani element se posle beleženja osobina ne vraća u populaciju (***izbor bez vraćanja***).

Reprezentativnost uzorka

- Krajnji zadatak statistike je određivanje raspodele obeležja na populaciji na osnovu obeležja na uzorku.
- Potrebno je obezbediti **reprezentativnost uzorka**, tj. obezbediti uzorak koji dobro reprezentuje populaciju.



Stratifikovani uzorak

- Pretpostavimo da se posmatrana populacija može podeliti na disjunktne podgrupe. Svaku takvu podelu nazivamo ***stratifikacija*** (raslojavanje), a dobijene podgrupe ***stratumima*** (slojevima).
- Neka je populacija obima N podeljena na L stratuma obima N_1, N_2, ..., N_L. Tada je

$$\sum_{j=1}^{L} N_j = N$$

- Uz oznaku $\omega_j = \frac{N_j}{N},\ j = 1,2,\ldots,L$

$$\sum_{j=1}^{L} \omega_j = 1$$

- Sa ω_j označavamo verovatnoću da element populacije bude u j-tom stratumu.

Stratifikovani uzorak, nastavak

- Iz populacije se dobija uzorak obima n, tako što se iz stratuma j dobija uzorak obima n_j, $j=1,2,\ldots,L$, pri čemu je

$$n_1 + n_2 + \ldots + n_L = n$$

- Izbor elemenata po stratumima može biti sa vraćanjem ili bez vraćanja. Postoje dva osnovna pristupa:
- (a) ***ravnomerni***, kad je obim uzorka n_j iz svakog stratuma isti i jednak:

$$n_j = \frac{n}{L}, \quad j = 1,2,\ldots,L$$

- (b) ***proporcionalni***, kad je obima uzorka n_j iz svakog stratuma proporcionalan obimu stratuma. Tada je

$$n_j = \omega_j n, \quad j = 1,2,\ldots,L$$

Stratifikovani uzorak, nastavak

- Kod izbora stratuma treba se rukovoditi principom da disperzija obeležja po stratumima bude što manja, jer je tada homogenost veća.

Primeri formiranja stratuma:

- Podela teritorije po reljefu;
- Podela stanovništva po zanimanju;
- Podela privrede po granama.

Grupni uzorak

- Pretpostavimo da se populacija može podeliti na disjunktne podgrupe.
- Od svih podgrupa na slučajan način se izabere nekoliko.
- Svi elementi izabranih grupa čine uzorak.
- Stratifikovani uzorak intuitivno više odgovara ideji reprezentativnog uzorka, ali se grupni uzorak može brže dobiti.
- Kod podele na grupe ne mora se poštovati zahtev o homogenosti po grupama.

Dvoetapni uzorak

- Pretpostavimo da se populacija može podeliti na grupe.
- Iz izabranih grupa se bira određeni broj elemenata.
- Dvoetapni uzorak je izvesna kombinacija grupnog i stratifikovanog uzorka.

Periodični uzorak

- Pretpostavimo da su elementi populacije na neki način poređani u niz. Elemente za uzorak izdvajamo periodično: npr. svaki četvrti.
- Periodični uzorak se može primeniti kod stratifikovanog uzorka pri izboru po stratumima, može se primeniti kod grupnog uzorka za izbor grupa, kao i kod dvoetapnog uzorka za izbor grupa ili elemenata iz izabranih grupa.

Prost slučajan uzorak

- **Definicija. Neka se u populaciji posmatra neko obeležje X. Prost slučajan uzorak obima n za posmatrano obeležje je n-dimenzionalna slučajna promenljiva $(X_1, X_2,..., X_n)$ pri čemu su slučajne promenljive $X_1, X_2,..., X_n$ nezavisne i sve imaju istu raspodelu kao posmatrano obeležje X.**
- Realizovan uzorak predstavlja konkretan niz vrednosti obeležja dobijenih na elementima populacije koji su izabrani u uzorak. To je jedna n-torka brojeva.

Tablično prikazivanje podataka

- Postoje dva načina predstavljanja podataka:
  - 1. brojčanim vrednostima
  - 2. po intervalima
- 1. Imamo realizovan uzorak obima n. Neka su sa x_1, x_2,..., x_k obeležene različite vrednosti (tzv. varijante) i zapisane u rastućem poretku. Tako se dobija **varijacioni niz**. Neka je n_1 puta dobijena vrednost x_1, ... Brojevi n_1, n_2,..., n_k su (**apsolutne**) **frekvencije** (**učestalosti**, **učestanosti**).

Vrednost obeležja	x_1	x_2	...	x_k
frekvencija	n_1	n_2	...	n_k

$n_1 + n_2 + ... + n_k = n$

U tabeli se mogu dati relativne frekvencije

$$\frac{n_j}{n}, \quad j = 1,..,k$$

Raspon i interval varijacije

- **Raspon varijacije** je razlika x_k-x_1= $x_{\max}$-$x_{\min}$
- **Interval varijacije** je interval $[x_1, x_k]=[x_{\min}, x_{\max}]$

Određivanje intervala

- Ako je n veliki broj, vrednosti obeležja se daju po intervalima.
- Određuju se najmanja i najveća vrednost u uzorku.
- Izabere se a_m (broj nešto manji od najmanje vrednosti u uzorku) i broj a_M (nešto veći od najveće vrednosti u uzorku).
- Interval $[a_m, a_M]$ se podeli na k intervala (klasa)
- $[a_1, a_2), [a_2, a_3), \ldots, [a_k, a_{k+1}]$,

  gde je $a_1 = a_m$, $a_{k+1} = a_M$.
- Najčešće se uzima između 6 i 20 intervala. Neki autori predlažu da broj klasa bude

$$k \leq 5 \log_{10} n$$

Podaci su dati po intervalima

- Kada su intervali (klase) određeni, određuje se koliko se elemenata uzorka nalazi u kojoj klasi i ti podaci se unose u tabelu

Vrednost obeležja	$[a_1, a_2)$	$[a_2, a_3)$	...	$[a_k, a_{k+1}]$
frekvencija	n_1	n_2	...	n_k

U tabeli se mogu dati relativne frekvencije

$$\frac{n_j}{n}, \quad j = 1,..,k$$

Najčešće su a_1 i a_{k+1} konačne vrednosti, ali mogu biti $-\infty$ i ∞.

Grafičko prikazivanje podataka

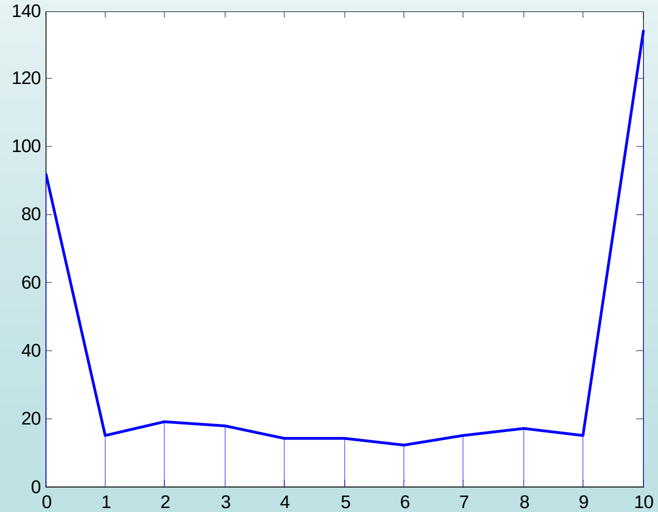
- Najčešće se koriste: poligon apsolutnih frekvencija, poligon relativnih frekvencija i empirijska funkcija raspodele.
- Neka je dat uzorak obima n. Kada u koord. sistemu redom predstavimo tačke $T_i\ (x_i, n_i),\ i=1,..,k$ i spojimo ih linijom, dobija se **poligon apsolutnih frekvencija**.
- Ako se tačke sa koordinatama $(x_i, n_i/n),\ i=1,..,k$ redom spoje linijom, dobija se **poligon relativnih frekvencija**.
- Neka je dat uzorak obima n. Ako se nad svakim intervalom konstruiše pravougaonik čija je površina jednaka apsolutnoj (relativnoj) frekvenciji za dati interval, dobija se **histogram apsolutnih** (relativnih) **frekvencija**.

Poligoni apsolutnih frekvencija

- Primer. U toku jedne godine dobijeni su podaci:

Stepen oblačnosti	0	1	2	3	4	5	6	7	8	9	10
Broj dana	92	15	19	18	14	14	12	15	17	15	134

Ovi podaci predstavljaju jedan uzorak obima 365.

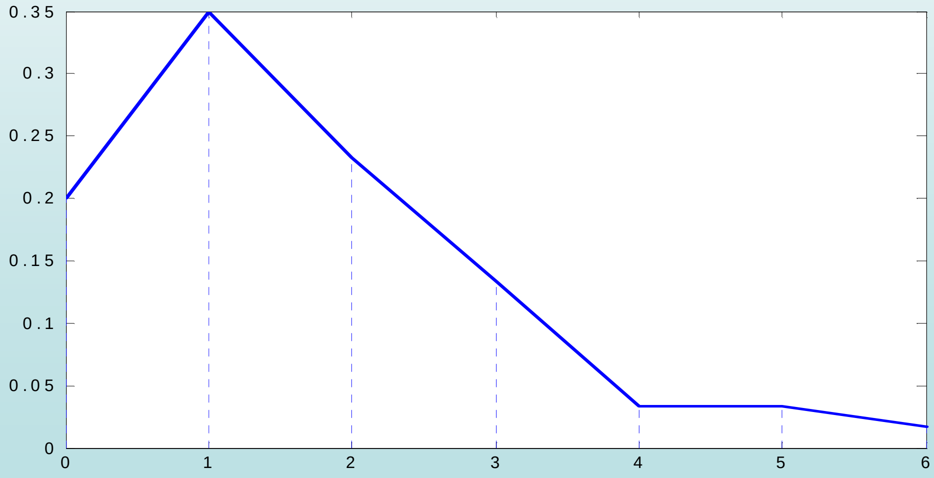


Poligon apsolutnih frekvencija

Poligoni relativnih frekvencija

- Primer. Posmatra se broj četvorki rođenih u toku jedne godine u jednoj oblasti, počev od 1800. god. Populaciju čine godine 1800, 1801, ... U periodu od 60 uzastopnih godina zabeleženo je

Broj četvorki	0	1	2	3	4	5	6
Broj godina	12	21	14	8	2	2	1

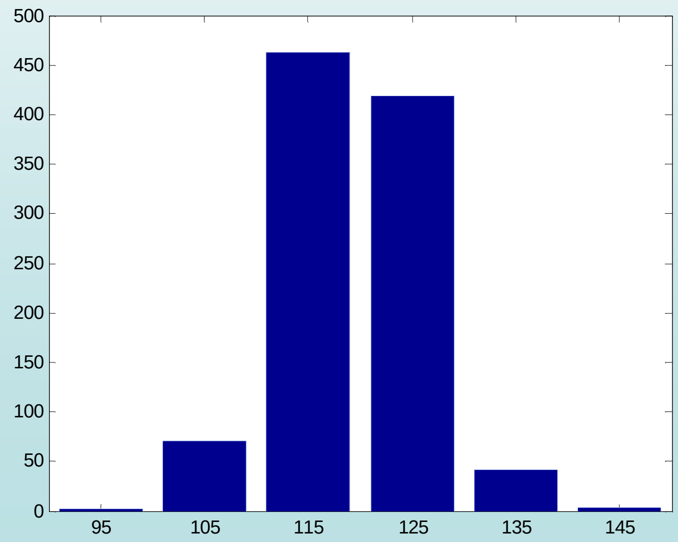


Poligon relativnih frekvencija

Histogram apsolutnih frekvencija

- Primer. Posmatra se broj artikala proizvedenih u toku jednog dana u jednoj fabrici. Broj artikala nije manji od 90 niti veći od 150. Populaciju čine radni dani, a obeležje je broj proizvedenih artikala po radnim danima.
  Proizvodnja u 1000 slučajno izabranih dana je data po intervalima

Broj artikala	[90,100)	[100, 110)	[110, 120)	[120, 130)	[130, 140)	[140, 150)
Broj dana	2	71	463	419	42	3

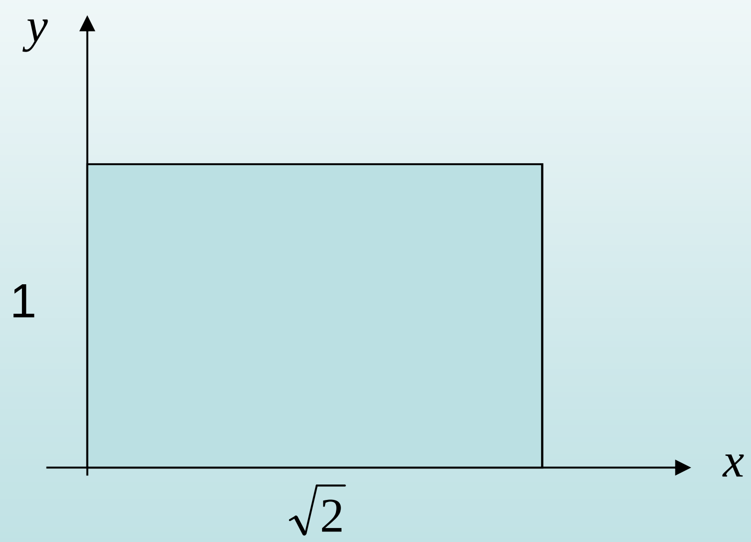


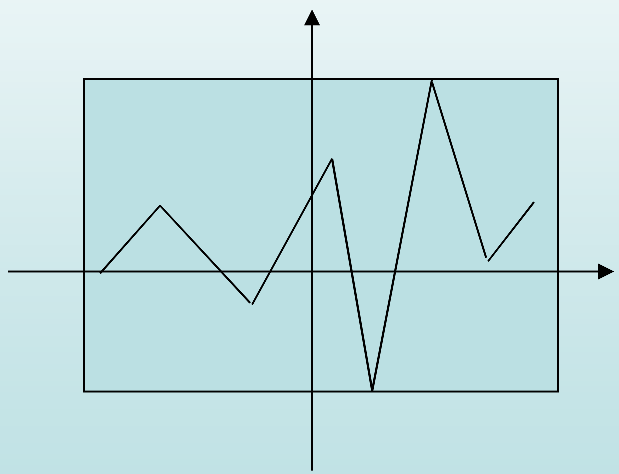
Histogram relativnih frekvencija

- Ukupna površina ispod grafika relativnih frekvencija je jednaka jedinici.
- Sličnost histograma relativnih frekvencija i gustine raspodele.

Predstavljanje podataka

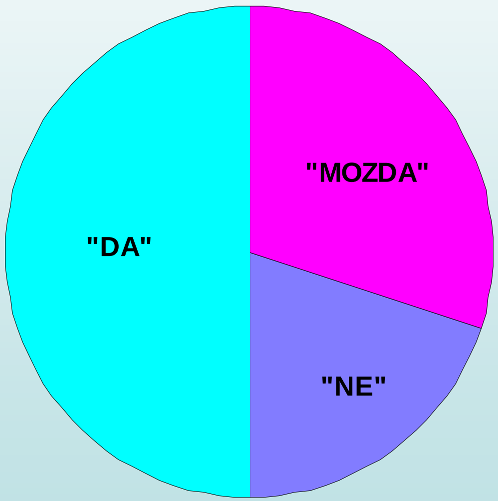
- Bitan je način predstavljanja podataka, jer može da dovede do pogrešnog zaključka, zbog moguće loše interpretacije (izbor merne jedinice).
- Trebalo bi da grafik bude predstavljen u odnosu 1:$\sqrt{2}$





Kružni dijagram

- Pomoću kružnih dijagrama se prikazuju odgovori ispitanika u anketama.
- Relativne frekvencije se predstavljaju kružnim isečcima odgovarajućih površina.



Empirijska funkcija raspodele

- Histogrami relativnih frekvencija su u vezi sa gustinom raspodele obeležja, a empirijska funkcija raspodele sa funkcijom rapodele obeležja.
- **Definicija. Neka je (X_1, ..., X_n) prost slučajan uzorak obima n za posmatrano obeležje. Funkcija**

$$F_n^* = \sum_{k=1}^{n} I[X_k < x]$$

I-Indikator događaja

je **empirijska funkcija raspodele**.

Empirijska funkcija raspodele obeležja X

- Neka je n_x broj elemenata uzorka za koje je vrednost obeležja X manja od realnog broja x. Tada se realizovana vrednost empirijske raspodele u tački x dobija po formuli:

$$F_n^*(x) = \frac{n_x}{n}$$

  i zove se empirijska funkcija raspodele obeležja X.

- Jednaka je relativnoj učestanosti događaja $[X < x]$.
- To je stepenasta funkcija koja uzima vrednosti sa [0,1], neopadajuća je za svako x, levo od najmanje varijante je nula, a desno od najveće varijante je jedan.

Empirijska funkcija raspodele

- Posmatrana kao sp, empirijska funkcija raspodele ima binomni zakon raspodele

$$P\left[F_n^*(x)=\frac{m}{n}\right]=\binom{n}{m}[F(x)]^m[1-F(x)]^{n-m} \quad (m=0,1,\dots,n)$$

$F(x)$ je teorijska funkcija raspodele za posmatrano obeležje.

- Iz Bernulijevog zakona sledi

$$F_n^*(x)\xrightarrow{p}F(x) \qquad n\to\infty$$

Centralna teorema mat. statistike

- Ako je $F(x)$ teorijska funkcija raspodele obeležja X, a $F_n^*(x)$ empirijska funkcija raspodele dobijena na osnovu prostog slučajnog uzorka obima n, tada, uniformno po x, funkcija $F_n^*(x)$ teži ka $F(x)$ sa verovatnoćom 1, tj.

$$P[\sup\left|F_n^*(x)-F(x)\right| \underset{n\to\infty}{\to} 0]=1$$

- Ova teorema nam daje opravdanje da zaključke koje donosimo o raspodeli obeležja na uzorku, prenesemo na celu populaciju, ako je obim uzorka dovoljno veliki.

Empirijska funkcija raspodele

- Primer. Rezultati bacanja tri numerisane kocke su dati u tabeli. Nacrtati empirijsku funkciju raspodele.

Redni broj bacanja	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Broj šestica	0	1	0	1	1	0	1	0	1	0	1	0	2	1	2	2

- Empirijska funkcija raspodele se dobija pomoću tabele

Broj šestica	0	1	2	3
Frekvencije	6	7	3	0

$$F_{16}^{*}=\begin{cases}0, & x<0\\ \frac{6}{16}, & 0\le x<1\\ \frac{13}{16}, & 1\le x<2\\ 1, & 2\le x\end{cases}$$