UNIWERSYTET WARSZAWSKI

Wydział Matematyki, Informatyki i Mechaniki
Uniwersytet Warszawski

Statystyczna Analiza Danych 1

Skrypt do ćwiczeń

Ewa Szczurek
Michał Ciach
Magda Grynkiewicz
Anna Macioszek
Błażej Miasojedow
Bartosz Piotrowski
Piotr Pokarowski
Grzegorz Skoraczyński
Paulina Urban

Sem. letni 2018/2019

Spis treści

1 Zmienne losowe

Niech dana będzie prestrzeń probabilistyczna (Ω, F, P). Funkcję X określoną na zbiorze zdarzeń elementarnych Ω, o wartościach będących podzbiorem zbioru liczb rzeczywistych

$$X : \Omega \to Y,$$

$Y \subset \mathbb{R}$ oraz taką, że dla każdego $t \in \mathbb{R}$ zbiór

$$\{\omega \in \Omega : X(\omega) < t\}$$

jest zdarzeniem, czyli należy do zbioru zdarzeń F, nazywamy zmienną losową.

Podstawowe typy zmiennych losowych to

- Dyskretna (skokowa): Przyjmuje wartości ze skończonego lub przeliczalnego zbioru, np. wyników rzutu kostką
- Ciągła: Przyjmuje ciągłe wartości ze zbioru $\mathbb{R}$, np. pomiarów ekspresji genów

Przykład 1.1. (Zmienna losowa ciągła)

- Eksperyment: rejestracja czasu X świecenia losowo wybranej żarówki, do przepalenia.
- X to zmienna losowa o wartościach rzeczywistych nieujemnych.
- Wylosowanie innej żarówki da na ogół inną wartość. Wartości te są przyjmowane z pewnym prawdopodobieństwem, które możemy oszacować, losując i rejestrując odpowiednio dużo razy.

$\boxtimes$

1.1 Rozkłady zmiennych losowych

Rozkład zmiennej losowej X to funkcja $P_X(A) = P(X^{-1}(A))$, gdzie A to dowolny podzbiór zbioru wartości Y, a P to miara prawdopodobieństwa określona na zdarzeniach losowych. ($A \in \mathcal{B}(Y)$, gdzie $\mathcal{B}(Y)$ sigma-ciało podzbiorów zbioru wartości zmiennej losowej X.)

Dystrybuanta zmiennej losowej X to funkcja dana wzorem

$$F_X(t) = P(\{\omega \in \Omega : X(\omega) \leq t\}) = P_X(\{x : x \leq t\}) = P_X((-\infty, t))$$

Rozkład zmiennej dyskretnej X zdefiniowany jest przez zbiór par $x, P_X(x)$, gdzie $x \in Y$: wartość zmiennej X, a $P_X(x) = P_X(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$: prawdopodobieństwo, że zmienna X przyjme wartość x. Inaczej ujmując

$$P_X(x) = P(X^{-1}(x))$$

Dla dyskretnego rozkładu prawdopodobieństwa mamy

- Dla dowolnego $A \subset Y$, $P_X(A) = \sum_{x_i \in A} P_X(x_i)$
- $\sum_i P_X(x_i) = 1$

Rozkład zmiennej ciągłej X, to rozkład P_X dla którego istnieje funkcja f, taka że

$$P_X(A) = \int\limits_A f(x)dx$$

zwana funkcją gęstości prawdopodobieństwa.

Dla ciągłego rozkładu prawdopodobieństwa mamy

- $\int\limits_{-\infty}^{\infty} f(x)dx = 1$
- Dystrybuanta X wyraża się wzorem $F_X(t) = \int_{-\infty}^{t} f(x)dx$

1.2 Wartość oczekiwana zmiennej losowej

Intuicyjnie ujmując, wartość oczekiwana to wartość określająca spodziewany wynik doświadczenia losowego.

Wartość oczekiwaną zmiennej dyskretnej X o rozkładzie P_X definiujemy jako

$$\mathbb{E}(X) = \sum_i x_i P_X(x_i)$$

Wartość oczekiwaną zmiennej ciągłej X o gęstości rozkladu f definiujemy zaś jako

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

Własności wartości oczekiwanej:

- $\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$.
- $\mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{E}[X_i]$

1.3 Wariancja zmiennej losowej

Intuicyjnie, jest to miara zmienności zmiennej losowej.

Wariancję zmiennej losowej X liczymy jako średnią kwadratów odchyleń od wartości oczekiwanej.

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

Zachodzi też

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

Własności wariancji:

- $\mathrm{Var}[aX + b] = \mathbb{E}[aX + b - a\,\mathbb{E}[X] - b]^2 = a^2\,\mathbb{E}[X - \mathbb{E}[X]]^2 = a^2\,\mathrm{Var}[X]$.
- dla niezależnych zmiennych losowych $X_1, \ldots, X_n$, $\mathrm{Var}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathrm{Var}[X_i]$

1.4 Kwantyle

Kwantylem rzędu p $(0 < p < 1)$ dla zmiennej losowej X typu ciągłego o dystrybuancie F i gęstości f nazywamy każdą liczbę x_p spełniającą którykolwiek z równoważnych warunków:

$$\begin{aligned} F(x_p) &= p \\ P(X \leq x_p) &= p \\ \int_{-\infty}^{x_p} f(x)dx &= p. \end{aligned}$$

Mediana to kwantyl rzędu $p = 0.5$

2 Ważne rozkłady

Rozkłady zmiennych losowych można podzielić na grupy

- Rozkłady jednostajne, dyskretny lub ciągły
- Rozkłady związane ze schematem Bernoulliego, lub losowaniem z bądz bez zwracania: rozkład zero-jedynkowy, dwumianowy, geometryczny i hipergeometryczny
- Rozkłady związane z rozkładem Poissona: Poissona i wykładniczy
- Rozkład normalny, występujący często w naturze, i inne związane z nim: t-Studenta, chi-kwadrat

2.1 Rozkład jednostajny dyskretny (Uniform(n))

Zmienna X o rozkładzie jednostajnym dyskretnym przyjmuje z tym samym prawdopodobieństwem wartości z dowolnego skończonego zbioru. Tutaj oznaczamy go $\{1, \ldots, n\}$, gdzie parametr n to liczba całkowita dodatnia.

$$P_X(x) = 1/n, \text{ dla } x = 1, 2, \ldots, n$$

$$\mathbb{E}(X) = \frac{n+1}{2}$$

$$\mathrm{Var}(X) = \frac{n^2-1}{12}$$

Przykład 2.1. (**Rzut kostką**) $n = 6$, $\mathbb{E}(X) = 3.5$, $\mathrm{Var}(X) = \frac{35}{12}$. ⊠

2.2 Rozkład jednostajny ciągły (Uniform(a, b))

Zmienna X o rozkładzie jednostajnym ciągłym przyjmuje wartości z danego przedziału $[a, b]$ z tym samym prawdopodobieństwem proporcjonalnym do jego długości.

$$P_X(x) = \frac{1}{b-a}, \text{ dla } x \in [a, b]$$

$$\mathbb{E}(X) = \frac{a+b}{2}$$

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Przykład 2.2. Błąd zaokrąglania do liczb całkowitych na przedziale $[-0.5, 0.5]$. ⊠

2.3 Rozkład zero-jedynkowy (Bernoulli(p))

Zmienna X o rozkładzie zero-jedynkowym odpowiada próbie Bernoulliego mogącej zakończyć się sukcesem (wartość 1) albo porażką (0). X przyjmuje wartośc 1 z prawdopodobieństwem p i wartość 0 z przwdopodobieństwem $1 - p$.

$$P_X(0) = 1 - p, P_X(1) = p$$

$$\mathbb{E}(X) = p$$

$$\mathrm{Var}(X) = p(1-p)$$

Przykład 2.3. (**Rzut monetą**) $p = 1 - p = \mathbb{E}(X) = 0.5$, $\mathrm{Var}(X) = 1/4$. ⊠

2.4 Rozkład dwumianowy (Binomial(n, p))

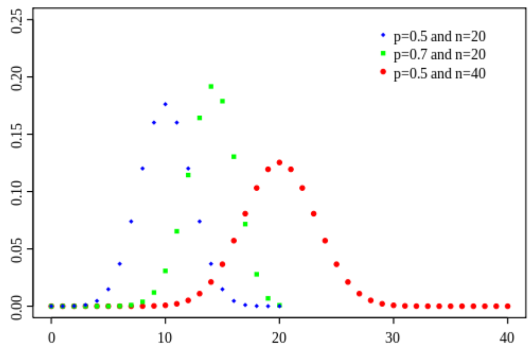
Zmienna X jest sumą wyników w n próbach Bernoulliego (schemacie Bernoulliego), czyli $X = x$ oznacza, że w n próbach było x sukcesów (Binomial(1,p)=Bernoulli(p))

$$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ dla } x = 0, 1, \ldots, n$$

$$\mathbb{E}(X) = np$$

$$\mathrm{Var}(X) = np(1-p)$$

Przykład 2.4. Losowanie ze zwracaniem (po losowaniu odkładam spowrotem) z N kul, gdzie M jest białych a $N - M$ czarnych. Sukces to kula biała. Prawdopodobieństwo sukcesu za każdym losowaniem to M/N. Losowanie n razy ze zwracaniem to schemat Bernoulliego. ⊠



Rysunek 1: Rozkład dwumianowy. Ilustracja z `https://commons.wikimedia.org/wiki/File:Binomial_distribution_pmf.svg`

2.5 Ujemny rozkład dwumianowy (NegativeBinomial(p, r))

Rozkład liczby prób Bernoulliego (z prawdopodobieństwem sukcesu p w jednej próbie) potrzebnych do uzyskania pierwszych r sukcesów.

$$P_X(x) = \binom{x-1}{r-1}(p)^r(1-p)^{x-r}, \text{ dla } x = r, r+1, \ldots$$

$$\mathbb{E}(X) = \frac{r}{p}$$

$$\text{Var}(X) = \frac{r(1-p)}{p^2}$$

2.6 Rozkład wielomianowy ($Multinomial(n, p_1, \ldots, p_k)$)

Uogólnienie rozkładu dwumianowego. Tutaj zmienna $X = [X_1, \ldots, X_k]^T$ jest k-wymiarową zmienną, zliczającą ile wypadło wyników i-tego rodzaju w n próbach, każdy z prawdopodobieństwem $p_1, \ldots, p_k$, gdzie $\sum_i p_i = 1$. Czyli wartość zmiennej losowej $x = [x_1, \ldots, x_k]^T$ oznacza, że w n próbach było x_i wyników typu i, $i \in 1, \ldots, k$, przy czym $\sum_i^k n_i = n$ oraz $x_i = 0, 1, \ldots, n$.

$$P_X([x_1, \ldots, x_k]^T) = \binom{n}{x_1 \ldots x_k} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

gdzie $\binom{n}{x_1 \ldots x_k} = n!/(x_1!x_2! \ldots x_k!)$.

$$\mathbb{E}(X_i) = np_i$$

$$\text{Var}(X_i) = np_i(1-p_i.$$

2.7 Rozkład hipergeometryczny (Hypergeometric(N, M, n))

Rozkład liczby sukcesów w pierwszych n próbach Bernoulliego, przy założeniu, że w N próbach będzie M sukcesów.

$$P_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \text{ dla } x = 0, 1, \ldots, n$$

$$\mathbb{E}(X) = np$$

$$\text{Var}(X) = (N-n)(N-1)np(1-p),$$

gdzie $p = \frac{M}{N}$.

Przykład 2.5. Losowanie bez zwracania (element wylosowany nie wraca z powrotem) z pudełka, w którym jest N kul, w tym M kul białych. Losujemy n kul. Zmienna losowa zliczająca, ile kul białych wylosujemy sięgając po n kul ma rozkład hipergeometryczny. ⊠

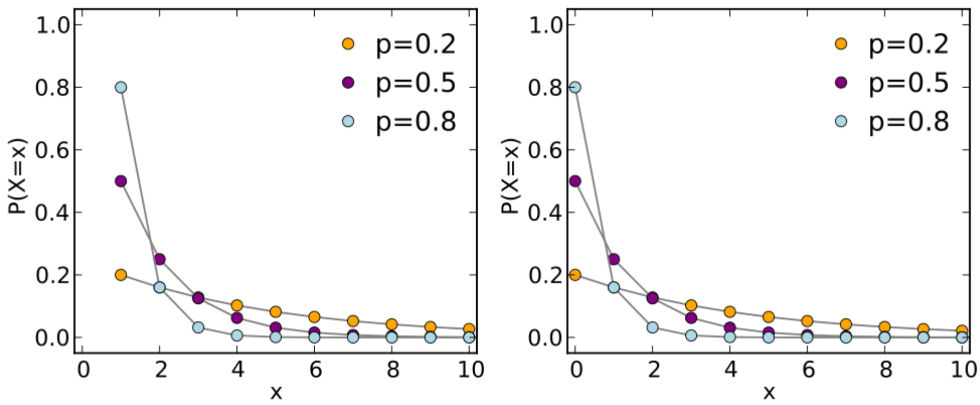
2.8 Rozkład geometryczny (Geometric(p))

Rozkład liczby prób Bernoulliego (z prawdopodobieństwem sukcesu p w jednej próbie) potrzebnych do uzyskania pierwszego sukcesu.

$$P_X(x) = p(1-p)^{x-1}, \text{ dla } x = 1, 2, \ldots$$

$$\mathbb{E}(X) = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$



Rysunek 2: Po lewej: rozkład geometryczny jak zdefiniowany powyżej. Po prawej: rozkład geometryczny w definicji liczby porażek przed pierwszą próbą (wówczas wartości zmiennej losowej geometrycznej to $x = 0, 1, \ldots$).

Przykład 2.6. Szansa na wygranie w loterii to 1 do 14mln. Jaka jest wartość oczekiwana liczby kuponów, które muszę nabyć, aby wygrać w loterii? ⊠

2.9 Rozkład Poissona (Poisson(λ))

Rozkład liczby zdarzeń w jednostce czasu przy założeniu że zdarzenia zachodzą jednostajnie w czasie z częstotliwością λ.

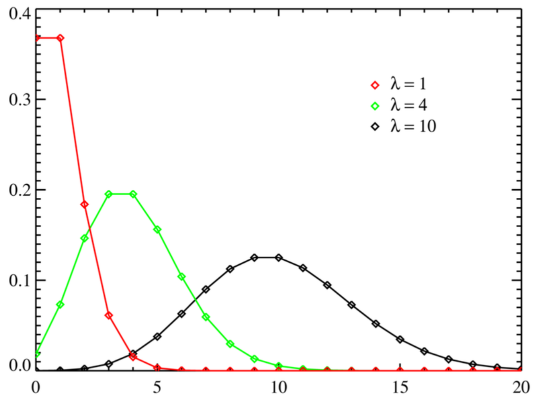
$$P_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \text{ dla } x = 0, 1, \ldots$$

$$\mathbb{E}(X) = \lambda$$

$$\mathrm{Var}(X) = \lambda$$

Mówi się o nim "Prawo małych liczb", gdyż określa rozkład prawdopodobieństwa liczby wystąpień zdarzenia, które zdarza się rzadko ale ma bardzo wiele możliwości, aby się zdarzyć. Rozkład Poissona można wyprowadzić z rozkładu dwumianowego, gdy liczba prób $n \to \infty$ i prawdopodobieństwo sukcesu $p \to 0$, tak, że $np = \lambda$.

- Niech X - liczba zdarzeń w danym przedziale czasowym
- Podzielmy ten przedział na bardzo wiele (n) bardzo małych podprzedziałów.
- Tak małych, że prawdopodobieństwo zajścia więcej niż jednego zdarzenia jest zaniedbywalne w porównaniu z prawdopodobieństwem p zajścia jednego zdarzenia (które samo jest bardzo małe).
- Załóżmy, że
  1. zdarzenia w podprzedziałach są niezależne
  2. prawdopodobieństwo zdarzenia jest identyczne w każdym podprzedziale
  3. zdarzenia nie zachodzą naraz
- Rozkład X będzie przybliżony przez rozkład dwumianowy, z liczbą prób n równej liczbie podprzedziałów i prawdopodobieństwem sukcesu p. A ten rozkład z kolei przybliża rozkład Poissona z parametrem np.



Rysunek 3: Rozkład Poissona. Ilustracja z `https://commons.wikimedia.org/wiki/File:Poisson_distribution_PMF.png`

Przykład 2.7. Rozkład Poissona: Policz w R Rozważmy dwie kostki rzucane razem 100 razy. Niech X zmienna losowa odpowiadająca liczbie razy, gdy wypadły dwie szóstki. Rozkład X to rozkład dwumianowy z $n = 100$ i $p = 1/36$.

a Policz w R prawdopodobieństwa $X = k$ dla $k = 0, \dots, 11$.

b Przybliżmy ten rozkład rozkładem Poissona z $\lambda = np$. Policz te same proawdopodobieństwa w przybliżonym rozkładzie.

⊠

2.10 Rozkład Poissona-alternatywna definicja (Poisson(λt))

Rozkład liczby zdarzeń w przedziale czasowym długości t przy założeniu że zdarzenia zachodzą jednostajnie w czasie z częstotlwością λ.

$$P_X(x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t}, \text{ dla } x = 0, 1, \dots$$

$$\mathbb{E}(X) = \lambda t$$

$$\mathrm{Var}(X) = \lambda t$$

Powyżej definiowaliśmy przyjmując $t = 1$.

2.11 Rozkład wykładniczy (Exp(λ))

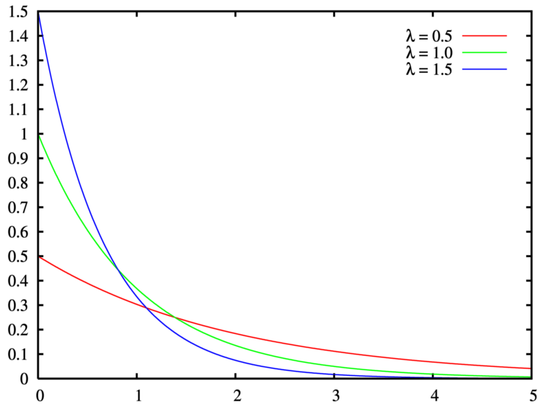
Rozkład czasu, który upłynie do pierwszego zdarzenia, przy założeniu że zdarzenia zachodzą jednostajnie w czasie z częstotlwością λ.

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

$$\mathbb{E}(X) = 1/\lambda$$

$$\mathrm{Var}(X) = 1/\lambda^2$$

Dystrybuanta rozkładu wykładniczego wyraża się wzorem $F(x) = \int_{-\infty}^{x} f(u)du = 1 - e^{-\lambda x}$ dla $x \leq 0$ i $F(x) = 0$ dla $x < 0$.



Rysunek 4: Rozkład wykładniczy. Ilustracja z `https://upload.wikimedia.org/wikipedia/commons/b/b1/Exponential_distribution_pdf.png`

Rozkład wykładniczy ma własność braku pamięci (patrz zadania).
Rozkład wykładniczy ściśle wiąże się z rozkładem Poissona.

- Wezmy rozkład Poissona z parametrem λ odpowiadający liczbie zdarzeń w jednostce czasu.
- Załóżmy, że zdarzenie następuje w chwili t_0.
- Niech zmienna T oznacza czas do następnego zdarzenia.
- Rozkład T można wyznaczyć z prawdopodobieństwa

$$P(T > t) = P(\text{ w przedziale } (t_0, t_0 + t) \text{ nie zaszło żadne zdarzenie}).$$

- Liczba zdarzeń w przedziale $(t_0, t_0 + t)$ długości t zadana jest rozkładem Poissona z parametrem λt
  - a zatem to prawdopodobieństwo wynosi $e^{-\lambda t} =$
  - $=$ 1 - dystrybuanta rozkładu wykładniczego z parametrem λ
  - czyli rozkład T to wykładniczy z parametrem λ.
- W ogólności, czasy pomiędzy kolejnymi zdarzeniami w rozkładzie Poissona(λ), to niezależnie i identycznie rozłożone zmienne losowe o rozkładzie wykładniczym z parametrem λ.

2.12 Rozkład Laplace'a (Laplace(μ, σ))

Funkcja gęstości prawdopodobieństwa dla rozkładu Laplace'a wyraża się wzorem

$$f(x;\mu,\sigma) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right),$$

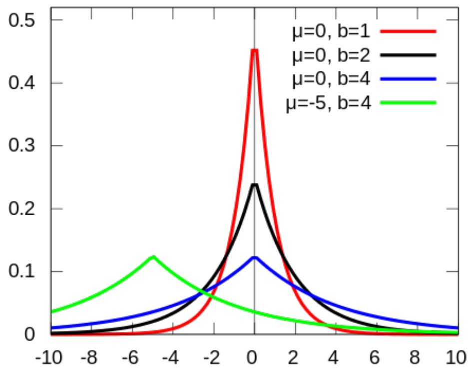
gdzie $\mu \in R$, $\sigma \in R$.

$$\mathbb{E}(X) = \mu$$

$$\mathrm{Var}(X) = 2\sigma^2$$

Dystrybuanta

$$F(x) = \begin{cases} 1 - \frac{1}{2}\exp\left(-\frac{x-\mu}{\sigma}\right), & \text{dla } x \geq \mu \\ \frac{1}{2}\exp\left(\frac{x-\mu}{\sigma}\right), & \text{dla } x < 0 \end{cases}$$



Rysunek 5: Rozkład Laplace'a-wykres gęstości. Ilustracja z `https://upload.wikimedia.org/wikipedia/commons/thumb/0/0a/Laplace_pdf_mod.svg/500px-Laplace_pdf_mod.svg.png`

2.13 Rozkład Normalny (Normal(μ, σ^2))

Zwany także rozkładem Gaussowskim. Parametry: średnia $\mu \in \mathbb{R}$, wariancja $\sigma^2 \in \mathbb{R}_+$, z dodatnim pierwiastkiem (odchylenie standardowe).

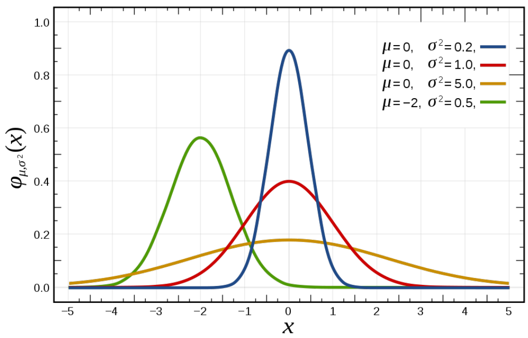
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ dla } x \in \mathbb{R}$$

$$\mathbb{E}(X) = \mu$$

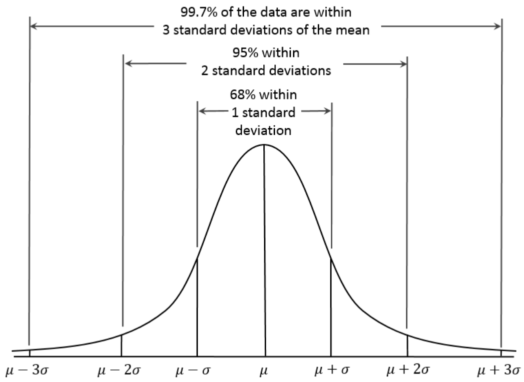
$$\mathrm{Var}(X) = \sigma^2$$

Rozkład normalny standardowy ma średnią $\mu = 0$ i wariancję $\sigma^2 = 1$. Stąd, jego funkcja gęstości to

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \text{ dla } x \in \mathbb{R}$$



Rysunek 6: Rozkład normalny-wykres gęstości. Ilustracja z `https://upload.wikimedia.org/wikipedia/commons/thumb/7/74/Normal_Distribution_PDF.svg/1000px-Normal_Distribution_PDF.svg.png`



Rysunek 7: Rozkład normalny-Reguła empiryczna. Ilustracja z `https://upload.wikimedia.org/wikipedia/commons/a/a9/Empirical_Rule.PNG`

2.14 Rozkład χ^2 (ChiSquared(ν).)

Rozkład sumy kwadratów ν niezależnych rozkładów standardowych normalnych. Parametr ν jest liczbą całkowitą dodatnią i zwany jest stopniem swobody.

$$f(x) = \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} \text{ dla } x \geq 0$$

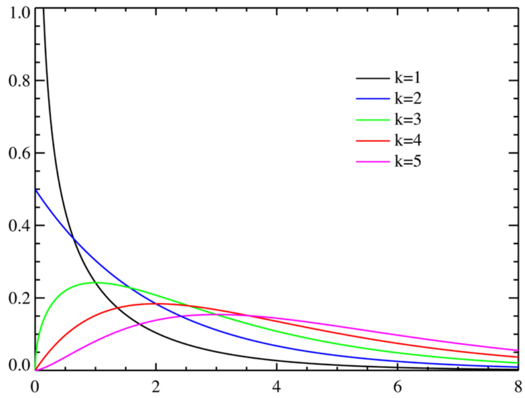
$$\mathbb{E}(X) = \nu$$

$$\mathrm{Var}(X) = 2\nu$$

2.15 Rozkład $t-$Studenta (t(ν).)

Rozkład wartości zmiennej X gdy $X = \frac{Y}{\sqrt{Z/\nu}}$ dla zmiennej Y standardowej normalnej $Y \sim Normal(0,1)$ i $Z \sim ChiSquared(\nu)$. Parametr ν jest liczbą całkowitą dodatnią i zwany jest stopniem swobody.

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})(1+x^2/\nu)^{(\nu+1)/2}} \text{ dla } x \in \mathbb{R}$$



Rysunek 8: Rozkład χ^2 - wykres gęstości. Ilustracja z `Wikipedia`

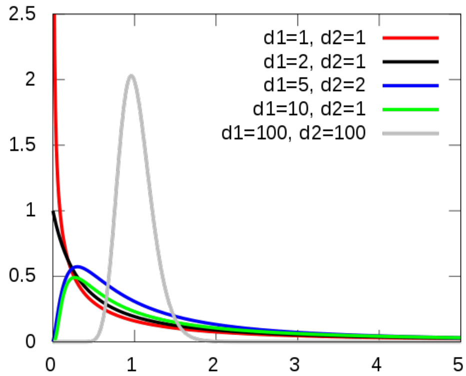
$$\mathbb{E}(X) = 0$$
$$\mathrm{Var}(X) = \frac{\nu}{\nu - 2}$$

Zastosowania rozkładów χ^2 i $t - Studenta$ w statystyce odkryjemy na następnych wykładach.

2.16 Rozkład Snedecora-Fishera F ($F(\nu_1, \nu 2)$)

Rozkład zmiennej $X = \frac{Y/\nu_1}{Z/\nu_2}$, dla Y i Z niezależnych zmiennych $\sim \chi^2$ ze stopniami swobody ν_1 i ν_2.

$$f(x) = \frac{1}{B(\nu_1/2, \nu_2/2)} \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} x^{\nu_1/2-1}}{(1 + \frac{\nu_1}{\nu_2}x)^{(\nu_1+\nu_2)/2}} \text{ dla } x > 0.$$
$$E(X) = \frac{\nu_2}{\nu_2 - 2} \text{ dla } \nu_2 > 2$$
$$Var(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \text{ dla } \nu_2 > 4$$



Rysunek 9: Rozkład Snedecora-Fishera F. Tutaj $d_1 = \nu_1$ i $d_2 = \nu_2$. `Wikipedia`

2.17 Przydatne funkcje: gamma i beta

O funkcjach gamma i beta można myśleć jak o odpowiednio ciągłych wersjach silni $n!$ i symbolu Newtona $\binom{n}{k}$.

Funkcja Beta $B(\alpha, \beta)$ zdefiniowana jest na liczbach dodatnich $\alpha, \beta > 0$ jako

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Dla α i β dodatnich całkowitych

$$B(\alpha, \beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$$

Funkcja Gamma $\Gamma(x)\colon \mathbb{R}_+ \to \mathbb{R}$ zdefiniowana jest na liczbach dodatnich $x > 0$

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

Ponieważ $\Gamma(x+1) = x\Gamma(x)$ i $\Gamma(1) = 1$, dla dodatniego całkowitego x zachodzi

$$\Gamma(x) = (x-1)!$$

2.18 Rozkład Gamma (Gamma(λ, r))

Rozkład czasu, który upłynie do pierwszych r zdarzeń, przy założeniu że zdarzenia zachodzą jednostajnie w czasie z częstotlwością λ.

$$f(x) = \frac{1}{\Gamma(r)}\lambda^r x^{r-1}e^{-\lambda x},\ \text{dla } x \in [0, \infty)$$

$$\mathbb{E}(X) = r/\lambda$$

$$\mathrm{Var}(X) = r/\lambda^2$$

2.19 Rozkład Beta (Beta(α, β))

Rozkład tej części jednostki czasu, który upłynie do pierwszych α zdarzeń, przy założeniu że zachodzi $\alpha + \beta$ zdarzeń na jednostkę czasu (jednostajnie tak jak w rozkładzie Poissona).

$$f(x) = \frac{x^\alpha(1-x)^{\beta-1}}{B(\alpha, \beta)},\ \text{dla } x \in [0, 1]$$

$$\mathbb{E}(X) = \frac{\alpha}{(\alpha+\beta)}$$

$$\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$