Metody statystyczne I
Organizacja zajęć
Wybrane rozkłady zmiennych losowych

Anna Sznajderska

Szkoła Główna Handlowa w Warszawie

2022/2023

Literatura

Podręczniki podane w Sylabusie:

- J.Jóźwiak, J.Podgórski, Statystyka od podstaw, PWE, Warszawa 2006.
- J. Krysicki, D. Bartos, W. Dyczka, K. Królikowska, M. Wasilewski, Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. I, II, PWN, Warszawa, 2007.
- G.S. Maddala, Ekonometria, PWN, 2006.

Podręcznik polecany dla osób zainteresowanych:

- N. Mukhopadhyay, Probability and Statistical Inference, Routledge, 2020.
- R. Zieliński, Siedem wykładów wprowadzających do statystyki matematycznej, PWN, 1990. **https:www.impan.pl∼rziel7ALL.pdf**
- W. Niemiro, Wykłady na stronie: **https:dydmat.mimuw.edu.plstatystyka-i**

Zaliczenie

- Egzamin - 80%
- Aktywność na zajęciach - 20%
  - prezentacja zadania z listy
  - możliwość zamiany zadań między studentami
- Dodatkowe punkty za aktywność podczas wykładu i ćwiczeń.
- KONSULTACJE: poniedziałek lub piątek, serdecznie zapraszam!
- Proszę o kontakt (Teams lub mail) przed przyjściem na konsultacje.

Metody statystyczne I

- „Matematyka jest królową wszystkich nauk, jej ulubieńcem jest prawda, a prostość i oczywistość jej strojem" Jan Śniadecki,
- Lepsze zrozumienie zagadnień poznanych na ekonometrii,
- Użyteczność programów statystycznych, możliwość praktycznego wykorzystania zdobytej wiedzy,
- Czas studiów jako czas rozwoju osobistego, wewnętrzna motywacja,
- Odkrycie satysfakcji jaką może dać studiowanie matematyki, odkrycie piękna matematyki.

Plan zajęć

Wykłady

1. Wybrane rozkłady zmiennych losowych. Funkcje zmiennych losowych.
2. Wielowymiarowe rozkłady zmiennych losowych: funkcje prawdopodobieństwa, rozkłady brzegowe i warunkowe, momenty wielowymiarowych zmiennych losowych.
3. Metody uzyskiwania estymatorów. Własności estymatorów: nieobciążoność, zgodność, efektywność.
4. Testy hipotez statystycznych: błędy pierwszego i drugiego rodzaju, moc testu, lemat Neymana-Pearsona, test ilorazu wiarygodności.
5. ANOVA, model regresji liniowej, model regresji logistycznej.

Oprogramowanie

- strona główna programu R: https://www.r-project.org/
- podręczniki: https://cran.r-project.org/manuals.html
- Kleiber Ch. i Zeileis A., Appied Econometrics with R
- Przemysław Biecek http://www.biecek.pl/R/#Przewodnik
- Michał Rubaszek https://web.sgh.waw.pl/ mrubas/

Zmienna dyskretna

Zmienna ta przyjmuje skończoną lub przeliczalną liczbę możliwych wartości z określonym prawdopodobieństwem przypisanym do każdej z tych wartości

- Przykład 1: rzucamy dwoma kostkami (żółtą i czerwoną) interesuje nas suma oczek:

x	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

  X jest zmienną losową, czyli funkcją $S \rightarrow \mathcal{X}$, gdzie S to zbiór zdarzeń elementarnych $S = \{11, \dots, 16, 21, \dots, 61, \dots 66\}$, $\mathcal{X}$ to podzbiór zbioru liczb rzeczywistych $\mathcal{X} = \{2, 3, 4, \dots, 12\}$
- Przykład 2: rzucamy monetą, interesuje nas liczba rzutów monetą potrzebnych do otrzymania pierwszej reszki (Y):

$$P(Y = y) = \left(\frac{1}{2}\right)^{y}, y = 1, 2, \dots$$

  zbiór możliwych wartości jakie przyjmuje Y jest przeliczalny. Zbiór równoliczny ze zbiorem liczb naturalnych nazywamy zbiorem **przeliczalnym**.

Zmienne dyskretne

Rozkład zmiennej dyskretnej:

$$f(x) = P(X = x) \text{ dla } x \in \mathcal{X} = \{x_1, x_2, \dots\}.$$

- $f(x) \geqslant 0$ dla każdego $x \in \mathcal{X}$,
- $\sum_{i=1}^{\infty} P(X = x_i) = 1$.

Dystrybuanta rozkładu prawdopodobieństwa (*distribution function or cumulative distribution function - cdf*):

$$F(x) = P(X \leqslant x) = P\{s : s \in S \text{ taki, że } X(s) \leqslant x\} x \in R.$$

Zmienna ciągła

Przykład: X czas oczekiwania na windę w budynku
$P(X = x) =?$, jakie jest prawdopodobieństwo, że będziemy czekać dokładnie 5 min?

Zmienne ciągłe

Funkcja gęstości rozkładu prawdopodobieństwa (*probability mass function - pmf*)

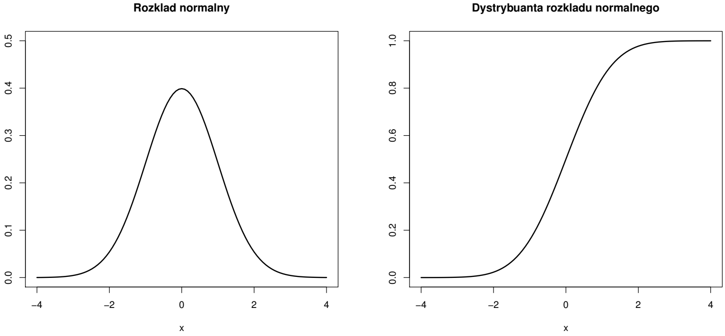
$$f(x) \geqslant 0 \text{ dla każdego } x \in R$$

$$\int_R f(x)dx = 1.$$

Dystrybuanta rozkładu prawdopodobieństwa (*distribution function or cumulative distribution function - cdf*)

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(y)dy, x \in R.$$

Gęstość i dystrybuanta rozkładu



Przykład - mediana

Mediana rozkładu (x_m) to taka wartość, że 50% możliwych wartości zmiennej X jest poniżej x_m oraz 50% możliwych wartości zmiennej X jest powyżej x_m. Inaczej $F(x_m) = \frac{1}{2}$, czyli $P(X \leqslant x_m) = P(X \geqslant x_m) = \frac{1}{2}$. Rozważmy zmienną X o funkcji gęstości:

$$f(x) = \begin{cases} \frac{3}{7} w^2 \text{ jeśli } 1 < w < 2 \\ 0 \text{ jeśli } w \leqslant 1 \text{ i } w \geqslant 2 \end{cases}$$

Wtedy:

$$F(x) = \begin{cases} 0 \text{ jeśli } w \leqslant 1 \\ \frac{1}{7}(w^3 - 1) \text{ jeśli } 1 < w < 2 \\ 1 \text{ jeśli } w \geqslant 2 \end{cases}$$

Sprawdź czy pole pod wykresem funkcji gęstości jest równe 1?
Obicz medianę w_m dla tego rozkładu.

Przykład - mediana

Mediana rozkładu (x_m) to taka wartość, że 50% możliwych wartości zmiennej X jest poniżej x_m oraz 50% możliwych wartości zmiennej X jest powyżej x_m. Inaczej $F(x_m) = \frac{1}{2}$, czyli $P(X \leqslant x_m) = P(X \geqslant x_m) = \frac{1}{2}$.
Rozważmy zmienną X o funkcji gęstości:

$$f(x) = \begin{cases} \frac{3}{7}w^2 \text{ jeśli } 1 < w < 2 \\ 0 \text{ jeśli } w \leqslant 1 \text{ i } w \geqslant 2 \end{cases}$$

Wtedy:

$$F(x) = \begin{cases} 0 \text{ jeśli } w \leqslant 1 \\ \frac{1}{7}(w^3 - 1) \text{ jeśli } 1 < w < 2 \\ 1 \text{ jeśli } w \geqslant 2 \end{cases}$$

Sprawdź czy pole pod wykresem funkcji gęstości jest równe 1?
Obicz medianę w_m dla tego rozkładu.

$$F(w_m) = \frac{1}{2}, w_m = (4.5)^{\frac{1}{3}}$$

Kwantyl i wartość krytyczna

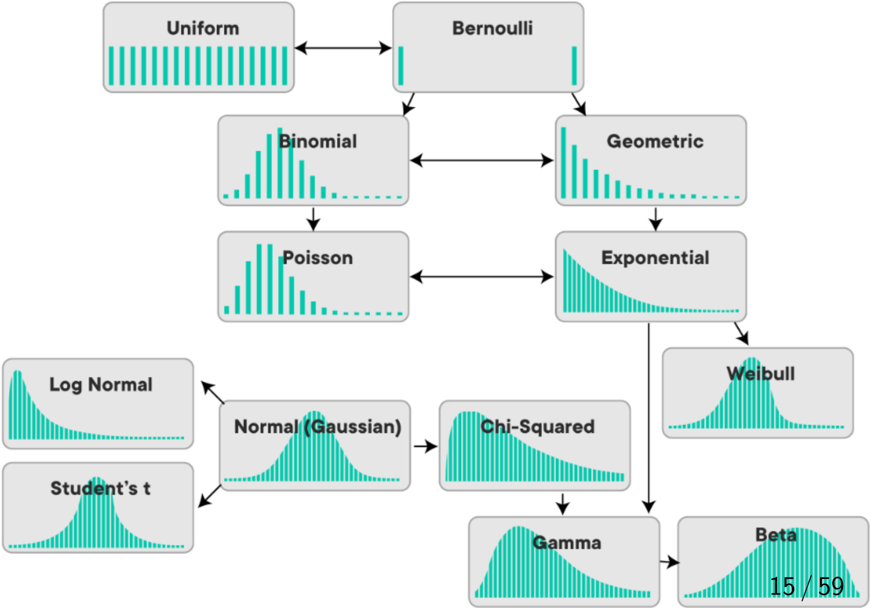
Jesli istnieje dokładnie jedna liczba ξ_q taka, że $P(X \leqslant \xi_q) = F(\xi_q) = q$ to ξ_q jest q-tym kwantylem.

Kwantyle - wskaż kwantyl rzędu 1/4, 1/5, 1/2, 3/4

1.1.3 Przykład (Waga noworodków). Powiedzmy, że wylosowano 114 noworodków[1] w celu poznania cech fizycznych dzieci urodzonych w Warszawie w roku 2009. Waga noworodków była taka:

3080	3650	3250	4000	3180	3480	4140	3930	3950	2700
3720	3520	3200	3700	3500	3790	3900	3760	3740	3200
3280	3960	3300	2490	3260	3780	3600	3060	2850	3490
2620	3690	3200	3070	4640	3760	3190	3180	3760	3670
3310	3770	2580	2700	3740	2700	3760	3960	2800	3500
3460	3800	3394	3640	2680	3490	3000	2900	4320	3450
3200	3530	3330	2680	2700	3580	2500	2660	3600	3114
3760	3640	2780	2760	3480	2420	2110	2930	3160	3012
2900	3750	4010	3230	2570	3480	3340	3420	3330	2030
3730	3640	3420	4330	3790	3120	3890	3070	3270	2750
2470	3620	2740	3800	3440	3160	3620	3190	2380	3100
2400	2500	2540	3270						

Przegląd rozkładów prawdopodobieństwa



Wybrane rozkłady zmiennych losowych

Zmienne dyskretne:

- rozkład jednopunktowy
- rozkład dwupunktowy (ang. Bernoulli distribution)
- rozkład Bernoulliego (inaczej zwany rozkładem dwumianowym)
- rozkład Poissona

Zmienne ciągłe:

- rozkład jednostajny
- rozkład Gamma, a w szczególności rozkład wykładniczy i rozkład χ^2,
- rozkład normalny, jak również rozkład t Studenta, rozkład χ^2, rozkład lognormalny,
- rozkład Beta,
- rozkład Cauchego.

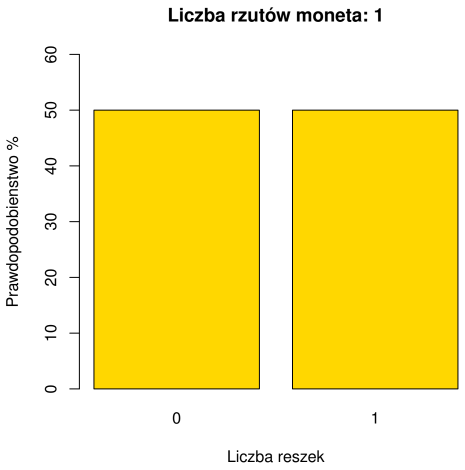
Rozkład Bernoulliego

Rozkład łącznej liczby sukcesów w n doświadczeniach Bernoulliego, gdy szansa sukcesu w pojedynczym doświadczeniu wynosi p.
Rozkład sumy $X_1 + X_2 + \ldots + X_n$ gdzie zmienne losowe X_i są niezależne i mają ten sam rozkład dwupunktowy
$P(X_i = 1) = p, P(X_i = 0) = 1 - p, i = 1, 2, \ldots, n.$

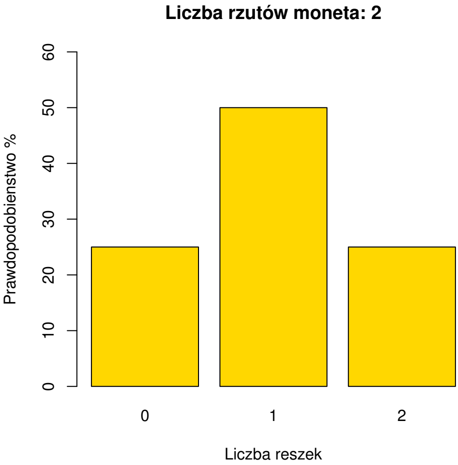
$$P(X = k) = B(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \ldots, n, p \in [0, 1],$$

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

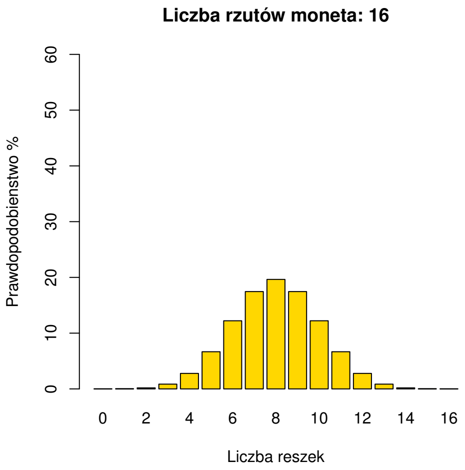
Rozkład Bernoulliego



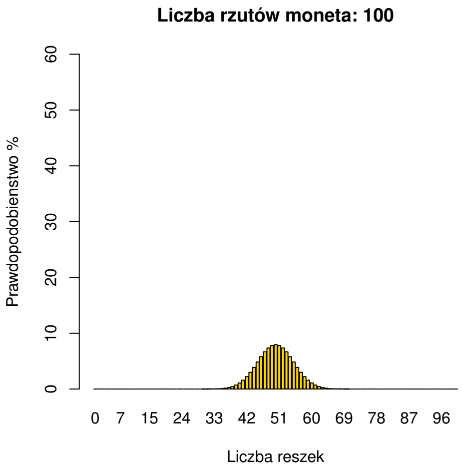
Rozkład Bernoulliego



Rozkład Bernoulliego



Rozkład Bernoulliego



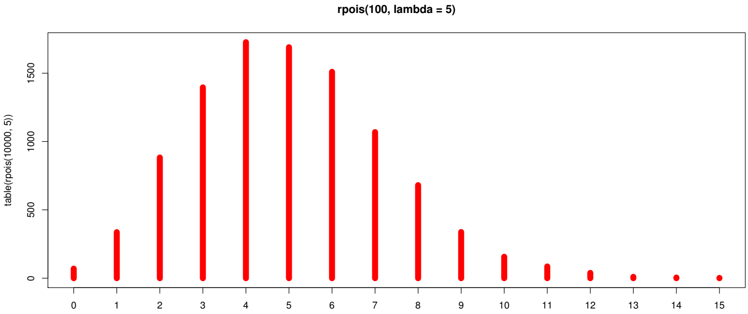
Jakie jest prawdopodobieństwo wyrzucenia 30 reszek w 100 rzutach symetryczną monetą?

Rozkład Poissona

Rozkład graniczny dla rozkładu Bernoulliego $B(n, p_n)$, gdy $n \to \infty$, $p_n \to 0$, $np_n \to \lambda$.
Rozkład Poissona pojawia się jako rozkład zdarzeń rzadkich - pożar, powódź, wygrana w Toto-Lotka.

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0, \dots, n,\ \lambda > 0.$$

Rozkład Poissona



Rozkład normalny

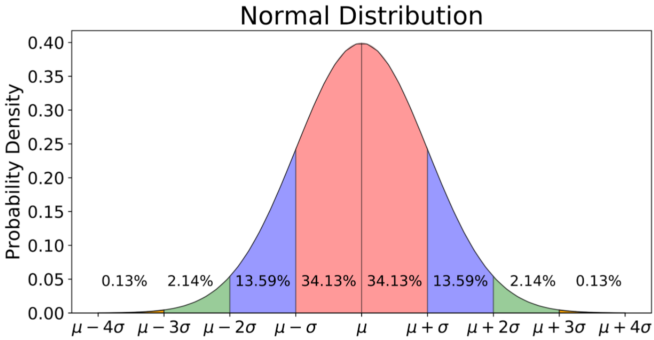
Jednowymiarowy standardowy rozkład normalny $N(0,1)$ ma gęstość:

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

Jeśli zmienna losowa X ma rozkład normalny $N(0,1)$ to zmienna losowa $\sigma X + m$ ma rozkład $N(m,\sigma^2)$ o gęstości:

$$\varphi_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Rozkład normalny



Rozkład t-Studenta

Jeśli zmienne X_i są niezależne i mają rozkład normalny $N(\mu, \sigma^2)$ oraz

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

to t ma rozkład t-Studenta z $n-1$ stopniami swobody

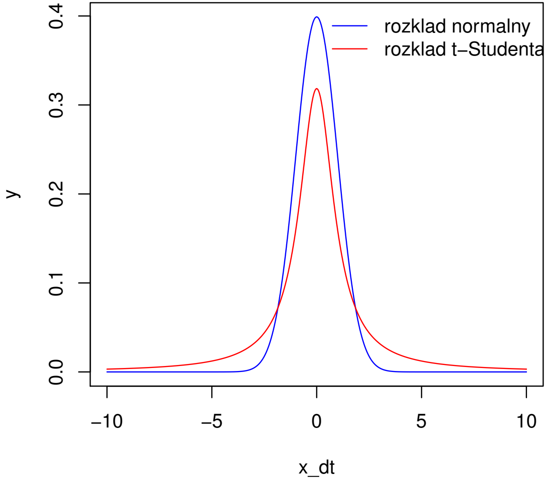
Rozkład t-Studenta

Funkcja gęstości zmiennej X o rozkładzie t-Studenta o v stopniach swobody:

$$f(x) = a(1 + \frac{1}{v}x^2)^{-\frac{1}{2}(v+1)} \text{ dla } -\infty < x < \infty,$$

$$a = a(v) = \{\sqrt{v\pi}\}^{-1}\Gamma(\frac{1}{2}(v+1))\{\Gamma(\frac{1}{2}v)\}^{-1}, v = 1, 2, 3, \ldots.$$

Rozkład t-Studenta



Rozkład χ^2

Jeśli zmienne X_i są niezależne i mają standardowy rozkład normalny $N(0,1)$ to zmienna

$$\chi^2 = \sum_{i=1}^{n} X_i^2$$

ma rozkład χ^2 o n stopniach swobody.

- Uwaga: Rozkład χ^2 to szczególny przypadek rozkładu Gamma. Zmienna o rozkładzie Gamma$(\frac{1}{2}v, 2)$ ma rozkład χ^2 z v stopniami swobody.

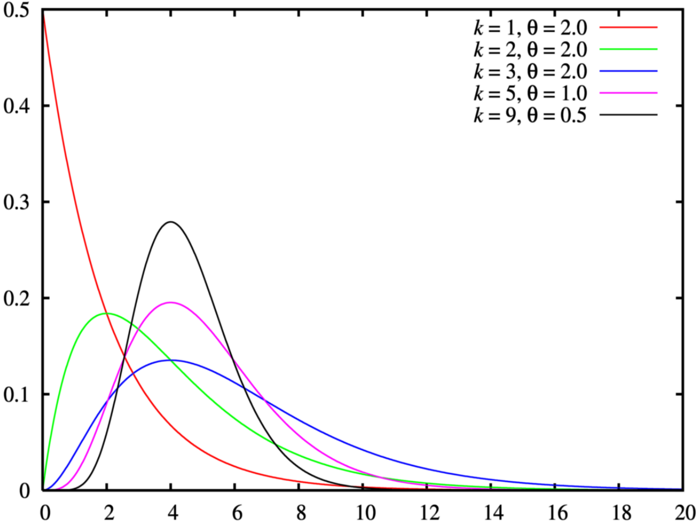
Rozkład Gamma $\Gamma(\alpha, \beta)$

$$\gamma_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \mathbb{1}_{(0,\infty)}(x),\ \alpha, \beta > 0$$

Pytanie: czy rozkład Gamma jest prawo- czy lewo- skośny?

Uwaga: często stosowanym rozkładem jest odwrotny rozkład Gamma (ang. inverse Gamma), oznaczany jako IG. Jeśli zmienna X ma rozkład $\Gamma(\alpha, \beta)$ to zmienna $1/X$ ma rozkład $IG(\alpha, 1/\beta)$.

Rozkład Gamma



Funkcja Gamma

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx \quad \text{dla } \alpha > 0.$$

Funkcja Gamma ma następujące własności:

- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$,
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$,
- $\Gamma(n) = (n-1)!$ dla $n = 1, 2, \ldots$,
- przybliżenie Stirlinga dla dużych wartości α, $\Gamma(\alpha) \sim \sqrt{2\pi} e^{-\alpha} \alpha^{\alpha - \frac{1}{2}}$.

Pytanie: ile w przybliżeniu jest równe $n!$?

Rozkład Fishera-Snedecora - rozkład F

Rozkład zmiennej losowej F o d_1 i d_2 stopniach swobody, jeśli $X_1 \sim \chi^2(d_1)$ oraz $X_2 \sim \chi^2(d_2)$ oraz X_1 i X_2 są niezależne

$$F = \frac{\frac{X_1}{d_1}}{\frac{X_2}{d_2}}.$$

Rozkład wykładniczy

Ciągły odpowiednik rozkładu geometrycznego, ma własność braku pamięci.

$$g(x) = \frac{1}{\lambda} e^{-\frac{1}{\lambda} x} \mathbb{1}_{(0,\infty)}(x), \lambda \in (0,\infty).$$

- rozkład wykładniczy dobrze modeluje czas oczekiwania na zdarzenie (np. nadejście nowego klienta)
- Uwaga: Rozkład wykładniczy to szczególny przypadek rozkładu Gamma. Zmienna o rozkładzie Gamma$(1,\lambda)$ ma rozkład wykładniczy o parametrze λ.
- Narysuj funkcję gęstości rozkładu wykładniczego w R dla $\lambda = 1$ oraz $\lambda = \frac{1}{2}$.
  - Co dzieje się z wykresem funkcji gęstości, gdy zmniejszymy λ?
  - Czy rozkład wykładniczy jest lewoskośny czy prawoskośny?
  - Jeśli zmienna X reprezentuje oczekiwaną długość życia to zmniejszenie parametru λ powoduje, że prawdopodobieństwo śmierci we wczesnych latach życia rośnie czy maleje?

Rozkład Beta

$$\beta_{a,b}(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x),$$

$$a, b > 0,$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Rozkład Cauchego

$$f(x) = \frac{h}{\pi(h^2 + (x-m)^2)}, h > 0.$$

Najczęściej $m = 0$ i $h = 1$. Uwaga

- h to parametr lokalizacji, a m nazywamy parametrem skali.
- Szczególny przypadek rozkładu t-Studenta dla $v = 1$, patrz slajd 25.
- Momenty zwykłe i centralne niezdefiniowane, np. nie istnieje wartość oczekiwana, zobacz Link.
- Suma niezależnych zmiennych losowych o rozkładzie Cauchego ma rozkład Cauchego.
- Jeśli $X_1, X_2, \dots X_n$ to niezależne zmienne losowe o rozkładzie Cauchego to nX_1 ma ten sam rozkład co $X_1 + X_2 + \dots + X_n$.

Twierdzenie o sumowaniu zmiennych

Niech $X_1, \ldots, X_n$ będą zmiennymi iid, oznaczmy $U = \sum_{i=1}^{n} X_i$ oraz $\overline{X} = U/n$, wtedy:

- Jeśli zmienne X_i mają rozkład normalny $N(\mu, \sigma^2)$ to zmienna U ma rozkład normalny $N(n\mu, n\sigma^2)$ a zmienna $\overline{X}$ ma rozkład $N(\mu, \frac{1}{n}\sigma^2)$.
- Jeśli zmienne X_i mają rozkład Gamma $\Gamma(\alpha, \beta)$ to zmienna U ma rozkład Gamma $\Gamma(n\alpha, \beta)$.
- Jeśli zmienne X_i mają rozkład χ^2 o υ_i stopniach swobody to zmienna U ma rozkład χ^2 o $\upsilon = \sum_{i=1}^{n} \upsilon_i$ stopniach swobody.

Twierdzenie o średniej i wariancji z próbki

Niech $X_1, \ldots, X_n$ będą zmiennymi iid o rozkładzie $N(\mu, \sigma^2)$, oznaczmy średnią z próby $\overline{X} = 1/n \sum_{i=1}^{n} X_i$ i wariancję z próby $S^2 = 1/(n-1) \sum_{i=1}^{n} (X_i - \overline{X})^2$.

- $\overline{X}$ i S^2 są niezależne,
- rozkład $\overline{X}$ to $N(\mu, \frac{1}{n}\sigma^2)$,
- rozkład $(n-1)S^2/\sigma^2$ to χ^2 z $(n-1)$ stopniami swobody.

Statystyka pozycyjna

Rozwazmy próbkę $X_1, \ldots, X_n$. Dla każdego $\omega \in \Omega$ niech $X_{1:n}(\omega) \leqslant X_{2:n}(\omega) \leqslant \ldots \leqslant X_{n:n}(\omega)$ będzie ciągiem liczb $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$ uporządkowanym w kolejności rosnącej. Określone w ten sposób zmienne losowe $X_{1:n}(\omega), X_{2:n}(\omega), \ldots, X_{n:n}(\omega)$ nazywamy statystykami pozycyjnymi.

Pierwsza i ostatnia statystyka pozycyjna to, odpowiednio, najmniejsza i największa obserwacja w próbce.

Obliczmy funkcję gęstości dla $X_{n:n}$ z próbki o rozkładzie jednostajnym.

Definicja

Moment zwykły rzędu r zmiennej losowej X definiujemy jako:

$$\eta_r = E[X^r] \text{ dla ustalonego } r = 1, \ldots.$$

Pierwszy moment η_1 to ...
Moment centralny rzędu r zmiennej losowej X o średniej μ definiujemy jako:

$$\mu_r = E[(X - \mu)^r] \text{ dla ustalonego } r = 1, \ldots.$$

Pierwszy moment centralny μ_1 to ...
Drugi moment centralny μ_2 to ...

Wartość oczekiwana

$$E(X) = \begin{cases} \sum_{i:x_i \in \mathcal{X}} x_i f(x_i) & \text{dla zmiennej dyskretnej} \\ \int_{\mathcal{X}} x f(x) dx & \text{dla zmiennej ciągłej} \end{cases}$$

- Uwaga 1.

$$E(g(X)) = \begin{cases} \sum_{i:x_i \in \mathcal{X}} g(x_i) f(x_i) & \text{dla zmiennej dyskretnej,} \\ \int_{\mathcal{X}} g(x) f(x) dx & \text{dla zmiennej ciągłej,} \end{cases}$$

- Uwaga 2.

$$E\{a_0 + \sum_{i=1}^{k} a_i g_i(X)\} = a_0 + \sum_{i=1}^{k} a_i E\{g_i(X)\}.$$

Wartość oczekiwana

Policz wartość oczekiwaną następujących zmiennych X, Y i Z:

wartości X	-1	1	3	5	7
$P(X = x)$	0.1	0.2	0.4	0.2	0.1

wartości Y	-1	1	3	5	7
$P(Y = y)$	0.2	0.15	0.3	0.15	0.2

- Funkcja gęstości zmiennej Z:

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{dla } 0 < x < 2, \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

Wartość oczekiwana dla zmiennej o rozkładzie normalnym $N(\mu, \sigma^2)$

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

podstawmy $y = \dfrac{x-\mu}{\sigma}, \dfrac{dy}{dx} = \dfrac{1}{\sigma}, dx = \sigma dy$

$$x = \sigma y + \mu$$

$$E(X) = \frac{1}{\cancel{\sigma}\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-\frac{y^2}{2}} \cancel{\sigma} dy$$

$$E(X) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{y^2}{2}} dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy =$$

$$= 0 + \frac{\mu}{\sqrt{2\pi}} \sqrt{2\pi} = \mu$$

Wariancja zmiennej losowej (ozn. σ^2)

$$D^2(X) = E[X - E(X)]^2 =$$
$$= \begin{cases} \sum_i [x_i - E(X)]^2 p_i & \text{dla zmiennej dyskretnej} \\ \int_{-\infty}^{\infty} [x_i - E(X)]^2 f(x) dx & \text{dla zmiennej ciągłej} \end{cases}$$

Uwaga. (Dowód - ćwiczenie)

$$D^2(X) = E(X^2) - E^2(X)$$

Inne charakterystyki liczbowe zmiennej losowej

- współczynnik skośności (asymetrii)

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$$

  - Jeśli $\gamma_1 > 0$ to mówimy o asymetrii dodatniej, prawostronnej, wtedy prawy ogon rozkładu jest dłuższy.
  - Jeśli $\gamma_1 < 0$ to mówimy o asymetrii ujemnej, lewostronnej, wtedy lewy ogon rozkładu jest dłuższy.
- współczynnik skupienia (kurtoza)

$$K = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^{4/2}}$$

  - Jeśli $K > 0$ to kurtoza jest dodatnia, intensywność wartości skrajnych jest większa niż dla rozkładu normalnego („ogony" rozkładu są „grubsze"). Rozkład leptokurtyczny.
  - Jeśli $K < 0$ to kurtoza jest ujemna, intensywność wartości skrajnych jest mniejsza niż w przypadku rozkładu normalnego („ogony" rozkładu są „węższe"). Rozkład platykurtyczny.

Wartość oczekiwana

Można pokazać, że:

1. dla nieujemnej ciągłej zmiennej losowej:

$$E(X) = \int_0^\infty \{1 - F(x)\}dx \text{ przy założeniu, że } \lim_{x \to \infty} x\{1 - F(x)\} = 0.$$

2. dla nieujemnej dyskretnej zmiennej losowej:

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

Rozkład jednostajny

Rozkład o gęstości:

$$f(x) = \frac{1}{b-a} \text{ dla } -\infty < a < x < b < \infty$$

nazywamy rozkładem jednostajnym na przedziale (a, b).
Narysuj funkcję gęstości rozkładu jednostajnego w programie R.

Zadanie. Załóżmy, że czas czekania na autobus mierzony w minutach jest zmienną o rozkładzie jednostajnym na przedziale (0,5). Jakie jest prawdopodobieństwo, ze będziemy czekać na autobus więcej niż 3,8 min?

Rozkład jednostajny - wartość oczekiwana i wariancja

$$E(X) = \int_a^b xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{x^2}{2(b-a)}|_a^b = \frac{(a+b)}{2}.$$

$$E(X^2) = \int_a^b x^2 f(x)dx = \frac{x^3}{3(b-a)}|_a^b = \frac{(b^2+ab+a^2)}{3}.$$

$$Var(X) = \frac{(b^2+ab+a^2)}{3} - (\frac{(a+b)}{2})^2 = \frac{1}{12}(b-a)^2.$$

Funkcja generująca momenty

Funkcja generująca momenty zmiennej X to:

$$M_X(t) = E[e^{tX}],$$

o ile wartość oczekiwana jest skończona dla $|t| < a$ dla ustalonego $a > 0$. Oznacza to, że:

$$M_X(t) = \sum_{i:x_i \in \mathcal{X}} e^{tx_i} f(x_i) \text{ dla zmiennej } X \text{ dyskretnej,}$$

$$M_X(t) = \int_{\mathcal{X}} e^{tx} f(x) dx \text{ dla zmiennej } X \text{ ciągłej.}$$

Twierdzenie. Jeśli zmienna losowa X ma skończoną funkcję generującą momenty to

$$\eta_r = \frac{d^r M_X(t)}{dt^r} \text{ w punkcie } t = 0.$$

Funkcja generująca momenty

Twierdzenie. Zmienna losowa X ma funkcję generującą momenty $M_X(t)$ dla $|t| < a$ dla pewnego $a > 0$. Niech $Y = cX + d$ będzie zmienną losową dla pewnych $c, d \in \mathcal{R}$. Wtedy:

$$M_Y(t) = e^{td} M_X(tc).$$

Porównaj z zadaniem 18. Wyprowadź wzór na funkcję generującą momenty dla rozkładu normalnego.

Przykład - rozkład Poissona

Funkcja generująca momenty dla zmiennej X o rozkładzie Poissona:

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} (\lambda e^t)^x \frac{1}{x!} = \dots$$

Przykład - rozkład Poissona

Funkcja generująca momenty dla zmiennej X o rozkładzie Poissona:

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} (\lambda e^t)^x \frac{1}{x!} = e^{\{-\lambda + \lambda e^t\}},$$

$$\frac{dM_X(t)}{dt} = \lambda e^t M_X(t),$$

$$\frac{d^2 M_X(t)}{dt^2} = \lambda e^t M_X(t) + (\lambda e^t)^2 M_X(t).$$

Zauważmy, że $M_X(0) = \sum_{i: x_i \in \chi} e^0 f(x_i) = 1$.
Zatem:

$$\eta_1 = \lambda,$$

$$\eta_2 = \lambda + \lambda^2, \text{ stąd}$$

$$Var(X) = E(X^2) - (E(X))^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

Przykład - rozkład Gamma

Można pokazać, że funkcja generująca momenty dla zmiennej X o rozkładzie Gamma ma następującą postać:

$$M_X(t) = \frac{1}{(1 - \beta t)^\alpha}.$$

- jaka jest wartość oczekiwana i wariancja zmiennej X?
- jaka jest wartość oczekiwana i wariancja zmiennej o rozkładzie wykładniczym?
- jaka jest wartość oczekiwana i wariancja zmiennej o rozkładzie χ^2?

Wskazówka. Potraktuj zmienną o rozkładzie wykładniczym i rozkładzie χ^2 jako szczególny przypadek zmiennej o rozkładzie Gamma.

Twierdzenie

Rozważmy zmienną losową X_i, która posiada funkcję generującą momenty $M_{X_i}(t)$ dla $i = 1, \ldots, n$. Załóżmy, że $X_1, \ldots, X_n$ są niezależne. Wtedy funkcja generująca momenty dla zmiennej $U = \sum_{i=1}^{n} a_i X_i$ jest dana wzorem $\prod_{i=1}^{n} M_{X_i}(ta_i)$, gdzie $a_1, \ldots, a_n \in \mathbb{R}$.

Uwaga. Skończona funkcja generująca momenty jednoznacznie determinuje rozkład zmiennej.

Dowód stwierdzenia ze slajdu 37

Niech zmienne $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o rozkładzie $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$. Zapiszmy $U = \sum_{i=1}^n X_i$, $\mu = \sum_{i=1}^n \mu_i$, $\sigma^2 = \sum_{i=1}^n \sigma_i^2$. Wtedy:

$$M_U(t) = \prod_{i=1}^n \{e^{t\mu_i + \frac{1}{2}t^2\sigma_i^2}\} = e^{t\sum_{i=1}^n \mu_i + \frac{1}{2}t^2 \sum_{i=1}^n \sigma_i^2} = e^{t\mu + \frac{1}{2}t^2\sigma^2},$$

zatem zmienna U ma rozkład normalny o parametrach μ i σ^2. Patrz Uwaga na poprzednim slajdzie.

Przypomnienie - Centralne twierdzenie graniczne

Niech $X_1, \ldots, X_n$ będą iid o średniej μ i wariancji σ^2, $-\infty < \mu < \infty$ i wariancji $0 < \sigma^2 < \infty$. Wtedy gdy $n \to \infty$:

$$Y_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{dyst} N(0,1).$$

dyst oznacza zbieżność do dystrybuanty.
Uwaga. zmienne losowe mogą mieć zarówno rozkład dyskretny jak i typu ciągłego.

$$P(y_1 < Y_n \leqslant y_2) = \phi(y_2) - \phi(y_1)$$

Przypomnienie - Transformacja zmiennej losowej

Jeśli X ma funkcję gęstości $f(x)$ oraz $Y = g(X)$, gdzie g jest funkcją różniczkowaną, wtedy $h(y)$ funkcja gęstości zmiennej Y dana jest wzorem:

$$h(y) = f(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|.$$

Praca domowa - Rachunek prawdopodobieństwa

Zadanie 1.

- Znajdź rozkład zmiennej $Z = \frac{X^b}{a}$ jeśli zmienna X ma rozkład Weibulla dany wzorem $f(x, a) = a^{-1} b x^{b-1} exp(-x^b/a) I(x > 0)$.
  - Jaki rozkład ma zmienna $2\sum_{i=1}^{n} Z_i$?

Zadanie 2.

- Znajdź rozkład zmiennej $Z = \frac{-ln(X)}{\theta}$ jeśli zmienna X ma rozkład dany wzorem $f(x, \theta) = \theta^{-1} x^{(1-\theta)/\theta} I(0 < x < 1)$.
  - Jaki rozkład ma zmienna $2\sum_{i=1}^{n} Z_i$?

Uwaga: skorzystaj z faktu, że rozkład wykładniczy to szczególny przypadek rozkładu Gamma (slajd 34), następnie skorzystaj z tw. o sumie zmiennych o rozkładzie Gamma (slajd 37), a na koniec z tego, że rozkład chi kwadrat to szczególny przypadek rozkładu Gamma (slajd 29).
Punktacja: 1 punkt za zadanie 1 i 1 punkt za zadanie 2

Dziękuję za uwagę.