## Atsitiktiniai dydžiai ir jų skirstiniai.
## Bernulio schemos tikimybių aproksimacijos

Prieš pradedant spręsti 7 užduotį, reikėtų prisiminti Bernulio schemos tikimybių apytikslio skaičiavimo formules, o prieš atliekant 8 ir 9 užduotis, praverstų peržiūrėti paskaitas apie atsitiktinius dydžius ir jų savybes.

Trumpai priminsime kai kurias sąvokas. Bet kokį atsitiktinį dydį $X$ nusako jo *pasiskirstymo funkcija* $F(x)$

$$F(x) := P(X < x)\,, \quad x \in \mathbb{R}\,.$$

Žinodami atsitiktino dydžio pasiskirstymo funkciją, galime nesunkiai rasti patekimo į intervalą $[a, b)$ tikimybę

$$P(a \leq X < b) = F(b) - F(a). \tag{1}$$

Paprasčiausios atsitiktinio dydžio skaitinės charakteristikos yra vidurkis $EX$ ir dispersija $DX = EX^2 - (EX)^2$. Jų skaičiavimo būdas priklauso nuo atsitiktinio dydžio tipo.

Aptarsime keletą klasikinių atsitiktinių dydžių, kartu nurodydami su jais susijusias *Excel* funkcijas. Skaičiavimams galima naudoti ir analogiškas statistinio paketo *R* arba *Python* bibliotekos *SciPy* funkcijas. Dažnai bus pakankama ir kokia nors "gudresnė" skaičiuoklė, pavyzdžiui:
https://www.statskingdom.com/index.html
https://www.123calculus.com/en

**Diskretieji atsitiktiniai dydžiai.** Jų reikšmių aibė yra baigtinė arba skaiti. Tokie atsitiktiniai dydžiai $X$ dažniausiai nusakomi *reikšmių skirstiniu*, nurodant galimas reikšmes $k$ ir jų tikimybes $P(X = k)$. Tada pasiskirstymo funkcija ir vidurkis yra:

$$F(x) = \sum_{k<x} P(X = k)\,, \quad EX = \sum_{k} k\, P(X = k)\,.$$

Pastebėsime, kad $F(x + 0) = F(x) + P(X = x)$.

- *Binominis atsitiktinis dydis* $X \sim \mathcal{B}(n; p)$. Jo galimos reikšmės $k = 0, 1, \ldots, n$.

$$\begin{aligned} P(X = k) &= C_n^k\, p^k (1-p)^{n-k} = \texttt{binom.dist}(k; n; p; \texttt{false}), \\ F(x+0) &= \sum_{k \leq x} C_n^k\, p^k (1-p)^{n-k} = \texttt{binom.dist}(x; n; p; \texttt{true}), \\ & EX = np, \quad DX = np(1-p), \end{aligned} \tag{2}$$

  čia $x \leq n, \quad 0 \leq p \leq 1$.

- *Geometrinis atsitiktinis dydis* $X \sim \mathcal{G}(p)$. Jo galimos reikšmės $k = 1, 2, \ldots$.

$$\begin{aligned} P(X = k) &= (1-p)^{k-1} p, \\ F(x+0) &= p \sum_{k \leq x} (1-p)^{k-1} = 1 - (1-p)^{[x]}, \\ & EX = \frac{1}{p}, \quad DX = \frac{1-p}{p^2}, \end{aligned} \tag{3}$$

  čia $x \geq 1, \quad 0 < p < 1$.

- *Puasono atsitiktinis dydis* $X \sim \mathcal{P}(\lambda), \quad \lambda > 0$. Jo galimos reikšmės $k = 0, 1, 2, \ldots$.

$$P(X = k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} = \texttt{poisson.dist}(k; \lambda; \texttt{false}), \tag{4}$$

$$F(x+0) = \mathrm{e}^{-\lambda} \sum_{k \leq x} \frac{\lambda^k}{k!} = \texttt{poisson.dist}(x; \lambda; \texttt{true}), \tag{5}$$

$$EX = DX = \lambda\,.$$

**Absoliučiai tolydūs atsitiktiniai dydžiai.** Tokio atsitiktinio dydžio $X$ patekimo į intervalą $[a, b]$ tikimybė yra

$$P(a \le X \le b) = \int_a^b p(x)\,dx\,, \quad p(x) \ge 0\,.$$

Funkcija $p(x)$ vadinama $X$ *tankio funkcija* arba tiesiog tankiu. Tada

$$EX = \int_{-\infty}^{\infty} xp(x)\,dx\,.$$

Pastebėsime, kad bet kokiam absoliučiai tolydžiam atsitiktiniam dydžiui $X$ ir realiajam skaičiui $a$ "taškinė" tikimybė $P(X = a)$ lygi 0, o tankio funkcija $p(x)$ tenkina sąlygą

$$\int_{-\infty}^{\infty} p(x)\,dx = 1\,. \tag{6}$$

Kita vertus, pasirodo, kad bet kuri neneigiama funkcija $p(x)$, tenkinanti pastarąją lygybę, gali būti laikoma kažkokio atsitiktinio dydžio tankiu. Absoliučiai tolydžiojo atsitiktinio dydžio pasiskirstymo funkciją ir tankį sieja lygybės:

$$F(x) = \int_{-\infty}^{x} p(u)\,du\,, \quad F'(x) = p(x)\,. \tag{7}$$

- *Tolygiai pasiskirstęs intervale* $[a, b]$ *atsitiktinis dydis* $X \sim \mathcal{T}([a, b])$.

$$\begin{aligned} p(x) &= \begin{cases} 0, & \text{jei } x \notin [a,b], \\ \frac{1}{b-a}, & \text{jei } x \in [a,b]. \end{cases} \\ F(x) &= \begin{cases} 0, & \text{jei } x < a, \\ \frac{x-a}{b-a}, & \text{jei } x \in [a,b], \\ 1, & \text{jei } x > b. \end{cases} \\ & EX = \frac{a+b}{2}\,, \quad DX = \frac{(b-a)^2}{12}\,. \end{aligned} \tag{8}$$

- *Eksponentinis atsitiktinis dydis* $X \sim \mathcal{E}(\alpha)$, $\alpha > 0$.

$$\begin{aligned} p(x) &= \begin{cases} 0, & \text{jei } x < 0, \\ \alpha \mathrm{e}^{-\alpha x}, & \text{jei } x \ge 0. \end{cases} \\ F(x) &= \begin{cases} 0, & \text{jei } x < 0, \\ 1 - \mathrm{e}^{-\alpha x}, & \text{jei } x \ge 0. \end{cases} \\ & EX = \alpha^{-1}\,, \quad DX = \alpha^{-2}\,. \end{aligned} \tag{9}$$

- *Normalusis atsitiktinis dydis* $X \sim \mathcal{N}(a; \sigma^2)$, $a \in \mathbb{R}\,,\ \sigma > 0$.

$$\begin{aligned} p(x) &= \varphi(x; a, \sigma) := \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-a)^2}{2\sigma^2}} = \texttt{norm.dist}(x; a; \sigma; \texttt{false})\,. \\ F(x) &= \Phi(x; a, \sigma) := \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \mathrm{e}^{-\frac{(u-a)^2}{2\sigma^2}} du = \texttt{norm.dist}(x; a; \sigma; \texttt{true})\,. \\ & EX = a\,, \quad DX = \sigma^2\,. \end{aligned} \tag{10}$$

Kai $a = 0$, $\sigma = 1$, atsitiktinis dydis $X$ vadinamas *standartiniu normaliuoju* $X \sim \mathcal{N}(0;1)$. Jo tankio ir pasiskirstymo funkcijos yra
$\varphi(x) = \varphi(x;0,1) = \texttt{norm.s.dist}(x;\texttt{false})$,
$\Phi(x) = \Phi(x;0,1) = \texttt{norm.s.dist}(x;\texttt{true})$.

**Bernulio schemos tikimybių skaičiavimas.** Tegu sėkmės tikimybė viename Bernulio schemos bandyme lygi $p$, $n$ - bandymų skaičius, o $S_n$ - atsitiktinis dydis lygus gautų sėkmių skaičiui. Tada $S_n \sim \mathcal{B}(n;p)$. Labiausiai tikėtinos $S_n$ reikšmės yra vienas arba du sveikieji skaičiai iš intervalo $[np+p-1,\ np+p]$. Nedidelėms $n$ reikšmėms tikimybės $P(S_n = k)$ randamos naudojantis (2) formule. Kai eksperimentų daug, skaičiuojant šias tikimybes, galima pasinaudoti apytikslėmis formulėmis:

$$P(S_n = k) \approx \frac{(np)^k}{k!}\mathrm{e}^{-np}\,, \tag{11}$$

$$P(S_n = k) \approx \frac{1}{\sqrt{np(1-p)}}\varphi\left(\frac{k-np}{\sqrt{np(1-p)}}\right), \tag{12}$$

$$P(b < S_n < c) \approx \Phi\left(\frac{c-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right). \tag{13}$$

Apytikslėje lygybėje (11) paklaida neviršija $np^2$. Kai $np^2$ reikšmė yra didelė, geriau naudoti (12) arba (13) lygybes. Formulė (11) vadinama Puasono, o (12) ir (13) Muavro-Laplaso įverčiais.

**1 pavyzdys.** Tegu $X \sim \mathcal{B}(5;0,8)$. Raskite atsitiktinio dydžio $Y := X^2 - 1$ pasiskirstymo funkcijos $F_Y(x)$ reikšmę taške $x = 2,7$ ir $P(1 \le Y \le 10)$.
Pagal apibrėžimą

$$F_Y(2,7) = P(Y < 2,7) = P(X^2 < 3,7) = P(-\sqrt{3,7} < X < \sqrt{3,7}) = P(X=0) + P(X=1),$$

nes $X \sim \mathcal{B}(5;0,8)$ ir jo galimos reikšmės yra $0,1,2,3,4,5$. Taigi

$$F_Y(2,7) = C_5^0\,0,8^0 0,2^5 + C_5^1\,0,8^1 0,2^4 = 0,00672.$$

Žinodami, kad $X$ reikšmės yra neneigiamos, analogiškai gausime:

$$P(1 \le Y \le 10) = P(2 \le X^2 \le 11) = P(\sqrt{2} \le X \le \sqrt{11}) = P(X=2) + P(X=3) = 0,256.$$

**2 pavyzdys.** Tarkime, $X$- normalusis atsitiktinis dydis, t. y. $X \sim \mathcal{N}(-2;9)$, o $Y$ - atsitiktinai pasirinktas intervalo $[0,3]$ skaičius. Raskite atsitiktinio dydžio $Z := \max(X,Y)$ pasiskirstymo funkcijos $F_Z(x)$ reikšmę taške $x = 2$.
Atsitiktiniai dydžiai yra nepriklausomi, be to $Y \sim \mathcal{T}([0,3])$. Todėl

$$F_Z(2) = P(\max(X,Y) < 2) = P(X < 2)P(Y < 2) = F_X(2)F_Y(2)\,.$$

Panaudoję (8) ir (10) lygybes, gausime $F_Y(2) = 2/3$ ir

$$F_X(2) = \Phi(2;-2,3) = \texttt{norm.dist}(2;-2;3;\texttt{true}) \approx 0,90879\,.$$

Taigi $F_Z(2) \approx 0,60589$.

**3 pavyzdys.** Atsitiktinio dydžio $X$ tankio funkcija yra

$$p(x) = \begin{cases} 0, & \text{jei } x \notin [0,\pi], \\ C\sin x, & \text{jei } x \in [0,\pi]. \end{cases}$$

Raskite pasiskirstymo funkcijos $F(x)$ reikšmę taške $x = \pi/2$ ir $P(X > \pi/3)$.

Pažymėkime $\mathbf{1}_A(x)$ aibės $A$ indikatorių

$$\mathbf{1}_A(x) := \begin{cases} 0, & \text{jei } x \notin A, \\ 1, & \text{jei } x \in A. \end{cases}$$

Tada $p(x) = \mathbf{1}_{[0,\pi]}(x) \cdot C \sin x$. Pirmiausiai rasime konstantą $C$. Pasinaudoję tankio funkcijos savybe (6), gausime

$$\int_{-\infty}^{\infty} p(x)\, dx = C \int_0^{\pi} \sin x\, dx = -C \cdot \cos x \Big|_0^{\pi} = 2C = 1.$$

Taigi $C = 1/2$ ir

$$F(x) = \int_{-\infty}^{x} p(u)\, du = \frac{\mathbf{1}_{[0,\pi]}(x)}{2}(1 - \cos x) + \mathbf{1}_{(\pi,\infty)}(x) = \begin{cases} 0, & \text{jei } x < 0, \\ \frac{1}{2}(1 - \cos x), & \text{jei } x \in [0, \pi], \\ 1, & \text{jei } x > \pi. \end{cases}$$

Dabar jau nesunkiai randame:

$$F\left(\frac{\pi}{2}\right) = \frac{1}{2}\left(1 - \cos\frac{\pi}{2}\right) = \frac{1}{2},$$

$$P(X > \pi/3) = 1 - P(X \le \pi/3) = 1 - (P(X < \pi/3) + P(X = \pi/3)) = 1 - \left(F\left(\frac{\pi}{3}\right) + 0\right) = 1 - \frac{1}{4} = \frac{3}{4},$$

nes $P(X = a) = 0$ bet kokiam $a \in \mathbb{R}$ ir absoliučiai tolydžiajam dydžiui $X$.

**4 pavyzdys.** Informacija perduodama triukšmingu kanalu, kuris vidutiniškai iškraipo 0,1% visų siunčiamų bitų. Siunčiamas 7 kB failas. Koks labiausiai tikėtinas iškraipytų bitų skaičius? Kokia tikimybė, kad iškraipytų bitų skaičius nuo labiausiai tikėtino skirsis ne daugiau kaip 10?

Šiuo atveju sėkmė - gauti iškraipytą bitą. Pagal sąlygą tokios "sėkmės" tikimybė kiekvienu atveju yra $p = 0,001$. Mums reikalingas labiausiai tikėtinas "sėkmių" skaičius po $n = 7 \cdot 1024 \cdot 8 = 57344$ bandymų. Tam, kad jį rastume, nėra būtina skaičiuoti tikimybes $P(S_{57344} = k)$ visoms 57345 galimoms $k$ reikšmėms. Kadangi $np+p = 57,345$, tai labiausiai tikėtinas iškraipytų bitų skaičius $k_0$ priklauso intervalui $[56,345; 57,345]$, taigi $k_0 = 57$. Be to, pritaikę (11) ir (4) lygybes, gausime

$$P(S_{57344} = 57) \approx \texttt{poisson.dist}(57; 57,344; \texttt{false}) \approx 0,0527\,.$$

Atsakymą į antrąjį klausimą sužinosime pasinaudoję (11) ir (5) lygybėmis. Taigi

$$P(47 \le S_{57344} \le 67) \approx \texttt{poisson.dist}(67; 57,344; \texttt{true}) - \texttt{poisson.dist}(46; 57,344; \texttt{true}) \approx 0,83515$$

**5 pavyzdys.** Tikimybė, kad į fakultetą įstojęs moksleivis sėkmingai baigs studijas, lygi 0,8. Kiek mažiausiai reikėtų priimti pirmakursių, kad su tikimybe ne mažesne kaip 90% studijas sėkmingai baigtų ne mažiau kaip 100 studentų?

Tegu $n \ge 100$ - priimtų pirmakursių skaičius. Sėkmės tikimybė kiekvienam iš jų yra $p = 0,8$. Sėkmingai studijas baigusiųjų skaičius $S_n \sim \mathcal{B}(n; p)$. Reikia rasti mažiausią $n$, kuriam

$$P(100 \le S_n \le n) \ge 0,9.$$

Pasinaudoję (13) lygybe, gausime

$$P(100 \le S_n \le n) \approx \Phi\left(\frac{\sqrt{n}}{2}\right) - \Phi\left(\frac{100 - 0,8n}{0,4\sqrt{n}}\right) \ge 0,9.$$

Kai $n \geq 100$, tai

$$\Phi\left(\frac{\sqrt{n}}{2}\right) \geq \Phi(5) \approx 1.$$

Todėl lieka rasti mažiausią nelygybės

$$\Phi\left(\frac{100 - 0,8n}{0,4\sqrt{n}}\right) \leq 0,1 \tag{14}$$

sprendinį. Pirmiausiai rasime standartinio normaliojo skirstinio $\alpha = 0,1$ lygio kvantilį $z_{0,1}$, t.y. lygties

$$\Phi(z_{0,1}) = 0,1$$

sprendinį. Čia praverčia dar viena Excel funkcija: $z_{0,1} = \texttt{norm.s.inv}(0,1) \approx -1,28$. Funkcija $\Phi(x)$ yra monotoniškai didėjanti. Todėl iš (14) nelygybės gauname, kad

$$\frac{100 - 0,8n}{0,4\sqrt{n}} \leq -1,28.$$

Nesunku patikrinti, kad mažiausias sveikas skaičius, tenkinantis pastarąją nelygybę, yra $n = 133$.

Lentelėje pateikiamos skaičiavimams skirtos funkcijos, susijusios su kai kurių klasikinių atsitiktinių dydžių skirstiniais.

| **Formulė** | *Excel* **funkcija** | *Python* `scipy.stats` **funkcija** |
|---|---|---|
| *Binominis atsitiktinis dydis* $X \sim \mathcal{B}(n;p)$ | | |
| $P(X=k) = C_n^k p^k (1-p)^{n-k}$ | `binom.dist`$(k;n;p;$`false`$)$ | `binom.pmf`$(\mathrm{k}=k, \mathrm{n}=n, \mathrm{p}=p)$ |
| $P(X \le x) = \sum_{k \le x} C_n^k p^k (1-p)^{n-k}$ | `binom.dist`$(x;n;p;$`true`$)$ | `binom.cdf`$(\mathrm{k}=x, \mathrm{n}=n, \mathrm{p}=p)$ |
| *Puasono atsitiktinis dydis* $X \sim \mathcal{P}(\lambda)$ | | |
| $P(X=k) = \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$ | `poisson.dist`$(k;\lambda;$`false`$)$ | `poisson.pmf`$(\mathrm{k}=k, \mathrm{mu}=\lambda)$ |
| $P(X \le x) = \mathrm{e}^{-\lambda} \sum_{k \le x} \frac{\lambda^k}{k!}$ | `poisson.dist`$(x;\lambda;$`true`$)$ | `poisson.cdf`$(\mathrm{k}=x, \mathrm{mu}=\lambda)$ |
| *Normalusis atsitiktinis dydis* $X \sim \mathcal{N}(a;\sigma^2)$ | | |
| $\varphi(x;a,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x-a)^2}{2\sigma^2}}$ | `norm.dist`$(x;a;\sigma;$`false`$)$ | `norm.pdf`$(\mathrm{x}=x, \mathrm{loc}=a, \mathrm{scale}=\sigma)$ |
| $\varphi(x) = \varphi(x;0,1)$ | `norm.s.dist`$(x;$`false`$)$ | `norm.pdf`$(\mathrm{x}=x)$ |
| $\Phi(x;a,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{x} \mathrm{e}^{-\frac{(u-a)^2}{2\sigma^2}} du$ | `norm.dist`$(x;a;\sigma;$`true`$)$ | `norm.cdf`$(\mathrm{x}=x, \mathrm{loc}=a, \mathrm{scale}=\sigma)$ |
| $\Phi(x) = \Phi(x;0,1)$ | `norm.s.dist`$(x;$`true`$)$ | `norm.cdf`$(\mathrm{x}=x)$ |
| $\alpha$ lygio kvantilis $z_\alpha = \Phi^{-1}(\alpha)$ | `norm.s.inv`$(\alpha)$ | `norm.ppf`$(\mathrm{q}=\alpha)$ |
| *Studento atsitiktinis dydis* $T_n \sim \mathcal{S}t(n)$ | | |
| Tankio funkcija $f(x;n) = F'(x;n)$ | `t.dist`$(x;n;$`false`$)$ | `t.pdf`$(\mathrm{x}=x, \mathrm{df}=n)$ |
| Pasisk. f-ja $F(x;n) = P(T_n < x)$ | `t.dist`$(x;n;$`true`$)$ | `t.cdf`$(\mathrm{x}=x, \mathrm{df}=n)$ |
| $\alpha$ lygio kvantilis $t_\alpha(n) = F^{-1}(\alpha;n)$ | `t.inv`$(\alpha;n)$ | `t.ppf`$(\mathrm{q}=\alpha, \mathrm{df}=n)$ |
| $\chi^2$ *atsitiktinis dydis* $\chi_n^2 \sim \chi^2(n)$ | | |
| Tankio funkcija $f(x;n) = F'(x;n)$ | `chisq.dist`$(x;n;$`false`$)$ | `chi2.pdf`$(\mathrm{x}=x, \mathrm{df}=n)$ |
| Pasisk. f-ja $F(x;n) = P(\chi_n^2 < x)$ | `chisq.dist`$(x;n;$`true`$)$ | `chi2.cdf`$(\mathrm{x}=x, \mathrm{df}=n)$ |
| $\alpha$ lygio kvantilis $\chi_\alpha^2(n) = F^{-1}(\alpha;n)$ | `chisq.inv`$(\alpha;n)$ | `chi2.ppf`$(\mathrm{q}=\alpha, \mathrm{df}=n)$ |