Notatki z ćwiczeń z Probabilistyki 19/20 [1]

Grzegorz Siudem & grupy T1, T2, T3

23 maja 2020

Notatki z ćwiczeń oraz inne materiały pomocne w rozwiązywaniu prac domowych z Probabilistyki. Plik na bieżąco aktualizowany. Zgłoszenie każdego błędu będzie bardzo mile widziane.

[1] Złożono w klasie `tufte-handout`, `www.ctan.org/pkg/tufte-latex`

Spis treści

C1: Wprowadzenie

Podczas ćwiczeń testowaliśmy klasyczną (częstościową) definicję prawdopodobieństwa na przykładach rzutów monetą, kostką i losowania kart.

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

gdzie A to zbiór realizacji sprzyjających analizowanemu zdarzeniu, a Ω zbiór wszystkich możliwych zdarzeń.

Uwaga! Powyższy wzór można stosować pod warunkiem, że wszystkie *zdarzenia elementarne* są równoprawdopodobne (por. przykład z wykładu).

Policzmy z klasycznego wzoru prawdopodobieństwo wyciągnięcia kilku wybranych układów pokerowych. Zakładamy, że losujemy 5 kart z talii o 52 kartach, a zatem przestrzenią stanów są zbiory 5-elementowe kart bez powtórzeń, a zatem

$$|\Omega| = \binom{52}{5} = \frac{52!}{5!47!}.$$

Zwróćmy uwagę, że wszystkie te wyniki nie zależą od kolejności losowania. Moglibyśmy, co jest jednak bardziej wymagające rachunkowo, rozważyć analogicznie piątki uporządkowane. Ważne, żeby tych dwóch metod nie mieszać.

Rozważmy teraz kilka przykładowych układów

- Para, układ P gdzie mamy **dokładnie** jedną parę

$$|P| = \binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}^3.$$

Co interpretujemy następująco: mamy 13 możliwości wyboru rangi pary (od dwójek do asów), następnie kart o wybranej randze mamy 4, a potrzebujemy 2. Trzeci czynnik oznacza wybór trzech kart o różnych rangach (i różnych od naszej wyjściowej) o dowolnych kolorach każda (ostatni czynnik).

- Kareta, czyli układ K, w których mamy cztery karty tej samej rangi

$$|K| = \binom{13}{1}\binom{48}{1}.$$

Gdzie pierwszy czynnik to liczba możliwych rang naszej karety, a czynnik drugi liczba możliwych wyborów piątej karty.

- Ful, czyli układy F składający się z trójki kart tej samej rangi i pary kart innej rangi

$$|F| = \binom{13}{1}\binom{12}{1}\binom{4}{3}\binom{4}{2}.$$

Najpierw wybieramy rangę trójki, potem rangę pary, potem kolory trójki i kolory pary.

C2: Rozkłady dyskretne

Najważniejsze wnioski z wykładu:

- Zadania domowe, spisane czytelnie i podpisane, oddajemy na **jednej kartce formatu A4**.
- Przypadkowy charakter zjawisk lub procesów może wymagać opisu statystycznego lub probabilistycznego, które, choć powiązane, nie są tożsame.
- W rozkładach dyskretnych prawdopodobieństwo zdarzenia jest proporcjonalne do liczby sprzyjających mu zdarzeń.

Przyjmując, jak na poprzednich zajęciach, że dla rozkładów dyskretnych mamy

$$\mathbb{P}(X = k) = \frac{1}{|\Omega|}|X_k| = C \cdot p_k.$$

Poszukujemy zatem wartości stałej C dla wybranych rozkładów:

- $p_k = 1$ dla $k = 1, \dots, N \Rightarrow \mathbb{P}(X = k) = 1/N$,
- $p_k = (1-q)^k$ dla $q \in (0,1)$, $k \in \mathbb{N} \Rightarrow \mathbb{P}(X = k) = q(1-q)^k$,
- $p_k = \lambda^k/k!$ dla $\lambda > 0$, $k \in \mathbb{N} \Rightarrow \mathbb{P}(X = k) = \lambda^k e^{-\lambda}/k!$,
- $p_k = \binom{N}{k} q^k (1-q)^{-k}$ dla $q \in (0,1)$, $k = 0, \dots, N \Rightarrow \mathbb{P}(X = k) = \binom{N}{k} q^k (1-q)^{N-k}$,

Dyskretny rozkład jednostajny.

Rozkład geometryczny.

Rozkład Poissona.

Rozkład Bernoulliego.

oraz zwartych postaci ich dystrybuant

$$F_X(t) = \mathbb{P}(X \leqslant t),$$

tam gdzie to możliwe

- Dla rozkładu jednostajnego dyskretnego $F(k) = k/N$,
- Dla rozkładu geometrycznego $F(k) = 1 - (1-p)^k$.

Dla uproszczenia stosujemy notację, w której argumentami dystrybuanty są tylko liczby z dziedziny funkcji masy prawdopodobieństwa, w ogólności funkcja miałaby postać

$$F(t) = \frac{\lfloor t \rfloor}{N} I_{[1,N]}(t)$$

C3: *Rozkłady ciągłe*

Najważniejsze wnioski z wykładu:

- Intuicyjne zrozumienie czym są zmienne losowe i niezależność.
- Opis zmiennych ciągłych – funkcja gęstości prawdopodobieństwa.
- Dystrybuanta ($F(x) = \mathbb{P}(X \leqslant x)$) jest ważną wielkością, pozwala także zdefiniować kwantyle.

Sprawdźmy normowanie do 1 i policzmy dystrybuantę (gdzie to możliwe) dla rozkładów

W razie wątpliwości proszę sprawdzić opisy na stronie 50.

- jednostajnego ciągłego,
- wykładniczego,
- Pareto,

Wyznaczmy kwantyle dla rozkładu prawdopodobieństwa o gęstości danej wzorem

$$f(x) = 2xI_{[0,1]}(x).$$

- Sprawdzenie normowania rozkładu jednostajnie ciągłego:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{dla} \quad x \in \langle a, b\rangle \\ 0 & \text{dla} \quad x \in \mathbb{R} \setminus \langle a, b\rangle \end{cases} \qquad a, b \in \mathbb{R} \quad ,$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} 0dx + \int_{a}^{b} \frac{1}{b-a}dx + \int_{b}^{\infty} 0dx = \int_{a}^{b} \frac{1}{b-a}dx =$$

$$= \frac{1}{b-a} \cdot x\Big|_{x=a}^{x=b} = \frac{b-a}{b-a} = 1 \quad \checkmark$$

- Obliczenie dystrybuanty rozkładu jednostajnie ciągłego $F(x)$:

$$\begin{array}{llcl} 1^\circ) & x \leq a & \Longrightarrow & F(x) = 0 \\ 2^\circ) & x \in (a, b\rangle & \Longrightarrow & F(x) = \int_a^x f(\omega)d\omega \\ 3^\circ) & x \geq b & \Longrightarrow & F(x) = 1 \end{array}$$

Ad 2° $F(x) = \int_{a}^{x} \frac{1}{b-a}d\omega = \frac{1}{b-a} \cdot \omega\Big|_{\omega=a}^{\omega=x} = \frac{1}{b-a}(x-a) = \frac{x-a}{b-a}$

$$F(x) = \begin{cases} 0 & \text{dla} \quad x \leq a \\ \frac{x-a}{b-a} & \text{dla} \quad x \in (a, b\rangle \\ 1 & \text{dla} \quad x > b \end{cases} \quad \checkmark$$

- Sprawdzenie normowania rozkładu wykładniczego:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \lambda e^{-\lambda x} & \text{dla } x \geq 0 \end{cases} \qquad \lambda > 0$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{\infty} \lambda e^{-\lambda x}dx = \int_{0}^{\infty} \lambda e^{-\lambda x}dx =$$

$$= \left| \begin{array}{c} -\lambda x = u \\ dx = \frac{du}{-\lambda} \end{array} \right| = \left| \begin{array}{ccc} x = 0 & \Rightarrow & u = 0 \\ x \to \infty & \Rightarrow & u \to -\infty \end{array} \right| = \lambda \int_{0}^{-\infty} e^{u} \cdot \frac{du}{-\lambda} =$$

$$= -\left[e^{u}\right]\Big|_{0}^{-\infty} = -\big(\lim_{u \to -\infty}[e^{u}] - e^{0}\big) = -(0-1) = 1 \quad \checkmark$$

◆ Obliczenie dystrybuanty rozkładu wykładniczego $F(x)$:

$$F(x) = \int_{-\infty}^{x} f(\psi)d\psi = \int_{0}^{x} \lambda e^{-\lambda\psi} d\psi = \left| \begin{array}{c} -\lambda\psi = u \\ d\psi = \frac{du}{-\lambda} \end{array} \right| = \left| \begin{array}{ccc} \psi = 0 & \Rightarrow & u = 0 \\ \psi = x & \Rightarrow & u = -\lambda x \end{array} \right| =$$

$$= \lambda \int_{0}^{-\lambda x} e^{u} \cdot \frac{du}{-\lambda} = -\left[e^{u}\right]\Big|_{0}^{-\lambda x} = -\left[e^{-\lambda x} - e^{0}\right] = 1 - e^{-\lambda x} \quad \checkmark$$

◆ Sprawdzenie normowania rozkładu Pareto:

$$f(x) = \begin{cases} 0 & \text{dla} \quad x \leq x_m \\ \frac{kx_m^k}{x^{k+1}} & \text{dla} \quad x > x_m \end{cases} \qquad x, x_0 \in \mathbb{R} \quad , k > 0$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{x_m} 0dx + \int_{x_m}^{\infty} \frac{kx_m^k}{x^{k+1}}dx = \int_{x_m}^{\infty} \frac{kx_m^k}{x^{k+1}}dx =$$

$$= kx_m^k \int_{x_m}^{\infty} x^{-(k+1)}dx = kx_m^k \cdot \frac{1}{-(k+1)+1} x^{-(k+1)+1}\Big|_{x=x_m}^{x=\infty} =$$

$$= -x_m^k \cdot \left(x^{-k}\right)\Big|_{x=x_m}^{x=\infty} = -x_m^k \cdot \left(\lim_{x\to\infty}\left[\frac{1}{x^k}\right] - \frac{1}{x_m^k}\right) = -x_0^k \cdot \frac{-1}{x_m^k} = 1 \quad \checkmark$$

◆ Obliczenie dystrybuanty rozkładu Pareto $F(x)$:

$$F(x) = \int_{x_m}^{x} \frac{kx_0^k}{\eta^{k+1}}d\eta = kx_m^k \int_{x_m}^{x} \eta^{-(k+1)}d\eta = \frac{kx_m^k}{-(k+1)+1}\left[\eta^{-(k+1)+1}\right]\Big|_{\eta=x_m}^{\eta=x} =$$

$$= \frac{kx_m^k}{-k}\left(x^{-k} - x_m^{-k}\right) = 1 - \left(\frac{x_m}{x}\right)^k \quad \checkmark$$

◆ Wyznaczenie kwantyli dla rozkładu prawdopodobieństwa danego wzorem: $f(x) = 2xI_{[0,1]}(x)$:

$$F(x) = \begin{cases} 0 & \text{dla} \quad x \leq 0 \\ \int_0^x 2\xi \, d\xi & \text{dla} \quad x \in (0, 1\rangle \\ 1 & \text{dla} \quad x > 1 \end{cases}$$

Dla dla$x \in (0, 1\rangle$:

$$F(x) = \int_0^x 2\xi \, d\xi = 2\frac{\xi^2}{2}\Big|_0^x = x^2$$

$$F(Q_p) = p \Rightarrow F^{-1}[F(Q_p)] = F^{-1}(p) \Longrightarrow Q_p = F^{-1}(p)$$

$$F^{-1}[F(x)] = \sqrt{F(x)} \Longrightarrow \underline{\underline{Q_p = \sqrt{p}}} \quad \checkmark$$

C4: Czy tego oczekiwaliśmy?

Najważniejsze wnioski z wykładu:

- Definicja i podstawowe własności momentów zmiennej losowej.
- Czym są estymatory i dlaczego ich potrzebujemy?
- Prezentowanie danych – wykresy skrzynkowe i histogramy.

Zadania do wykonania:

- Należy wyznaczyć wybrane momenty (albo uzasadnić, że to niemozliwe) dla rozkładów umieszczonych na stronach 49-50.
- Należy narysować histogram dla wyników rzutów kością oraz wykres skrzynkowy dla przygotowanych do tego danych

Jeśli to tylko nie utrudnia drastycznie rachunków proponuję liczyć dla dowolnego n, a tam gdzie sprawa jest nieoczywista pozostaje nam sprawdzanie dla $n = 1, 2, \ldots$.

Oba pliki dostępne na stronie przedmiotu.

- **Momenty rozkładu ciągłego wykładniczego**: stosujemy podstawienie: $u = \lambda x$, $du = \lambda dx$

$$f(x) = \lambda e^{-\lambda x}, \quad x \in [0, \infty) \quad \Rightarrow \quad \mathbb{E}X^n = \int_0^\infty x^n \lambda e^{-\lambda x} dx$$

$$\mathbb{E}X^n = \int_0^\infty \left(\frac{u}{\lambda}\right)^n e^{-u} du = \frac{1}{\lambda^n}\int_0^\infty u^n e^{-u} du = \frac{1}{\lambda^n}\int_0^\infty u^{(n+1)-1} e^{-u} du.$$

Korzystamy z podstawowych własności funkcji Γ Eulera

$$\mathbb{E}X^n = \frac{\Gamma(n+1)}{\lambda^n} = \frac{n!}{\lambda^n}.$$

- **Momenty rozkładu ciągłego jednostajnego**: wykonujemy elementarne całkowanie

$$f(x) = \frac{1}{b-a}, \quad x \in [a, b] \quad \Rightarrow \quad \mathbb{E}X^n = \int_a^b \frac{x^n}{b-a} dx$$

$$\mathbb{E}X^n = \frac{1}{b-a}\left[\frac{x^{n+1}}{n+1}\right]_a^b = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)} \tag{1}$$

można równoważnie zapisać (przypadek $m = n+1$):

korzystając ze wzoru skróconego mnożenia

$$b^m - a^m = (b-a)(b^{m-1} + b^{m-2}a + \ldots + ba^{m-2} + a^{m-1})$$

$$\mathbb{E}X^n = \frac{1}{n+1}\sum_{k=0}^n b^{n-k}a^k$$

Czy ten krok upraszcza wzór? Wręcz przeciwnie, ale pozwala na uprawiedliwienie intuicyjnego faktu: średnia $\mathbb{E}X^1 = \frac{b+a}{2}$.

Na podstawie wzoru (1) pierwsze trzy momenty tego rozkładu

$$\mathbb{E}X = \frac{a+b}{2}, \ \mathbb{E}X^2 = \frac{a^2+ab+b^2}{3}, \ \mathbb{E}X^3 = \frac{(a+b)\left(a^2+b^2\right)}{4}.$$

- **Momenty rozkładu wykładniczego metodą momentów**: n-ty moment rozkładu jest n-tą pochodną funkcji tworzącej momenty po t w punkcie $t = 0$

Uwaga! Metoda momentów wykracza poza naszą aktualną wiedzę z wykładu, poznamy ją na późniejszych zajęciach.

$$\mathbb{E}X^n = \frac{d^n}{dt^n}M(t)\Big|_{t=0}.$$

Funkcją tworzącą momenty jest transformatą Laplace'a funkcji rozkładu prawdopodobieństwa, dla rozkładu wykładniczego:

$$M(t) = \int_0^\infty e^{tx} f(x)dx = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \int_0^\infty \lambda e^{x(t-\lambda)} dx$$
$$= \frac{\lambda e^{x(t-\lambda)}}{t-\lambda}\Big|_0^\infty = \frac{-\lambda}{-\lambda+t} = \frac{\lambda}{\lambda - t}$$

dla $t < \lambda$.

Pierwszy moment rozkładu wykładniczego:

$$\mathbb{E}X = \frac{d}{dt}M(t)\Big|_{t=0} = \frac{d}{dt}\frac{\lambda}{\lambda-t}\Big|_{t=0} = \frac{\lambda}{(\lambda-0)^2} = \frac{1}{\lambda}.$$

Drugi moment rozkładu wykładniczego:

$$\mathbb{E}X^2 = \frac{d^2}{dt^2}M(t)\Big|_{t=0} = \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} = \frac{2}{\lambda^2}.$$

- **Pierwszy moment rozkładu Gaussa**:

$$\mathbb{E}X = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} x \exp\left[\frac{-(x-\mu)}{\sqrt{2}\sigma}\right]^2.$$

Wprowadza się nową zmienną $t = \frac{x-\mu}{\sigma\sqrt{2}}$, wówczas równanie sprowadza się do postaci:

$$x = \sigma\sqrt{2}t + \mu$$
$$dt = \frac{1}{\sqrt{2}\sigma}dx$$

$$\mathbb{E}X = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)e^{-t^2}\sqrt{2}\sigma dt =$$
$$= \frac{1}{\sqrt{\pi}}\left[\int_{-\infty}^{\infty} \mu e^{-t^2} dt + \sqrt{2}\sigma \int_{-\infty}^{\infty} te^{-t^2} dt\right] = \frac{\mu}{\sqrt{\pi}}\sqrt{\pi} + 0 = \mu$$

- **Rozkład Pareto:** indykator ogranicza całkowanie, tak więc:

$$\mathbb{E}X^n = \int_{x_m}^{\infty} x^n \frac{\alpha x_m^\alpha}{x^{\alpha+1}} dx = \int_{x_m}^{\infty} x^{n-\alpha-1}\alpha x_m^\alpha dx = \frac{x^{n-\alpha}}{n-\alpha}\alpha x_m^\alpha \Big|_{x_m}^\infty$$

Jeżeli przyjmie się, że $\alpha > n$ to wynik całki można zapisać jako

$$\mathbb{E}X^n = \frac{x_m^{n-\alpha}}{\alpha - n}\alpha x_m^\alpha = \frac{\alpha}{\alpha-n}x_m^n$$

Zatem pierwszy moment rozkładu Pareto, o ile istnieje, przyjmuje postać $\frac{\alpha x_m}{\alpha-1}$, a drugi $\frac{\alpha x_m^2}{\alpha-2}$.

Jeżeli nie został spełniony warunek $\alpha > n$ to całka jest rozbieżna i momenty nie istnieją.

- **Momenty rozkładu dwupunktowego**:

$p(X = 0) = 1 - p,\ p(X = 1) = p\ dla\ p \in [0, 1]$

$$\mathbb{E}X^n = \sum_k x_k^n p_k = 0^n \cdot (1-p) + 1^n \cdot p = p$$

- **Momenty rozkładu gamma**:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^\alpha \cdot e^{-\beta x},$$

gdzie gdzie $\alpha, \beta \in \mathbb{R}_+$, a Γ to funkcja gamma Eulera.

Całkujemy od 0, bo dla rozkładu gamma $I_{(0,\inf)}(x)$

$$\begin{aligned}\mathbb{E}X^n &= \int_0^\infty x^n \cdot f(x)dx = \int_0^\infty x^n \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^\alpha \cdot e^{-\beta x}dx = \\ &= \int_0^\infty x^{n+\alpha-1} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot e^{-\beta x}dx = \\ &= \frac{1}{\Gamma(\alpha)}\int_0^\infty x^{n+\alpha-1} \cdot \frac{\beta^{n+\alpha-1}}{\beta^{n-1}} e^{-\beta x}dx = \\ &= \frac{1}{\Gamma(\alpha)\cdot\beta^{n-1}}\int_0^\infty (\beta x)^{n+\alpha-1} \cdot e^{-\beta x}dx.\end{aligned}$$

Stosujemy podstawienie: $u = \beta x$, $du = \frac{dx}{\beta}$

$$EX^n = \frac{1}{\Gamma(\alpha)\cdot\beta^n}\int_0^\infty u^{n+\alpha-1}\cdot e^{-u}du = \frac{1}{\beta^n}\cdot\frac{\Gamma(n+\alpha)}{\Gamma(\alpha)}.$$

Korzystamy też z funkcji Γ Eulera

$$\Gamma(n) = \int_0^\infty v^{n-1}\cdot e^{-v}dv$$

◆ **Momenty rozkładu Zipfa**

$$Ex = \sum_{k=1}^{N} k\frac{1}{k^s H_s(n)} = \frac{1}{H_s(n)}\sum_{k=1}^{N} k^{-(s-1)} = \frac{1}{H_s(n)}H_{s-1}(n)$$

$$Ex^2 = \sum_{k=1}^{N} k^2\frac{1}{k^s H_s(n)} = \frac{1}{H_s(n)}\sum_{k=1}^{N} k^{-(s-2)} = \frac{1}{H_s(n)}H_{s-2}(n)$$

Rozkład Zipfa:

$$p_k = \frac{1}{k^s H_s(n)},\ H_s(n) = \sum_{k=1}^{N}\frac{1}{k^s}$$

◆ **Momenty rozkładu Poissona**

$$\mathbb{E}X = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = \lambda$$

Rozkład Poissona wyraża się wzorem:

$$\mathbb{P}(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}, k\in\mathbb{N}\cup 0$$

gdyż $\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = e^\lambda$.

$$\begin{aligned}\mathbb{E}X^2 &= \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=1}^{\infty}(k-1+1)\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \\ &= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = (\lambda^2+\lambda)e^{-\lambda}\sum_{l=0}^{\infty}\frac{\lambda^l}{l!} = \\ &= \lambda^2+\lambda.\end{aligned}$$

więc wariancja rozkładu wynosi

$$\mathbb{D}^2(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda^2+\lambda-\lambda^2 = \lambda.$$

◆ **Momenty rozkładu Bernoulliego**

$$\begin{aligned}\mathbb{E}X &= \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n}\frac{n!}{k!\,(n-k)!}p^{k(1-p)^{n-k}} = \\ &= np\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!\,(n-1-(k-1))!}p^{k-1}(1-p)^{n-1-(k-1)} = \\ &= np\sum_{l=0}^{n-1}\frac{(n-1)!}{l!\,(n-1-l)!}p^l(1-p)^{n-1-l} = np.\end{aligned}$$

Rozkład Bernoulliego wyraża się wzorem:

$$P(X=k) = \binom{n}{k}p^k q^{n-k},$$

gdzie: $k = 0, 1, ..., n$ i $q = 1-p$.

Ponieważ suma po prawej stronie przedostatniej równości jest równa $(p+1-p)^{n-1} = 1$.

$$\mathbb{E}X^2 = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k} =$$

$$= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!\,(n-2-(k-2))!} p^{k-2} (1-p)^{n-2-(k-2)} +$$

$$+ np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!\,(n-1-(k-1))!} p^{k-1} (1-p)^{n-1-(k-1)} =$$

$$= n(n-1)p^2 \sum_{l=0}^{n-2} \binom{n-2}{l} p^l (1-p)^{n-2-l} +$$

$$+ np \sum_{l=0}^{n-1} \binom{n-1}{l} p^l (1-p)^{n-1-l} = n(n-1)p^2 + np.$$

Uwzględniając, że $EX = np$ otrzymujemy:

$$\mathbb{D}^2(X) = EX^2 - (EX)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

- **Momenty rozkładu Cauchy'ego**

$$\mathbb{E}X = \frac{1}{\pi} \int_{-\infty}^{+\infty} x \frac{1}{\gamma} \frac{1}{1 + (\frac{x-x_0}{\gamma})^2} dx = \left| \begin{array}{c} \frac{x-x_0}{\gamma} = u \\ \frac{1}{\gamma} dx = du \\ x = \gamma u + x_0 \end{array} \right| =$$

$$= \frac{1}{\pi} \int_{-\infty}^{+\infty} (\gamma u + x_0) \frac{1}{1+u^2} du =$$

$$= \frac{\gamma}{2\pi} \ln(u^2+1) + \frac{x_0}{\pi} \operatorname{arctg}(u) \Big|_{u=-\infty}^{u=+\infty} \to \infty$$

Dalsze momenty także są rozbieżne (zawierają te same wyrażenia z dokładnością do stałej + kolejne potęgi u). Momenty rozkładu Cauchy'ego nie istnieją.

- **Momenty rozkładu geometrycznego:**

$p_k = (1-q)^k q$

$$\mathbb{E}X = \sum_{x=0}^{\infty} x(1-q)^x q = q(1-q) \sum_{x=0}^{\infty} x(1-q)^{x-1} =$$

$$= q(1-q) \sum_{x=0}^{\infty} \frac{d}{dq} \Big[-(1-q)^x \Big] = -q(1-q) \frac{d}{dq} \frac{1}{q} = \frac{1-q}{q}.$$

$$\mathbb{E}X^2 = \sum_{x=0}^{\infty} x^2 (1-q)^x q = q(1-q)^2 \sum_{x=0}^{\infty} x^2 (1-q)^{x-2} =$$

$$= q(1-q)^2 \sum_{x=0}^{\infty} (1-q)^{x-2} x(x-1+1) =$$

$$= q(1-q)^2 \sum_{x=0}^{\infty} (1-q)^{x-2} x(x-1) + \mathbb{E}X.$$

$$\mathbb{E}X^2 = q(1-q)^2 \frac{2}{q^3} + \frac{1-q}{q} = \frac{q^2 - 3q + 2}{q^2}.$$

$\sum_{x=0}^{\infty} (1-q)^{x-2} x(x-1) = \sum_{x=0}^{\infty} \frac{d^2}{dq^2} (1-q)^x = \frac{2}{q^3}.$

◆ **Momenty rozkładu zeta**

$$\mathbb{E}X^n = \sum_{x=0}^{\infty} x^n \frac{1}{x^s \zeta(s)} = \frac{1}{\zeta(s)} \sum_{x=0}^{\infty} \frac{1}{x^{s-n}}$$

zmienna X posiada w takim razie n-ty moment tylko w przypadku kiedy $n < s - 1$, moment wynosi wtedy:

$$\mathbb{E}X^n = \frac{\zeta(s-n)}{\zeta(s)}$$

◆ **Momenty rozkładu χ^2**

$f(x) = \frac{1}{2^{(k/2)-1}\Gamma(k/2)} x^{k-1} e^{-x^2/2} I_{[0,\infty)}(x)$

Podstawiamy $\frac{x^2}{2} = u$, $dx = \frac{du}{\sqrt{2u}}$

$$\begin{aligned}\mathbb{E}X^n &= \int_0^\infty \frac{1}{2^{(k/2)-1}\Gamma(k/2)} x^{k+n-1} e^{-x^2/2} dx = \\ &= \int_0^\infty \frac{1}{2^{(k/2)-1}\Gamma(k/2)} \sqrt{2u}^{k+n-1} e^{-u} \cdot \frac{1}{\sqrt{2u}} du = \\ &= \frac{2^{(k/2)+(n/2)-(1/2)}}{2^{(k/2)-(1/2)}} \frac{1}{\Gamma(k/2)} \int_0^\infty u^{(k/2)+(n/2)-1} e^{-u} du = \\ &= 2^{n/2} \frac{1}{\Gamma(k/2)} \int_0^\infty u^{(k/2)+(n/2)-1} e^{-u} du.\end{aligned}$$

Z definicji Gammy Eulera:

$$EX^n = 2^{n/2} \frac{\Gamma[(k+n)/2]}{\Gamma(k/2)}.$$

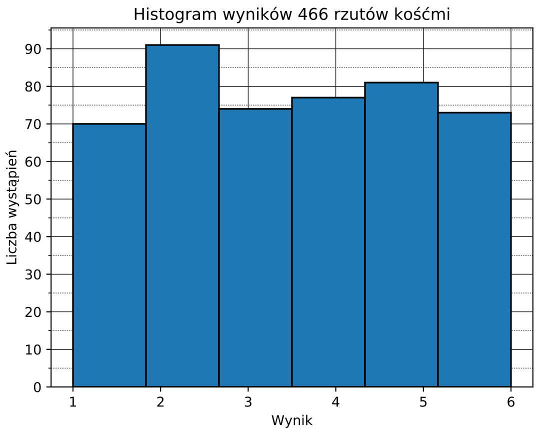
◆ **Histogram rzutów kośćmi**: `pyplot`, patrz Rys. 1 na stronie 11.

Wykonano w Pythonie 3.8.1 przy użyciu biblioteki matplotlib 3.1.2

```
from matplotlib import pyplot as plt
from matplotlib.ticker import MultipleLocator
with open("kostki.csv", "r") as f:
    data = [int(letter) for letter in f.read() if letter in "123456"]
    bins = len(set(data))
    _, ax = plt.subplots()
    plt.grid(which='major', linestyle='-', linewidth='0.5', color='black')
    plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
    ax.set_axisbelow(True)
    plt.title(f"Histogram {len(data)} rzutow koscmi")
    plt.xlabel("Wynik")
    plt.ylabel("Liczba wystapien")
    ax.yaxis.set_major_locator(MultipleLocator(10))
    ax.yaxis.set_minor_locator(MultipleLocator(5))
    ax.hist(data, bins, edgecolor='black', linewidth=1.2)
    plt.show()
```

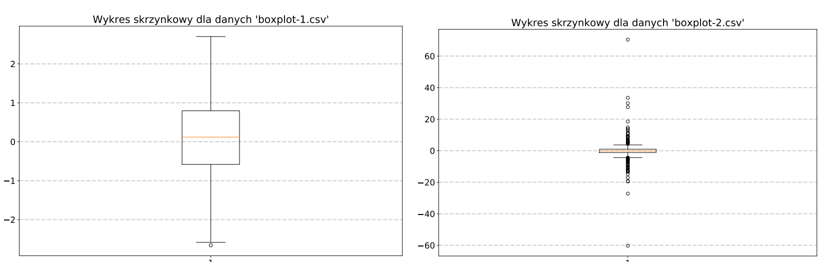
◆ **Wykresy skrzynkowe**: `python`, patrz Rys. 2 na stronie 11.

dla 'boxplot-1.csv' oraz 'boxplot-2.csv'



Rysunek 1: Histogram rzutów kostką narysowany w Pythonie.

```
import pandas as pd
from matplotlib import pyplot as plt
import numpy as np
df = np.array(pd.read_csv('boxplot-1.csv',header=None))
plt.figure(figsize=(13,8))
plt.rcParams.update({'font.size': 16})
plt.boxplot(df)
plt.grid(True)
plt.title('Wykres skrzynkowy dla danych \'boxplot-1.csv\'')
plt.grid(color='#95a5a6', linestyle='--', linewidth=2, axis='y', alpha=0.5)
plt.grid(axis='x',alpha=0)
plt.savefig('boxlot-1_lotnicy.pdf',bbox_inches='tight', pad_inches=0.2)
```



Rysunek 2: Wykresy skrzynkowe wykonane w Pythonie.

◆ **Wykresy skrzynkowe**: `Excel`

Formuły podane są dla polskiej wersji językowej Excela pakietu Office 365

Utworzenie tablicy z fomułami:

```
=MIN(dane)  [B1]
=KWARTYL(dane,1)  [B2]
=KWARTYL(dane,2)  [B3]
=KWARTYL(dane,3)  [B4]
=KWARTYL(dane,4)  [B5]
=MAX(dane)  [B6]
```

Utworzenie kolejnej tablicy z formułami:

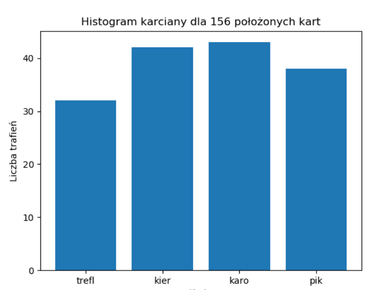
=B2–B1 [C1]
=B3–B2 [C2]
=B4–B3 [C3]
=B5–B4 [C4]

Utworzenie wykresu skumulowanego kolumnowego z danych C1:C4 i odwrócenie danych/serii. Następnie zamiana wypełnienia danych z najwyższej i najniższej serii na linię błędów.

- **Histogram kart**: `python`, patrz Rys. 3 na stronie 12.

Kod w dla wersji Python 3.7.4, powinien działać dla każdego powyżej 3.

```
import matplotlib.pyplot as plt
with open("karty.txt", "r") as f:
    r = f.read().split(',')
karty = ['trefl', 'kier', 'karo', 'pik']

counter = [0, 0, 0, 0]
for i in range(len(r)):
    if r[i] == karty[0]:
        counter[0] += 1
    if r[i] == karty[1]:
        counter[1] += 1
    if r[i] == karty[2]:
        counter[2] += 1
    if r[i] == karty[3]:
        counter[3] += 1
plt.title("Histogram␣karciany␣dla␣156␣polozonych␣kart")
plt.xlabel("Karta")
plt.ylabel("Liczba␣trafien")
plt.bar(karty, counter)
plt.show()
```



Rysunek 3: Histogram wylosowanych kart wykonany w języku Python.

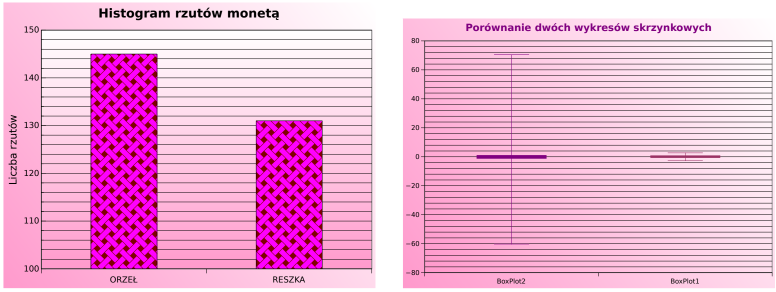
- **Histogram i boxplot**: `GNUMERIC`, patrz Rys 4 na stronie 13.

Typ wykresu to "Statystyczny", a podtyp to "Pionowy wykres skrzynkowy".

- **Histogram** w języku `Wolfram`, patrz Rys. 5 na stronie 13.

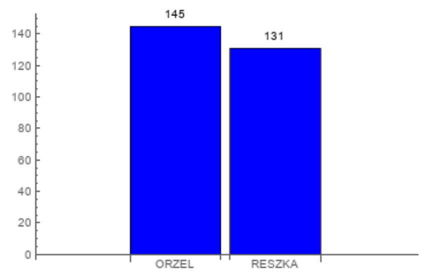
Histogram powstał w WolframCloud

```
data = {...};
o = Count[data, O];
r = Count[data, R];
BarChart[{{o, r}},
```



Rysunek 4: Histogram i wykres skrzynkowy narysowane w środowisku GNUMERIC. Umiejętna żonglerka dobranymi ze smakiem kolorami z niebanalnej palety barw zapewnia tak spójny jak i charakterystyczny sposób przedstawienia danych.

```
ChartLabels -> {"ORZEL", "RESZKA"},
LabelingFunction->Above,
ChartStyle->{Blue, Blue}]
```



Rysunek 5: Histogram rzutow monetą wykonany w chmurze `Wolfram Cloud`.

- **Histogram** w środowisku `matlab` , patrz Rys. 6 na stronie 14.

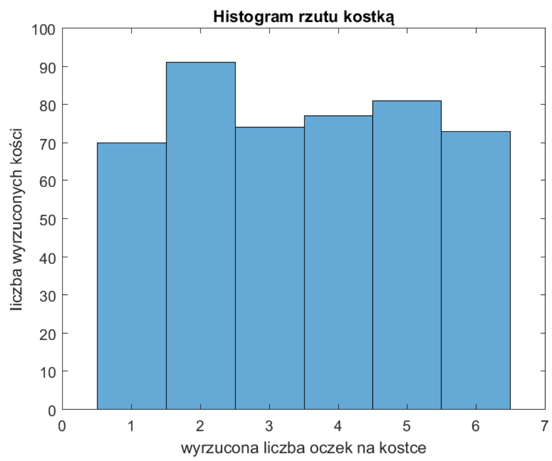
  Wykonano w matlabie R2017a.

  W tym celu najpierw zaimportowano dane z komputera, korzystając z zakładek `HOME -> Import Data`, wybrano ścieżkę na komputerze prowadzącą do pliku `kostki.csv`, a w następnej zakładce: `Import` wybrano w `Output Type: Numeric Matrix`, po czym zaimportowano dane, które były widoczne w workspace jako: kostki (o wartości 1x466 double)

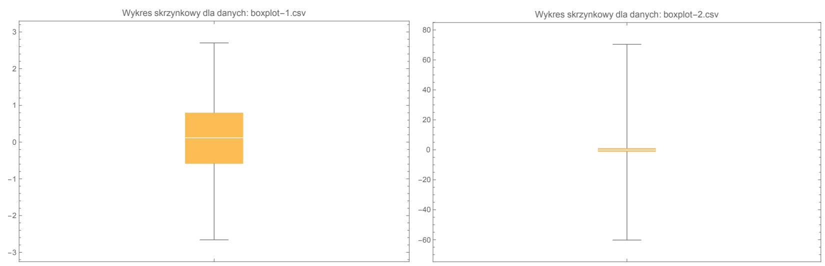
```
histogram(kostki)
title('Histogram rzutu kostka')
xlabel('wyrzucona liczba oczek na kostce')
ylabel('liczba wyrzuconych kosci')
```

- **Wykresy skrzynkowe** w `Mathematice`, patrz Rys. 7 na stronie 14

```
boxplot = Import["...\boxplot-1.csv"];
boxplot2 = Import["...\boxplot-2.csv"];
BoxWhiskerChart[boxplot, PlotLabel -> "Wykres skrzynkowy dla boxplot-1.csv"]
BoxWhiskerChart[boxplot2, PlotLabel -> "Wykres skrzynkowy dla boxplot-2.csv"]
```



Rysunek 6: Histogram rzutów kostką narysowany w środowisku matlab.



Rysunek 7: Wykresy skrzynkowe narysowane w środowisku Wolfram Mathematica.

C5: Po jednym warunkiem

Najważniejsze wnioski z wykładu:

- Prawdopodobieństwo warunkowe.
- Wzór Bayesa.
- Wnioskowanie Bayesowskie.

Zadania do wykonania:

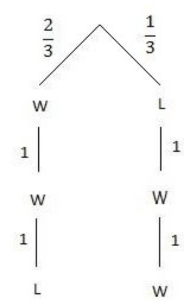
Zachęcam do rozwiazywania każdego z zadań różnymi metodami: ropisując p-stwa, przez analizę drzewek, symulacje numeryczne etc.

- Rozważmy grę, w której gracz wybiera jedną z trzech kart, dwie z znich skutkują nagrodą, a jedna nie. Gracz wybrał jedną z kart, po czym rozdający odsłania jedną z pozostałych dwóch, która kryje nagrodę. Gracz może zmienić swój pierwotny wybór na drugą nieodsłoniętą kartę. Czy i o ile taka zmiana zwiększa szanse zwycięstwa?

- Oba zadania to przyklady zagadki logicznej Monty'ego Hall'a. Gdy losujemy ze zbioru 3-elemtnowego w którym jeden element jest inny niz dwa pozostale, mamy szanse $\frac{1}{3}$ na wylosowanie go. Iluzja tutaj jest wyjecie ze zbioru jednego z dwoch takich samych elementów po tym jak już wybraliśmy. W pierwszym zadaniu trafienie pierwszym strzałem wygranej wynosi $\frac{2}{3}$. Jeżeli trafiliśmy wygrana i zmienimy swój wybór - wybierając opcje inna (po odsłonieciu jednej nagrody zostaje nam tylko nagroda - nie-nagroda) czeka nas na pewno porażka. Zmieniając kartę wygramy tylko w przypadku, kiedy początkowo wybierzemy przegrana kartę (czyli prob. $\frac{1}{3}$). Taka zmiana wiec zmniejsza nam szanse zwycięstwa - z $\frac{2}{3}$ do $\frac{1}{3}$.

Wybierając kartę na początku mamy szansę 2/3 na wygraną. Odsłonięcie jednej z kart z nagrodą nie ma wpływu na to początkowe prawdopodobieństwo. Prawdopodobieństwo przypisane odsłoniętej karcie jest równe 0 (nie ma szansy żeby była tam też druga nagroda), a suma prawdopodobieństw musi być równa 1, więc niewybrana nieodsłonięta karta ma szansę na zawieranie nagrody 1/3.

Metoda drzewa: oznaczenia: W-Winner, karta wygrywająca, L-Looser, karta przegrywająca. Z założenia zadania wiemy, że druga karta jest kartą wygraną. Prawdopodobieństwo wygranej bez zamiany karty:



$$P = \frac{2}{3} \tag{2}$$

Prawdopodobieństwo wygranej z zamianą karty:

$$P = \frac{1}{3} * 1 * 1 = \frac{1}{3} \tag{3}$$

Widać, że większe prawdopodobieństwo wygranej jest bez zamiany karty.

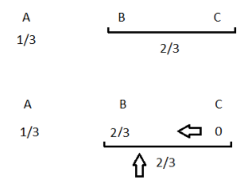
Oscylator harmoniczny: (Komentarz w celu niedublowania rozwiązań) Nasze rozwiązanie również wykorzystuje drzewko stochastyczne, jednak uwzględnia również teoretyczny opis zagadnienia, jako sytuację odwrotną do tzn. paradoksu Monty'ego Halla. Prowadzący, odkrywając kartę z wygraną, zmniejsza szansę na sukces z 2/3 do 1/3. (Spowodowane zmniejszeniem liczebności zbioru kart wygranych). Wniosek jest ten sam: nie opłaca się zmieniać karty w sytuacji, podanej w zadaniu.

- Alicja, Barbara i Cezary po napisaniu egzaminu w pewien sposób dowiedzieli się, że zdało go tylko jedno z nich. Alicja pyta

Złośliwego Wykładowcę o to kim jest ta osoba, ale w odpowiedzi dostaje tylko informację, że egzaminu na pewno nie zaliczył Cezary, cieszy się więc, oceniając, że jej szanse na zaliczenie wzrosły z $1/3$ do $1/2$. Gdzie popełnia błąd?

- Oba zadania to przyklady zagadki logicznej Monty'ego Hall'a. Gdy losujemy ze zbioru 3-elemtnowego w którym jeden element jest inny niz dwa pozostale, mamy szanse $\frac{1}{3}$ na wylosowanie go. Iluzja tutaj jest wyjecie ze zbioru jednego z dwoch takich samych elementów po tym jak już wybraliśmy. W drugim zadaniu mamy dość analogiczny problem. Szanse Alicji na zdanie zawsze wynosiły $\frac{1}{3}$. Wiedząc, że z prawdopodobieństwem $\frac{2}{3}$ któreś z przyjaciół Alicji zdało, po odkryciu wyniku Cezarego wiemy z prawdopodobieństwem $\frac{2}{3}$ była to Barbara. Alicja popełnia błąd sądząc, że ujawnienie niezdania któregoś z jej przyjaciół wpływa na prawdopodobieństwo jej zdania.

Rysunek do zadania 2 (egzamin)



Alternatywnie: początkowo prawdopodobieństwo zdania egzaminu przez każdą z osób wynosi $1/3$. Możemy trójkę znajomych podzielić na dwie grupy: Alicja (jest w tej grupie sama, prawdopodobieństwo zdania wynosi $1/3$) oraz grupę reszty znajomych (prawdopodobieństwo, że ktoś z tej grupy zdał wynosi $2/3$, z czego po $1/3$ na osobę). Dowiadując się, że Cezary nie zaliczył egzaminu, prawdopodobieństwo zdania każdej z grup się nie zmienia. Prawdopodobieństwo zmienia się jedynie wewnątrz drugiej grupy, czyli w efekcie Barbara ma $2/3$ szans na zdanie a Alicja ciągle $1/3$. Alicja popełniła błąd, myśląc, że prawdopodobieństwo równo "przejdzie" z Cezarego na nią i Barbarę.

Rozwiązanie z drzewkiem stochastycznym

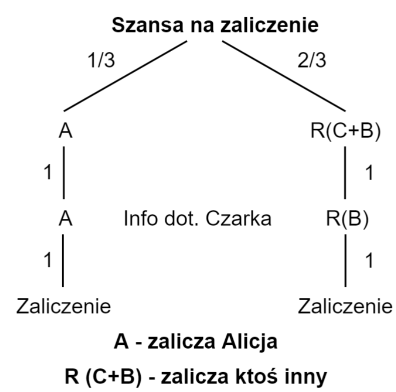
Widzimy, iż na początku, skoro mamy 3 osoby, to szansa Alicji wynosi

$$P(A) = \frac{1}{3} \tag{4}$$

Szansa na zaliczenie pozostałych jest dopełnieniem do 1, czyli:

$$P(R) = \frac{2}{3} \tag{5}$$

Po informacji od prowadzącego, Czarek wypadł z rozważań, natomiast zauważmy, iż w ogólności nie zwiększa to szans Alicji. Dzieje się tak, ponieważ dla całej trójki, prawdopodobieństwo zaliczenia jest takie samo i nie zmienia się z faktem, iż Czarek nie zaliczył (z punktu widzenia naszych rozważań nadal reszta ma większe p-stwo niż Alicja, nieistotne też, czy Czarek wie o swoim niepowodzeniu). Jej szansa na zaliczenie nadal wynosi $1/3$, myliła się.



- Rozważmy n monet, z których k jest asymetrycznych i reszka wypada na nich średnio 2 razy częściej niż orzeł. Wybieramy losowo monetę i w wyniku rzutu widzimy orła. Jaka jest szansa, że była to moneta asymetryczna?

- Prawdopodobieństwo wyrzucenia monety asymetrycznej:

$$P(asym) = \frac{k}{n}$$

Prawdopodobieństwo wyrzucenia monety symetrycznej:

$$P(sym) = \frac{n-k}{n}$$

Prawdopodobieństwo wyrzucenia orła na monecie symetrycznej:

$$P(o|sym) = \frac{1}{2}$$

Prawdopodobieństwo wyrzucenia orła na monecie asymetrycznej:

$$P(o|asym) = \frac{1}{3}$$

Prawdopodobieństwo wyrzucenia orła:

$$P(o) = \frac{k}{n} \cdot \frac{1}{3} + \frac{n-k}{n}\frac{1}{2} = \frac{k}{3n} + \frac{n-k}{2n} = \frac{3n-k}{6n}$$

Szukane prawdopodobieństwo na podstawie wzoru Bayesa:

$$P(asym|o) = \frac{P(o|asym)P(asym)}{P(o)} = \frac{\frac{1}{3} \cdot \frac{k}{n}}{\frac{3n-k}{6n}} = \frac{2k}{3n-k}.$$

Można również zbadać zagadnienie numerycznie, komputerowo generując liczby pseudolosowe i porównując wynik z teorią.

Poniższy program napisano w Pythonie 3.8.1.

```
import random as rand
try:
    k = int(input('Podaj liczbe asymetrycznych monet: '))
    s = int(input('Podaj liczbe symetrycznych monet: '))
    n = k + s
    print(f'Calkowita liczba monet: {n}')
    tries = int(input('Podaj liczbe prob: '))
    if tries <= 1:
        raise ValueError
except ValueError:
    print('Podano nieprawidlowe dane.')
    exit()
sym = ['r', 'o']
asym = ['r', 'r', 'o']

count_sym_o = 0
count_asym = 0
count_asym_o = 0

for _ in range(tries):
    coin = rand.randint(1, n)
```

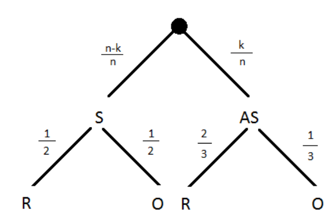
Monety asymetryczne przy losowym wyborze strony dwa razy częściej zwaracją reszkę.

```
        if coin > k:
            if rand.choice(sym) == 'o':

                count_sym_o += 1
            else:
                pass
        else:
            count_asym += 1
            if rand.choice(asym) == 'o':

                count_asym_o += 1
            else:
                pass
p_a = count_asym / tries
p_o = (count_sym_o + count_asym_o) / tries
p_oa = count_asym_o/count_asym
print(f'Przewidywane␣P(A):␣{k/n:␣.4f}\tOtrzymane␣P(A):␣{p_a:␣.4f}')
print(f'Przewidywane␣P(O):␣{(3*n−k)/(6*n):␣.4f}\tOtrzymane␣P(O):␣{p_o:␣.4f}')
print(f'Przewidywane␣P(O|A):␣{1/3:␣.4f}\tOtrzymane␣P(O|A):␣{p_oa:␣.4f}')
print(f'\nZe␣wz.␣Bayesa:␣P(A|O)␣=␣P(O|A)*P(A)/P(O)␣=␣', end='')
print(f'{p_oa␣*␣p_a␣/␣p_o:␣.6f}')
print(f'Przewidywane␣P(A|O)␣=␣␣2k/(3n−k)␣=␣{2*k/(3*n−k):␣.6f}')
```

Wybierz losowy element z asymetrycznej monety - szansa na $o = \frac{1}{3}$

Przykładowy wynik działania tego programu:

```
Podaj liczbe asymetrycznych monet: 50
Podaj liczbe symetrycznych monet: 100
Calkowita liczba monet: 150
Podaj liczbe prob: 100000
Przewidywane P(A):   0.3333     Otrzymane P(A):   0.3310
Przewidywane P(O):   0.4444     Otrzymane P(O):   0.4463
Przewidywane P(O|A):   0.3333   Otrzymane P(O|A):
0.3363

Ze wz. Bayesa: P(A|O) = P(O|A)*P(A)/P(O) =  0.249367
Przewidywane P(A|O) =  2k/(3n−k) =  0.250000
```



Metoda drzewa: Oznaczenia: S-symetryczna, AS-asymetryczna, O-orzeł, R-reszka, A-wybrano monetę asymetryczną, B-wyrzucono orła, A ∩ B -wybrano monetę asymetryczną i wyrzucono orła

$$P(B) = \frac{n-k}{n} * \frac{1}{2} + \frac{k}{n} * \frac{1}{3} = \frac{3n-k}{6n} \tag{6}$$

$$P(A \cap B) = \frac{k}{3n} \tag{7}$$

Ze wzoru na prawdopodobieństwo warunkowe obliczono jakie jest prawdopodobieństwo wybrania monetyy asymetrycznej pod warunkiem, że wyrzucono orła:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{k}{3n}}{\frac{3n-k}{6n}} = \frac{2k}{3n-k} \tag{8}$$

- W fabryce butów na $2n$ par k składa się z dwóch lewych butów, a pozostałe są dobre. Jak zapakować te buty do dwóch paczek po n par każda, żeby szansa na wykrycie błędu przez kontrolera była jak najmniejsza?

Kontroler sprawdza jedną parę z losowo wybranej paczki.

- Do pierwszej paczki wkładamy a par lewych butów. Wtedy w drugiej paczce będzie $k - a$ par lewych butów.

  Wtedy szansa na wykrycie błędu w pierwszej paczce wynosi $P(W_1) = \frac{a}{n}$, a w drugiej $P(W_2) = \frac{k-a}{n}$.

  Więc łączna szansa na wykrycie błędu wynosi $P(W) = \frac{1}{2} \cdot P(W_1) + \frac{1}{2} \cdot P(W_2) = \frac{k}{2n}$, więc niezależnie od podziału zawsze jest takie samo.

Prawdopodobieństwo kontroli p(K) (pKontroli) wyraża się sumą prawdopodobieństwa wylosowania pierwszej paczki i z niej „lewego” pudełka oraz wylosowania drugiej paczki i z niej „lewego” pudełka. Prawdopodobieństwo wylosowania każdej paczki jest jednakowe i wynosi 0.5. Prawdopodobieństwo kontroli jeśli do jednej z paczek włożyliśmy x „lewych” pudełek (do drugiej odpowiednio k-x „lewych”) wynosi

$$P(K) = 0.5 * x/n + 0.5 * (k - x)/n; \tag{9}$$

gdzie n to liczba pudełek w każdej paczce, k liczba „lewych” pudełek w sumie, a x liczba „lewych” pudełek w jednej z paczek.

Łatwo zauważyć, że dla każdego k, niezależnie od x P(K) będzie takie samo

```
using namespace std;

int main()
{
    int check =1;
    int n =100;
    float pKontroli[n][n];
        for(int k=0; k<n; k++)
        {
            pKontroli[k][0]=(0.5*0/n)+(0.5*(k-0)/n);\
```

```
            for(int x=1; x<k; x++)
            {
                pKontroli[k][x]=(0.5*x/n)+(0.5*(k-x)/n);\
                if(pKontroli[k][x-1]!=pKontroli[k][x])
                 check=0;
            }
        }
    cout<<check;
    return 0;
}
```

- Po rzucie 3 kostkami $k10$ wiemy tylko, że na każdej z nich wypadła inna liczba oczek. Jakie są prawdopodobieństwa, że: a) na żadnej nie wypadła 9, b) na pewnej kostce wypadła 9?

- Zbiór wszystkich możliwych zdarzeń, w których na kostkach są różne liczby oczek będzie miał liczność (kombinacje bez powtórzeń)

$$|\Omega| = \binom{10}{3} = \frac{10!}{3! \cdot 7!} = \frac{8 \cdot 9 \cdot 10}{6}$$

Zakładamy dla ułatwienia, że kostki są nierozróżnialne (nie interesuje mnie kolejność rzutów). a) Zbiór możliwych zdarzeń, w których nie wypadła 9 to po prostu taka sama kombinacja, tylko bez jednej możliwej cyfry.

$$|A| = \binom{9}{3} = \frac{9!}{3! \cdot 6!} = \frac{7 \cdot 8 \cdot 9}{6}$$

Korzystając ze wzoru z pierwszych zajęć:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{7}{10}$$

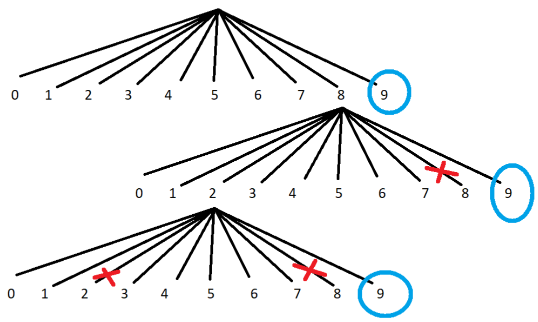
b) Jeśli na jednej kostce wypadła 9, na dwóch pozostałych wypadło coś dowolnego, innego od 9 i różnego:

$$|B| = 1 \cdot \binom{9}{2} = \frac{9!}{2! \cdot 7!} = \frac{9 \cdot 8}{2}$$

Zatem

$$P(B) = \frac{|B|}{|\Omega|} = \frac{3}{10}$$

Dla zachowania jasności rysunku dokładnie przedstawiona jest tylko jedna część. Reszta jest analogiczna



Bez warunku o różnej wartości oczek na każdej z kostek widzimy, że ilość możliwości wynosi 1000 ($10 \cdot 10 \cdot 10$). Warunek ten przy rozważaniu drugiego piętra drzewka odejmuje nam po jednej możliwej gałęzi (co skutkuje 10 gałęziami na trzecim piętrze). Natomiast na trzecim piętrze mamy o dwie gałązki mniej.

Biorąc to pod uwagę mamy $10 \cdot 9 \cdot 8 = 720$ gałęzi lub licząc inaczej $1000 - 10 \cdot 10 - 10 \cdot 9 \cdot 2 = 720$ Jakie jest prawdopodobieństwo, że na jednej z kości wystąpi 9?

Wystąpienie 9 na piętrze daje nam $9 \cdot 8 = 72$ wystąpień, na drugim piętrze daje nam 8 wystąpień, ale na 8 piętrze jest aż dziewięć dziewiątek, czyli też $9 \cdot 8 = 72$ wystąpień. Na trzecim piętrze wystąpienie 9 jest pojedyncze dla każdej gałęzi z piętra drugiego (oprócz tej gałęzi idącej od dziewiątki) to znaczy, że 9 jest $1 \cdot 8 \cdot 9 = 72$ Daje to nam w sumie 216 jednakowo prawdopodobnych kombinacji gdzie występuje „9".

$$P(''9'') = \frac{216}{720} = 0.3 \tag{10}$$

Prawdopodobieństwo, że na żadnej kostce nie wypadła „9" To 9 opcji na pierwszy piętrze drzewka,a dla każdej opcji to 8 gałązek na drugim, dla której każdej to 7 na trzecim. Co daje w sumie $9 \cdot 8 \cdot 7 = 504$ kombinacji bez „9"

$$P(bez''9'') = \frac{504}{720} = 0.7 \tag{11}$$

Zbiór kombinacji z 9 i zbiór kombinacji bez 9 to dla rozłączne zbiory, więc ich prawdopodobieństwo powinno sumować się do 1, co też się dzieje.

C6: Porozmawiajmy o Centralnym

Najważniejsze wnioski z wykładu:

- Opis testów diagnostycznych.
- Prawa wielkich liczb.

Zadania do wykonania:

- Napisz skrypt, który w automatyczny sposób wygeneruje LaTeXową tablicę wartości dystrybuanty wystandaryzowanego rozkładu normalnego.
- [2] Skrypt ten jest zaczerpnięty (garściami) z wykładu numer 4 i został wykonany w Mathematice.

```
TableForm[
  Table[NumberForm[
    CDF[NormalDistribution[], a + b] - 0.5, {5, 5}], {a, 0, 1,
    0.1}, {b, 0, 0.1, 0.01}],
  TableHeadings -> {Table[NumberForm[#, {3, 3}] &@a, {a, 0, 1, 0.1}],
    Table[NumberForm[#, {3, 3}] &@b, {b, 0, 0.1, 0.01}]}] // TeXForm
```

[2] Niewielka lecz kluczowa dla tego zadania różnica polega na dodaniu na końcu komendy TeXForm, która zamienia wyjście programu na jego LaTeXową wersję.

- Skrypt w Pythonie do generowania LaTeXowej tablicy wartości dystrybuanty wystandaryzowanego rozkładu normalnego

```python
from math import erf, sqrt

def phi(x):
    return (1.0 + erf(x / sqrt(2.0))) / 2.0

output = "\\begin{tabular}{ c | c c c c c c c c c c }\n"
output += "\t\\hline\n"
output += "\tx"
for i in range(0, 10):
    output += " & " + str(i / 100.0)
output += "\\\\\n\t\\hline\n"
for x in range(40):
    output += "\t " + str(x / 10.0)
    for i in range(0, 10):
        output += " & " + '{:.4f}'.format(phi(x / 10.0 + i / 100.0))
    output += "\\\\\n"
output += "\t\\hline\n"
output += "\end{tabular}\n"
print(output)
```

- Dwóch korektorów przeczytało książkę. Pierwszy znalazł 151 błędów, drugi 182, przy czym błędów znalazionych przez obu było 94. Czy byli to dobrzy korektorzy (odpowiedz ilościowo)?

- Pierwszy korektor znalazł 151 błędów, a drugi 182 błędy. Błędów znalezionych przez obu było 94. Na podstawie tego wyznaczamy ilość błędów wykrytych jako:

$$|A| + |B| - |A \cap B| = 151 + 182 - 94 = 239$$

Wiemy więc, że w tekście jest przynajmniej 239. Czułość testu wykonanego przez pierwszego korektora wynosi więc (maksymalnie) $\frac{151}{239} = 63\%$. Czułość testu drugiego wynosi (maksymalnie) $\frac{182}{239} = 76\%$. Z uwagi na specyfikę takiego zadania, czułość na takim poziomie nie jest wystarczająca. Nie byli to więc dobrzy korektorzy.

- Stosując twierdzenie Poissona[3] odpowiedz na pytanie jakie jest prawdopodobieństwa, że słodka bułka ma w środku (1) 0 rodzynek oraz (2) 15 rodzynków, jeśli średnio zawiera 5 rodzynków.

[3] Dlaczego możemy je tu stosować?

- Twierdzenie Poissona:

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \xrightarrow{n\to\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$

Dla $p_n \to 0$, $np_n \to \lambda > 0$
1) Prawdopodobnieństwo znalezienia 0 rodzynków:

$$P_k(\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \lambda = 5$$

$$P_0(5) = e^{-5} \frac{5^0}{0!} = e^{-5} = 6.74 \cdot 10^{-3}$$

2)Prawdopodobnieństwo znalezienia 15 rodzynków:

$$P_15(5) = e^{-5} \frac{5^15}{15!} = e^{-5} \cdot 0.023 = 1.57 \cdot 10^{-4}$$

W zadaniach 3 i 4 możemy w pewnym przybliżeniu stosować tw. Poissona, ponieważ:

- rodzynek ma mały rozmiar w stosunku do bułki, przez co rodzynki w jednej bułce nie "wpływają" na siebie i w bułce zmieści się dużo rodzynek
- z tego samego powodu prawdopodobieństwo, że losowo wybrana objętość bułki (rozmiaru rodzynka) jest rodzynkiem, jest małe
- średnia liczba rodzynek $\lambda = p_n n$

Te założenia to jednak tylko przybliżenia, ponieważ w bułce nie zmieści się nieskończona ilość rodzynek. Przybliżenie to jest jednak bardzo dobre

---

- Stosując twierdzenie Poissona[4] odpowiedz na pytanie ile średnio rodzynków powinna zawierać słodka bułka, żeby z prawdopodobieństwem 0.99 losowe ciastko zawierało co najwyżej 5 rodzynków?

[4] Dlaczego możemy je tu stosować?

- Twierdzenie Poissona:

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \stackrel{n\to\infty}{\to} \frac{\lambda^k}{k!} e^{-\lambda} \tag{12}$$

Dla $p_n \to 0$, $np_n \to \lambda > 0$.
Możemy przyjąć, że liczba rodzynków w pączku ma rozkład Poissona z parametrem λ.
Rozkład Poissona:

$$f(k,\lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \tag{13}$$

Rozwiązanie zadania: $P(X \leq 5)$

$$P(X=5)+P(X=4)+P(X=3)+P(X=2)+P(X=1)+P(X=0) = 0,99 \tag{14}$$

$$\frac{\lambda^5 e^{-\lambda}}{5!} + \frac{\lambda^4 e^{-\lambda}}{4!} + \frac{\lambda^3 e^{-\lambda}}{3!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^0 e^{-\lambda}}{0!} = 0,99 \tag{15}$$

$$\lambda \approx -1,17323 \text{ lub } \lambda \approx 1,78528 \tag{16}$$

Równanie zostało rozwiązane za pomocą programu WolfhramAlpha. Wyszły dwa rozwiązania, dodatnie i ujemne. Za prawidłową odpowiedź przyjęto wartość dodatnią, ponieważ rodzynki nie mogą być "ujemne". Dlatego: $\lambda \approx 1,79$
Średnio słodka bułka powinna zawierać $\approx 1,79$ rodzynków.

- Średnią ilość rodzynków spełniającą założenia można wyznaczyć numerycznie, wyznaczając prawdopodobieństwo zdarzenia dla rónych średnich do osiągnięcia wartości zbliżonej $P = 0.99$. Poniższy kod w języku Java pozwala na wyznaczenie z dużą dokładnością szukanej średniej. ($P = 0.99$ z dokładnością do 4 cyfr po przecinku). Otrzymujemy $\lambda = 1.7852$.

```java
public static void main(String[] args) {
        double avg = 5;          //srednia rodzynkow
        double s = 0.0001;       //rozdzielczosc sprawdzania
        double Pmax = 0.99;      //prawdopodobienstwo szukane
        double P;

        do {
        double P0 = Math.pow(avg, 0) / 1 * Math.pow(Math.E, -avg);
        double P1 = Math.pow(avg, 1) / 1 * Math.pow(Math.E, -avg);
        double P2 = Math.pow(avg, 2) / 2 * Math.pow(Math.E, -avg);
        double P3 = Math.pow(avg, 3) / 6 * Math.pow(Math.E, -avg);
        double P4 = Math.pow(avg, 4) / 24 * Math.pow(Math.E, -avg);
        double P5 = Math.pow(avg, 5) / 120 * Math.pow(Math.E, -avg);
```

```
P = P0 + P1 + P2 + P3 + P4 + P5;

 if (P<Pmax)
         avg -= s;
}while (P<Pmax);
System.out.println(avg + " " + P);
}
```

- Wyznacz w przybliżeniu (z twierdzenia Poissona) jakie jest prawdopodobieństwo, że wśród losowo wybranych 400 osób 10 urodziło się w 5 sierpnia.

- Prawdopodobieństwo, że osoba urodziła się 5 sierpnia $p = \frac{1}{365}$, $q = 1 - p = \frac{364}{365}$, $n = 400$. Wartość oczekiwana

$$EX = \lambda = n \cdot p = 1,096$$

Odchylenie standardowe

$$s = \sqrt{c \cdot p \cdot q} = 1,045$$

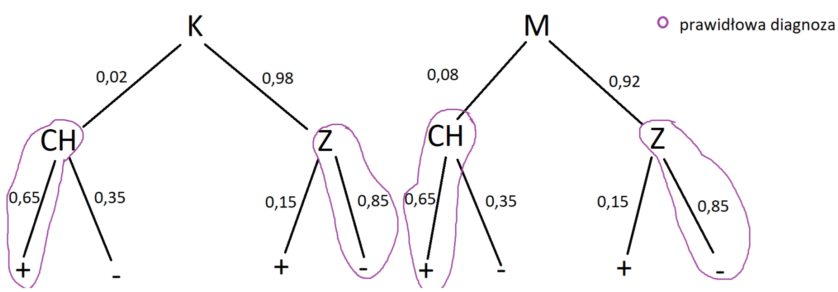
Prawdopodobieństwo, że wśród losowo wybranych 400 osób 10 urodziło się 5 sierpnia ($k = 10$):

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$
$$P(X = 10) = 2,301 * 10^{-7}$$

- W pewnej grupie wiekowej na chorobę wieńcową cierpi 8% mężczyzn i 2% kobiet. Wylosowano z tej grupy mężczyznę i kobietę, by poddać ich próbie wysiłkowej[5]. Oblicz dla każdego z nich
  - Prawdopodobieństwo, że próba wysiłkowa doprowadziła do prawidłowej diagnozy.
  - Prawdopodobieństwo, że osoba z wynikiem dodatnim jest chora.
  - Prawdopodobieństwo, że osoba z wynikiem ujemnym jest zdrowa.

[5] Tak jak w zadaniu z wykładu, czułość testu to 65%, a jego swoistość 85%.

- Metoda drzewek:
Drzewko odnosi się do wszystkich podpunktów, zaznaczone fragmenty odnoszą się do podpunktu pierwszego.
Legenda do drzewek:



- M - mężczyzna, K - kobieta,
- CH - chory, Z - zdrowy,
- + - pozytywny wynik, - - negatywny wynik.

Tabelka jest taka sama dla kobiet i mężczyzn: Prawdopodobień-

	wynik pozytywny	wynik negatywny
osoba zdrowa	15%	85%
osoba chora	65%	35%

stwo, że diagnoza jest prawidłowa u kobiet:

$$P = 0,02 \cdot 0,65 + 0,98 \cdot 0,85 = 0,013 + 0,833 = 0,846 \qquad (17)$$

Prawdopodobieństwo, że diagnoza jest prawidłowa u mężczyzn:

$$P = 0,08 \cdot 0,65 + 0,92 \cdot 0,85 = 0,052 + 0,782 = 0,834 \qquad (18)$$

Prawdopodobieństwo, że kobieta z wynikiem dodatnim jest chora:

- A- jest chora
- B - wynik dodatni

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad (19)$$

odczytane z drzewka:

$$P(A|B) = \frac{0,02 \cdot 0,65}{0,02 \cdot 0,65 + 0,98 \cdot 0,15} = 0,08125 \qquad (20)$$

Prawdopodobieństwo, że mężczyzna z wynikiem dodatnim jest chory (analogicznie wykonano analizę dla mężczyzn):

$$P(A|B) = \frac{0,08 \cdot 0,65}{0,08 \cdot 0,65 + 0,92 \cdot 0,15} = 0,27 \qquad (21)$$

Prawdopodobieństwo, że kobieta z wynikiem ujemnym jest zdrowa:

- A - jest zdrowa
- B - wynik ujemny

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{22}$$

odczytane z drzewka:

$$P(A|B) = \frac{0,98 \cdot 0,85}{0,02 \cdot 0,35 + 0,98 \cdot 0,85} = 0,992 \tag{23}$$

Prawdopodobieństwo, że mężczyzna z wynikiem ujemnym jest zdrowy: (analogicznie wykonano analizę dla mężczyzn):

$$P(A|B) = \frac{0,92 \cdot 0,85}{0,92 \cdot 0,85 + 0,08 \cdot 0,35} = 0,965 \tag{24}$$

◆ Metoda numeryczna:

```
import random as rd
N = 1e7
test_sw = 0.85
test_cz = 0.65
zdrowiplus_k, zdrowiminus_k, chorzyplus_k, chorzyminus_k = (0, 0, 0, 0)
zdrowiplus_m, zdrowiminus_m, chorzyplus_m, chorzyminus_m = (0, 0, 0, 0)

# Bierzemy N kobiet i N mezczyzn (populacja 2N)
for i in range(int(N)):
    # Dla kobiet
    if rd.random() <= 0.02:
        # Kobieta jest chora, przeprowadzamy test
        if rd.random() <= test_cz:
            chorzyplus_k += 1
        else:
            chorzyminus_k += 1
    else:
        # Kobieta jest zdrowa, przeprowadzamy test
        if rd.random() <= test_sw:
            zdrowiminus_k += 1
        else:
            zdrowiplus_k += 1

    # Dla mezczyzn
    if rd.random() <= 0.08:
        # Mezczyzna jest chory, przeprowadzamy test
```

```
        if rd.random() <= test_cz:
            chorzyplus_m += 1
        else:
            chorzyminus_m += 1
    else:
        # Mezczyzna jest zdrowy, przeprowadzamy test
        if rd.random() <= test_sw:
            zdrowiminus_m += 1
        else:
            zdrowiplus_m += 1

print(f'Kobiety: \n \
    P(CH i + v Z i -): {(zdrowiminus_k + chorzyplus_k)/N * 100 :.3}%\n \
    P(CH|+): {chorzyplus_k/(zdrowiplus_k+chorzyplus_k) * 100 :.3}% \n \
    P(Z|-): {zdrowiminus_k/(chorzyminus_k+zdrowiminus_k) * 100 :.3}%\n')
print(f'MEZCZYZNI: \n \
    P(CH i + v Z i -): {(zdrowiminus_m + chorzyplus_m)/N * 100:.3f}%\n \
    P(CH|+): {chorzyplus_m/(zdrowiplus_m+chorzyplus_m) * 100 :.3f}% \n \
    P(Z|-): {zdrowiminus_m/(chorzyminus_m+zdrowiminus_m)*100 :.3f}%\n')
```

Środowisko *lstlisting* nie pozwala na wstawienie symbolu ∩ jako unicode, dlatego spójnik 'i'

Przykładowy wynik działania programu dla $N = 10^7$:

```
Kobiety:
    P(CH i + v Z i -): 84.6%
    P(CH|+): 8.13%
    P(Z|-): 99.2%

MEZCZYZNI:
    P(CH i + v Z i -): 83.398%
    P(CH|+): 27.343%
    P(Z|-): 96.536%
```

◆ Udowodnij twierdzenie Poissona, przeliczając występującą w jego tezie granicę przy założeniu upraszczającym $np_n = \lambda$.

Twierdzenie Poissona mówi, że rozkład Poissona jest granicznym przypadkiem rozkładu dwumianowego. Oznaczamy je $X \sim Poiss(\lambda)$. Wówczas

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Zakładamy:

$$np_n = \lambda$$

Obliczamy granicę rozkładu dwumianowego przy $n \to \infty$

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} p_n^k \frac{(1-p_n)^n}{(1-p_n)^k} =$$
$$= \frac{1}{k!} \lim_{n\to\infty} n(n-1)...(n-k+1) p_n^k (\frac{1}{1-p_n})^k (1-\frac{\lambda}{n})^n =$$
$$= \frac{1}{k!} \lim_{n\to\infty} n^k \frac{n}{n} \frac{n-1}{n} ... \frac{n-k+1}{n} p_n^k (\frac{1}{1-p_n})^k (1-\frac{\lambda}{n})^n =$$
$$= \frac{1}{k!} \lim_{n\to\infty} (\lambda)^k \frac{n}{n} \frac{n-1}{n} ... \frac{n-k}{n} (\frac{1}{1-p_n})^k (1-\frac{\lambda}{n})^n =$$
$$= \frac{1}{k!} \lambda^k e^{-\lambda}$$

- Otrzymany wynik potwierdziliśmy również prostą operacją w Mathematice:

$$p = \lambda / n \tag{25}$$

$$P = n!/(((n-k)!) * (k!)) * ((p)^k) * ((1-p))^{(n-k)}; \tag{26}$$

$$Limit[P, n- > Infinity]//Simplify \tag{27}$$

C7: O estymie estymatora.

Najważniejsze wnioski z wykładu:

- zastosowania Mocnego Prawa Wielkich Liczb.
- Centralne Twierdzenie Graniczne.
- Definicje estymatorów i ich najwazniejsze cechy.

Zadania do wykonania:

- W Polsce jest $n = 24.6 \cdot 10^6$ podatników[6] i każdy z nich może pomylić się przy wypełnianiu zeznania podatkowego, przy czym wartość pomyłki opisuje zmienna losowa X_i, dla której $\mathbb{E}X_i = 0$ oraz $\sigma = \sqrt{\mathcal{D}^2 X_i} = 10^2 PLN$. Jaka jest szansa, że straty państwa w wyniku tych błędów przekroczą jeden grosz na podatnika? A 5 groszy?

[6] Dane za rok 1998.

- Liczba podatników:

$$\begin{aligned} n &= 24,6 * 10^6 \\ \sqrt{n} &= 4960 \\ E(X) &= 0 \end{aligned}$$

Odchylenie standardowe:

$$\sigma = \sqrt{D^2 X_i} = 100PLN$$

Korzystamy z Centralnego Twierdzenia Granicznego (n jest bardzo dużą liczbą):

$$p(X_1 + X_2 + X_3 + ... + X_n > xn) = p(\frac{X_1 + X_2 + ... + X_n}{\sigma\sqrt{n}} > \frac{xn}{\sigma\sqrt{n}}) = 1 - \phi(\frac{x\sqrt{n}}{\sigma})$$

Gdzie $\phi(t)$ to wartość dystrybuanty rozkładu normalnego w punkcie t.

Za x podstawiamy 0,01 bo szukamy jaka jest szansa, że państwo straci 0,01 PLN na jednego podatnika. Prawdopodobieństwo to wynosi:

$$p = 1 - \phi(\frac{0,01 * 4960}{100}) = 1 - \phi(0,496) = 1 - 0,69 = 0,31 = 31\%$$

Dla x=0,05

$$p = 1 - \phi(\frac{0,05 * 4960}{100}) = 1 - \phi(2,48) = 1 - 0,9934 = 0,0066 = 0,66\%$$

◆ Metoda numeryczna w języku java.

```
import java.util.Random;

public class BledyPodatkowe {
        static double stdev=100;
        static double n=24.6*1000000;
        public static void main(String[] args) {
                Random r=new Random();
                double bladNaPodatnika;
                double sum;
                int k=0;int m=0;
                int liczbaProb=100;
                for(int j=1;j<liczbaProb;j++) {
                        sum=0;
                        for(int i=0;i<n;i++) {
                        sum+=r.nextGaussian()*stdev;
                        }
                        bladNaPodatnika =sum/n;
                if(bladNaPodatnika>0.01) {k++;}
                if(bladNaPodatnika>0.05) {m++;}
                }
        System.out.println((double)k/liczbaProb);
        System.out.println((double)m/liczbaProb);

        }

}
```

Dla 100 iteracji prawdopodobieństwo strat państwa większych niż 1 grosz na podatnika wynosi 0,31, a większych niż 5 groszy wynosi 0,01. Zrobienie symulacji dla większych n wymagałoby większej mocy obliczeniowej.

◆ W ubiegłym roku studenci Probabilistyki wramach prac domowych rzucili kostkami $k6$ ok. $n = 10^4$ razy. Ile wynosi prawdopodobieństwo, że suma wyrzuconych oczek mieści się między 34950 a 35050?

◆ Skorzystano z Centralnego Twierdzenia Granicznego:

$$\frac{\sum_{i=1}^{n} x_i - n\mu}{\sigma\sqrt{n}} \tag{28}$$

X_i - wynik rzutu kostką

$$n = 10^4 = 10000 \tag{29}$$

$$EX_i = 3,5 \tag{30}$$

$$D^2 = \frac{91}{6} - (3,5)^2 = 2,9 \tag{31}$$

Ponieważ:

$$\frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{1}{6})1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} \tag{32}$$

$$EX^2 = \sum_{i=1}^{6} k^2 P(X = k) = \sum_{i=1}^{6} k^2 \frac{1}{6} = \frac{7}{2} \tag{33}$$

$$P\left(\frac{34950 - n \cdot 3,5}{\sqrt{2,9 \cdot n}} < \frac{\sum_{i=1}^{6} - n \cdot 3,5}{\sqrt{2,9 \cdot n}} < \frac{35050 - n \cdot 3,5}{\sqrt{2,9 \cdot n}}\right) = \tag{34}$$

$$= P(-0,29 < N(0,1) < 0,29) = F(0,29) - F(-0,29) = \tag{35}$$

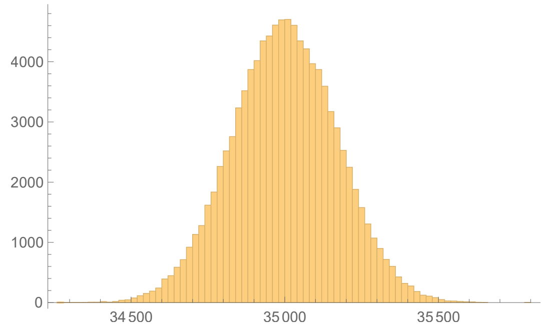
$$= 2 \cdot 0,61 - 1 = 1,22 - 1 = 0,22 \tag{36}$$

Prawdopodobieństwo wynosi 0,22.

- Metoda symulacyjna z wykorzystaniem Wolfram Mathematica 12.1.

```
sum = Table[Total[RandomInteger[5, 10000] + 1], 100000];
sum2 = Select[sum, # < 35500 &];
sum3 = Select[sum2, # > 34950 &];
Length[sum3]
Histogram[sum, 50]
```

Zaprezentowany fragment kodu kolejno:
-generuje listę 100 000 sum 10 000 liczb od 1 do 6 (od 0 do 5 +1)(sum);
-przekształca listę sum na listę spełniającą pierwszy warunek < 35050(sum2);
-przekształca listę sum2 na listę spełniającą drugi warunek < 35500(sum3);
-zlicza ilość elementów elementów tej listy;
-dodatkowo generuje histogram dla pierwszej listy;



Rysunek 8: Histogram sum

Wynikiem Zliczeń w tym przypadku było 22 733 co przy podzieleniu przez liczbę wszystkich przypadków daję szacowane prawdopodobieństwo zdarzenia równe 22,733%

- Narysuj zamkniętą krzywą mieszczącą się w kwadracie jednostkowym. Oblicz przy pomocy metody Monte Carlo pole powierzchni ograniczone tą krzywą. Spróbuj oszacować dokładność obliczeń.

W wersji podstawowej możecie Państwo policzyć ile prób *wpadło* do zbioru "ręcznie".

- Symulacja wykonana w języku Java wyznacza pole koła wpisanego w kwadrat o boku 1:

```
public static void main(String[] args) {
                Random r = new Random();

                int a = 1;              //dlugosc boku kwadratu o srodku w (1/2, 1/2)
                int n = 1000000;        //ilosc punktow losowanych
                int ctr = 0;            //zliczanie wystapien
                double [] X = new double [n];
                double [] Y = new double [n];
                for (int i=0; i<n; i++) {
                        X[i] = r.nextDouble();
                        Y[i] = r.nextDouble();

                        if (Math.sqrt((X[i]-0.5)*(X[i]-0.5)+(Y[i]-0.5)*(Y[i]-0.5))<0.5)
                                ctr++;
                }
                System.out.println("Pole kola wpisanego w kwadrat o boku 1: "
                + (ctr/(double)n));
        }
```

Kilka przykładowych wyników działania programu: 0.785028, 0.785754, 0.785059, 0.785717, 0.785501. Obliczając pole na podstawie wzoru konwencjonalnego przybliżając $\pi = 3.14159$ otrzymujemy pole równe 0,78540, co pozwala zauważyć że metoda ta jest dokładna (dla 10^6 próbek i kwadratu o polu 1) do trzeciego miejsca po przecinku.

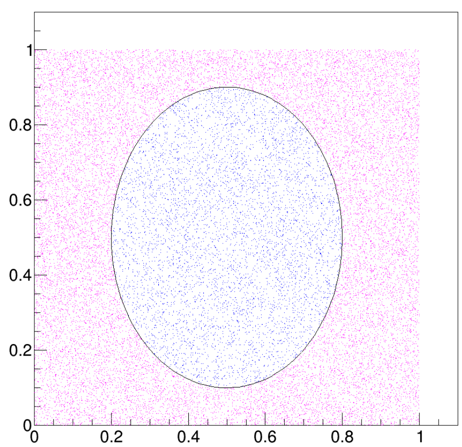
- Poniższy kod napisany w CERN ROOT 6.08 pozwala na obliczenie pola powierzchni dowolnej elipsy za pomocą metody Monte Carlo. Aby obliczać pole poza kwadratem jednostkowym, należy zamienić definicję limitów.

Aby nie zmniejszyć czytelności kodu, pozwolono aby wchodził na margines(oczywiście niektóre fragmenty nadal należało przenieść do następnej linii

```
void elipsa(){
        TRandom *gen = new TRandom();
        //ustawienie parametrow polosi a i b;
        double osA = 0.3;
        double osB = 0.4;
        //ustawienie polozenia srodka elpisy;
        double polozenieX = 0.5;
        double polozenieY = 0.5;
        //ustawienie ilosci wylosowanych punktow;
        const int rozmiar = 30000;
        //ilosc punktow wewnatrz elipsy;
        int num1 = 0;
        //ilosc punktow na zewnatrz elipsy;
        int num2 = 0;
        TGraph *graf1 = new TGraph();
        TGraph *graf2 = new TGraph();
        double x = 0;
        double y = 0;
        double gornylimitX = 1; //polozenieX + 1.1*osA;
        double dolnylimitX = 0; //polozenieX - 1.1*osA;
        double gornylimitY = 1; //polozenieY + 1.1*osB;
        double dolnylimitY = 0; //polozenieY - 1.1*osB;

        for(int i = 0; i < rozmiar; i++){
                x = gen->Uniform(dolnylimitX, gornylimitX);
                y = gen->Uniform(dolnylimitY, gornylimitY);
                if(TMath::Sqrt(TMath::Power((x - polozenieX)/osA,2) +
                TMath::Power((y - polozenieY)/osB,2)) <= 1 && num1 < rozmiar){
                        graf1->SetPoint(num1,x,y);
                        num1++;
                }
                if(TMath::Sqrt(TMath::Power((x - polozenieX)/osA,2) +
                TMath::Power((y - polozenieY)/osB,2)) >1 && num2 < rozmiar){
                        graf2->SetPoint(num2,x,y);
                        num2++;
                }
```

```
        }
        TCanvas *kanwa1 =
        new TCanvas("kanwa1" , "Plotting Canvas 1", 150,10,990,990);
        graf2->Draw("PA");
        graf1->Draw("Psame");
        graf1->SetMarkerColor(4);
        graf2->SetMarkerColor(6);
        TEllipse *kolo = new TEllipse(polozenieX,polozenieY,osA,osB);
        kolo->Draw("same");
        kolo->SetFillStyle(3008);
        double Pole = (1.0*num1/(num1 + num2))*
        ((gornylimitX - dolnylimitX)*(gornylimitY - dolnylimitY));
        cout<<"Pole figury wynosi: "<<Pole<<endl;
        double niepewnosc = 1.0/TMath::Sqrt(num1 + num2);
        cout<<"Niepewnosc pola figury wynosi: "<<niepewnosc<<endl;
}
```

Poniżej znajduje się przedstawienie sposobu działania tego programu. Można zauważyć, że graf nie jest wypełniony całkowicie. Ponieważ nie wpływa to na wynik, postanowiono nie dodawać zbędnych linijek kodu do programu.



Rysunek 9: Przykładowy graf

Pole figury zostało obliczone zgodnie ze wzorem

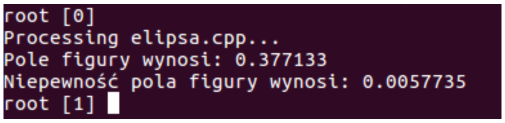
$$P_f = \frac{n_a}{n_{all}} * P_k$$

n_a to wszystkie punkty znajdujące się wewnątrz figury, n_all to wszystkie punkty, P_k to pole wygenerowanego kwadratu.

Niepewność pola została przyjęta jako względna dokładność obliczenia całki

$$\Delta P_f = \frac{1}{\sqrt{n_{all}}}$$

Wartości zwrócone przez program:



Rysunek 10: przykładowe wartości zwracane przez program

- Wyznaczanie przy pomocy metody Monte Carlo pola koła P_k wpisanego w kwadrat jednostkowy o polu $P = 1$ opiera się na poniższym równaniu:

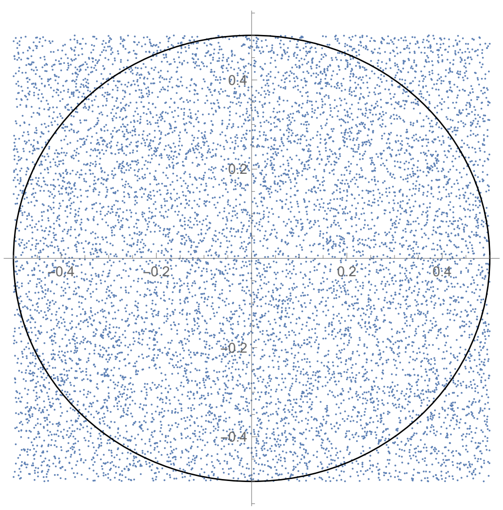
$$P_k = P\frac{k}{n} \tag{37}$$

gdzie k to liczba punktów w okręgu i na okręgu, a n to liczba wszystkich punktów.

```
points = RandomReal[{-0.5, 0.5}, {10000, 2}];
plot = ListPlot[points, AspectRatio -> 1];
circle = Graphics[Circle[{0, 0}, 0.5]];
Show[plot, circle]
Count[Map[Norm, points], _?(# <= 0.5 &)]/10000.
```

7

[7] Kod zapisany w Mathematice



Rysunek 11: Wykres okręgu o promieniu $R = 0.5$ wpisanego w kwadrat z wylosowanymi 10 000 punktów.

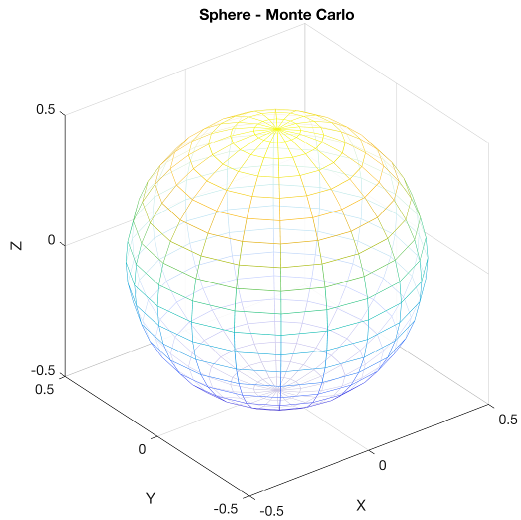
Dla $n = 10000$ obliczone pole wynosiło $P_k = 0.7854$, dla $n = 100000$ $P_k = 0.7851$, a dla $n = 1000000$ $P_k = 0.7855$. Gdy przyjmiemy $\pi = 3.1415926535$ pole tego koła obliczone z klasycznego wzoru z dokładnością pięciu miejsc po przecinku wynosi $P_k = \pi r^2 = 0.78539$

- Pomimo, iż w zadaniu mowa o krzywej zamkniętej na jednostkowej płaszczyźnie, człowiek nie płaszczak 3-ci wymiar lubi, więc

Płaszczak - popularnonaukowa nazwa oznaczająca istoty istniejące w przestrzeni dwuwymiarowej. Zpolszczona nazwa stworzeń z książki *Flatlandia czyli Kraina Płaszczaków* autorstwa *Edwina Abbotta*.

postanowiliśmy odwzorować zadanie z użyciem metody Monte Carlo w trójwymiarowej przestrzeni.

Stworzono, więc sferę jednostkową umieszczoną w jednostkowym sześcianie.



- Objętość sześcianu $V_c = a^3 = 1$
- Objętość sfery $V_s = \frac{4}{3}\pi r^3 = \frac{4}{3}\pi(0,5)^3 = 0,5235987756$

Metoda Monte Carlo

$$\frac{V_c}{V_s} \approx \frac{k}{N}$$

gdzie N liczba losowo wybranych punktów w sześcianie, zaś k liczba punktów które znalazły się wewnątrz sfery.

Wizualizacja symulacji dla $N = 10000$:

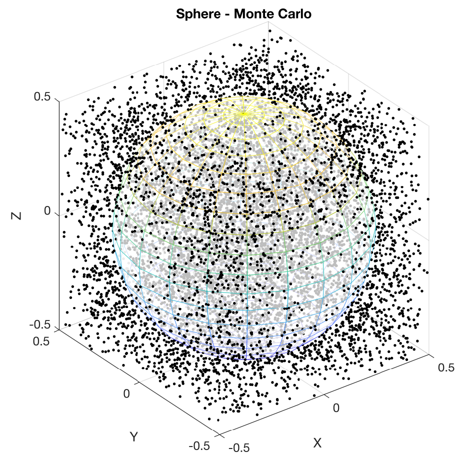
Wyniki dla symulacji przy $N = 10000$:

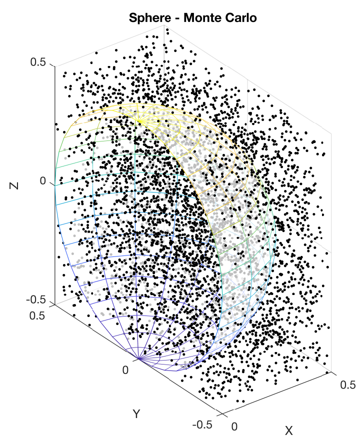
$$k = 5206$$

$$V_s \approx \frac{k}{N} * V_c = \frac{5206}{10000} * 1 = 0,5206$$

Widać różnice wynoszące $\frac{1}{100}$ realnej wartości objętości.

$$V_s = 0,5235987756 \approx 0,5206$$





W teorii jeśli $N \to \infty$ to $\left(\frac{k}{N} * V_c\right) \to V_s$, spróbujmy więc podnieść N.

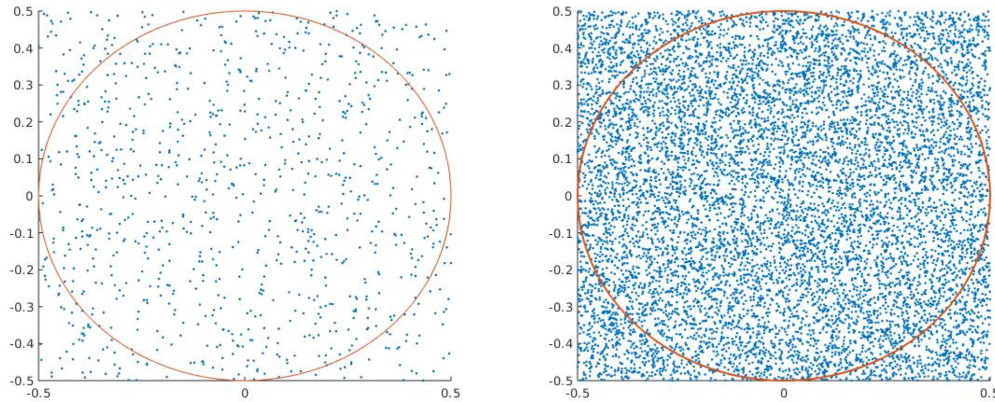
$$N = 1000000$$
$$k = 523473$$
$$V_s \approx 0,523473$$

Dla tak dużego N różnica objętości wynosi około $\frac{1}{1000}$.

Objętość wyznaczoną można na wykorzystać do oszacowania π:

$$\pi \approx \frac{3}{4}\frac{V}{r_3} = \frac{3}{4}\frac{0,523473}{(0,5)^3} = 3,140838$$

- Zadanie zostało wykonane w Matlabie.



Wykres z lewej przedstawia 1000 punktów, a wykres z prawej 10000.

Można wyznaczyć pole koła wpisanego w kwadrat jednostkowy (o polu 1) na podstawie liczby punktów, które zawarte są w tymże kształcie. Stanowią one jakąś część całego kwadratu, zatem pole kształtu możemy wyznaczyć według takiego wzoru:

$$P_{ko} = P_k \frac{n_{in}}{n_{all}}$$

gdzie:

Pko - pole koła, Pk - pole kwadratu, n in - liczba punktów w kole, n all - liczba wszystkich punktów

Narysowane koło ma pole:

$$P_{ko} = \pi r^2 = \pi \cdot (0,5)^2 = \mathbf{0,785}$$

Pole koła wyznaczone na podstawie liczby punktów:

Dla 1000 punktów:

$$P_{ko} = 1 \frac{805}{1000} = 0,805$$

Dla 10000 punktów:

$$P_{ko} = 1 \frac{7837}{10000} = 0,7837$$

Dla 100000 punktów:

$$P_{ko} = 1 \frac{78521}{100000} = 0,78521$$

Dla 1000000 punktów:

$$P_{ko} = 1\frac{785412}{1000000} = 0,7854$$

Wyznaczenie pola staje się dokładniejsze wraz ze wzrostem liczby punktów. W przybliżeniu staje się równe wartości teoretycznej przy liczbie 100 000 punktów.

Skrypt w Matlabie:

```
%ta funkcja znajduje sie w oddzielnym skrypcie
function h = circle(x,y,r)
hold on
th = 0:pi/50:2*pi;
xunit = r * cos(th) + x;
yunit = r * sin(th) + y;
h = plot(xunit, yunit, 'linewidth', 1.5);
hold off
end

rng(0,'twister')
a = -0.5;
b = 0.5;
x = (b-a).*rand(10000,1) + a;
y = (b-a).*rand(10000,1) + a;

counter = 0;

hold on
scatter(x,y,'.');
h = circle(0,0,0.5);
hold off

for i=1:length(x)
    if(hypot(x(i),y(i)) <= 0.5)
        counter = counter + 1;
    end
end
```

Zachęcam do rozwiązań analitycznych, jak również symulacji.

◆ Dodajemy 10 000 liczb, zaokrąglonych z dokładnością do 10^{-m}. Błędy zaokrąglenia są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym w przedziale $(-10^{-m}/2, 10^{-m}/2)$. Wyznaczyć przedział, w którym z prawdopodobieństwem 0.99 będzie się zawierał błąd sumy.

- n = 10000 liczb zaokrąglonych do ok. 10^{-m}

  X_k- niepewność k-tej liczby

  $$S_n = \sum_{k=1}^{n} X_k$$

  dla P(X)=0,99 wyznaczamy przedział w którym będzie zawierał się błąd sumy

  dla każdego k $\quad E(X_k) = \int_{-\infty}^{\infty} xg(x)dx = \frac{1}{10^{-m}} \int_{-\frac{1}{2}10^{-m}}^{\frac{1}{2}10^{-m}} xdx = 0 = M$

  $$V(X_k) = \int_{-\infty}^{\infty} x^2 g(x)dx = E(X^2) - (E(X))^2 = \frac{1}{10^{-m}} \int_{-\frac{1}{2}10^{-m}}^{\frac{1}{2}10^{-m}} x^2 dx = \frac{10^{-2m}}{12} = \sigma^2$$

  $$\sigma = \frac{10^{-m}}{2\sqrt{3}}$$

  $$P(|\frac{S_n - Mn}{\sigma\sqrt{n}}| < y) = P(|\frac{S_n}{\frac{10^{-m}}{2\sqrt{3}} * 100}| < y) = P(|U_n| < y)$$

  Chcemy, żeby $P(|U_n| < y) = 0,99$

  $$P(|U_n| < y) = P(U_n < y) - P(U_n < -y) = F(y) - F(-y) = (\frac{1}{2} + \phi(y)) - (\frac{1}{2} + \phi(-y)) = \phi(y) - \phi(-y)$$

  $$P(|U_n| < y) = 2\phi(y) = 0,99$$

  $$\phi(y) = 0,495 \quad => \quad y = 2,58...$$

  z prawdopodobieństwem 0,99:

  $$\frac{S_n}{\frac{10^{2-m}}{2\sqrt{3}}} < 2,58$$

  $$|S_n| = |\sum_{k=1}^{n} X_k| < \frac{10^{2-m}}{2\sqrt{3}} * 2,58$$

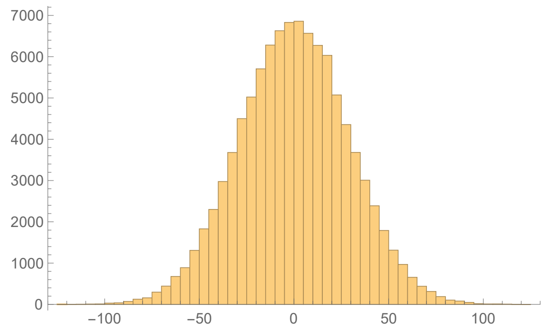
  Otrzymujemy przedział:

  $$-\frac{10^{2-m}}{2\sqrt{3}} * 2,58 < \sum_{k=1}^{n} X_k < \frac{10^{2-m}}{2\sqrt{3}} * 2,58$$

- Metoda symulacyjna z wykorzystaniem Wolfram Mathematica 12.1.

```
sum = Table[Total[(RandomReal[1, 10000] - 1/2)], 100000];
sum2 = Sort[sum]
Part[sum2, 500]
Part[sum2, 99500]
Histogram[sum, 50]
```

Zaprezentowany fragment kodu kolejno:
-generuje listę 100 000 sum 10 000 liczb od -1/2 do 1/2 (od 0 do 1 -1/2)(sum);
-sortuję tą listę;
-wypisuje 500 i 99 500 element tej listy;
-dodatkowo generuje histogram dla pierwszej listy;



Rysunek 12: Histogram sum

Rozkład wyników powinien być symetryczny względem zera, dlatego biorąc pod uwagę wygenerowanie 100 000 wyników odrzucamy po 500 z każdego skaju wybierając element znajdujący się krawędzi naszego zbioru w którym znajduje się 99% wyników. Otrzymane w ten sposób wyniki wynoszą w tym przypadku odpowiednio -74.1079 oraz 74.2388. Moduły wyników są do siebie zbliżone co zgadza się z intuicją o symetrii rozkładu. Z całości wynika że po zsumowaniu 10000 liczb w ten sposób nasze zaokrąglenie spowoduje błąd mieszczący się, w 99% przypadków, w przedziale od około $-74.1079 * 10^{-m}$ do $74.2388 * 10^{-m}$.

C8: Teoretycznie tak.

Najważniejsze wnioski z wykładu:

- Rodzaje zbieżności.
- Aksjomatyczna definicja prawdopodobieństwa.
- Metody estymacji.

Zadania do wykonania:

- Należy wyznaczyć wybraną metodą parametry rozkładów umieszczonych na stronach 49-50.

C9: Znaj rozkłady, nie ma rady!

Najważniejsze wnioski z wykładu:

- wzór właczeń-wyłączeń oraz ciągłość prawdopodobieństwa.
- funkcje tworzące, charakterystyczne i generujące momenty.
- najważniejsze relacje pomiędzy typowymi rozkładami prawdopodobienstwa.
- funkcje zmiennych losowych.

Zadania do wykonania:

- Należy wyznaczyć funkcje tworzące, funkcje generujące momenty i funkcje charakterystyczne (albo uzasadnić, że to niemozliwe) dla rozkładów umieszczonych na stronach 49-50.

C10: Postawmy sobie hipotezę.

Najważniejsze wnioski z wykładu:

- Na czym polega testowanie hipotez statystycznych.
- Jak je rozstrzygać?

Zadania do wykonania:

- Wyjaśnij (oraz podaj przykład) czym są i czym się różnią między sobą błędy pierwszego i drugiego rodzaju. Dlaczego nie traktujemy ich tak samo?
- Stosując funkcje tworzące albo charakterystyczne albo generujące momenty znajdź rozkład sumy n niezależnych zmiennych losowych o ustalonym rozkładzie spośród tych opisanych na stronach 49-50.
- Niech $X \sim \mathcal{N}(0, 1)$. Wyznacz rozkład zmiennej losowej X^n.
- Niech X będzie zmienną losową z ciągłego rozkładu wykładniczego. Wyznacz rozkład zmiennej losowej e^{-X}.
- Niech X będzie zmienną losową z rozkładu jednostajnego na $[0, 2\pi]$. Wyznacz rozkład zmiennej $\sin(X)$.

C11: Czasami trzeba się dopasować.

Najważniejsze wnioski z wykładu:

- korelacja i metody regresji to wspaniałe narzędzia, ale trzeba używać ich rozważnie.

Zadania do wykonania:

- Opierając się na twierdzeniu z wykładu (lub z podręcznika [2]) wyprowadź wzory na współczynniki a i b równania prostej w przypadku kiedy to prosta minimalizuje odpowiednią wartość oczekiwaną.
- Czy i dlaczego niezależność dwóch zmiennych losowych implikuje ich nieskorelowanie?
- Czy i dlaczego brak niezależności dwóch zmiennych losowych informuje o wartości współczynnika korelacji?
- Czy i dlaczego nieskorelowanie dwóch zmiennych losowych implikuje ich niezależność?
- Czy i dlaczego zerowanie współczynnika korelacji dla dwóch zmiennych losowych implikuje ich niezależność?
- Podaj przykład dwóch zmiennych losowych, które choć posiadają współczynnik korelacji 0 są ewidentnie skorelowane (niekoniecznie liniowo).
- Podaj (minimum trzy na zespół, bez powtarzania) przykłady rzeczywistych danych[8] dla których wysoka korelacja oznacza wynikania w jedną lub drugą stronę oraz brak wynikania

[8] Polecam narzędzia i strony prezentowane podczas wykładu.

C12: Tyle o sobie wiemy ile nas sprawdzono.

Najważniejsze wnioski z wykładu:

- poznaliśmy typowe rodzaje testów statystycznych.

Zadania do wykonania:

- Wiedząc, że dane umieszczone na stronie przedmiotu pochodzą z rozkładu o nieznanych wariancji i wartości oczekiwanej sprawdź, czy na poziomie istotności $\alpha = 0.05$ mamy podstawę do odrzucenia hipotezy o tym, że wartość oczekiwana wyjściowego rozkładu była równa zero na rzecz hipotezy przeciwnej.
- Wiedząc, że dane umieszczone na stronie przedmiotu pochodzą z rozkładu o odchyleniu standardowym równym $\sigma = 1$ sprawdź, czy na poziomie istotności $\alpha = 0.05$ mamy podstawę do odrzucenia hipotezy o tym, że wartość oczekiwana wyjściowego rozkładu była równa zero na rzecz hipotezy przeciwnej.
- Udowodnij, że dla rzutów uczciwą monetą długości nieprzerwanych pasm zer albo jedynek mają rozkład geometryczny.
- Przetestuj hipotezę, że długości ciagów zer i jedynek w wyrzuconej przez Ciebie próbce pochodzą z rozkładu geometrycznego. W oparciu o poprzednie zadanie wyciągnij z tego wnioski o uczciwości Twojej monety. Opisz dokładnie wszystkie kroki.
- Przetestuj hipotezę o równomiernym rozkładzie liczb w losowaniach lotto[9]. Opisz dokładnie wszystkie kroki.

[9] `www.lotto.pl/lotto/wyniki-i-wygrane/statystyki`

Notacje i oznaczenia

σ-ciało $\mathcal{F}$

- $\Omega \in \mathcal{F}$,
- Jeśli $A \in \mathcal{F}$ to $A' \in \mathcal{F}$,
- Jeśli $A_i \in \mathcal{F}$ dla $i = 1, 2, \ldots$ to $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Aksjomatyczna definicja prawdopodobieństwa Prawdopodobieństwo to funkcja $\mathbb{P}$ o wartościach w $\mathbb{R}$, określona na σ-ciele zdarzeń $\mathcal{F} \subset 2^{\Omega}$, która spełnia następujące warunki

- $\mathbb{P}(A) \geqslant 0$ dla każdego $A \in \mathcal{F}$,
- $\mathbb{P}(\Omega) = 1$,
- Jeśli $A_i \in \mathcal{F}$ $i = 1, 2, \ldots$ oraz $A_i \cap A_j = \emptyset$ dla $i \neq j$ to

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Przestrzeń probabilistyczna Przestrzenią probabilistyczną nazywamy trójkę $(\Omega, \mathcal{F}, \mathbb{P})$, gdzie Ω to zbiór zdarzeń elementarnych, $\mathcal{F} \subset 2^{\Omega}$ σ-ciało, a $\mathbb{P}$ p-stwo na $\mathcal{F}$.

Zbieżność według prawdopodobienstwa Ciąg zmiennych losowych (X_n) jest zbieżny według p-stwa do zmiennej losowej X gdy dla każdego $\varepsilon > 0$

$$\left(\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0\right) \equiv \left(X_n \xrightarrow{P} X\right).$$

Zbieżność z prawdopodobieństwem 1 (p.n.) Ciąg zmiennych losowych (X_n) jest zbieżny prawie na pewno (z p-stwem 1) do zmiennej losowej X gdy

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) \equiv \left(X_n \xrightarrow{p.n.} X\right).$$

Indykatorem zbioru $A \subseteq \mathbb{R}$ nazwiemy funkcję[10]

$$I_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

O dwóch zdarzeniach A i B powiemy, że są **niezależne**[11] gdy

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Funkcją masy prawdopodobieństwa[12] zmiennej[13] X nazwiemy

$$p_k = \mathbb{P}(X = x_k).$$

[10] W skrócie indykator odpowiada na pytanie czy dany x należy do zbioru A.

[11] Oczywiście zawsze dotyczy to pewnego szczególnego rozkładu prawdopodobieństwa, nie jest samodzielną cechą zdarzeń A i B.

[12] Wielkość ta ma sens jedynie dla rozkładów dyskretnych.

[13] Tam gdzie nie prowadzi to do nieporozumień pomijamy indeks X.

Gęstością prawdopodobieństwa nazwiemy funkcję $f_X = f$

$$\mathbb{P}(X \in A) = \int_A f(x)dx.$$

Dystrybuantą zmiennej losowej[14] X nazywamy funkcję

$$F(x) = F_X(x) = \mathbb{P}(X \leqslant x). \tag{38}$$

Wartość oczekiwaną[15] funkcji $G : \mathbb{R} \to \mathbb{R}$ dla zmiennej X zapiszemy[16]

$$\mathbb{E}[G(X)] = \int_{\mathbb{R}} G(x)f(x)dx,$$
$$\mathbb{E}[G(X)] = \sum_k G(x_k)p_k.$$

Kwantylem rzędu $p \in (0,1)$ zmiennej losowej nazywamy

$$Q(p) = \inf \{x \in \mathbb{R} : p \leq F(x)\}.$$

[15] O ile istnieje.

[16] W wariancie ciągłym i dyskretnym.

Rozkłady dyskretne

- Rozkład dwupunktowy[17]

$$\mathbb{P}(X = 0) = 1 - p, \ \mathbb{P}(X = 1) = p.$$

- Rozkład jednostajny (dyskretny)[18]

$$p_k = \frac{1}{N},$$

- Rozkład geometryczny[19]

$$p_k = (1-q)^k q.$$

- Rozkład Zipfa[20]

$$p_k = \frac{1}{H_{N,s}k^s}$$

- Rozkład Poissona[21]

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!},$$

- Rozkład Bernoulliego[22]

$$p_k = \binom{N}{k} p^k (1-p)^{N-k},$$

- Rozkład zeta[23]

$$p_k = \frac{1}{\zeta(s)k^s}$$

[17] gdzie $p \in [0,1]$.

[18] gdzie $N \in \mathbb{N}$, a $k = 1, 2, \ldots, N$.

[19] gdzie $q \in (0,1]$, a $k \in \mathbb{N}$.

[20] gdzie $s \in \mathbb{R}_+$, a $k = 1, 2, \ldots, N$, a stała normowania dana jest przez uogólnione liczby harmoniczne $H_{N,s} = \sum_{k=1}^{N} k^{-s}$.

[21] gdzie $\lambda \in \mathbb{R}_+$, $k \in \mathbb{N}$.

[22] gdzie $p \in (0,1)$, $N \in \mathbb{N}$, $k = 1, 2, \ldots, N$.

[23] gdzie $s \in \mathbb{R}_+$, $k \in \mathbb{N}_+$, a ζ to funkcja zeta Riemanna.

Rozkłady ciągłe

- Rozkład jednostajny (ciągły)[24]

$$f(x) = \frac{1}{b-a} I_{[a,b]}(x),$$

- Rozkład normalny (Gaussa)[25]

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

- Rozkład log-normalny [26]

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

- Rozkład wykładniczy[27]

$$f(x) = \lambda e^{-\lambda x} I_{[0,\infty)}(x),$$

- Rozkład Pareto[28]

$$f(x) = \frac{\alpha x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}} I_{[x_m,\infty)}(x).$$

- Rozkład Cauchy'ego[29]

$$f(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]},$$

- Rozkład gamma[30]

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} I_{(0,\infty)}(x),$$

- Rozkład χ[31]

$$f(x) = \frac{1}{2^{(k/2)-1}\Gamma(k/2)}\, x^{k-1} e^{-x^2/2} I_{[0,\infty)}(x),$$

Dystrybuanta

Najważniejsze własności dystrybuanty:

- jest funkcją niemalejącą i prawostronnie ciągłą.
- $\lim_{x\to-\infty} F(x) = 0$ oraz $\lim_{x\to\infty} F(x) = 1$.

Wzór (38) dla przypadku ciągłego przyjmuje postać[32]

$$F(x) = \int_{-\infty}^{x} f(t)\,dt,$$

natomiast dla przypadku dyskretnego

$$F(x) = \sum_{x_i \le x} \mathbb{P}(X = x_i).$$

[25] gdzie $\mu \in \mathbb{R}$, a $\sigma \in \mathbb{R}_+$.

[26] gdzie $\mu \in \mathbb{R}$, a $\sigma \in \mathbb{R}_+$.

[27] gdzie $\lambda \in \mathbb{R}$.

[28] gdzie $\alpha, x_m \in \mathbb{R}_+$.

[29] gdzie $x_0 \in \mathbb{R}$, $\gamma \in \mathbb{R}_+$.

[30] gdzie $\alpha, \beta \in \mathbb{R}_+$, a Γ to funkcja gamma Eulera.

[31] gdzie $k \in \mathbb{R}_+$, a Γ to funkcja gamma Eulera.

[32] Dla przejrzystości pomijamy indeksy opisujące zmienną losową i przyjmujemy, że f to funkcja gęstości prawdopodobieństwa.

Momenty

Dla zmiennej o rozkładzie dyskretnym, mówimy, że posiada n-ty moment gdy $\sum_k |x_k|^n p_k < \infty$ i oznaczamy go przez

$$\mathbb{E}X^n = \sum_k x_k^n p_k.$$

Dla zmiennej o rozkładzie ciągłym, mówimy, że posiada n-ty moment gdy $\int |x|^n f(x)dx < \infty$ i oznaczamy go przez

$$\mathbb{E}X^n = \int_{\mathbb{R}} x^n f(x)dx.$$

Twierdzenia graniczne

Twierdzenie 1 *Poissona*[33]

Jeśli $n \to \infty$, $p_n \to 0$ oraz $np_n \to \lambda > 0$ to

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} \xrightarrow{n\to\infty} \frac{\lambda^k}{k!}e^{-\lambda}.$$

[33] Por. rozdział 7.4 w [1].

Twierdzenie 2 [34]

Niech zmienna losowa S_n ma rozkład Bernoulliego z parametrami n i p oraz $\lambda = np$. Wówczas dla każdego $B \subseteq \mathbb{N}$ mamy

$$\left| \mathbb{P}(S_m \in B) - \sum_{k\in B} \frac{\lambda^k}{k!}e^{-\lambda} \right| \leqslant \frac{\lambda^2}{n}.$$

[34] Por. rozdział 7.4 w [1].

Twierdzenie 3 *SPWL Bernoulliego*[35]

Jeżeli S_n jest liczbą sukcesów w schemacie Bernoulliego z n prób z prawdopodobieństwem sukcesu w pojedynczej próbie równym p to dla każdego $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n}{n} - p \right| \leqslant \varepsilon \right) = 1$$

[35] Por. rozdział 7.2 w [1].

Twierdzenie 4 *SPWL Markowa*[36]

Niech $(X_n)_n$ będzie ciągiem zmiennych losowych takim, że

- $\lim_{n\to\infty} \frac{\mathcal{D}^2 S_n}{n^2} = 0$
  LUB
- *X_n są parami nieskorelowane i mają wspólnie ograniczone wariancje.*

wówczas dla każdego $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n - \mathbb{E}S_n}{n} \right| > \varepsilon \right) = 0.$$

[36] Por. rozdział 7.2 w [1].

Twierdzenie 5 *MPWL Bernoulliego*[37]

Niech S_n oznacza tym razem liczbe sukcesów w n próbach Bernoulliego z p-stwem sukceu p. Wtedy dla każdego $\varepsilon > 0$ mamy

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geqslant n} \left| \frac{S_k}{k} - p \right| \leqslant \varepsilon \right) = 1.$$

[37] Por. rozdział 7.2 w [1].

Twierdzenie 6 *MPWL Kołmogorowa*[38]

Jeżeli $(X_n)_{n\in\mathbb{N}}$ *jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie* $\mathbb{E}|X_1| < \infty$ *to* $\lim_{n\to\infty} \frac{S_n}{n} = \mathbb{E}X_1$.

[38] Por. rozdział 7.2 w [1].

Centralne Twierdzenie Graniczne

Twierdzenie 7 *CTG*[39]

Niech (X_n) *będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie oraz niech* $\mathbb{E}X_1 = \mu$ *i* $\mathcal{D}^2 X_1 = \sigma^2 > 0$ *Wówczas dla dowolnego* t

$$\mathbb{P}\left(\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leqslant t\right) \xrightarrow{n\to\infty} \Phi(t).$$

[39] Por. rozdział 7.5 w [1].

Literatura

[1] Jacek Jakubowski and Rafał Sztencel. *Rachunek prawdopodobieństwa dla (prawie) każdego*. SCRIPT, 2006.

[2] Agnieszka Plucińska and Edmund Pluciński. *Probabilistyka*. WNT, 2000.