Linearna kombinacija *n* slučajnih varijabli

Za bilo koji skup slučajnih varijabli X_1, X_2,..., X_n i konstanti a_1, a_2,..., a_n, slučajna varijabla
$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n$
zove se **linearna kombinacija** X_i-ova.

Očekivanje linearne kombinacije

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n)$$

Varijanca linearne kombinacije

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) + \sum_i \sum_{j \neq i} a_i a_j Cov(X_i, X_j)$$

za nezavisne varijable:

$$V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n)$$

Slučajni uzorak

U mnogim statističkim problemima vrijednosti u uzorku od n elemenata $\{x_1, x_2, \ldots, x_n\}$ iz neke populacije mogu se zamisliti kao opažene vrijednosti niza od n jednakih slučajnih varijabli.
Neka su X_1, X_2,...,X_n takve nove slučajne varijable.

Def:
Neka su X_1, X_2,...,X_n slučajne varijable. Kažemo da one čine **slučajni uzorak** veličine n, ako su varijable X_i
(a) nezavisne i
(b) imaju iste raspodjele vjerojatnosti.

Kažemo da su X_i **nezavisne i identično raspodijeljene.**

primjena:
- mjerenje fizikalne veličine (populacija je beskonačna)
- uzimanje uzorka s povratom
- uzimanje uzorka iz velike populacije

Slučajni uzorak $\{X_1, X_2, \ldots, X_n\}$

= skup od n identičnih i nezavisnih slučajnih varijabli X_i

Sve X_i imaju istu raspodjelu vjerojatnosti pa dakle

isto očekivanje μ

istu varijancu σ^2.

Prosjek $\overline{X}$ slučajnog uzorka

= slučajna varijabla

= linearna kombinacija slučajnih varijabli X_i

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

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

Varijanca prosjeka $V(\overline{X}) = \frac{\sigma^2}{n}$ $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Total T_0 slučajnog uzorka

= slučajna varijabla

= linearna kombinacija slučajnih varijabli X_i

$$T_0 = \sum_{i=1}^{n} X_i$$

Očekivanje totala $E(T_0) = n\mu$

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

Raspodjela prosjeka uzorka

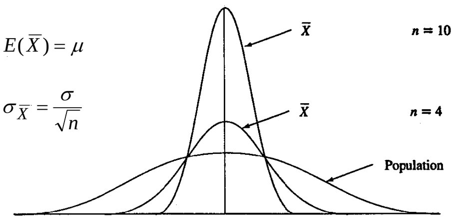
Neka je uzorak sastavljen od slučajnih varijabli proizvoljna oblika raspodjele!

Kakav će biti *oblik raspodjele* prosjeka (ili totala)?

Ako je uzorak sastavljen od normalnih slučajnih varijabli, onda je i prosjek normalno raspodijeljen

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

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



Raspodjela prosjeka i totala u općenitom slučaju

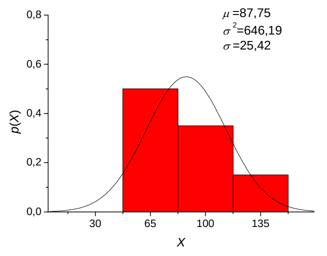
- Izračun pomoću pravila vjerojatnosti *ili*
- Simulacijski eksperiment

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



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

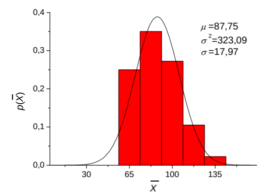
X_1=prihod od prvog kupca
X_2=prihod od drugog kupca

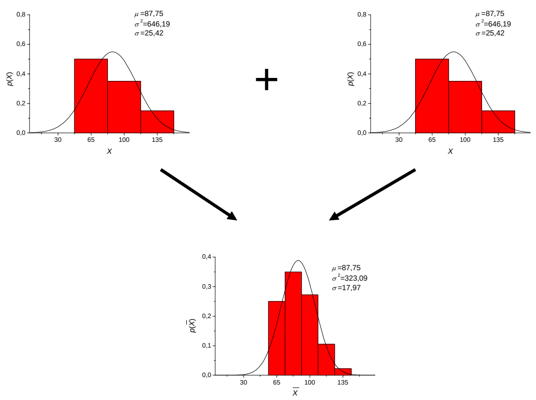
Tablica mogućih prihoda:

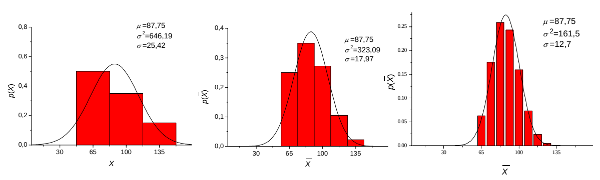
x_1	x_2	$p(x_1,x_2)$	x_1+x_2	$\bar{x}$
65	65	0,25	130	65
65	100	0,175	165	82,5
65	135	0,075	200	100
100	65	0,175	165	82,5
100	100	0,1225	200	100
100	135	0,0525	235	117,5
135	65	0,075	200	100
135	100	0,0525	235	117,5
135	135	0,0225	270	135

Slučajna varijabla $\bar{X}$ 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







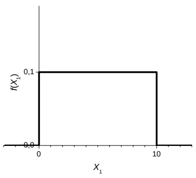
n = 1 n = 2 n = 4

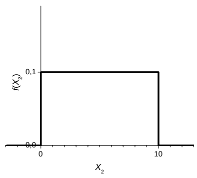
Primjer 2 (kontinuirana raspodjela):

Na putu do posla čekam autobus koji vozi svakih 10 minuta, a zatim tramvaj koji također vozi svakih 10 minuta.
Neka su slučajne varijable:

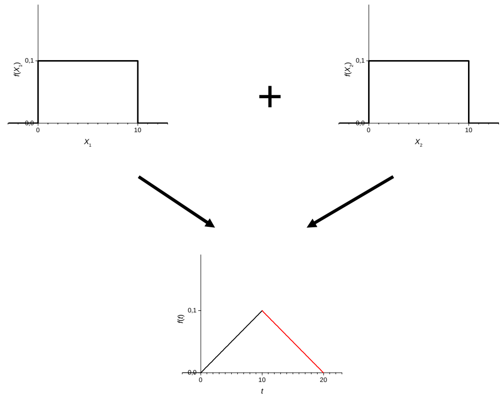
X_1 = vrijeme čekanja autobusa
X_2 = vrijeme čekanja tramvaja

$$f(x_1)=\begin{cases}1/10 & , \quad 0 \le x_1 \le 10 \\ 0 & , \quad x_1 < 0 \quad ili \quad x_1 > 10\end{cases} \qquad f(x_2)=\begin{cases}1/10 & , \quad 0 \le x_2 \le 10 \\ 0 & , \quad x_2 < 0 \quad ili \quad x_2 > 10\end{cases}$$





T_0 = ukupno vrijeme čekanja $=X_1+X_2$

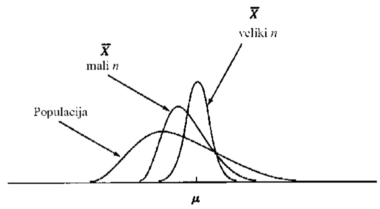


http://www.stat.sc.edu/~west/javahtml/CLT.html

Središnji granični teorem (CLT)

Neka je $X_1, X_2,\dots,X_n$ slučajni uzorak bilo koje raspodjele s očekivanjem μ i standardnom devijacijom σ. Ako je n dovoljno velik, $\overline{X}$ ima približno **normalnu** raspodjelu s očekivanjem $\mu_{\overline{X}} = \mu$ i standardnom devijacijom $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

T_0 ima također približno normalnu raspodjelu s očekivanjem $\mu_{T_0} = n\mu$ i standardnom devijacijom $\sigma_{T_0} = \sqrt{n}\cdot\sigma$.



Procjena parametara populacije na osnovi uzorka

- Populacija koja nas zanima ima karakteristične parametre μ, σ, σ^2, momenti, proporcije itd.
- Nismo u mogućnosti odrediti ih na cijelom skupu.
- Uzimamo uzorak
- Na osnovi uzorka želimo procijeniti parametre osnovnog skupa (populacije)

Neka je θ neki parametar raspodjele osnovnog skupa (μ, σ,...). Da bismo procjenili θ, uzimamo uzorak $\{X_1, X_2, \dots X_n\}$ i definiramo slučajnu varijablu $\hat{\theta}$ kao funkciju slučajnih varijabli $X_1, X_2, \dots X_n$.
$\hat{\theta}$ nazivamo **procjenjitelj** parametra θ.

Za procjenjitelj $\hat{\theta}$ kažemo da je **nepristrani procjenjitelj** parametra θ ako je njegovo očekivanje

$$E(\hat{\theta}) = \theta$$

Propozicija:
Ako je $\{X_1, X_2, \dots X_n\}$ slučajni uzorak neke populacije, onda je slučajna varijabla $\overline{X}$ nepristrani procjenjitelj očekivanja te populacije (μ).

Propozicija:
Ako je $\{X_1, X_2, \ldots X_n\}$ slučajni uzorak neke populacije, onda je slučajna varijabla

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$$

nepristrani procjenjitelj varijance osnovnog skupa σ^2.

Mogli smo za procjenjitelja odabrati $\hat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$

Ali ! $E(\hat{\sigma}_u^2) = \sigma^2 - \frac{\sigma^2}{n}$

Interval pouzdanosti prosjeka populacije

Najčešće želimo procijeniti μ populacije
(npr., pravu vrijednost mjerene veličine)

Uzimamo uzorak od n elemenata (velik n) i izračunamo njegov prosjek. Tako smo nepristrano procijenili μ.
Zanima nas koliko pouzdano, tj., u kojem intervalu se pouzdano nalazi prosjek populacije μ.

Neka je varijanca populacije σ^2.

Prema središnjem graničnom teoremu (CLT), $\overline{X}$ je normalno raspodijeljen:

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

Kad odaberemo uzorak, slučajna varijabla $\overline{X}$ poprimit će vrijednost $\overline{x}$.

Za normalnu raspodjelu vrijedi:

$$P\left(-1 \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le 1\right) = 68\% \qquad P\left(-1{,}96 \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le 1{,}96\right) = 95\% \qquad P\left(-2{,}575 \le \frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \le 2{,}575\right) = 99\%$$

To možemo pisati:
Prava vrijednost prosjeka populacije nalazi se u intervalu

$\mu \in \overline{x} \pm \frac{\sigma}{\sqrt{n}}$ s vjerojatnošću 68%.

Obično ne znamo varijancu osnovnog skupa pa ju moramo procijeniti.

Nepristrani procjenjitelj varijance je $\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$

Za procjenjitelja standardne devijacije najčešće uzimamo slučajnu varijablu

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

koja je vrlo bliska nepristranom procjenjitelju standardne devijacije.

Interval 68% pouzdanosti prosjeka je tada $\overline{x} \pm \frac{s}{\sqrt{n}} = \overline{x} \pm \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n(n-1)}}$

U fizici se dogovorno svi rezultati mjerenja pišu s navođenjem iznosa 68-postotne nepouzdanosti:

$$\overline{x} \pm M$$

gdje je $M = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n(n-1)}}$ **nepouzdanost** izračunata iz n nezavisnih mjerenja.

STANDARDNA POGREŠKA

Također se upotrebljava veličina

$$m = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}} = s$$

koja označava **preciznost** svakog pojedinog mjerenja.

i	x_i (ms)	$(x_i - \overline{x})^2$ (ms²)
1	584	14.5785
2	588	0.03306
3	591	10.12398
4	587	0.66942
5	590	4.76034
6	585	7.94214
7	585	7.94214
8	592	17.48762
9	588	0.03306
10	589	1.3967
11	587	0.66942

$\sum x_i = 6466\,\text{ms}$

$\overline{x} = \frac{1}{11}\sum x_i = 587.8182\,\text{ms}$

$\sum (x_i - \overline{x})^2 = 65.6364\,\text{ms}^2$

$m = s = \sqrt{\frac{\sum (x_i - \overline{x})^2}{11-1}} = 2.56196\,\text{ms}$

$M = \frac{m}{\sqrt{11}} = 0.7725\,\text{ms}$

Pisanje rezultata mjerenja

$$x = (587.8 \pm 0.8)\,\text{ms}$$

relativna pogreška: $R = \frac{M}{\overline{x}} = 0.14\%$