3.4 Binomni model

Neka je X indikatorska varijabla $\rightarrow$ indikator:

- nekog događaja (primjer 3.2)
- neke klase ili neke vrijednosti kategorijalne varijable (primjer 3.1)

Populacijska razdioba od X:

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

Interpretacija parametra p?

Neka je

$$X_1, X_2, \ldots, X_n$$

slučajni uzorak za X (Bernoullijeva).

Tada:

$$Y := X_1 + X_2 + \cdots + X_n \sim B(n, p),$$

$$\mathsf{E}Y = np, \;\; \mathsf{Var}Y = np(1-p)$$

Interpretacija: Y je frekvencija, a Y/n relativna frekvencija događaja/pripadnika klase u uzorku (duljine n).

Zapis pomoću funkcije gustoće:

$$
\begin{aligned}
f_{X_i}(x) &:= \mathsf{P}(X_i = x) = \\
&= p^x(1-p)^{1-x} \cdot \mathbb{1}_{\{0,1\}}(x),\ x \in \mathsf{R}; \\
f_Y(x) &:= \mathsf{P}(Y = x) = \\
&= \binom{n}{x} p^x(1-p)^{n-x} \cdot \mathbb{1}_{\{0,1,\dots,n\}}(x),\ x \in \mathsf{R}.
\end{aligned}
$$

3.5 Polinomijalni model

Neka je X diskretna varijabla (kategorijalna ili konačna numerička) s populacijskom distribucijom:

$$X \sim \begin{pmatrix} a_1 & a_2 & \cdots & a_k \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}.$$

Neka je

$$X_1, X_2, \ldots, X_n$$

slučajni uzorak za X.

Definiramo:

$$\mathbf{Y}_i := (\mathbb{1}_{\{X_i=a_1\}}, \mathbb{1}_{\{X_i=a_2\}}, \ldots, \mathbb{1}_{\{X_i=a_k\}}), \;\; 1 \leq i \leq n.$$

Tada su

$$\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$$

nezavisni jednako distribuirani slučajni vektori i

$$\mathbf{Y} = (N_1, N_2, \ldots, N_k) := \mathbf{Y}_1 + \mathbf{Y}_2 + \cdots + \mathbf{Y}_n$$

ima **polinomijalnu distribuciju** s parametrima $(n; p_1, p_2, \ldots, p_k)$.

Interpretacija?

Funkcija gustoće od $\mathbf{Y}$:

$$
\begin{aligned}
& f_{\mathbf{Y}}(x_1, \ldots, x_k) := \mathsf{P}(\mathbf{Y} = (x_1, \ldots, x_k)) = \\
= \; & \mathsf{P}(N_1 = x_1, N_2 = x_2, \ldots, N_k = x_k) = \\
= \; & \begin{cases} \frac{n!}{x_1! \cdots x_k!} \cdot p_1^{x_1} \cdots p_k^{x_k}, & x_1 + \cdots + x_k = n, \\ & x_1, \ldots, x_k \in \mathsf{N}_0; \\ 0 & \text{inače} \end{cases}
\end{aligned}
$$

Zadatak 1. Dokažite:

(a) $\mathsf{E}\mathbf{Y} = (np_1, np_2, \ldots, np_k)$.
(b) Kovarijacijska matrica od $\mathbf{Y}$ je:

$$\mathrm{cov}\mathbf{Y} = n(P - \mathbf{p}\mathbf{p}^T)$$

gdje je P dijagonalna matrica

$$P = \mathrm{diag}\,(p_1, p_2, \cdots, p_k)$$

a $\mathbf{p}$ vektor-stupac

$$\mathbf{p} = (p_1, p_2, \cdots, p_k)^T.$$

Primjer 3.5 Simetričnu igraću kocku bacamo 25 puta.

Model za frekvencije elementarnih ishoda:
neka je N_i frekvencija broja i u nizu od 25 bacanja.
Tada:

$$(N_1, N_2, \ldots, N_6) \sim \text{polinomijalna}\left(25; \frac{1}{6}, \cdots, \frac{1}{6}\right).$$

3.6 Ostali diskretni modeli

– Poissonov
– geometrijski
– hipergeometrijski

$\vdots$

3.7 Neprekidne slučajne varijable i vektori

Za s.v. X definiranu na nekom vjerojatnosnom prostoru kažemo da je *neprekidna*, ako vrijedi sljedeće:

(i) $\operatorname{Im} X$ je interval u $\mathbf{R}$,
(ii) postoji nenegativna funkcija $f_X : \mathbf{R} \to \mathbf{R}$ t.d. je za svaka dva broja $a, b \in \mathbf{R}$, $a < b$,

$$\mathsf{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx.$$

Funkciju f_X zovemo *gustoćom* s.v. X.

Primijetimo:

1. Funkcija distribucije od X je:

$$F_X(x) = \mathsf{P}(X \leq x) \stackrel{(ii)}{=} \int_{-\infty}^{x} f_X(t)dt, \quad x \in \mathbf{R}$$

2. Za svaki broj a je:

$$\mathsf{P}(X = a) = \lim_{b \to a} \mathsf{P}(a \leq X \leq b) \stackrel{(ii)}{=} \int_{a}^{a} f_X(t)dt = 0.$$

3. $\displaystyle\int_{-\infty}^{+\infty} f_X(t)\,dt = 1.$

Ako vrijedi

$$\int\limits_{-\infty}^{+\infty} |t| f_X(t)\, dt < +\infty$$

tada postoji mat. očekivanje neprekidne s.v. X i

$$\mathsf{E}[X] := \int\limits_{-\infty}^{+\infty} t f_X(t)\, dt.$$

Propozicija 3.3. Za izmjerivu funkciju g i neprekidnu s.v. X vrijedi

$$\mathsf{E}[g(X)] = \int_{-\infty}^{+\infty} g(t) f_X(t)\, dt$$

ukoliko integral zdesna apsolutno konvergira.

$$\begin{aligned}\mathsf{Var}[X] &= \mathsf{E}[(X - \mathsf{E}[X])^2] = \mathsf{E}[X^2] - \mathsf{E}[X]^2 = \\ &= \int_{-\infty}^{+\infty} t^2 f_X(t)\, dt - \mathsf{E}[X]^2.\end{aligned}$$

Za sl. vektor $\mathbf{X}$ definiran na nekom vjerojatnosnom prostoru kažemo da je *neprekidan*, ako vrijedi sljedeće:

(i) $\mathrm{Im}\,\mathbf{X}$ sadrži neprazni otvoreni skup u $\mathbf{R}^d$,
(ii) postoji nenegativna funkcija $f_{\mathbf{X}} : \mathbf{R}^d \to \mathbf{R}$ t.d. je za svaka dva vektora $\mathbf{a}, \mathbf{b} \in \mathbf{R}^d$, $\mathbf{a} < \mathbf{b}$,

$$\mathsf{P}(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = \int_{a_1}^{b_1} dx_1 \cdots \int_{a_d}^{b_d} dx_d\, f_{\mathbf{X}}(x_1, \ldots, x_d).$$

Funkciju $f_{\mathbf{X}}$ zovemo *gustoćom* s.ve. $\mathbf{X}$.

Vrijedi:

1. Marginalne distribucije komponenti neprekidnog s.ve. su neprekidne s.v. s gustoćom:

$$f_{X_i}(x) = \int\limits_{\mathsf{R}^{d-1}} f_{\mathbf{X}}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d) dx_1 \cdots dx_d.$$

Ista tvrdnja vrijedi i za marginalne distribucije s.ve. sastavljenih od komponenti danog s.ve.

2. Uvjetne distribucije komponenti neprekidnog s.ve., uz dano

$$X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1}, \ldots, X_d = x_d$$

su neprekidne s.v. s gustoćom

$$f_{X_i|X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_d}(x|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) :=$$
$$= \frac{f_{\mathbf{X}}(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d)}{f_{X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_d}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)}.$$

Ista tvrdnja vrijedi i za uvjetne distribucije s.ve. sastavljenih od komponenti danog s.ve.

3. Ako su X_1,..., X_k nezavisne neprekidne s.v., tada je s. ve. $\mathbf{X} = (X_1, \ldots, X_d)$ neprekidan s gustoćom:

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_d) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_d}(x_d)$$

za $x_1, \ldots, x_d \in \mathbf{R}$.

Obratno: komponente neprekidnog s. ve. $\mathbf{X} = (X_1, \ldots, X_d)$ su nezavisne s.v. ako vrijedi da je

$$f_{\mathbf{X}}(x_1, x_2, \ldots, x_d) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_d}(x_d)$$

za sve $(x_1, \ldots, x_d) \in \mathbf{R}^d$ osim eventualno na skupu Lebesguove mjere nula.

4. Ako je D izmjeriv podskup od $\mathbf{R}^d$ i $f_{\mathbf{X}}$ gustoća neprekidnog s. ve. $\mathbf{X}$, tada je

$$\mathsf{P}(\mathbf{X} \in D) = \int_D f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}.$$

Propozicija 3.4.
(a) Ako je (X,Y) neprekidan s. vektor, tada je neprekidna s.v. $Z := X + Y$ i vrijedi

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z - x)\, dx.$$

(b) Ako su X i Y nezavisne neprekidne s.v., tada je i $Z = X + Y$ neprekidna s.v. i

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x)\, dx.$$

Propozicija 3.5.
(a) Ako je (X,Y) neprekidan s. vektor, tada je neprekidna s.v. $Z := Y - X$ i vrijedi

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, z + x)\, dx.$$

(b) Ako su X i Y nezavisne neprekidne s.v., tada je i $Z = Y - X$ neprekidna s.v. i

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z + x)\, dx.$$

Propozicija 3.6.
(a) Ako je (X, Y) neprekidan s. vektor, tada je neprekidna s.v. $Z := XY$ i vrijedi

$$f_Z(z) = \int\limits_{-\infty}^{+\infty} f_{X,Y}\left(x, \frac{z}{x}\right) \frac{1}{|x|}\, dx.$$

(b) Ako su X i Y nezavisne neprekidne s.v., tada je i $Z = XY$ neprekidna s.v. i

$$f_Z(z) = \int\limits_{-\infty}^{+\infty} f_X(x) f_Y\left(\frac{z}{x}\right) \frac{1}{|x|}\, dx.$$

Propozicija 3.7.
(a) Ako je (X, Y) neprekidan s. vektor, tada je neprekidna s.v. $Z := Y/X$ i vrijedi

$$f_Z(z) = \int_{-\infty}^{+\infty} f_{X,Y}(x, zx)|x|\,dx.$$

(b) Ako su X i Y nezavisne neprekidne s.v., tada je i $Z = Y/X$ neprekidna s.v. i

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(zx)|x|\,dx.$$

Zadatak 2. Dokažite propozicije 3.4, 3.5, 3.6, i 3.7.

Neprekidni modeli:

- normalni
- eksponencijalni
- uniformni
- gama
- beta

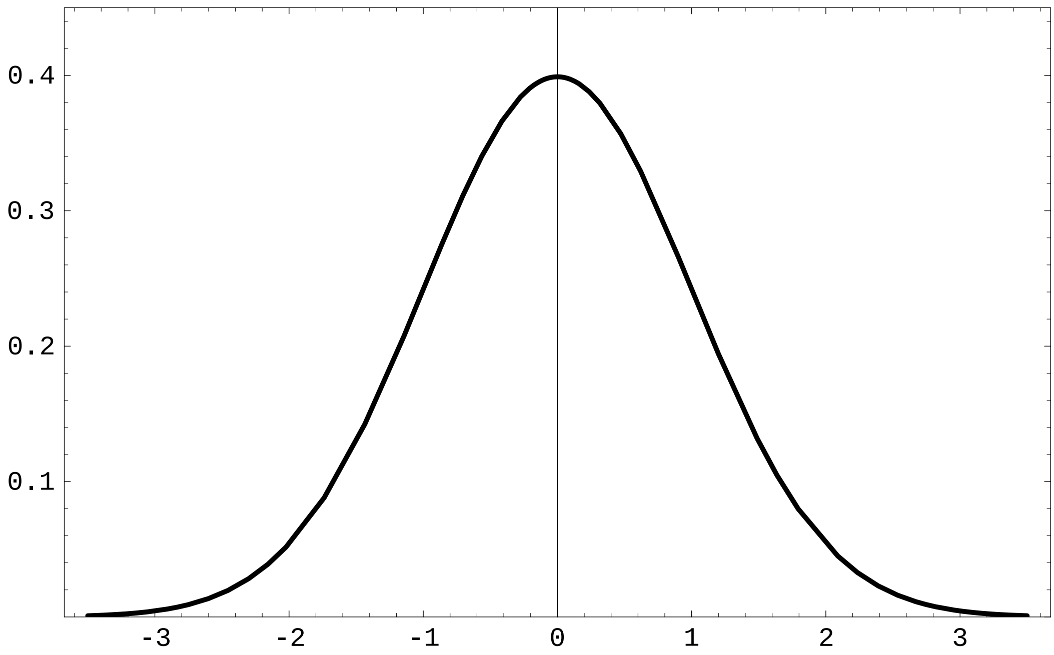
$\vdots$

3.8 Normalna razdioba

Definicija. Neprekidna s.v. X je *normalna* s.v. s parametrima μ i σ^2 ukoliko joj je funkcija gustoće

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

Pišemo: $X \sim N(\mu, \sigma^2)$



0.4
0.3
0.2
0.1
-3
-2
-1
0
1
2
3

Normalna razdioba je važna jer:

1. dobar je model za veliku većinu fizikalnih mjerenja
2. dobra je aproksimacija velike klase drugih distribucija (na primjer, binomne)
3. dobar je model za uzoračku razdiobu raznih statistika
4. zaključivanje na osnovi velikih uzoraka i neki statistički postupci zasnivaju se na pretpostavci normalnosti
5. pomoću nje se izvode mnoge druge distribucije

Primijetimo:

1. $\operatorname{Im} X = \mathbf{R}$ jer je $f_X(x) > 0$ za sve $x \in \mathbf{R}$

2. Postoje mat. očekivanje i varijanca:

$$
\begin{aligned}
\mathsf{E}[X] &= \int\limits_{-\infty}^{+\infty} t f_X(t)\,dt = \\
&= \int\limits_{-\infty}^{+\infty} (t-\mu) f_X(t)\,dt + \mu \int\limits_{-\infty}^{+\infty} f_X(t)\,dt = \\
&= \int\limits_{-\infty}^{+\infty} (t-\mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt + \mu = \\
&= \int\limits_{-\infty}^{+\infty} y \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}\,dy + \mu = \mu
\end{aligned}
$$

$$\begin{aligned}
\mathsf{Var}[X] &= \mathsf{E}[(X-\mu)^2] = \\
&= \int\limits_{-\infty}^{+\infty} (t-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt = \\
&= \int\limits_{-\infty}^{+\infty} y^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}\, dy = \\
&= \sigma^2 \quad \text{(parcijalna integracija)}
\end{aligned}$$

3. Za $k = 1, 2, 3$ izračunajmo

$$\begin{aligned}
\mathsf{P}(|X-\mu| \leq k\sigma) &= \mathsf{P}(\mu - k\sigma \leq X \leq \mu + k\sigma) = \\
&= \int\limits_{\mu-k\sigma}^{\mu+k\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt = \\
&= \int\limits_{-k}^{k} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \\
&= 2\Phi_0(k)
\end{aligned}$$

gdje je funkcija Φ_0, definirana sa

$$\Phi_0(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt, \ \ \text{za} \ \ x > 0,$$

$$\Phi_0(x) := -\Phi_0(-x), \ \ \text{za} \ \ x < 0, \qquad \Phi_0(0) = 0$$

tabelirana.

k	$\mathsf{P}(\lvert X-\mu\rvert \le k\sigma)$
1	0.6826
2	0.9544
3	0.9974

4. Normalna razdioba je invarijantna na afine transformacije, tj. ako je $X \sim N(\mu, \sigma^2)$ i $a, b \in \mathbf{R}$, $a \neq 0$, tada je

$$Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$$

5. Ako su X i Y nezavisne normalne s.v., tada je i $X + Y$ normalna s.v.

6. Standardizirana normalna s.v. je, prema 4., također normalno distribuirana s mat. očekivanjem 0 i varijancom 1:

$$X \sim N(\mu, \sigma^2) \;\;\Rightarrow\;\; X^* = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

S.v. $Z \sim N(0,1)$ zovemo *jedinična* ili *standardna* normalna s.v.

Funkcija distribucija standardne normalne s.v.:

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt$$

Vrijedi:

$$\Phi(x) = \frac{1}{2} + \Phi_0(x)$$

Zadatak 3. Dokažite tvrdnje 4 i 5.

Primjer 3.6

Neka je X visina na slučajan način odabranog čovjeka. Pretpostavimo: $X \sim N(170 \text{ cm}, 8^2 \text{ cm}^2)$.

Koliki je postotak ljudi čije su visine:

(a) veće od 190 cm
(b) manje od 160 cm
(c) u intervalu 165 - 180 cm?

(a):

$$
\begin{aligned}
\mathsf{P}(X > 190) &= \mathsf{P}\left(\frac{X-170}{8} > \frac{190-170}{8}\right) = \\
&= \mathsf{P}(X^* > 2.5) = 1 - \mathsf{P}(X^* \leq 2.5) = \\
&= 1 - \Phi(2.5) = \frac{1}{2} - \Phi_0(2.5) = \\
&= 0.0062 = 0.62\%
\end{aligned}
$$

(b):

$$\begin{aligned}
\mathsf{P}(X < 160) &= \mathsf{P}(X \leq 160) = \mathsf{P}\left(X^* \leq \frac{160 - 170}{8}\right) = \\
&= \Phi(-1.25) = \frac{1}{2} - \Phi_0(1.25) = \\
&= 0.1056 = 10.56\%
\end{aligned}$$

(c):

$$\begin{aligned}
\mathsf{P}(165 \le X \le 180) &= \mathsf{P}\left(\frac{165-170}{8} \le X^* \le \frac{180-170}{8}\right) \\
&= \mathsf{P}(-0.4 \le X^* \le 1.25) = \\
&= \Phi(1.25) - \Phi(-0.4) = \\
&= \Phi_0(1.25) + \Phi_0(0.4) = \\
&= 0.5498 = 54.98\%
\end{aligned}$$

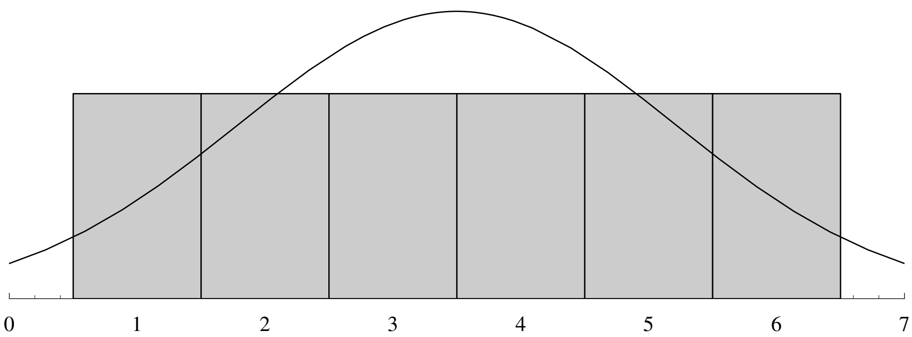
Primjer 3.7 Nacrtati histograme razdioba s.v.

$$Z_n := X_1 + X_2 + \cdots + X_n,$$

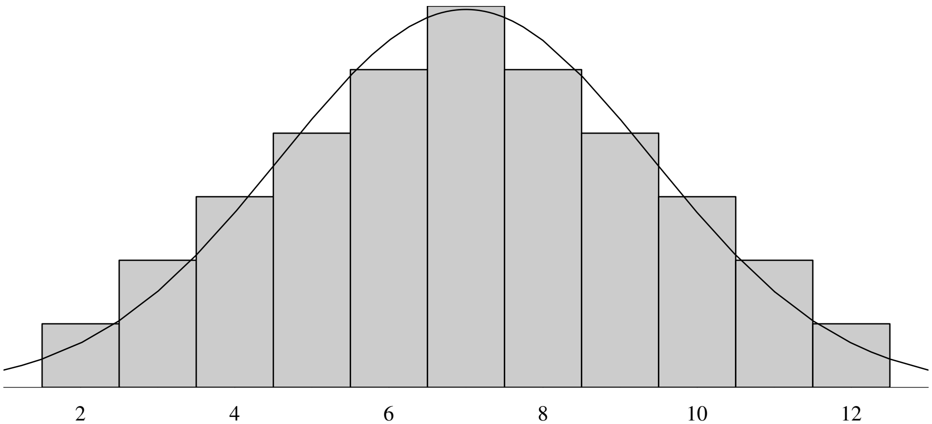
za razne $n = 1, 2, \ldots$ i usporediti ih s gustoćama normalnih razdioba s istim očekivanjem i varijancama kao Z_n, gdje su $X_1, X_2, \ldots$ n.j.d. s.v. definirane kako slijedi:

(a): Bacamo simetričnu kocku n puta $(n = 1, 2, \ldots)$.

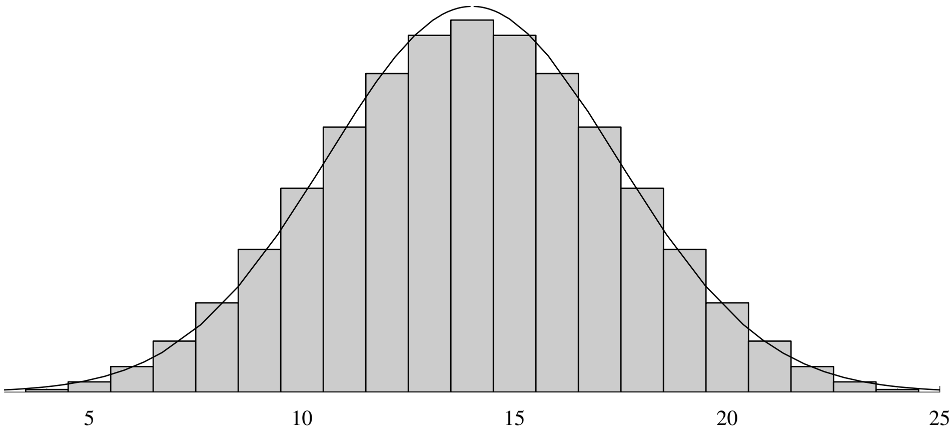
$X_i =$ broj koji se okrenuo u i-tom bacanju.



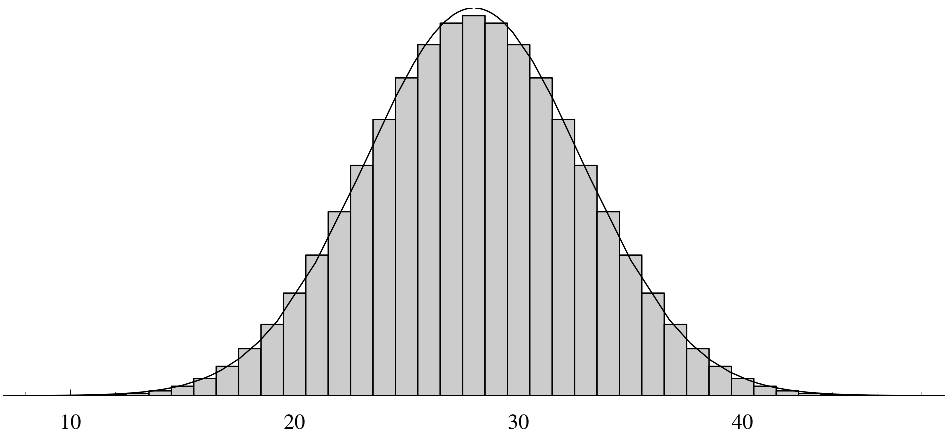
Razdioba od Z_2:



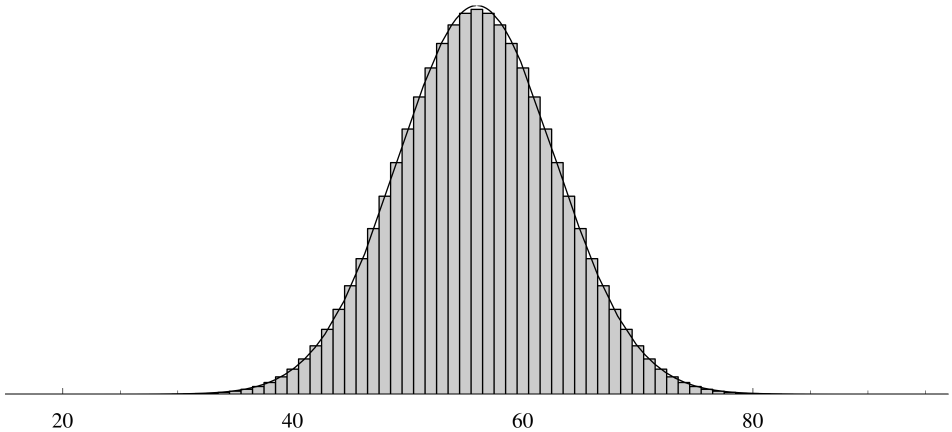
Razdioba od Z_4:



Razdioba od Z_8:



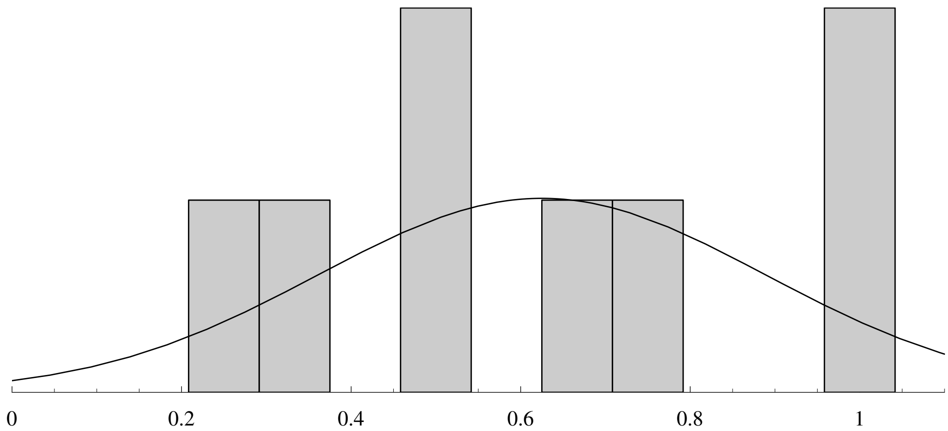
Razdioba od Z_{16}:



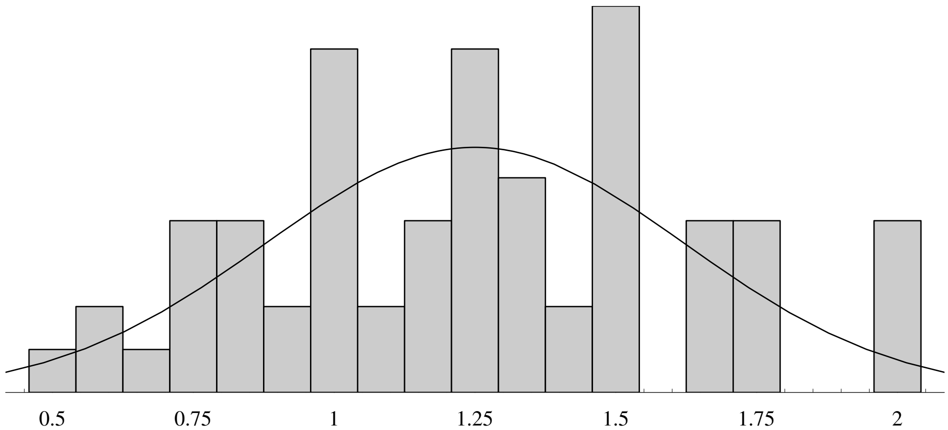
(b): Bacanje dva simetrična tetraedra.

$X_i =$ omjer minimuma i maksimuma ishoda bacanja

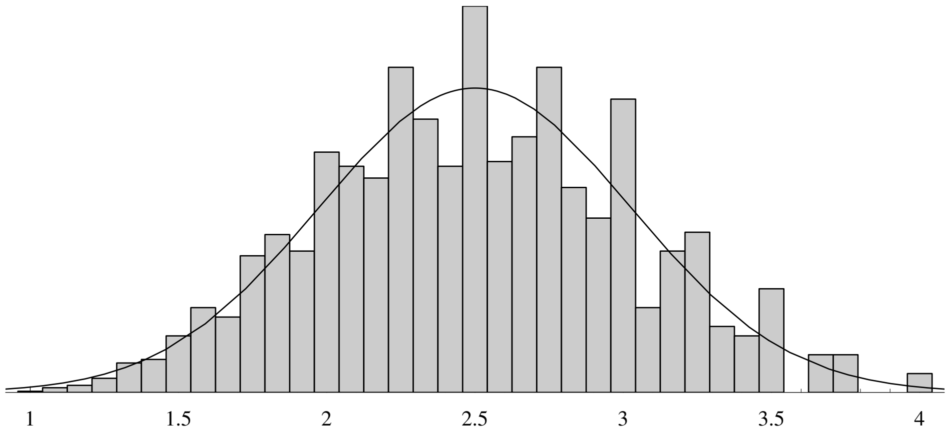
$$X_i \sim \begin{pmatrix} \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & \frac{2}{3} & \frac{3}{4} & 1 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \end{pmatrix}$$



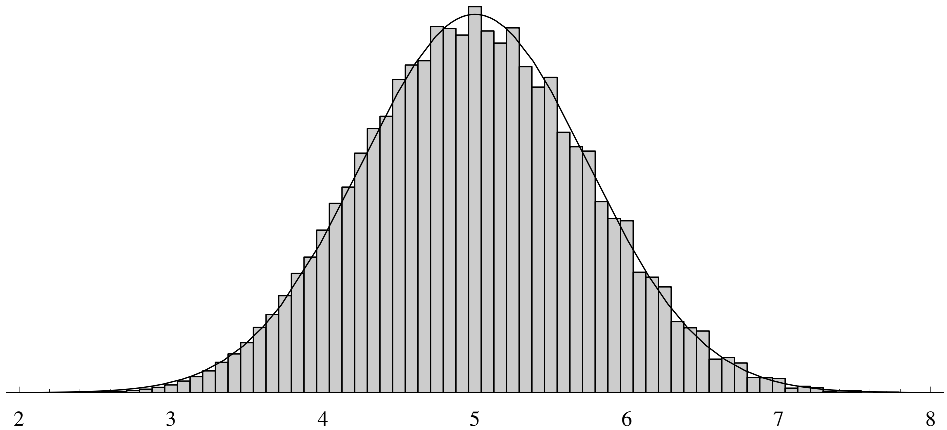
Razdioba od Z_2:



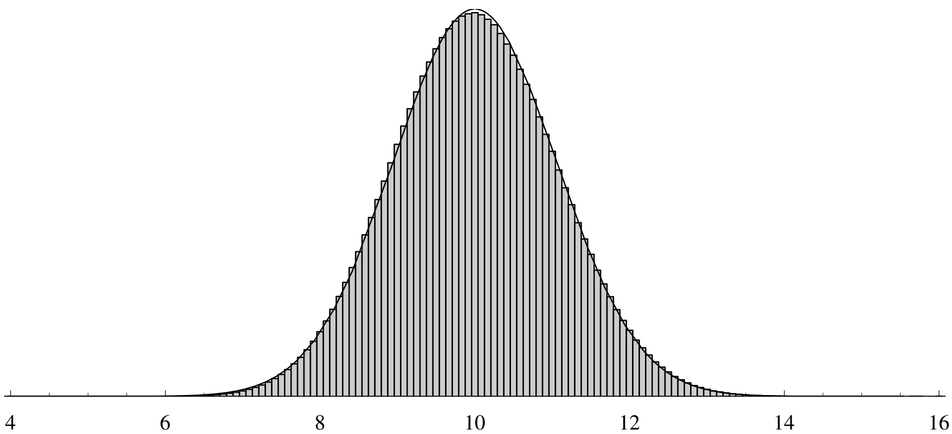
Razdioba od Z_4:



Razdioba od Z_8:



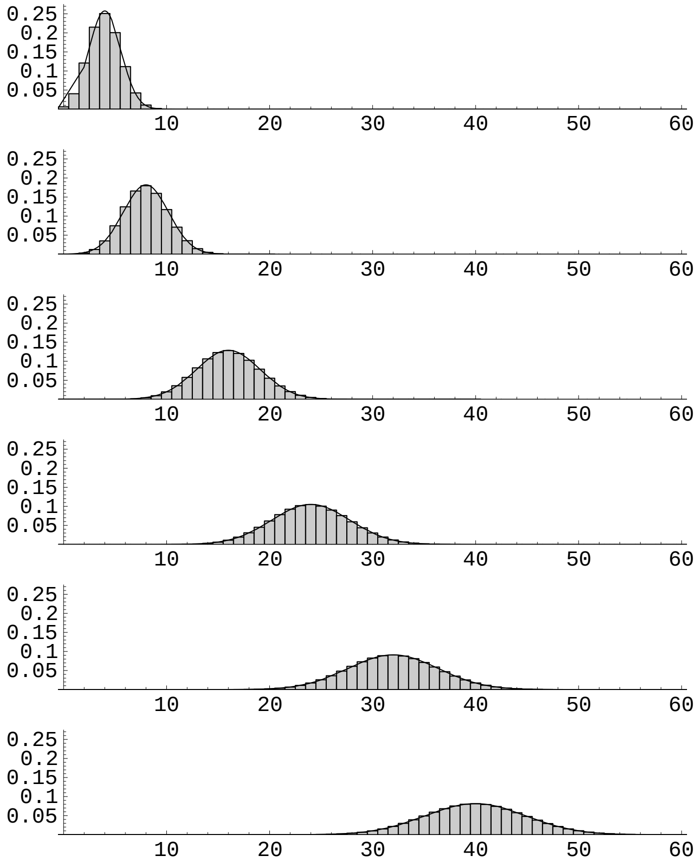
Razdioba od Z_{16}:

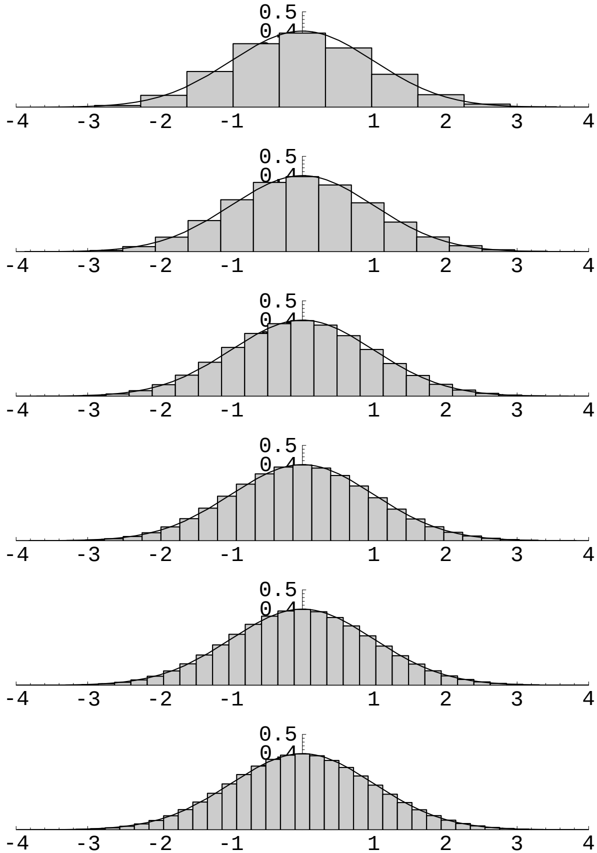


(c): Bernoullijevi pokusi

X_i je Bernoullijeva s.v. s vjerojat. uspjeha $p = \frac{2}{5}$

$\Rightarrow Z_n \sim B(n, \frac{2}{5})$, $n = 10, 20, 40, 60, 80, 100$ (slike!)


0.25
0.2
0.15
0.1
0.05
10
20
30
40
50
60


0.5
0.4
-4
-3
-2
-1
1
2
3
4

Što je zajedničko primjerima (a), (b) i (c)?

S.v. Z_n su dobivene kao *zbroj* od n

- nezavisnih
- jednako distribuiranih s.v.
- *konačne varijance*.

Teorem 3.8. (CGT) Neka je $(X_i, i = 1, 2, \ldots)$ niz n.j.d. s.v. konačne varijance, i

$$Z_n := X_1 + X_2 + \cdots + X_n, \;\; n = 1, 2, \ldots$$

Tada za niz standardiziranih s.v.

$$Z_n^* := \frac{Z_n - \mathsf{E}[Z_n]}{\sqrt{\mathsf{Var}[Z_n]}}, \;\; n = 1, 2, \ldots$$

i bilo koja dva realna broja a, b t.d. je $a < b$, vrijedi:

$$\lim_{n \to +\infty} \mathsf{P}(a \le Z_n^* \le b) = \Phi(b) - \Phi(a).$$

Kažemo da niz standardiziranih s.v. Z_n^* *po distribuciji konvergira* jediničnoj normalnoj razdiobi, i pišemo

$$Z_n^* \xrightarrow{D} N(0,1), \quad n \to +\infty.$$

Interpretacija: za *velike* n, aproksimativno je

$$\mathsf{P}(a \le Z_n^* \le b) \approx \Phi(b) - \Phi(a),$$

odn.

$$\mathsf{P}(A \le Z_n \le B) \approx \Phi\left(\frac{B - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{A - n\mu}{\sigma\sqrt{n}}\right),$$

gdje su $\mu = \mathsf{E}[X_i]$, $\sigma^2 = \mathrm{Var}[X_i]$ i $A < B$ brojevi.

Drugi zapis CGT-a:

Ako je $(X_n; n \in \mathrm{N})$ niz n.j.d. s.v. s konačnom varijancom, tada ($\mu = \mathrm{E}X_i$, $\sigma^2 = \mathrm{Var}X_i$):

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \sim AN(0,1),\ n \to +\infty$$

ili ekvivalentno

$$\frac{\overline{X}_n - \mu}{\sigma} \cdot \sqrt{n} \sim AN(0,1),\ n \to +\infty. \tag{1}$$

Primjer 3.8 (Aproksimacija binomne pomoću normalne razdiobe)

Vjerojatnost uspjeha u svakom od 900 nezavisnih pokusa je 0.8. Izračunajte vjerojatnost da je ukupan broj uspjeha

(a) jednak 710
(b) barem 710 i ne više od 740
(c) jednak 750.

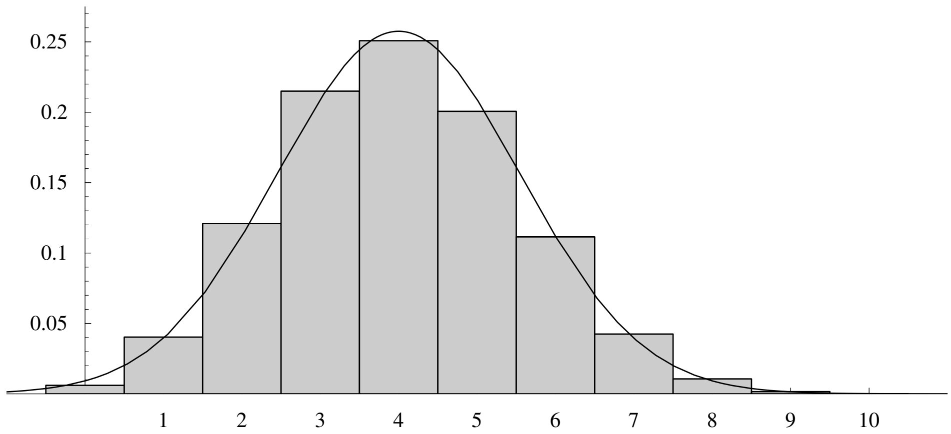
$X =$ ukupan broj uspjeha u 900 pokusa

$X \sim B(900, 0.8)$

$\mathsf{E}[X] = 900 \cdot 0.8 = 720$

$\mathsf{Var}[X] = 900 \cdot 0.8 \cdot 0.2 = 144 = 12^2$

$Z_{10} \sim B(10, \frac{2}{5})$ iz Primjera 3.15 (c):



(a):

$$
\begin{aligned}
& \mathsf{P}(X = 710) = \\
= \quad & \mathsf{P}(709.5 < X < 710.5) = \\
= \quad & \mathsf{P}\left(\frac{709.5 - 720}{12} < X^* < \frac{710.5 - 720}{12}\right) = \\
= \quad & \mathsf{P}(-0.875 < X^* < -0.792) \overset{\text{CGT}}{\approx} \\
\overset{\text{CGT}}{\approx} \quad & \Phi(-0.792) - \Phi(-0.875) = \\
= \quad & 0.0254
\end{aligned}
$$

(b):

$$
\begin{aligned}
& \mathsf{P}(710 \leq X \leq 740) = \\
= \quad & \mathsf{P}(709.5 < X < 740.5) = \\
= \quad & \mathsf{P}\left(\frac{709.5 - 720}{12} < X^* < \frac{740.5 - 720}{12}\right) = \\
= \quad & \mathsf{P}(-0.875 < X^* < 1.71) \overset{\text{CGT}}{\approx} \\
\overset{\text{CGT}}{\approx} \quad & \Phi(1.71) - \Phi(-0.875) = \\
= \quad & 0.7670
\end{aligned}
$$

(c):

$$\begin{aligned} & \mathsf{P}(X = 750) = \\ = \; & \mathsf{P}(709.5 < X < 710.5) \approx \\ \approx \; & 1 \cdot \frac{1}{12 \cdot \sqrt{2\pi}} e^{-\frac{1}{2\cdot 144}(750-720)^2} = \\ = \; & 0.0014 \end{aligned}$$

Primjer 3.9 Sumnjamo da novčić nije simetričan. Koliko puta ga treba bacati da bi sa 95% sigurnosti procijenili vjerojatnost padanja pisma na dvije točne decimale?

$X_n =$ ukupan broj pisama u n bacanja novčića

$p =$ vjerojatnost pisma (u jednom bacanju)

$X_n \sim B(n, p)$

Procjena za p: relativna frekvencija pisma $= \frac{X_n}{n}$.

Iz uvjeta zadatka za n treba biti ispunjeno:

$$\mathsf{P}\left(\left|\frac{X_n}{n} - p\right| \leq 0.5 \cdot 10^{-2}\right) \geq 0.95$$

Za veliki n je po CGT-u:

$$\mathsf{P}\left(\left|\frac{X_n - np}{\sqrt{npq}}\right| \leq \varepsilon\right) \approx \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\Phi_0(\varepsilon)$$

Slijedi:

$$\begin{aligned} 2\Phi_0(\varepsilon) &\approx \mathsf{P}(|X_n - np| \leq \varepsilon\sqrt{npq}) = \\ &= \mathsf{P}\left(\left|\frac{X_n}{n} - p\right| \leq \frac{\varepsilon}{\sqrt{n}} \cdot \sqrt{p(1-p)}\right) \leq \\ (*) \qquad &\leq \mathsf{P}\left(\left|\frac{X_n}{n} - p\right| \leq \frac{\varepsilon}{2\sqrt{n}}\right) \end{aligned}$$

Odaberimo ε t.d. je $\varepsilon/(2\sqrt{n}) = 0.5 \cdot 10^{-2}$
$\Rightarrow \varepsilon = 0.01\sqrt{n}$.

Ako je n t.d. je $0.95 \leq 2\Phi_0(0.01\sqrt{n})$, tada:

$$0.95 \leq 2\Phi_0(0.01\sqrt{n}) \overset{(*)}{\leq} \mathsf{P}\left(\left|\frac{X_n}{n} - p\right| \leq 0.5 \cdot 10^{-2}\right).$$

Dakle, nađimo n t.d. je:

$$0.95 \leq 2\Phi_0(0.01\sqrt{n}) \quad \Rightarrow \quad 1.96 \leq 0.01\sqrt{n}.$$

Slijedi: $n \geq 196^2 = 38416$