Statistika

Vanja Wagner, Ivan Biočić, Hrvoje Planinić
na temelju materijala B. Basraka i M. Marušića

4. Razdiobe

Podsjetnik: binomna razdioba

Ponavljamo isti pokus n puta **nezavisno**.

- U svakom pokusu bilježimo uspjeh ili "neuspjeh".
- Vjerojatnost uspjeha u svakom je pokusu jednaka p.

Bitno je u danom problemu prepoznati što je "pokus" i "uspjeh", tj. prepoznati da se problem može svesti na gornji slučaj.

Ako je X **ukupan broj uspjeha**, pišemo

$$X \sim B(n, p), \qquad X \in \{0, 1, \ldots, n\}.$$

$$\boxed{\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}, \qquad k = 0, 1, \ldots, n.$$

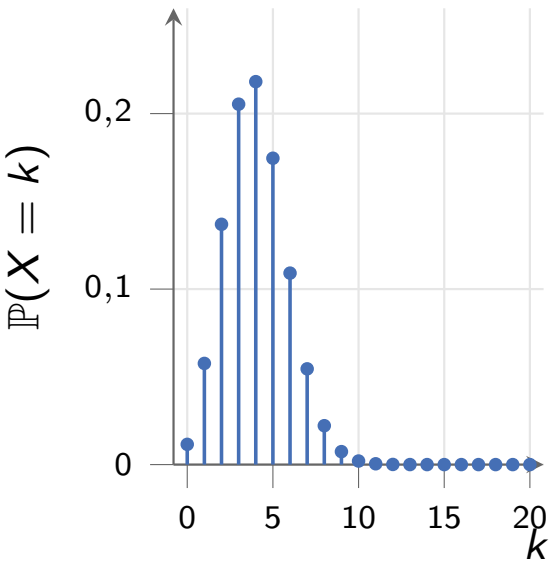
Vrijedi

$$\mathbb{E}[X] = np, \qquad \mathrm{Var}(X) = np(1-p), \quad \sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Binomna razdioba: promjena vjerojatnosti uspjeha

Broj pokusa je isti: $n = 20$. Mijenjamo vjerojatnost uspjeha p.

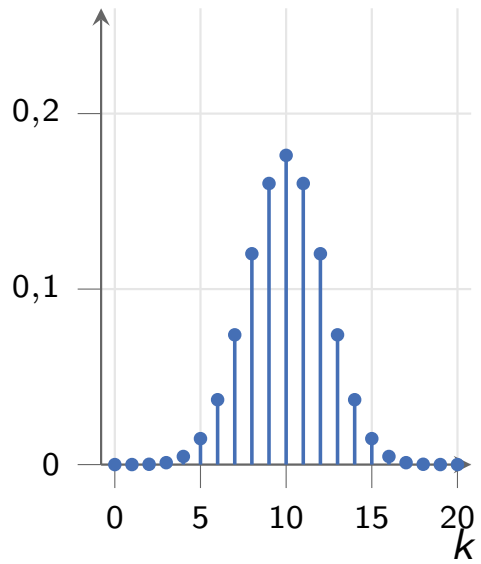
$n = 20,\ p = 0{,}2$



$\mathbb{E}[X] = 4$

$\sigma_X \approx 1{,}79$

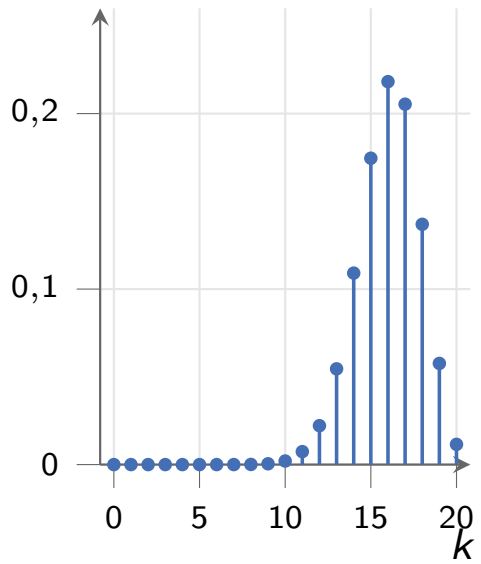
$n = 20,\ p = 0{,}5$



$\mathbb{E}[X] = 10$

$\sigma_X \approx 2{,}24$

$n = 20,\ p = 0{,}8$



$\mathbb{E}[X] = 16$

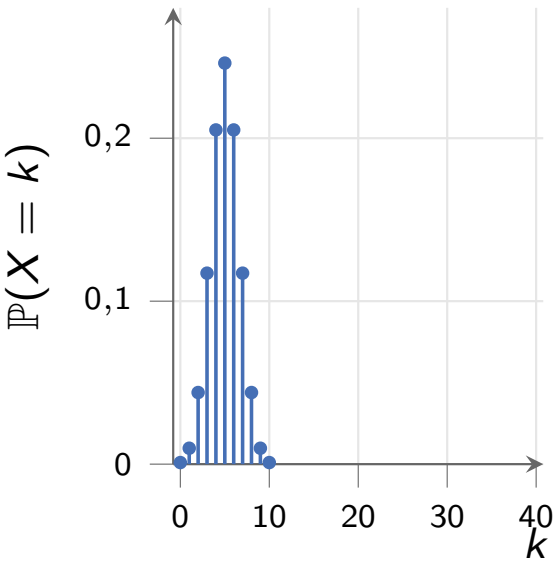
$\sigma_X \approx 1{,}79$

- Veći p "pomiče" razdiobu prema većem broju uspjeha.

Binomna razdioba: promjena broja pokusa

Vjerojatnost uspjeha je ista: $p = 1/2$. Mijenjamo broj pokusa n.

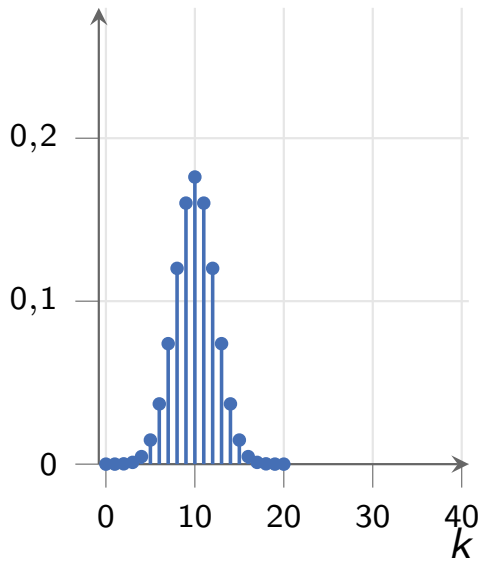
$n = 10,\ p = 0{,}5$



$\mathbb{E}[X] = 5$

$\sigma_X \approx 1{,}58$

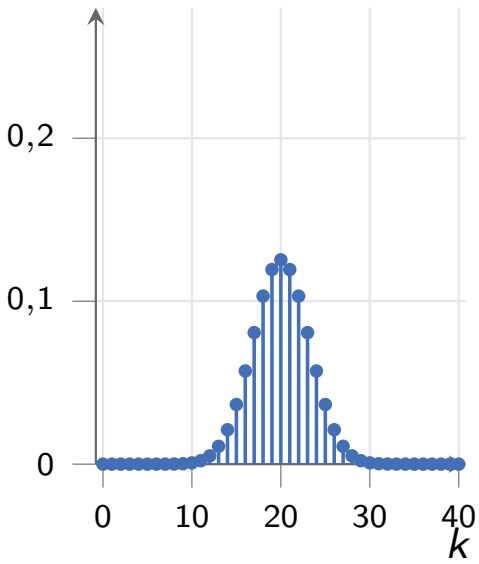
$n = 20,\ p = 0{,}5$



$\mathbb{E}[X] = 10$

$\sigma_X \approx 2{,}24$

$n = 40,\ p = 0{,}5$



$\mathbb{E}[X] = 20$

$\sigma_X \approx 3{,}16$

S porastom n rastu i očekivanje i standardna devijacija:

$$\mathbb{E}[X] = \frac{n}{2}, \qquad \sigma_X = \frac{\sqrt{n}}{2}.$$

Geometrijska razdioba

Nezavisno ponavljamo pokus s vjerojatnošću uspjeha $p \in (0, 1]$.

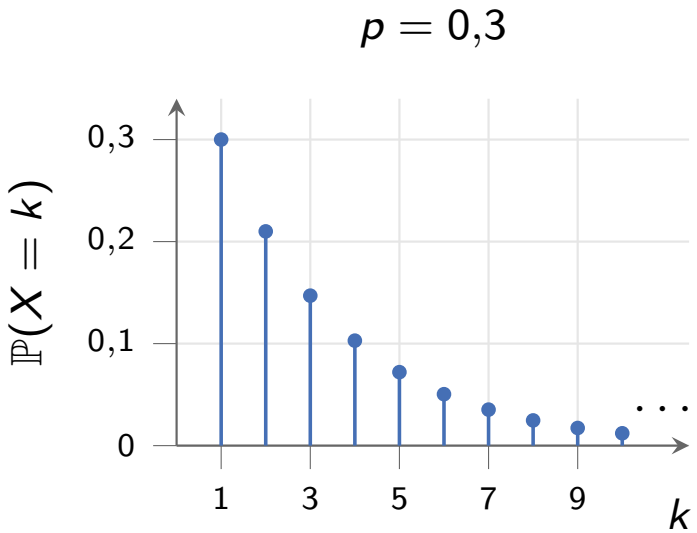
Neka je X **redni broj pokusa u kojem se dogodio prvi uspjeh**.

Za $k = 1, 2, 3, \ldots$ vrijedi

$$\boxed{\mathbb{P}(X = k) = (1 - p)^{k-1} p}.$$

[Prvih $k - 1$ pokusa su neuspjesi, a k-ti je uspjeh.]

Kažemo da X ima **geometrijsku razdiobu** s parametrom p (oznaka $X \sim G(p)$).



Vrijedi

$$\mathbb{E}[X] = \frac{1}{p}, \qquad \mathrm{Var}(X) = \frac{1 - p}{p^2}.$$

[Manji p, veće $\mathbb{E}[X]$, tj. u **prosjeku dulje** čekamo na prvi uspjeh.]

Zadatak

Bacamo po dvije kocke istovrmeno sve dok prvi put ne dobijemo šesticu na barem jednoj kocki. Odredite vjerojatnost da ćemo kocke bacati više od 3 puta, te očekivani broj bacanja.

Rješenje na ploči. (vjerojatnost je $\left(\frac{25}{36}\right)^3 \approx 0.3349$, očekivanje $\frac{36}{11} \approx 3.27$)

Poissonova razdioba

Za **velik** n i **malen** p, za $X_n \sim B(n,p)$ vrijedi $\mathbb{E}[X_n] = np$, te se može pokazati da vrijedi

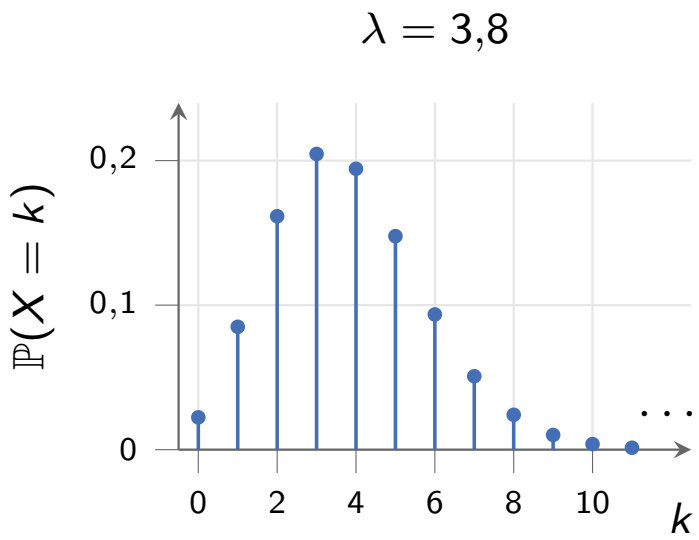
$$\mathbb{P}(X_n = k) \approx \frac{(np)^k}{k!} e^{-np}, \quad k \geqslant 0\,.$$

Za $\lambda > 0$ kažemo da X ima **Poissonovu** razdiobu s parametrom λ ($X \sim P(\lambda)$) ako je

$$\boxed{\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}} \quad k = 0, 1, 2, \ldots$$

Vrijedi

$$\mathbb{E}[X] = \lambda = \mathrm{Var}(X)\,.$$



Primjer 4.2.9 Broj jedinki **rijetke** vrste planktona u uzorku od 1 l vode možemo modelirati Poissonovom razdiobom. Ako je očekivani broj jedinki 3.8, uzimamo $\lambda = 3.8$.

Ako stignemo: provjera nezavisnosti za dvije diskretne varijable

U kutiji su 2 crvene i 2 plave kuglice. Izvlačimo dvije jednu za drugom, bez vraćanja. Neka je $X = 1$ ako je prva kuglica crvena, a $Y = 1$ ako je druga kuglica crvena; inače je vrijednost odgovarajuće varijable 0.

Želimo izračuati $\mathbb{P}(X = a, Y = b)$ za $a, b \in \{0, 1\}$ te provjeriti jesu li X i Y **nezavisne**.

(X, Y)	Ishod izvlačenja	Vjerojatnost
$(0, 0)$	plava, plava	$\frac{2}{4} \cdot \frac{1}{3} = \frac{1}{6}$
$(0, 1)$	plava, crvena	$\frac{2}{4} \cdot \frac{2}{3} = \frac{1}{3}$
$(1, 0)$	crvena, plava	$\frac{2}{4} \cdot \frac{2}{3} = \frac{1}{3}$
$(1, 1)$	crvena, crvena	$\frac{2}{4} \cdot \frac{1}{3} = \frac{1}{6}$

Marginalne razdiobe

Dobivene vjerojatnosti možemo zapisati u tablicu koja predstavlja distribuciju **slučajnog vektora** (X, Y) ili **zajedničku razdiobu** od X i Y:

X/Y	$y=0$	$y=1$	$\mathbb{P}(X=x)$
$x=0$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
$x=1$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$
$\mathbb{P}(Y=y)$	$\frac{1}{2}$	$\frac{1}{2}$	1

Da bismo dobili razdiobu od X, zbrajamo vjerojatnosti u retku, preko mogućih vrijednosti od Y. Za marginalnu razdiobu od Y zbrajamo vjerojatnosti u stupcu, preko mogućih vrijednosti od X: npr.

$$\mathbb{P}(X=0) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}, \qquad \mathbb{P}(Y=1) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

Vidimo da i X i Y imaju $B(1/2)$ razdiobu. Jeste li to očekivali za Y?

Nezavisnost

X/Y	$y=0$	$y=1$	$\mathbb{P}(X=x)$
$x=0$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$
$x=1$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$
$\mathbb{P}(Y=y)$	$\frac{1}{2}$	$\frac{1}{2}$	1

Za nezavisnost bi moralo vrijediti $\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$ za svaki par (x, y). No,

$$\mathbb{P}(X=1, Y=1) = \frac{1}{6} \neq \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(X=1)\mathbb{P}(Y=1),$$

pa X i Y **nisu nezavisne** (što bismo intuitivno i očekivali).

4.3. Neprekidne slučajne varijable

Neka je X visina slučajno odabrane odrasle osobe, izražena u centimetrima.

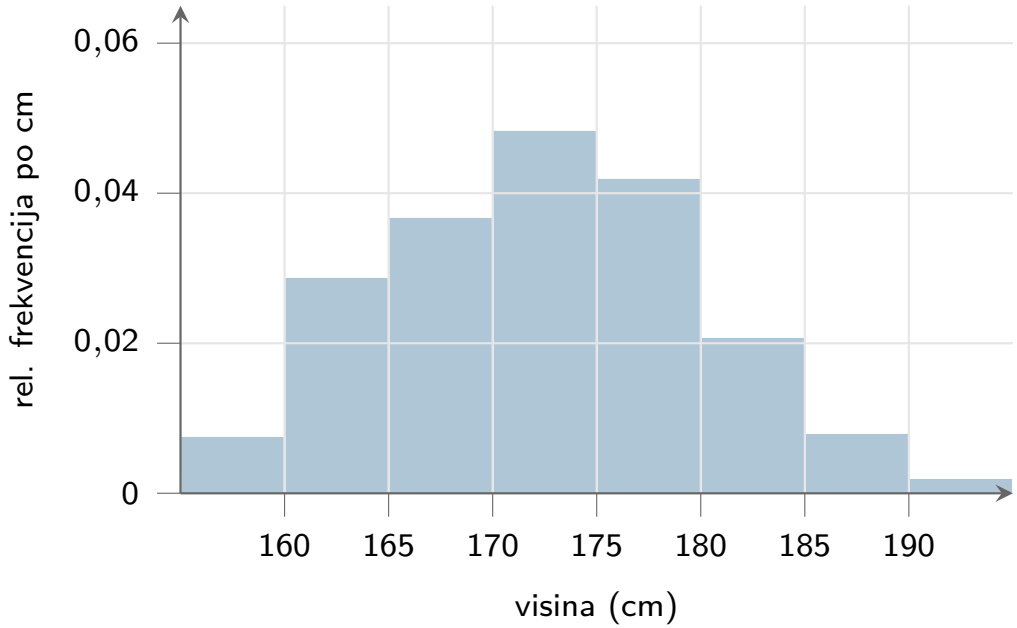
Kako bismo pomoću podataka procijenili

$$\mathbb{P}(177 \leqslant X \leqslant 182) \qquad \text{ili} \qquad \mathbb{P}(179 \leqslant X \leqslant 183)?$$

U histogramu je **površina** svakog stupca jednaka udjelu podataka u tom razredu.

No granice gornjih intervala padaju **unutar razreda** širine 5 cm. Kako iz histograma procijeniti udjele za takve intervale?

Histogram visina: razredi širine 5 cm

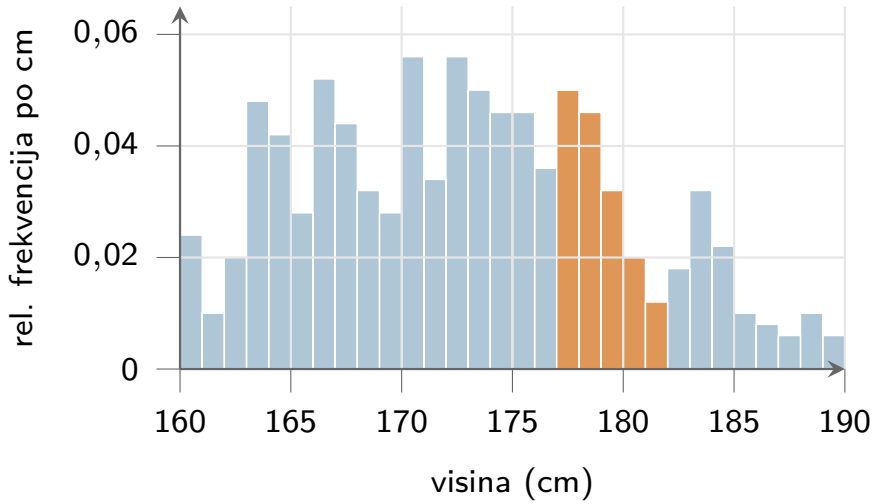


Ilustracija: 500 simuliranih visina.

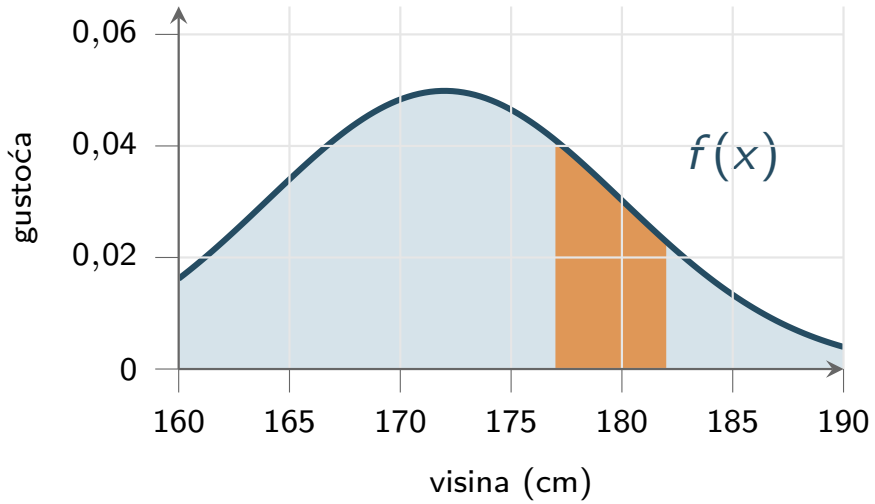
Od histograma do neprekidnog modela

Histogram s razredima širine 1 cm prikazuje isti uzorak od 500 visina.

Histogram: razredi širine 1 cm



Neprekidni model



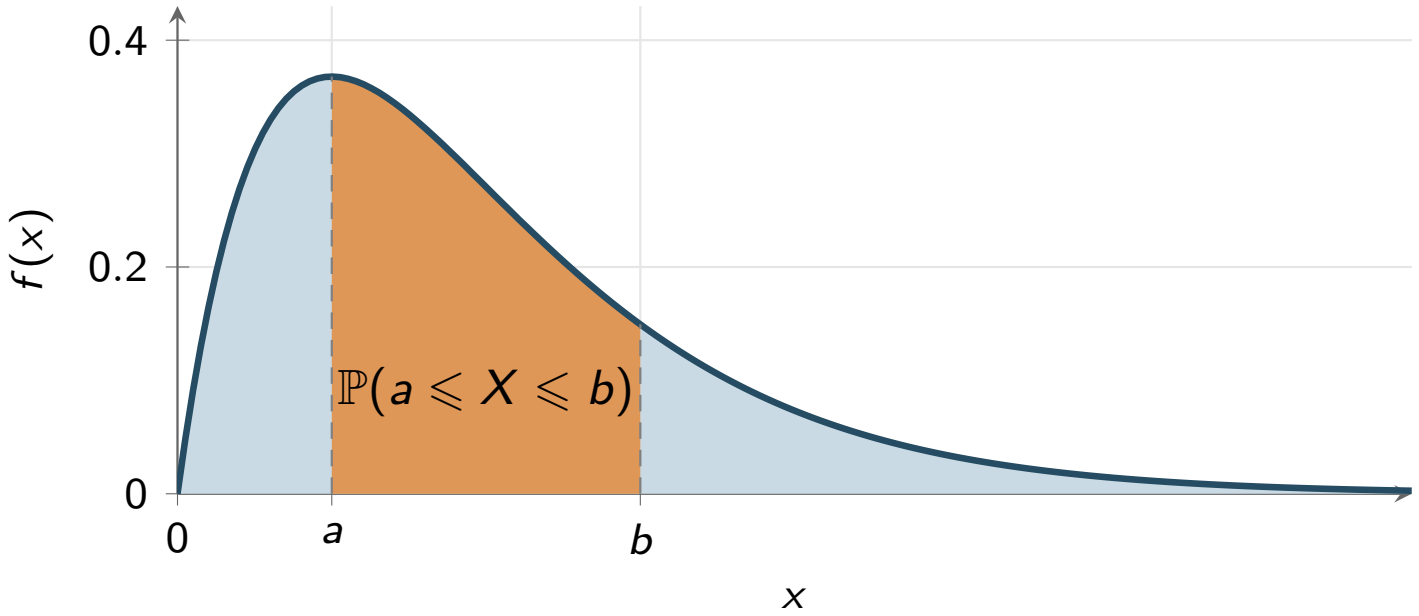
Ipak, kod **preuskih razreda** susjedni stupci mogu jako varirati, tj. previše ovise o konkretnom uzorku kojeg smo izabrali.

Zato za model umjesto "stepenastog" histograma često promatramo neprekidnu funkciju f. Vjerojatnost koju pridružujemo intervalu jednaka je **površini ispod njezina grafa**.

Neprekidna slučajna varijabla i funkcija gustoće

Kažemo da je X neprekidna slučajna varijabla s **funkcijom gustoće** f ako za svaki interval $[a, b]$ vrijedi

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f(x)\,dx = \text{površina ispod grafa } f \text{ na intervalu } [a, b]$$



Funkcija gustoće zadovoljava

$$f(x) \geqslant 0, \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

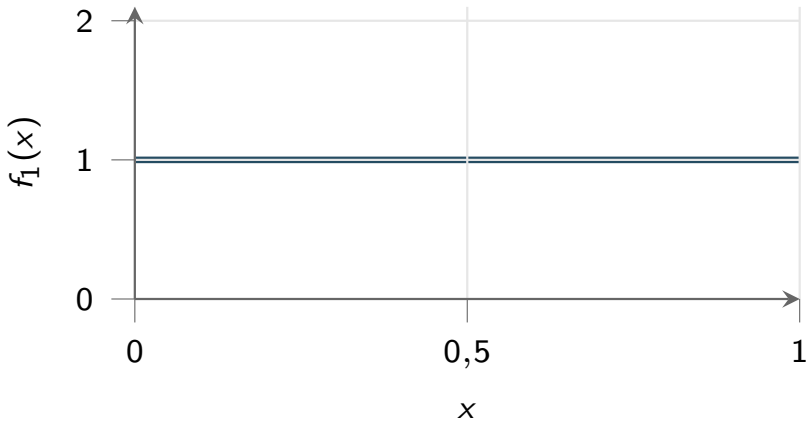
Primjer

Na intervalu $[0, 1]$ promatramo dvije gustoće:

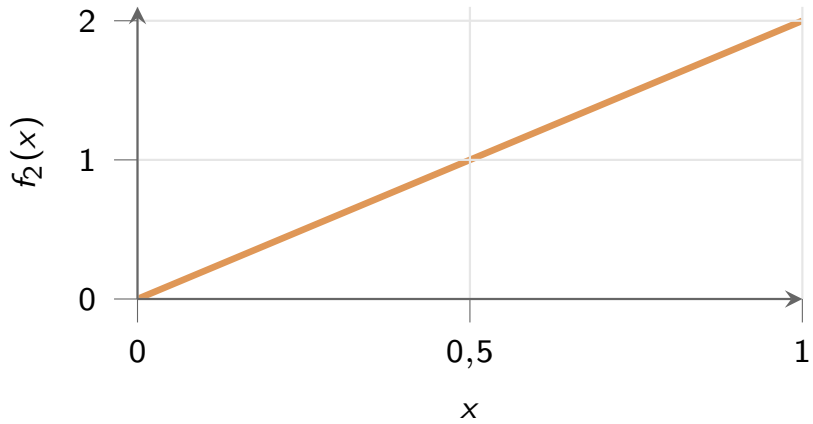
$$f_1(x) = 1, \qquad f_2(x) = 2x\,, \quad x \in [0, 1].$$

Obje gustoće su 0 van segmenta [0,1]. Uočite da je ukupna površina ispod grafa obje gustoće zaista jednaka 1.

$f_1(x) = 1$



$f_2(x) = 2x$



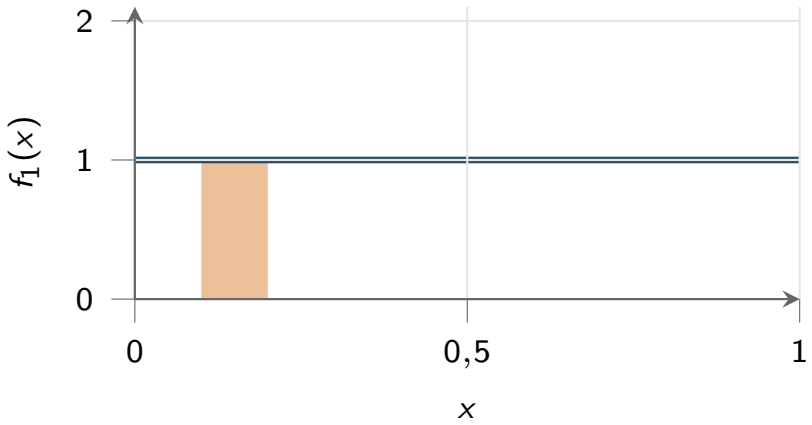
- slučajna varijabla s gustoćom f_1 odgovara **uniformnom** izboru točke iz $[0, 1]$;
- slučajna varijabla s gustoćom f_2 s većom vjerojatnosti poprima vrijednosti **bliže 1**.

Primjer

Izračunajte za obje gustoće

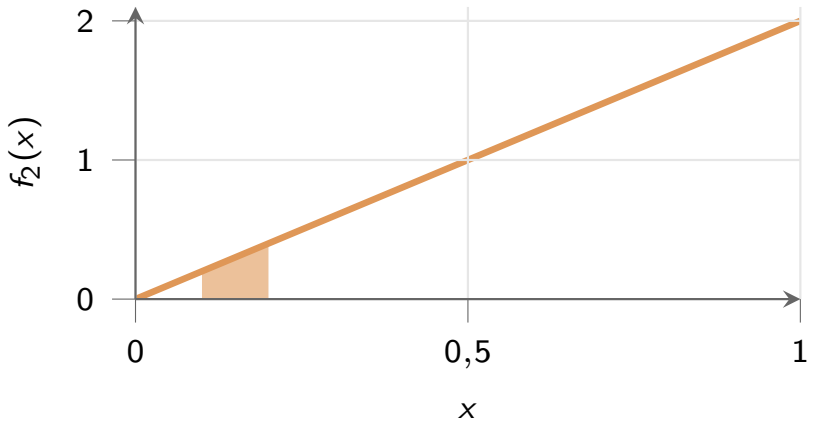
$$\mathbb{P}(0.1 \leqslant X \leqslant 0.2)\,.$$

$f_1(x) = 1$



$$\mathbb{P}(0.1 \leqslant X \leqslant 0.2) = \int_{0.1}^{0.2} 1\,dx = (0.2 - 0.1) \cdot 1 = 0.1$$

$f_2(x) = 2x$



$$\mathbb{P}(0.1 \leqslant X \leqslant 0.2) = \int_{0.1}^{0.2} 2x\,dx = \left[x^2\right]_{0.1}^{0.2} = 0.04 - 0.01 = 0.03.$$

Očekivanje i varijanca neprekidne slučajne varijable

Ako X ima gustoću f, definiramo

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx,$$

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x)\, dx.$$

Opet vrijedi

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

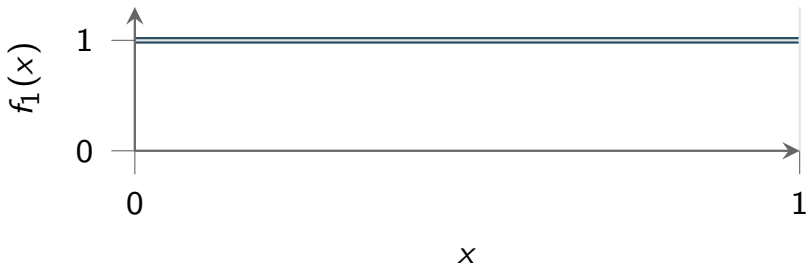
pri čemu je

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx\,.$$

Napomena: Ako je f pozitivna samo na segmentu $[a, b]$ (tj. X poprima vrijednosti samo u $[a, b]$) možemo u gornjim formulama zamijeniti $\int_{-\infty}^{\infty}$ sa $\int_a^b$.

Nastavak primjera: očekivanje i varijanca

$f_1(x) = 1$

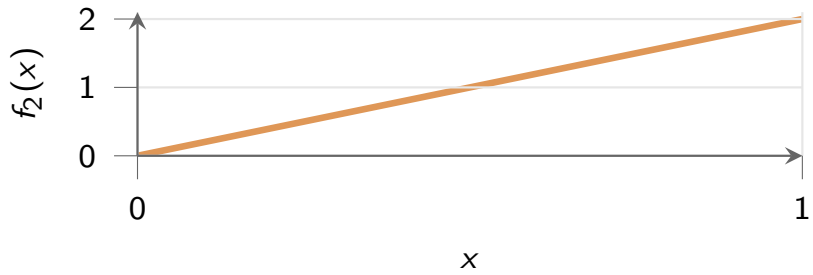


$$\mathbb{E}[X] = \int_0^1 x \cdot 1 \, dx = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2} - 0 = 0.5 \quad \text{(očekivano?)},$$

$$\mathbb{E}[X^2] = \int_0^1 x^2 \cdot 1 \, dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3} - 0 \approx 0.3333,$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \approx 0.0833$$

$f_2(x) = 2x$



$$\mathbb{E}[X] = \int_0^1 x \cdot 2x \, dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3} - 0 \approx 0.6667,$$

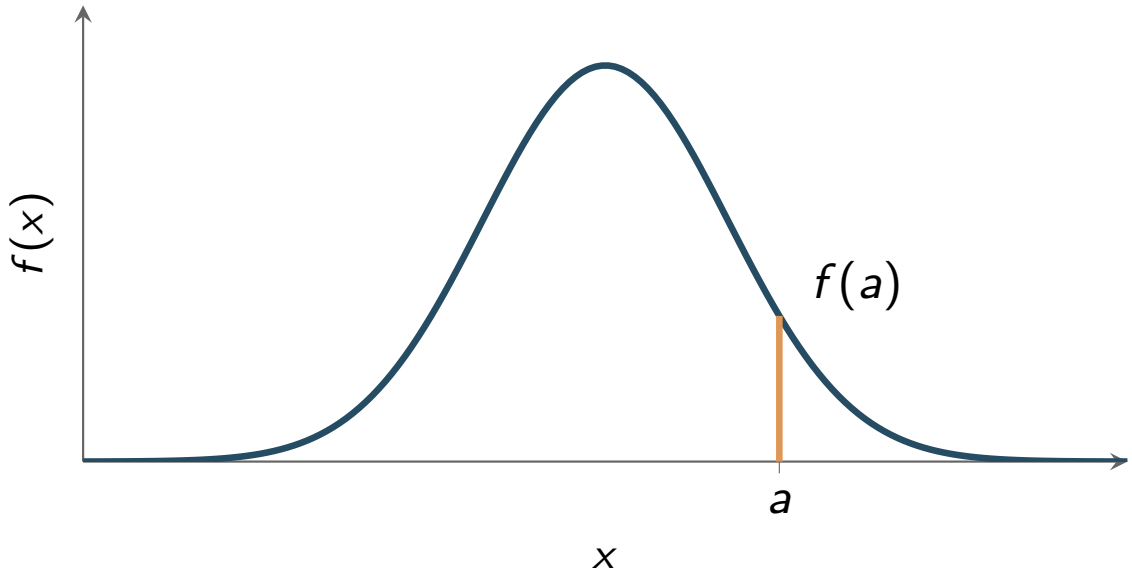
$$\mathbb{E}[X^2] = \int_0^1 x^2 (2x) \, dx = \left[\frac{x^4}{2}\right]_0^1 = \frac{1}{2} - 0 = 0.5,$$

$$\mathrm{Var}(X) \approx 0.0556\,.$$

Napomena – razlika u odnosu na diskretne varijable

Ako je X proizvoljna neprekidna slučajna varijabla, za bilo koju **unaprijed odabranu i fiksnu** vrijednost $a \in \mathbf{R}$ vrijedi

$$\mathbb{P}(X = a) = \int_a^a f(x)\,dx = 0.$$

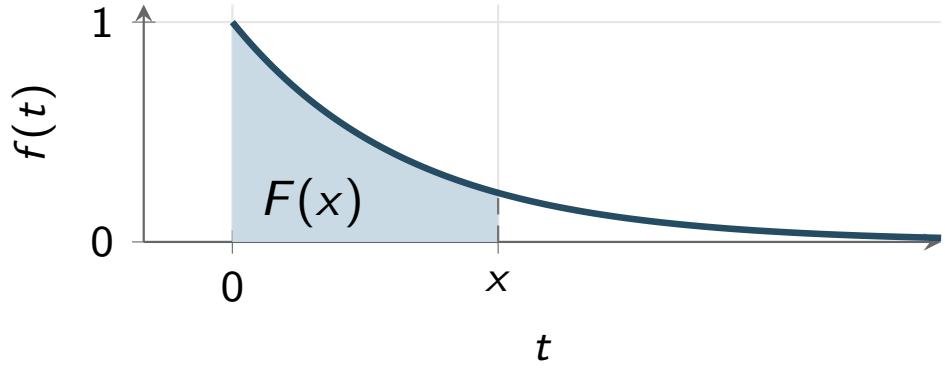


Funkcija distribucije

Za neprekidnu slučajnu varijablu X s gustoćom f **funkciju distribucije** F definiramo s

$$F(x) := \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(t)\,dt.$$

Primjer. Dolje je gustoća pozitivna samo na $[0, \infty)$.



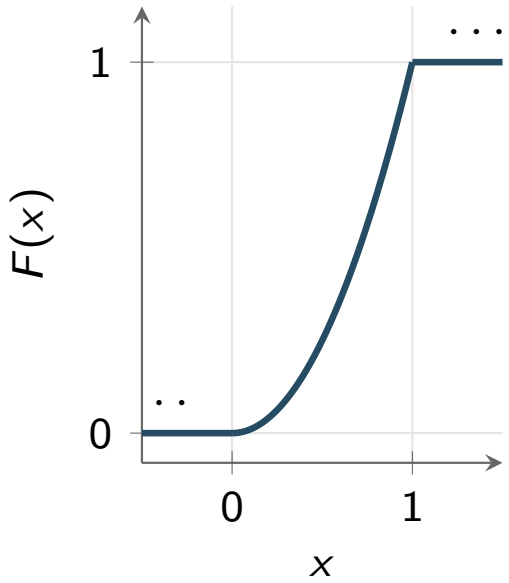
Za $a < b$ vrijedi

$$\begin{aligned}\mathbb{P}(a \leqslant X \leqslant b) &= \int_{a}^{b} f(t)\,dt \\ &= \int_{-\infty}^{b} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt \\ &= F(b) - F(a).\end{aligned}$$

Funkcija distribucije za gustoću f_2

Za gustoću $f_2(x) = 2x$ na $[0, 1]$ funkcija distribucije je

$$F(x) = \mathbb{P}(X \leqslant x) = \begin{cases} 0, & x < 0, \\ \int_0^x 2t\,dt = x^2, & 0 \leqslant x \leqslant 1, \\ 1, & x > 1. \end{cases}$$



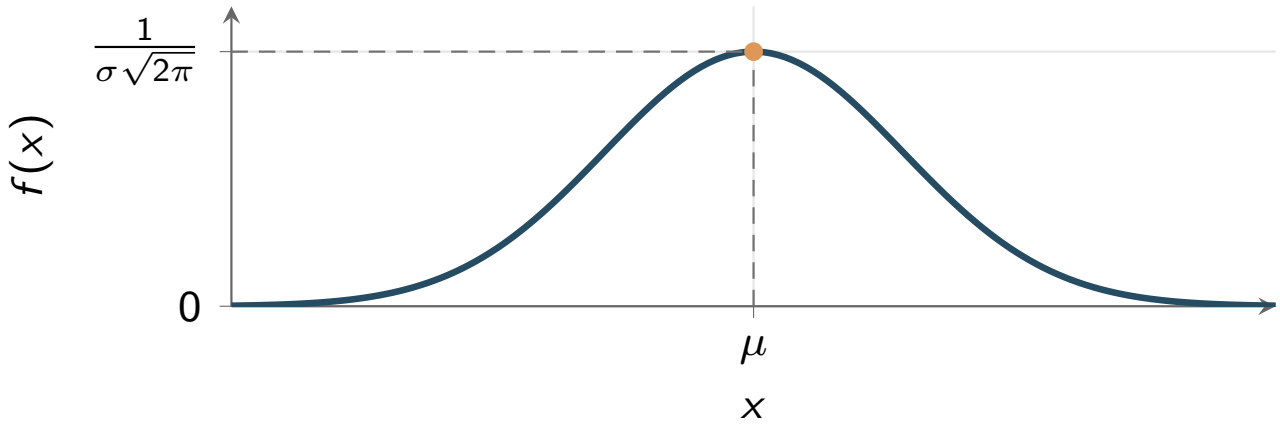
Sada možemo npr. izračunati

$$\begin{aligned} \mathbb{P}(0.1 \leqslant X \leqslant 0.2) &= F(0.2) - F_2(0.1) \\ &= 0.2^2 - 0.1^2 \\ &= 0.04 - 0.01 = 0.03. \end{aligned}$$

Normalna ili Gaussova razdioba

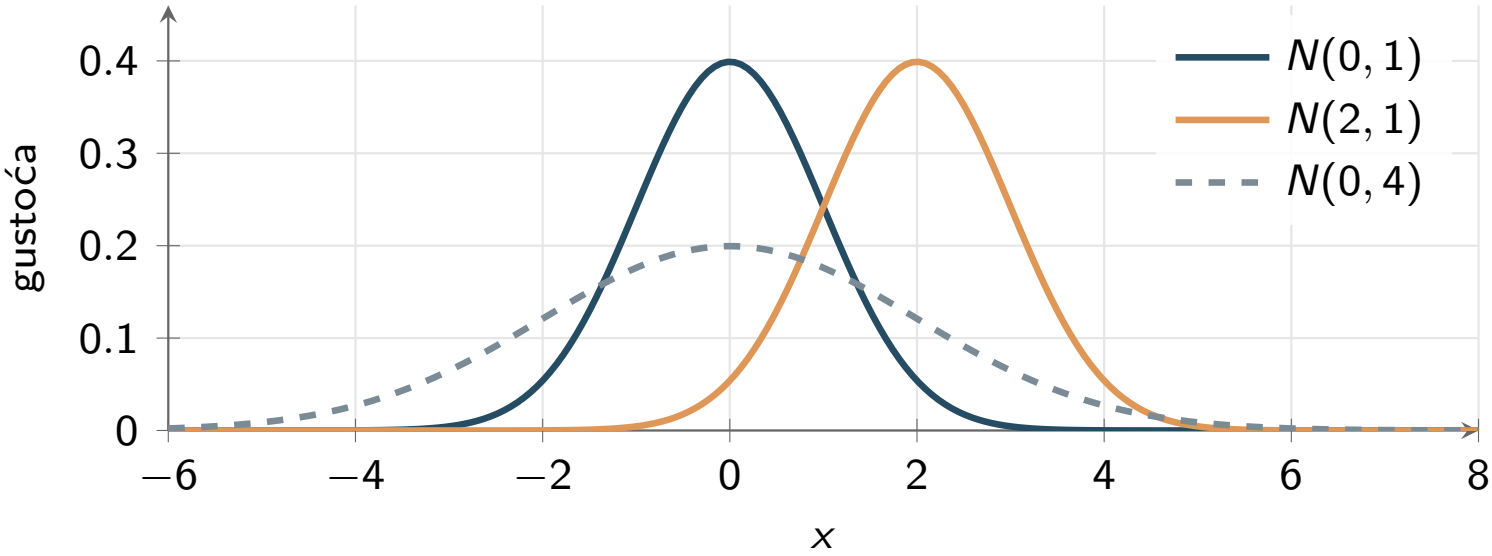
Za $\mu \in \mathbb{R}$ i $\sigma > 0$, kažemo da X ima **normalnu razdiobu** s parametrima μ i σ^2 (oznaka $X \sim N(\mu, \sigma^2)$) ako neprekdna s gustoćom

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



- Vrijedi $\mathbb{E}[X] = \mu$ (je li to očito?) te $\mathrm{Var}(X) = \sigma^2$.
- Budući da je f **svugdje pozitivna**, normalna slučajna varijabla može poprimiti **bilo koju** vrijednost iz $\mathbb{R}$.

Tri normalne gustoće



- gustoća je uvijek simetrična oko μ;
- parametar σ^2 kontrolira *raširenost* gustoće.

Linearne transformacije ostaju normalne

- Ako je $X \sim N(\mu, \sigma^2)$, onda je za sve $a, b \in \mathbf{R}$, $a \neq 0$ i $aX + b$ također **normalno distribuirana** – njeni parametri su

$$\mu_{novi} = \mathbb{E}[aX + b] = a\mu + b, \quad \sigma^2_{novi} = \mathrm{Var}(aX + b) = a^2\sigma^2 .$$

- Specijalno, ako je $X \sim N(\mu, \sigma^2)$ vrijedi

$$\mathbb{E}\left[\frac{X - \mu}{\sigma}\right] = 0 , \quad \mathrm{Var}\left(\frac{X - \mu}{\sigma}\right) = 1$$

pa dakle "standardizacirana" varijabla X zadovoljava

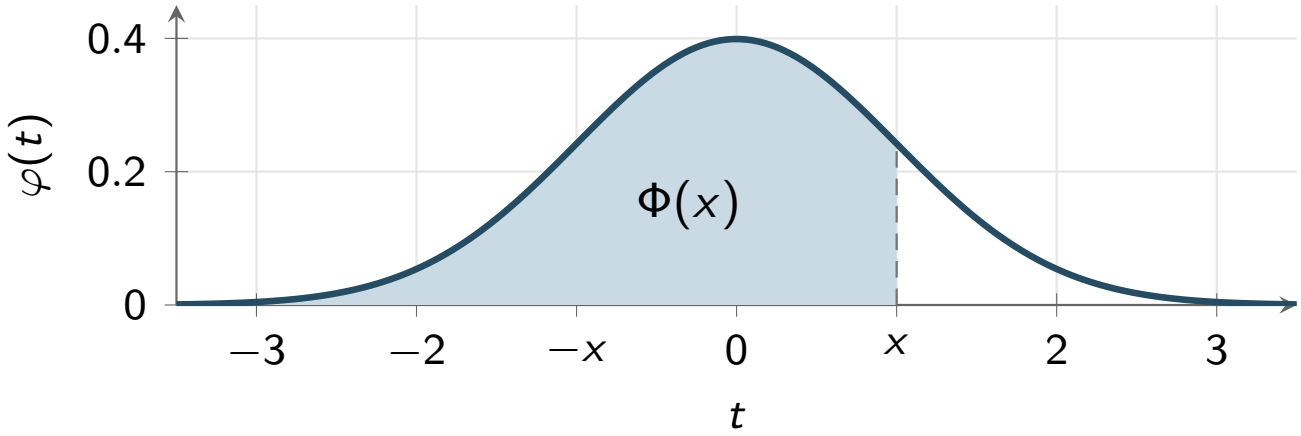
$$\frac{X - \mu}{\sigma} \sim N(0, 1).$$

- $N(0, 1)$ nazivamo **standardna normalna razdioba**.

Standardna normalna gustoća

Gustoća standardne normalne razdiobe jest

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$



Funkciju distribucije standardne normalne razdiobe označavamo s Φ. Zbog simetrije gustoće oko nule vrijedi $\Phi(-x) = 1 - \Phi(x)$.

Svođenje na standardnu normalnu

- preko Φ možemo odrediti vjerojatnosti $\mathbb{P}(a \leqslant X \leqslant b)$ za proizvoljnu $X \sim N(\mu, \sigma^2)$;
- zaista, vrijedi
$$Z := \frac{X - \mu}{\sigma} \sim N(0, 1);$$
- specijalno, $\mathbb{P}(c \leqslant Z \leqslant d) = \Phi(d) - \Phi(c)$;
- sada dobivamo
$$\begin{aligned} \mathbb{P}(a \leqslant X \leqslant b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{b-\mu}{\sigma}\right) \\ &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leqslant Z \leqslant \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \end{aligned}$$

Zadatak

Pretpostavimo da je visina slučajno odabrane osobe X ima $N(172, 8^2)$ razdiobu. Skicirajte površinu koja odgovara vjerojatnosti

$$\mathbb{P}(164 \leqslant X \leqslant 180)$$

te ju odredite koristeći vrijednosti funkcije Φ i identitet $\Phi(-x) = 1 - \Phi(x)$. Dane su neke vrijednosti funkcije Φ:

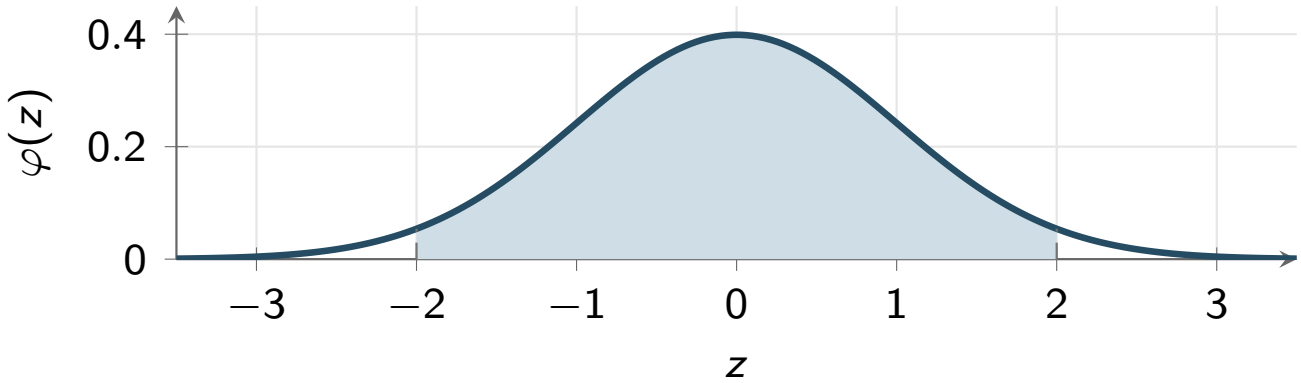
$$\Phi(0.5) = 0.6915, \qquad \Phi(1) = 0.8413, \qquad \Phi(1.5) = 0.9332, \qquad \Phi(2) = 0.9772.$$

Rješenje na ploči. (vjerojatnost je 0.6826)

"68–95–99.7 rule"

Za $X \sim N(\mu, \sigma^2)$ neka je $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$. Vrijedi

$$\begin{aligned}
\mathbb{P}(\mu - \sigma \leqslant X \leqslant \mu + \sigma) &= \mathbb{P}(-1 \leqslant Z \leqslant 1) \approx 0.6827, \\
\mathbb{P}(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) &= \mathbb{P}(-2 \leqslant Z \leqslant 2) \approx 0.9545, \\
\mathbb{P}(\mu - 3\sigma \leqslant X \leqslant \mu + 3\sigma) &= \mathbb{P}(-3 \leqslant Z \leqslant 3) \approx 0.9973.
\end{aligned}$$



Dakle, za podatke koji se mogu modelirati nekom normalnom razdiobom, približno 95% podataka biti će **unutar 2 standardne devijacije** od njenog očekivanja.

4.4. Normalna aproksimacija (za binomnu razdiobu)

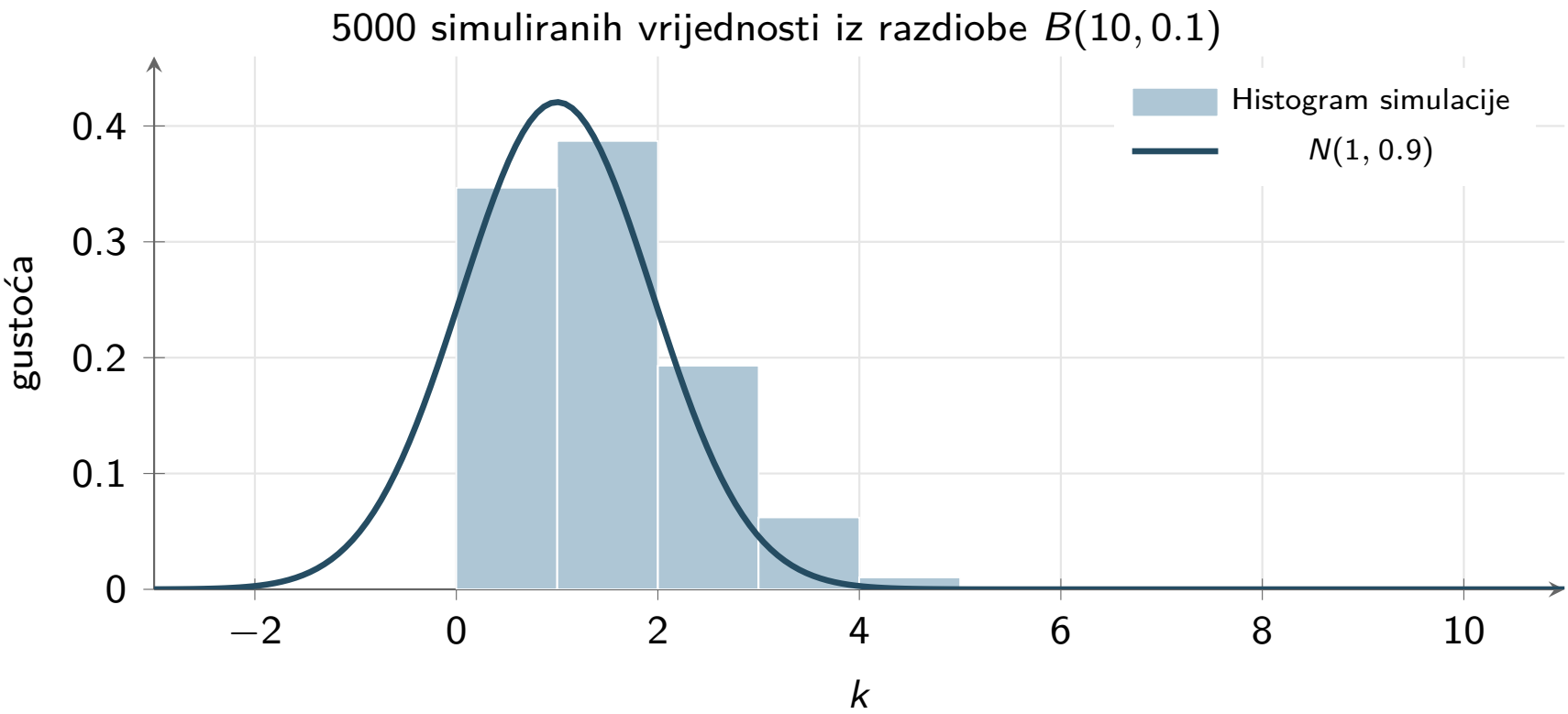
- Neka je $X \sim B(n, p)$ i $q = 1 - p$.
- Za fiksni $p \in (0, 1)$ i veliki n, tzv. **centralni granični teorem** (CGT) kaže da vrijedi

$$X \approx N(\mathbf{E}[X], \mathrm{Var}(X)) = N(np, npq).$$

- preciznije,

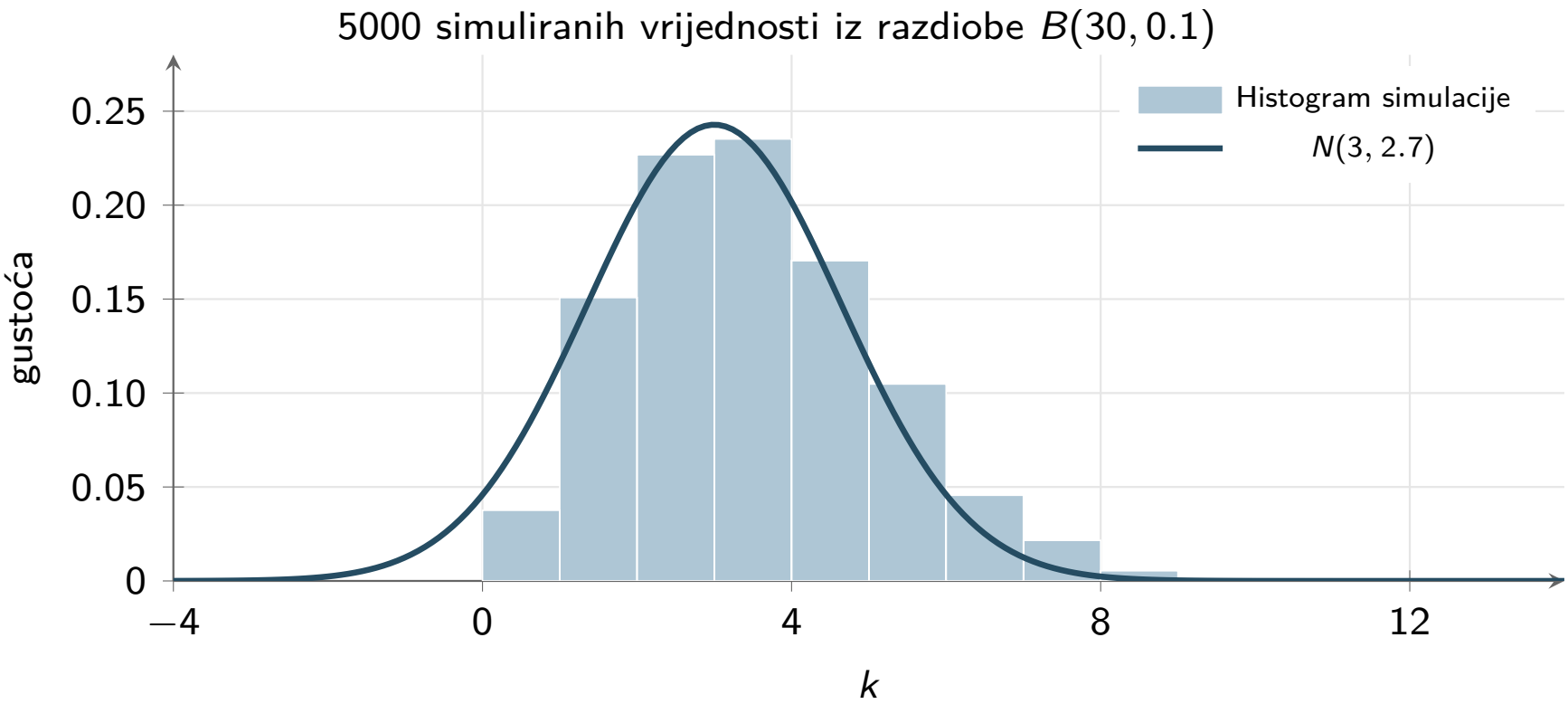
$$\begin{aligned}\mathbb{P}(a \leqslant X \leqslant b) &\approx \mathbb{P}(a \leqslant N(np, npq) \leqslant b) \\ &= \mathbb{P}\left(\frac{a - np}{\sqrt{npq}} \leqslant N(0, 1) \leqslant \frac{b - np}{\sqrt{npq}}\right) \\ &= \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right).\end{aligned}$$

Ilustracija CGT-a: $n = 10$



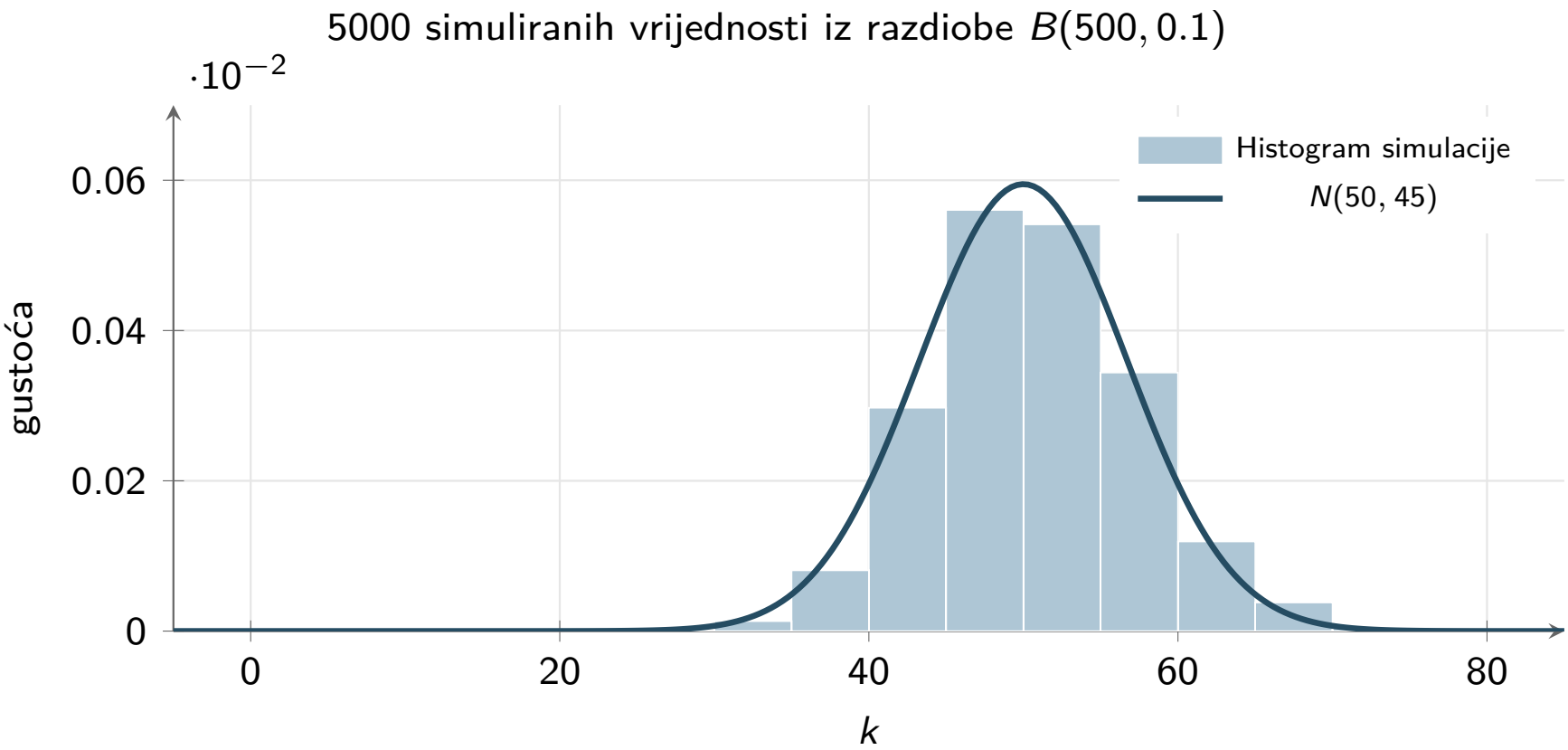
Ovdje je $np = 1$, $npq = 0.9$.

Ilustracija CGT-a: $n = 30$



Ovdje je $np = 3$, $npq = 2.7$.

Ilustracija CGT-a: $n = 500$



Ovdje je $np = 50$, $npq = 40$.

Korekcija po neprekidnosti

- ako je $X \sim B(n,p)$ i n velik, CGT dakle kaže da vrijede aproksimacije

$$\mathbb{P}(a \leqslant X \leqslant b) \approx \mathbb{P}(a \leqslant N(np, npq) \leqslant b)$$
$$\mathbb{P}(X \geqslant a) \approx \mathbb{P}(N(np, npq) \geqslant a)$$

  i slično;

- ako su $a, b \in \mathbf{N}_0$, često koristimo sljedeću korekciju koja uglavnom daje **bolju aproksimaciju**:

$$\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a - 1/2 \leqslant X \leqslant b + 1/2) \approx \mathbb{P}(a - 1/2 \leqslant N(np, npq) \leqslant b + 1/2)$$
$$\mathbb{P}(X \geqslant a) = \mathbb{P}(X \geqslant a - 1/2) \approx \mathbb{P}(N(np, npq) \geqslant a - 1/2)$$

  i slično. Prve jednakosti vrijede jer X poprima samo vrijednosti u $\{0, 1, 2, \ldots, n\}$.

Zadatak: bacanje kocke

Simetričnu kocku bacamo 600 puta. Neka je X broj dobivenih šestica (dakle $\mathbb{E}[X] = 100$). Koristeći CGT, tj. normalnu aproksimaciju, za binomnu razdiobu (sa i bez korekcije po neprekidnosti) procijenite

$$\mathbb{P}(90 \leqslant X \leqslant 110).$$

Rješenje

- Vrijedi $X \sim B(n,p)$ za $n = 600$, $p = 1/6$. Dakle,

$$\mathbb{E}[X] = np = 100, \qquad \mathrm{Var}(X) = npq = 600 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{250}{3} \approx 83.33,$$

  pa je $\sigma = \sqrt{83.33} \approx 9.13$.
- Neka je $Y \sim N(np, npq) = N(100, 83.33)$.

Bez korekcije po neprekidnosti

$$\begin{aligned}
\mathbb{P}(90 \leqslant X \leqslant 110) &\approx \mathbb{P}(90 \leqslant Y \leqslant 110) \\
&= [\text{kao prije}] \\
&= \Phi\left(\frac{10}{9.13}\right) - \Phi\left(-\frac{10}{9.13}\right) \\
&= 2\Phi(1.10) - 1 \\
&= 2(0.8643) - 1 \\
&= \boxed{0.7286}.
\end{aligned}$$

S korekcijom po neprekidnosti

$$\begin{aligned}
\mathbb{P}(90 \leqslant X \leqslant 110) &= \mathbb{P}(89.5 \leqslant X \leqslant 110.5) \\
&\approx \mathbb{P}(89.5 \leqslant Y \leqslant 110.5) \\
&= [\text{kao prije}] \\
&= \Phi\left(\frac{10.5}{9.13}\right) - \Phi\left(-\frac{10.5}{9.13}\right) \\
&= 2\Phi(1.15) - 1 \\
&= 2 \cdot (0.8749) - 1 \\
&= \boxed{0.7498}.
\end{aligned}$$

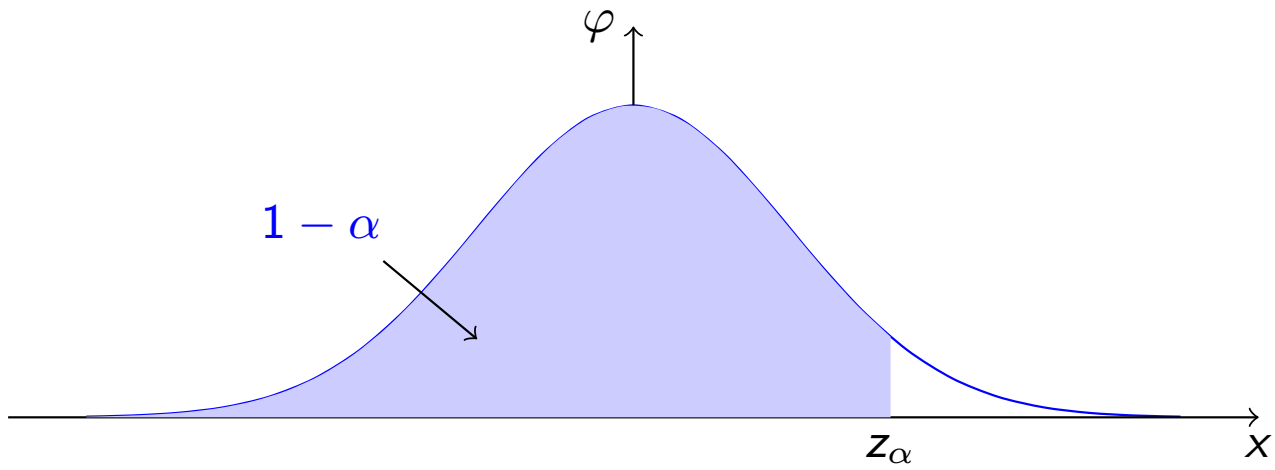
Napomena. Točna vrijednost za vjerojatnost je 0.7501.

Kvantili standardne normalne distribucije

Općenito, za $\alpha \in (0,1)$, broj z_α za koji vrijedi da je

$$\mathbb{P}(N(0,1) \leqslant z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$$

zovemo $(1-\alpha)$-kvantil standardne normalne razdiobe.

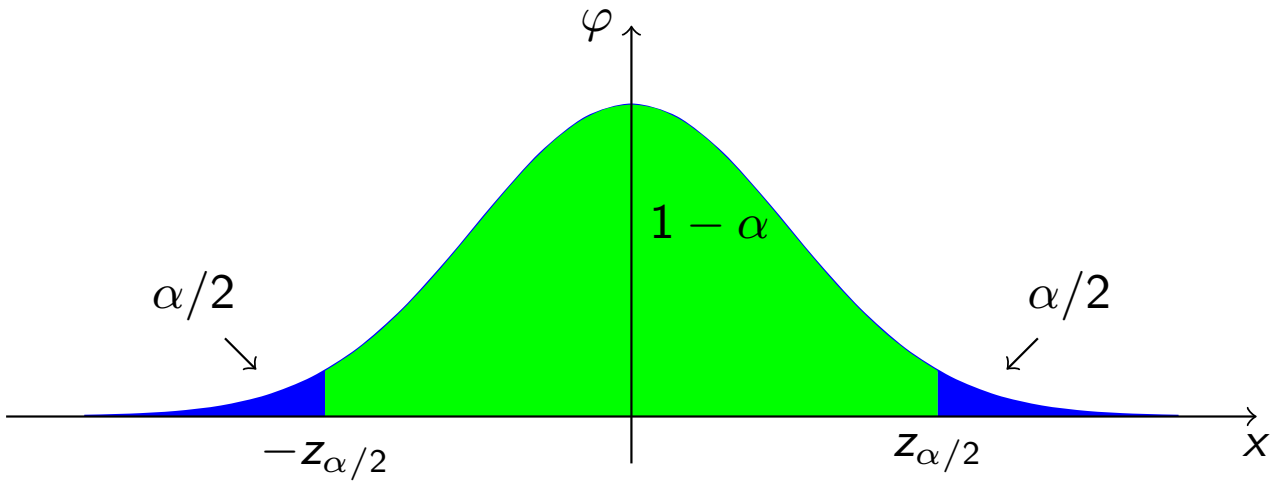


- vrijedi $\mathbb{P}(N(0,1) > z_\alpha) = 1 - \Phi(z_\alpha) = \alpha$;
- npr. $z_{0.05} \approx 1.645$ i $z_{0.025} \approx 1.96$.

Kvantili standardne normalne distribucije

Neka je $Z \sim N(0,1)$. Uočimo da je zbog simetrije standardne normalne razdiobe

$$\mathbf{P}(-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}) = 1-\alpha,$$



Za općenitu $X \sim N(\mu, \sigma^2)$ vrijedi

$$\mathbf{P}(\mu - z_{\alpha/2}\sigma \leqslant X \leqslant \mu + z_{\alpha/2}\sigma) = 1-\alpha\,.$$

Zadatak: klijanje sjemenki

U pokusu je posijano 400 sjemenki. Svaka sjemenka, nezavisno od ostalih, proklija s vjerojatnošću 0.6. Neka je X broj sjemenki koje proklijaju.
Uz normalnu aproksimaciju binomne razdiobe (bez korekcije po neprekidnosti) odredite **najmanji prirodni broj** k takav da vrijedi

$$\mathbb{P}(X \geqslant k) < 0.10.$$

Rješenje: klijanje sjemenki

- Vrijedi $X \sim B(400, 0.6)$, pa
$$\mathbb{E}[X] = 240, \qquad \mathrm{Var}(X) = 96, \qquad X \approx Y \sim N(240, 96).$$
- Po CGT-u,
$$\mathbb{P}(X \geqslant k) \approx \mathbb{P}(Y \geqslant k) = 1 - \mathbb{P}(Y \leqslant k) = 1 - \Phi\left(\frac{k - 240}{\sqrt{96}}\right).$$
- Dakle, imamo $\mathbb{P}(X \geqslant k) < 0.1$ ako
$$1 - \Phi\left(\frac{k - 240}{\sqrt{96}}\right) < 0.1\,, \text{tj.} \quad \Phi\left(\frac{k - 240}{\sqrt{96}}\right) > 0.9$$
- ako je z takav da je $\Phi(z) = \mathbb{P}(N(0, 1) \leqslant z) \approx 0.9$, zaključujemo da tražimo k-ove za koje je
$$\frac{k - 240}{\sqrt{96}} > z\,.$$

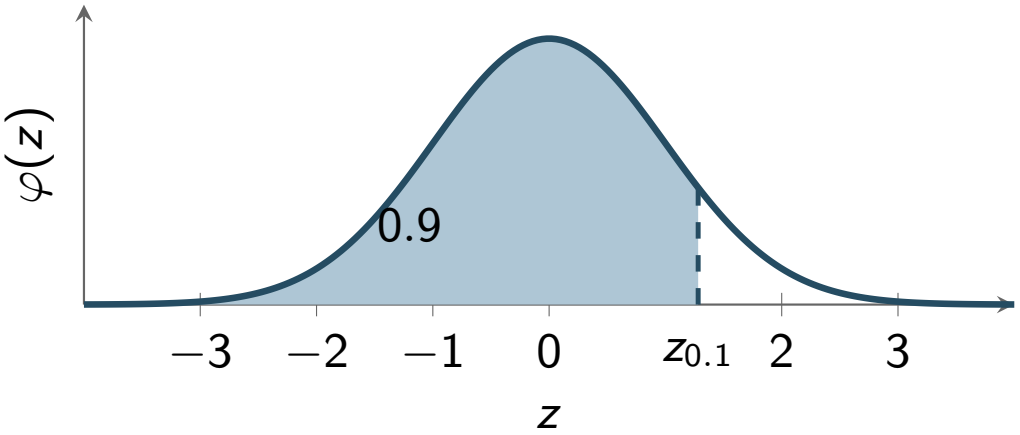
Rješenje: klijanje sjemenki

- broj z je zapravo kvantil z_α za $\alpha = 0.1$ jer je u tom slučaju $\Phi(z_\alpha) = 1 - \alpha = 0.9$;
- budući da je $z_{0.1} \approx 1.28$ tražimo najmanji k takav da vrijedi

$$k > 240 + 1.28\sqrt{96} \approx 252.54.$$

Dakle, najmanji takav cijeli broj je

$$\boxed{k = 253}.$$



Provjera: za $k = 252$ aproksimacija je oko 0.111, a za $k = 253$ oko 0.092.