

## 3. Binomna slučajna varijabla

### 3.1. Osnovno o binomnoj slučajnoj varijabli

Binomnu slučajnu varijablu dobivamo kad  $n$  puta ponavljamo isti pokus, kojem je vjerojatnost uspjeha  $p$ , a neuspjeha  $q = 1 - p$ .

Promatrajmo pokus koji izvodimo i označimo s  $U$  uspjeh u izvođenju, a s  $N$  neuspjeh. Označimo s  $X$  broj uspješno izvedenih pokusa. U literaturi se takva slučajna varijabla zove indikatorska ili Bernoullijeva. Imamo

$\Omega$	Pr	$X$
$U$	$p$	1
$N$	$q$	0

Za slučajnu varijablu  $X$  je

$$E(X) = p,$$

$$V(X) = (1 - p)^2 p + (0 - p)^2 q = pq^2 + p^2 q = pq(p + q) = pq.$$

Označimo s  $B_{np}$  slučajnu varijablu koja mjeri broj uspjeha u  $n$  ponavljanja prethodnog pokusa. Primijetimo da se slučajna varijabla  $B_{np}$  sastoji od zbroja  $n$  istih slučajnih varijabli  $X$ , tj. da je

$$B_{np} = X_1 + X_2 + \dots + X_n, \quad X_i = X \quad \text{za } i = 1, \dots, n.$$

Koristimo li znanje kako se očekivanje i varijanca ponašaju obzirom na zbrajanje slučajnih varijabli (vidjeti prethodno predavanje), imamo

$$E(B_{np}) = np,$$

$$V(B_{np}) = npq, \quad \sigma(B_{np}) = \sqrt{npq}.$$

## 3.2. Funkcija vjerojatnosti binomne slučajne varijable $B_{np}$

Svaki ishod ponavljanja  $n$  pokusa ima oblik

$$\underbrace{UUNUNNNU \dots NNU}_{n \text{ puta}}.$$

Pretpostavimo da se u prethodnom nizu uspjeh pojavio točno  $k$  puta. Onda se neuspjeh pojavio točno  $n - k$  puta. Koliko ima različitih mogućnosti da se uspjeh pojavi točno  $k$  puta?

Moramo izabrati mjesta na kojima ćemo staviti uspješan pokus. Drugim riječima, moramo izabrati  $k$ -člani podskup  $n$ -članog skupa. To se može napraviti na  $\binom{n}{k}$  načina. Pritom je

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Ako je  $k = 0$ , onda je po definiciji  $0! = 1$ , pa je i  $\binom{n}{k} = 1$ .

Primijetimo još jednu stvar, umjesto mjesta gdje smo stavljali uspješne pokuse, mogli smo gledati i sve moguće načine na koje ćemo postaviti neuspješne pokuse. Ako biramo mjesta na koja ćemo postaviti  $n - k$  neuspješnih pokusa, onda to možemo obaviti na  $\binom{n}{n-k}$  načina. Primijetite da je to isto kao da smo birali mjesta gdje ćemo staviti  $k$  uspješnih pokusa, pa mora biti

$$\binom{n}{k} = \binom{n}{n-k}.$$

Puno bolji način računanja faktoriijela (pokazat će se u kompleksnoj analizi i

onaj koji se može proširiti na negativne  $n$ ) je da brojnik i nazivnik skratimo, tj.

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k!}.$$

Prema tome, ukupna vjerojatnost da se uspjeh dogodi točno  $k$  puta je

$$\Pr(B_{np} = k) = \binom{n}{k} p^k q^{n-k}.$$

**Primjer 3.2.1.** *Nađite vjerojatnost da se u 4 bacanja kocke šestice pojave točno dva puta.*

Očito se radi o binomnoj slučajnoj varijabli kojoj je  $n = 4$ , a  $p = \frac{1}{6}$  i  $k = 2$ .  
Prema tome, imamo

$$\Pr(B_{4, \frac{1}{6}} = 2) = \binom{4}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = \frac{4 \cdot 3}{1 \cdot 2} \cdot \frac{1}{36} \cdot \frac{25}{36} = \frac{150}{1296}.$$

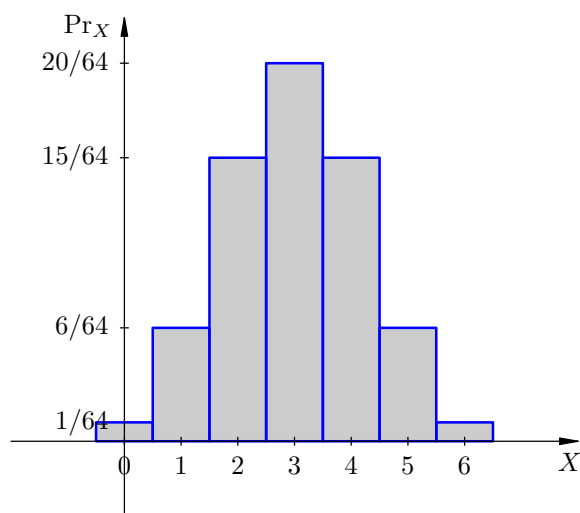
**Primjer 3.2.2.** *Nađite broj glava u šest bacanja novčića i njihove vjerojatnosti.*

Označimo s  $X$  broj glava u šest bacanja novčića. Očito se radi o binomnoj

slučajnoj varijabli kojoj je  $n = 6$ , a  $p = \frac{1}{2}$ . Prema tome, imamo

$X$	Pr
0	$\binom{6}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^6 = \binom{6}{0} \left(\frac{1}{2}\right)^6 = \frac{1}{64}$
1	$\binom{6}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^5 = \binom{6}{1} \left(\frac{1}{2}\right)^6 = \frac{6}{64}$
2	$\binom{6}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^4 = \binom{6}{2} \left(\frac{1}{2}\right)^6 = \frac{15}{64}$
3	$\binom{6}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^3 = \binom{6}{3} \left(\frac{1}{2}\right)^6 = \frac{20}{64}$
4	$\binom{6}{4} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = \binom{6}{4} \left(\frac{1}{2}\right)^6 = \frac{15}{64}$
5	$\binom{6}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^1 = \binom{6}{5} \left(\frac{1}{2}\right)^6 = \frac{6}{64}$
6	$\binom{6}{6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^0 = \binom{6}{6} \left(\frac{1}{2}\right)^6 = \frac{1}{64}$

Grafički to izgleda ovako:



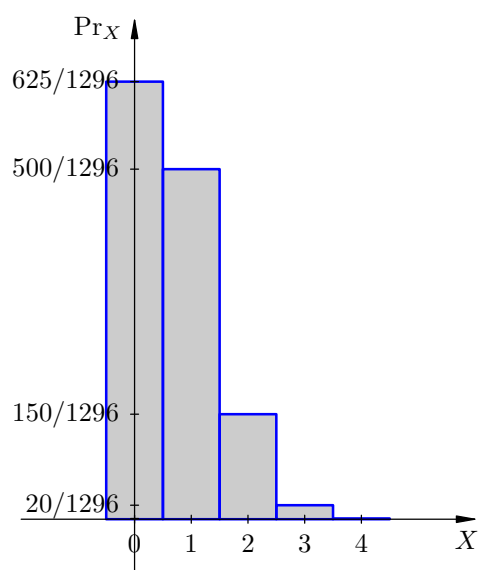
**Primjer 3.2.3.** Nađite broj šestica palih u 4 bacanja kocke, te očekivanje te slučajne varijable.

Označimo s  $X$  broj šestica u 4 bacanje kocke. Očito se radi o binomnoj

slučajnoj varijabli kojoj je  $n = 4$ , a  $p = \frac{1}{6}$ . Prema tome, imamo

$X$	Pr
0	$\binom{4}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^4 = \frac{625}{1296}$
1	$\binom{4}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^3 = \frac{500}{1296}$
2	$\binom{4}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = \frac{150}{1296}$
3	$\binom{4}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^1 = \frac{20}{1296}$
4	$\binom{4}{4} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^0 = \frac{1}{1296}$

Grafički to izgleda ovako:



Sada je lako pročitati da je

$$\Pr(\text{pala bar jedna šestica}) = 1 - \Pr(X = 0) = 1 - \frac{625}{1296} = \frac{671}{1296} < \frac{1}{2}.$$

Za očekivanje vrijedi:

$$E(X) = \frac{0 \cdot 625 + 1 \cdot 500 + 2 \cdot 150 + 3 \cdot 20 + 4 \cdot 1}{1296} = \frac{864}{1296}.$$



## 4. Normalna slučajna varijabla

### 4.1. Osnovno o normalnoj slučajnoj varijabli

U 18. stoljeću je De Moivre korištenjem novootkrivenog diferencijalnog i integralnog računa pokazao da se binomna slučajna varijabla dobro aproksimira jednom neprekidnom slučajnom varijablom i da je ta aproksimacija bolja što je broj ponavljanja pokusa veći. Ta nova slučajna varijabla je neprekidna slučajna varijabla i zove se normalna slučajna varijabla. Dakle,

$$B_{np} \approx N(\mu, \sigma), \quad \mu = np, \quad \sigma = \sqrt{npq}.$$

Funkcija gustoće te normalne slučajne varijable dana je s

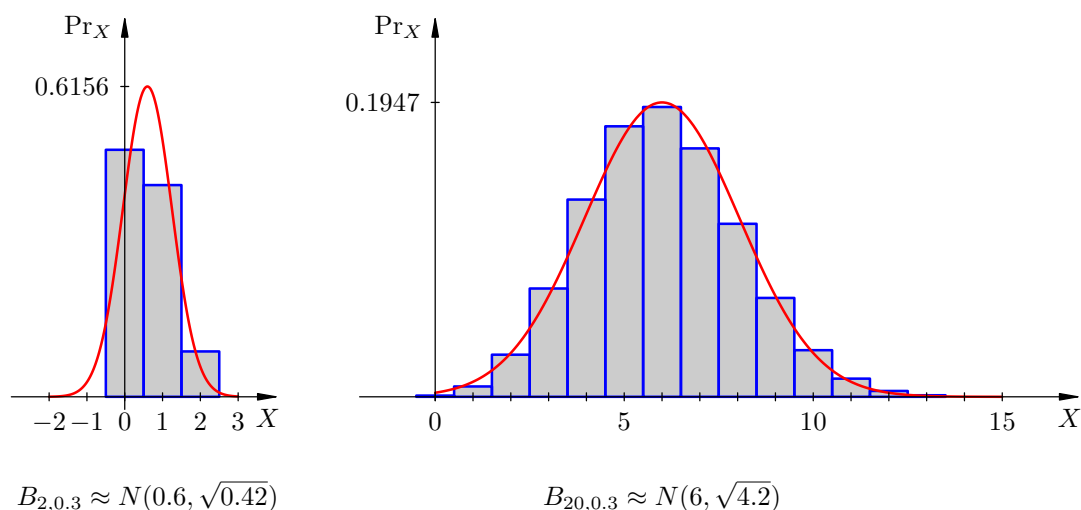
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Često se eksponencijalna funkcija  $e^x$  označava se kao  $\exp(x)$  da se komplicirani eksponenti lakše čitaju. U toj oznaci, prethodnu gustoću možemo zapisati kao

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

Usporedimo što se događa s porastom  $n$  pri aproksimaciji binomne varijable

normalnom i to na primjeru kad se  $n$  promijeni s 2 na 20.



Ova aproksimacija posljedica je centralnog graničnog teorema, koji u “slobodnoj interpretaciji” kaže

**Teorem 4.1.1. (Centralni granični teorem)** *Zbroj velikog broja jednakih nezavisnih slučajnih varijabli jednak je normalnoj slučajnoj varijabli  $N(\mu, \sigma)$ .*

Ako dobroj primjećujete, nigdje se ne govori samo o binomnoj razdiobi, jer teorem vrijedi i šire.

## 4.2. Tablice za normalnu slučajnu varijablu

Dosad smo vidjeli da je gustoća normalne slučajne varijable komplicirana eksponencijalna funkcija. Želimo li izračunati vjerojatnost da je normalna slučajna varijabla  $N(\mu, \sigma)$  manja ili jednaka nekom broju  $a$ , onda ćemo morati poznavati funkciju distribucije slučajne varijable  $N$ , tj. morat ćemo računati integral

$$\Pr(X \leq a) = \int_{-\infty}^a f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx,$$

što je netrivialni integral (ne postoji analitička formula za njegov zapis), iako se on može numerički izračunati.

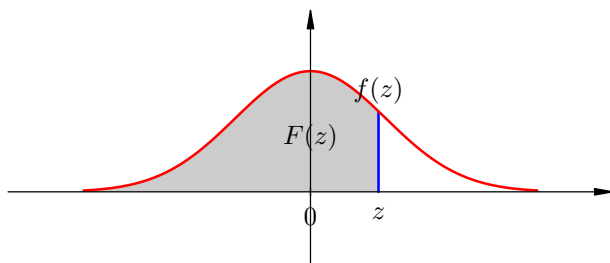
Primijetite da prethodni integral ovisi o tri parametra,  $a$ ,  $\mu$  i  $\sigma$ , pa bi tablice za to bile komplicirane. Umjesto toga, tabelira se samo funkcija distribucije normalne slučajne varijable  $N(0, 1)$ , a sve ostale normalne slučajne varijable jednostavnom

transformacijom svode se na tu normalnu slučajnu varijablu.

Normalna slučajna varijabla  $N(0, 1)$  ima funkciju gustoće

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

Primijetite da smo u prethodnoj formuli varijablu pisali slovom  $z$ , što će nam kasnije koristiti kod zamjene varijabli.



Želimo li izračunati

$$\Pr(a \leq Z \leq b) = \int_a^b f(z) dz = F(b) - F(a),$$

pri čemu veličine  $F(a)$  i  $F(b)$  nalazimo u tablicama.

Bilo koja normalna slučajna varijabla  $N(\mu, \sigma)$  transformira se na standardnu normalnu varijablu  $N(0, 1)$  tzv.  $Z$ -transformacijom

$$Z = \frac{X - \mu}{\sigma}.$$

**Primjer 4.2.1.** Neka je  $X \sim N(150, 20)$ . Koliko je  $\Pr(100 \leq X \leq 170)$ ?

Prevedimo ovu normalnu slučajnu varijablu na  $N(0, 1)$  i rezultat očitajmo iz tablica:

$$\begin{aligned} \Pr(100 \leq X \leq 170) &= \Pr\left(\frac{100 - 150}{20} \leq \frac{X - \mu}{\sigma} \leq \frac{170 - 150}{20}\right) \\ &= \Pr(-2.5 \leq Z \leq 1) = F(1) - F(-2.5) \\ &= 0.841 - 0.006 = 0.835. \end{aligned}$$

Možemo očitavati i iz točnije napisanih tablica, samo moramo paziti jer su one integrali od nule, a ne od  $-\infty$ , tj. tabelirano je  $F(a) - F(0)$ ! Dakle, imamo

$$\begin{aligned} F(1) &= 0.5 + 0.34134 = 0.84134 \\ F(-2.5) &= 0.5 - F(2.5) = 0.5 - 0.49379 = 0.00621, \end{aligned}$$

pa je  $F(b) - F(a) = 0.84134 - 0.00621 = 0.83513$ .

**Primjer 4.2.2.** Neka je  $X$  broj glava u 100 bacanja novčića. Nađite

$$\Pr(X > 60).$$

Točna vrijednost bila bi

$$\Pr(X > 60) = \binom{100}{61} \left(\frac{1}{2}\right)^{100} + \binom{100}{62} \left(\frac{1}{2}\right)^{100} + \dots + \binom{100}{100} \left(\frac{1}{2}\right)^{100},$$

što je poprilično dugotrajan posao.

Želimo li to približno izračunati, onda je

$$X = B_{100, \frac{1}{2}} \approx N(np, \sqrt{npq}) = N(50, 5),$$

pa je

$$\begin{aligned} \Pr(X > 60) &\approx \Pr(N > 60) = \Pr\left(\frac{N - \mu}{\sigma} > \frac{60 - 50}{5}\right) \\ &= \Pr(Z > 2) = 1 - F(2) = 1 - 0.977 = 0.023. \end{aligned}$$

(Po preciznijim tablicama  $F(2) = 0.5 + 0.47725 = 0.97725$ .)

*Bolje* je uzeti malo drugačiju aproksimaciju binomne varijable normalnom, tako je je  $X = 61$  jednako vjerojatan kao i svi drugi  $X$ -ovi

$$\begin{aligned} \Pr(X > 60) &\approx \Pr(N > 60.5) = \Pr\left(\frac{N - \mu}{\sigma} > \frac{60.5 - 50}{5}\right) \\ &= \Pr(Z > 2.1) = 1 - F(2.1) = 1 - 0.98213 = 0.01787. \end{aligned}$$

Na sličan način možemo izračunati i

$$\begin{aligned} \Pr(X = 60) &\approx \Pr(59.5 < N < 60.5) = \Pr\left(\frac{59.5 - 50}{5} < \frac{N - \mu}{\sigma} < \frac{60.5 - 50}{5}\right) \\ &= \Pr(1.9 < Z < 2.1) = F(2.1) - F(1.9) \\ &= 0.98213 - 0.97128 = 0.01085. \end{aligned}$$

Za normalnu varijablu  $N(\mu, \sigma)$ , katkad se traže sve vrijednosti  $x$  za koje je

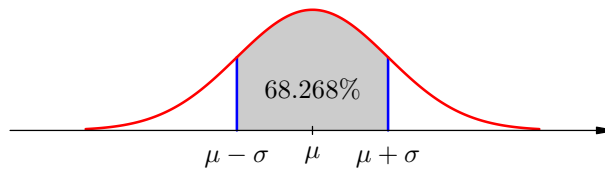
$$\Pr(|X - \mu| < k\sigma), \quad k = 1, 2, 3.$$

Prevedimo problem na standardnu normalnu razdiobu:

$$\begin{aligned} \Pr(|X - \mu| < k\sigma) &= \Pr(-k\sigma < X - \mu < k\sigma) = \Pr\left(-k < \frac{X - \mu}{\sigma} < k\right) \\ &= F(k) - F(-k) = 2(F(k) - F(0)). \end{aligned}$$

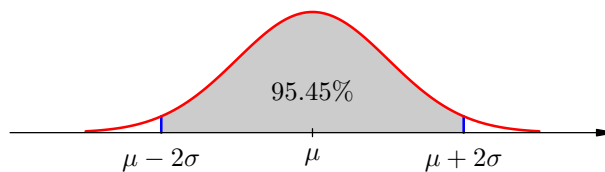
Za  $k = 1$  imamo

$$\Pr(|X - \mu| < \sigma) = 2 \cdot 0.34134 = 0.68268.$$



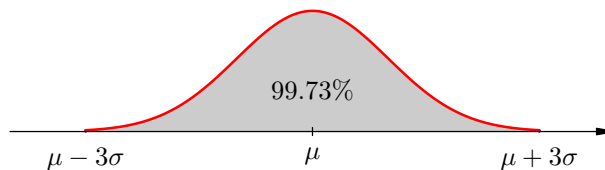
Za  $k = 2$  imamo

$$\Pr(|X - \mu| < 2\sigma) = 2 \cdot 0.47725 = 0.9545.$$



Za  $k = 3$  imamo

$$\Pr(|X - \mu| < 3\sigma) = 2 \cdot 0.49865 = 0.9973.$$



Ova posljednja činjenica često se naziva i pravilo  $3\sigma$ , jer vrlo visokom vjerojatnošću normalna razdioba ima vrijednosti u rasponu  $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$ .

## 5. Neki dodatni izvodi i pojmovi

### 5.1. Varijanca, kovarijanca, korelacija

U eksperimentima možemo promatrati i više od jedne slučajne varijable. Recimo, za osobu te slučajne varijable mogu biti  $X =$  visina osobe,  $Y =$  masa osobe. Naravno i  $X$  i  $Y$  distribuirane su prema nekom zakonu razdiobe. Ako želimo znati u kojoj su vezi visina i težina kao slučajne varijable, trebat će nam neki novi pojmovi, kao što su to kovarijanca i korelacija.

Naravno, možemo promatrati i više od dvije slučajne varijable. Nazovimo tada slučajne varijable  $X_1, X_2, \dots, X_n$ .

Prvo, za slučajnu varijablu  $X$  izvedimo vezu njezine varijance i očekivanja.

$$\begin{aligned}\text{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) = E(X^2) - (E(X))^2.\end{aligned}$$

Drugim riječima, prebacivanjem članova s jedne strane na drugu, ako varijancu označimo sa  $\sigma_X$ , a očekivanje s  $\mu_X$ , dobivamo:

$$E(X^2) = \mu_X^2 + \sigma_X^2.$$

Ako imamo dvije slučajne varijable  $X$  i  $Y$ , onda možemo definirati kovarijancu

s:

$$\text{Cov}(X, Y) = E((X - E(X)) \cdot (Y - E(Y))) = E(XY) - E(X)E(Y),$$

i korelaciju s

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}.$$

Iz te dvije relacije, kombiniranjem izlazi

$$E(XY) = \mu_X \mu_Y + \rho(X, Y) \sigma_X \sigma_Y.$$

Zako bismo imali više od dvije slučajne varijable  $X_1, \dots, X_n$ , onda kovarijanca i korelacija nisu skalari, nego matrice

$$\text{Cov}(X_1, \dots, X_n) = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix},$$

$$\rho(X_1, \dots, X_n) = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{nn} \end{bmatrix},$$

pri čemu je

$$c_{ij} = \text{Cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j),$$

$$r_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i) \text{Var}(X_j)}}.$$

Primijetite, da ako su slučajne varijable  $X_i$  i  $X_j$  nezavisne, onda je odgovarajući  $c_{ij}$  i  $r_{ij}$  jednak 0.

Prirodno se koeficijent korelacije pojavljuje u sljedećem problemu:

Za dane slučajne varijable  $X$  i  $Y$  treba naći najbolju aproksimaciju slučajne varijable  $Y$  varijablom  $Z$  oblika  $Z = \alpha X + \beta$ .

Po definiciji, najbolja aproksimacija slučajne varijable  $Y$  nekom familijom

slučajnih varijabli  $g(X)$  koje ovise o  $X$ , je onaj  $g^*(X)$  za koji vrijedi da je

$$\Delta^* = E((Y - g^*(X))^2)$$

minimalan mogući među svim funkcijama u familiji  $g(X)$ . Broj  $\Delta^*$  zovemo i srednje-kvadratna greška.

Pogledajmo, koliko je to za familiju  $g(X) = \alpha X + \beta$ . Iskoristimo pritom formule za  $E(X^2)$  i  $E(XY)$ .

$$E((Y - (\alpha X + \beta))^2) = (E(Y - (\alpha X + \beta)))^2 + \text{Var}(Y - \alpha X - \beta).$$

Raspišimo sad drugi član

$$\begin{aligned} \text{Var}(Y - \alpha X - \beta) &= \text{Var}(Y - \alpha X) \\ &= E((Y - \alpha X)^2) - (E(Y - \alpha X))^2 \\ &= E(Y^2 - 2\alpha XY + \alpha^2 X^2) - (\mu_Y - \alpha\mu_X)^2 \\ &= \mu_Y^2 + \sigma_Y^2 - 2\alpha(\mu_X\mu_Y + \rho(X, Y)\sigma_X\sigma_Y) \\ &\quad + \alpha^2(\mu_X^2 + \sigma_X^2) - (\mu_Y - \alpha\mu_X)^2 \\ &= \sigma_Y^2 - 2\alpha\rho(X, Y)\sigma_X\sigma_Y + \alpha^2\sigma_X^2 \\ &= (\alpha\sigma_X - \rho(X, Y)\sigma_Y)^2 + \sigma_Y^2(1 - \rho^2(X, Y)). \end{aligned}$$

Sada imamo

$$\begin{aligned} f(\alpha, \beta) &= E((Y - (\alpha X + \beta))^2) \\ &= (\mu_Y - \alpha\mu_X - \beta)^2 + (\alpha\sigma_X - \rho(X, Y)\sigma_Y)^2 + \sigma_Y^2(1 - \rho^2(X, Y)). \end{aligned}$$

Nađimo sada minimum funkcije dvije varijable  $\alpha$  i  $\beta$ .

$$\begin{aligned} \frac{\partial f(\alpha, \beta)}{\partial \alpha} &= -2(\mu_Y - \alpha\mu_X - \beta)\mu_X + 2(\alpha\sigma_X - \rho(X, Y)\sigma_Y)\sigma_X = 0, \\ \frac{\partial f(\alpha, \beta)}{\partial \beta} &= -2(\mu_Y - \alpha\mu_X - \beta) = 0. \end{aligned}$$

Iz druge jednadžbe odmah izlazi da je

$$\beta = \mu_Y - \alpha\mu_X,$$

pa uvrštavanjem u prvu imamo da mora biti

$$(\alpha\sigma_X - \rho(X, Y)\sigma_Y)\sigma_X = 0,$$

odakle odmah slijedi da je

$$\alpha = \rho(X, Y)\frac{\sigma_Y}{\sigma_X}.$$

Dakle

$$g^*(X) = \rho(X, Y) \frac{\sigma_Y}{\sigma_X} X + \mu_Y - \rho(X, Y) \frac{\sigma_Y}{\sigma_X} \mu_X,$$

a srednje kvadratna greška je

$$\Delta^* = E((Y - g^*(X))^2) = \sigma_Y^2(1 - \rho(X, Y)^2).$$

Primijetite da je za  $\rho(X, Y) = \pm 1$   $\Delta^* = 0$ , pa je  $Y = \alpha X + \beta$  s vjerojatnošću 1.

Pravac  $\alpha X + \beta$  katkad zovemo pravac regresije varijable  $Y$  prema varijabli  $X$ .

**Primjer 5.1.1.** *Nađite najbolju konstantnu aproksimaciju slučajne varijable  $Y$ .*

*Tražimo  $\alpha$  takav da je*

$$\begin{aligned} E(Y - \alpha)^2 &= E(Y^2 - 2\alpha Y + \alpha^2) \\ &= \mu_Y^2 + \sigma_Y^2 - 2\alpha\mu_Y + \alpha^2 \\ &= \sigma_Y^2 + (\mu_Y - \alpha)^2. \end{aligned}$$

*Očit je minimum kad je drugi član jednak 0, tj. kada je*

$$\alpha = \mu_Y = E(Y).$$

## 5.2. Čebiševljeva nejednakost

Neka je  $X$  slučajna varijabla,  $\mu$  njeno očekivane, a  $\varepsilon$  proizvoljni realni broj,  $\varepsilon > 0$ . Tada vrijedi

$$\Pr(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Prethodna nejednakost zove se Čebiševljeva nejednakost.

Ona se lako dokazuje. Označimo s  $D$  događaj

$$D = \{\omega : |X(\omega) - \mu| \geq \varepsilon\},$$

pri čemu je  $\omega$  elementarni ishod, a  $X$  slučajna varijabla na  $\Omega$ . Kvadriranjem (i korištenjem indicatorske varijable  $I_D$ ) dobivamo

$$(X - \mu)^2 \geq \varepsilon^2 I_D, \quad I_D = \begin{cases} 1, & \omega \in D, \\ 0, & \omega \notin D. \end{cases}$$

Onda je

$$\text{Var}(X) = E((X - \mu)^2) \geq E(\varepsilon^2 I_D) = \varepsilon^2 E(I_D) = \varepsilon^2 \Pr(D) = \varepsilon^2 \Pr(|X(\omega) - \mu| \geq \varepsilon).$$

Dijeljenjem dobivamo traženu nejednakost.

### 5.3. Slabi zakon velikih brojeva

Neka je zadan niz jednakih, nezavisnih slučajnih varijabli  $X_1, X_2, \dots, X_n$ . Definirajmo

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Korištenjem formule za očekivanje i varijancu nezavisnih slučajnih varijabli, dobivamo

$$\mu_{\bar{X}} = \mu\sigma_{\bar{X}}^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

pri čemu su  $\mu$  i  $\sigma$  očekivanje, odnosno standardna devijacija svake od slučajnih varijabli  $X_i$ . Sada možemo korištenjem Čebiševljeve nejednakosti lako pokazati da je

$$\Pr(|\bar{X} - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

Ako pustimo  $n \rightarrow \infty$ , onda

$$\Pr(|\bar{X} - \mu| > \varepsilon) \rightarrow 0.$$

Prethodna činjenica u teoriji vjerojatnosti zove se još i slabi zakon velikih brojeva.

Neka su sada  $Y_n$  binomna slučajna varijabla,  $Y_n = B_{n,p}$ , koja predstavlja broj uspjeha u  $n$  slučajnih pokusa. Označimo s

$$\bar{X} = \frac{Y_n}{n}$$

relativnu frekvenciju uspjeha u  $n$  jednostavnih pokusa

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

Slabi zakon velikih brojeva onda kaže da za  $n \rightarrow \infty$  vrijedi

$$\lim_{n \rightarrow \infty} \bar{X} = p,$$

tj. za velike  $n$  se relativna frekvencija uspjeha i  $p$  razlikuju za manje od  $\varepsilon$ .