Očekivanja

$$\mu_X = \sum_{x\in D_X} x\cdot p_X(x) = \sum_{x\in D_X}\sum_{y\in D_Y} x\cdot p(x,y) \qquad ili$$

$$\mu_X = \int_{-\infty}^{\infty} x\cdot f_X(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\cdot f(x,y)dxdy$$

$$\mu_Y = \sum_{y\in D_Y} y\cdot p_Y(y) = \sum_{x\in D_X}\sum_{y\in D_Y} y\cdot p(x,y) \qquad ili$$

$$\mu_Y = \int_{-\infty}^{\infty} y\cdot f_Y(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\cdot f(x,y)dxdy$$

Momenti 2D raspodjele

Pomoćni momenti

$$m_{rs} = \sum_{x\in D_X}\sum_{y\in D_Y} x^r y^s p(x,y) \qquad ili$$

$$m_{rs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^r y^s f(x,y)dxdy$$

Središnji momenti

$$M_{rs} = \sum_{x\in D_X}\sum_{y\in D_Y} (x-\mu_X)^r (y-\mu_Y)^s p(x,y) \qquad ili$$

$$M_{rs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^r (y-\mu_Y)^s f(x,y)dxdy$$

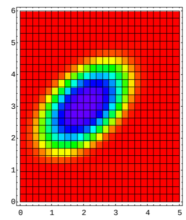
$m_{10} = \mu_X$

$m_{01} = \mu_Y$ očekivanja

$M_{20} = V(X) = \sigma_X^2$

$M_{02} = V(Y) = \sigma_Y^2$ varijance

$M_{11} = \sigma_{XY} = Cov(X,Y)$ kovarijanca



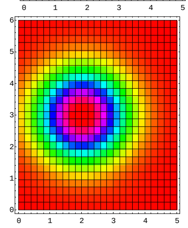
$m_{10} = 2 = \mu_X$

$m_{01} = 3 = \mu_Y$ očekivanja

$M_{20} = V(X) = \frac{4}{3}$

$M_{02} = V(Y) = \frac{4}{3}$ varijance

$M_{11} = \sigma_{XY} = \frac{2}{3}$ kovarijanca



$m_{10} = 2 = \mu_X$

$m_{01} = 3 = \mu_Y$ očekivanja

$M_{20} = V(X) = 1$

$M_{02} = V(Y) = 1$ varijance

$M_{11} = \sigma_{XY} = 0$ kovarijanca

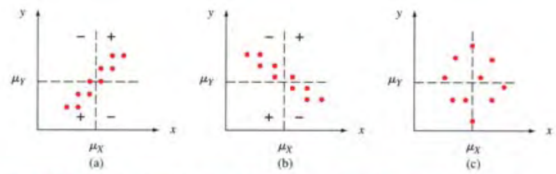


Figure 5.4 $p(x, y)$ = 1/10 for each of ten pairs corresponding to indicated points; (a) positive covariance; (b) negative covariance; (c) covariance near zero

Pomoćna formula

$$Cov(X,Y) = m_{11} - m_{10}\cdot m_{01} = E(X\cdot Y) - E(X)E(Y)$$

Korelacija

koeficijent korelacije: $$Corr(X,Y) = \rho_{X,Y} = \rho = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

Svojstva:

1. $Corr(aX+b, cY+d) = Corr(X,Y)$
2. $-1 \le \rho \le 1$
3. Ako su X i Y nezavisne varijable, onda je ρ=0. (Obrat ne vrijedi)
4. Ako je $Y=aX+b$ onda i samo onda je ρ=1 ili ρ= -1.

Linearna kombinacija dviju slučajnih varijabli

2D raspodjela
Slučajne varijable X_1 i X_2 koje se mogu zbrajati
Združena raspodjela vjerojatnosti $p(X_1,X_2)$

$a_1, a_2 \in \mathbf{R}$

Def: $Y = a_1X_1 + a_2X_2$ (linearna kombinacija)

Očekivanje linearne kombinacije

$$E(Y) = a_1E(X_1) + a_2E(X_2)$$

Varijanca linearne kombinacije

$$V(Y) = a_1^2V(X_1) + a_2^2V(X_2) + 2a_1a_2Cov(X_1, X_2)$$

Varijanca razlike nezavisnih varijabli

$$V(X_1 - X_2) = V(X_1) + V(X_2) = \sigma_1^2 + \sigma_2^2$$

Varijanca razlike je zbroj varijanci, a ne razlika!

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
$Y = a_1X_1 + a_2X_2 + ... + a_nX_n$
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

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

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 $\{x_1, x_2, ..., x_n\}$ 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

Prosjek $\bar{X}$ slučajnog uzorka

= slučajna varijabla
= linearna kombinacija slučajnih varijabli X_i

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

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

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

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

Raspodjela prosjeka uzorka

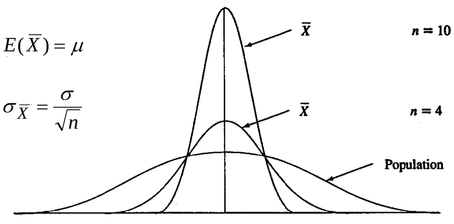
Neka je uzorak sastavljen od slučajnih varijabli proizvoljna oblika raspodjele!

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

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

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

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



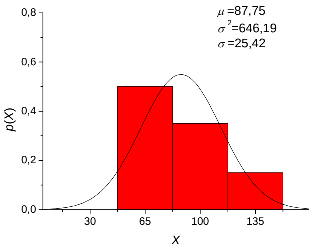
Raspodjela prosjeka i totala u općenitom slučaju

- 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



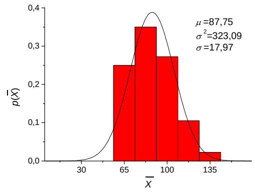
Određenog dana najavila su se dva kupca. Neka su slučajne varijable:
X_1=prihod od prvog kupca
X_2=prihod od drugog kupca

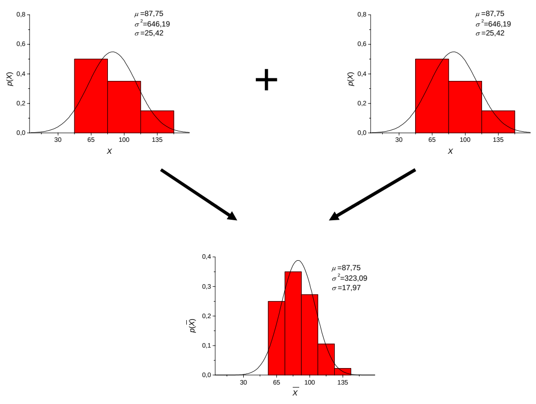
Tablica mogućih prihoda:

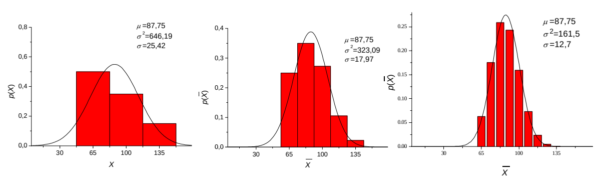
x_1	x_2	$p(x_1,x_2)$	x_1+x_2	$\bar{x}$
65	65	0,25	130	65
65	100	0,175	165	82,5
65	135	0,075	200	100
100	65	0,175	165	82,5
100	100	0,1225	200	100
100	135	0,0525	235	117,5
135	65	0,075	200	100
135	100	0,0525	235	117,5
135	135	0,0225	270	135

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







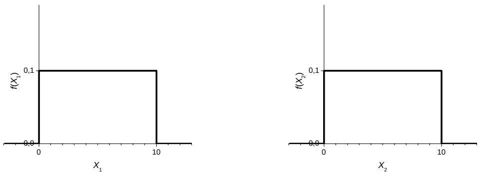
$n = 1$ $n = 2$ $n = 4$

Primjer 2 (kontinuirana raspodjela):

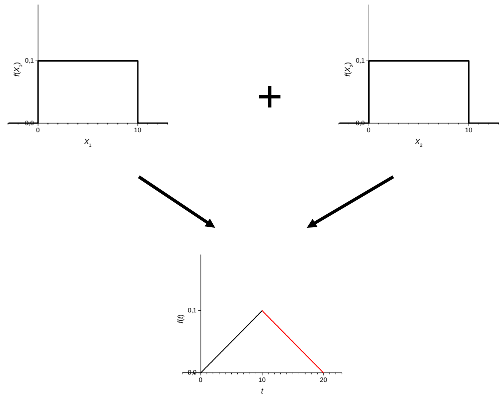
Na putu do posla čekam autobus koji vozi svakih 10 minuta, a zatim tramvaj koji također vozi svakih 10 minuta.
Neka su slučajne varijable:

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 $X_1, X_2,...,X_n$ 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 $\sigma_{T_0}=\sqrt{n}\cdot\sigma$.