Rachunek Prawdopodobieństwa

Anna Janicka

wykład XV, 25.01.2022

Centralne twierdzenie graniczne – przykłady cd

Różne ciekawe rozkłady

Plan na dzisiaj

24. Centralne Twierdzenie Graniczne – przykłady cd.

19. Trochę info o interesujących rozkładach

Centralne twierdzenie graniczne

Przykłady cd:

- *Liczba dziewczynek i chłopców*
- *Ilu studentów przyjąć?*
- Sumowanie błędów
- Przedziały ufności

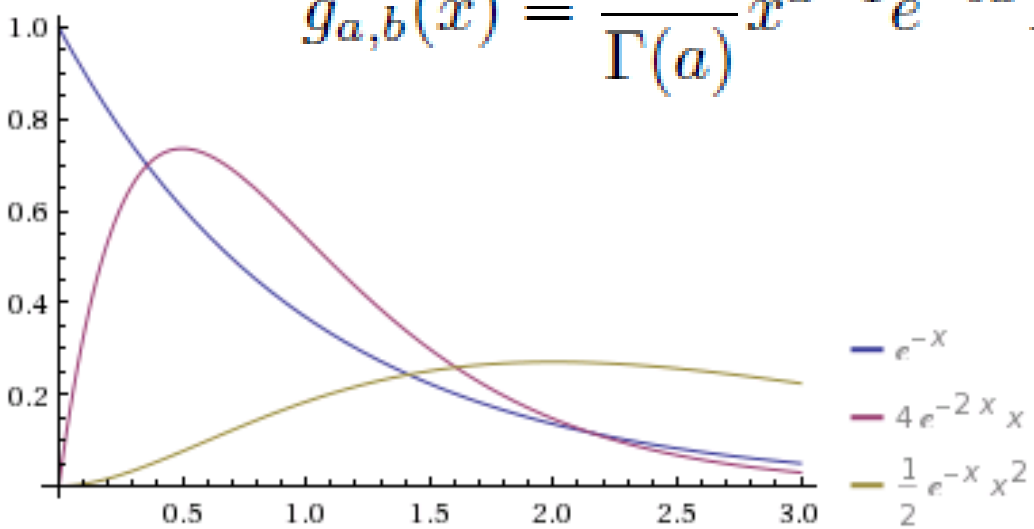
Krótki przegląd wykorzystywanych w praktyce rozkładów

1. Modelowanie różnego rodzaju zjawisk
2. Rozkłady „pojawiające się” w statystyce przy pewnych założeniach dotyczących modeli

Rozkład Gamma $\Gamma(a,b)$, $a,b>0$

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \qquad \Gamma(n) = (n-1)! \text{ dla } n = 1, 2, \ldots$$

$$g_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} 1_{(0,\infty)}(x)$$



$$\mathbb{E}X = a/b$$

$$\mathrm{Var}X = a/b^2$$

Własności rozkładu Gamma

- dla a=1 – rozkład wykładniczy exp(b)
- dla a całkowitego – tzw. rozkład Erlanga
- Γ(n/2, 1/2) – tzw. **rozkład chi-kwadrat** $\chi^2(n)$

Twierdzenie:

Suma niezależnych zmiennych losowych z rozkładów $\Gamma(a_i, b)$ ma rozkład $\Gamma(\Sigma a_i, b)$

$$g_{a,b}(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} 1_{(0,\infty)}(x)$$

Wykorzystanie rozkładu Gamma

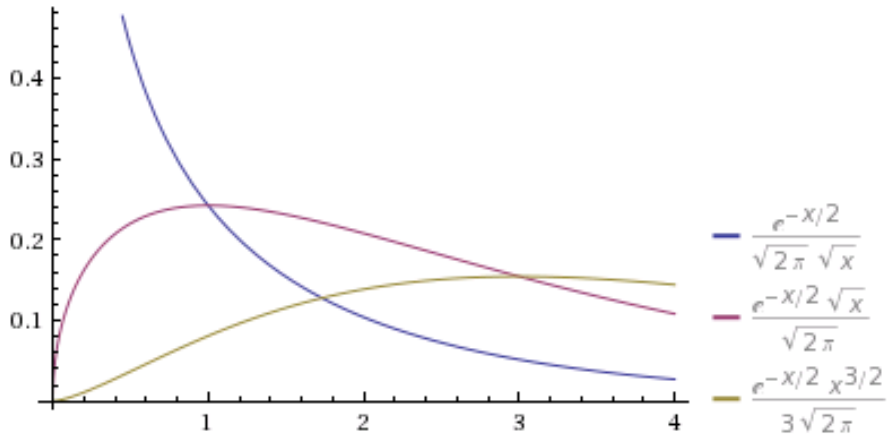
W ekonometrii: do modelowania czasu trwania (zjawisk, życia), ew. wielkości

w modelach regresji Poissona jako rozkład reszt

W statystyce: w statystyce Bayesowskiej

Opisuje czas oczekiwania na zdarzenie o numerze *a* w procesie Poissona, w którym na jednostkę czasu przypada 1/b zdarzenia

Rozkład chi-kwadrat $\chi^2(n)$



Twierdzenie: suma kwadratów *n* niezależnych zmiennych losowych o rozkładach N(0,1) ma rozkład $\chi^2(n)$

$$\mathbb{E}X = n, \quad \mathrm{Var}X = 2n$$

dla dużych n jak rozkład normalny

Rozkład średniej i wariancji z próby

Twierdzenie:

Jeżeli $X_1, X_2, \ldots, X_n$ *są niezależnymi zmiennymi losowymi o rozkładzie* $\mathcal{N}(0,1)$ *oraz* m, s^2 *oznaczają odpowiednio średnią i wariancję z próby:*

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

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - m)^2,$$

to $\sqrt{n}m$ *ma rozkład* $\mathcal{N}(0,1)$, ns^2 *ma rozkład* χ^2_{n-1} *oraz zmienne* m, s^2 *są niezależne.*

Wykorzystanie rozkładu chi-kwadrat

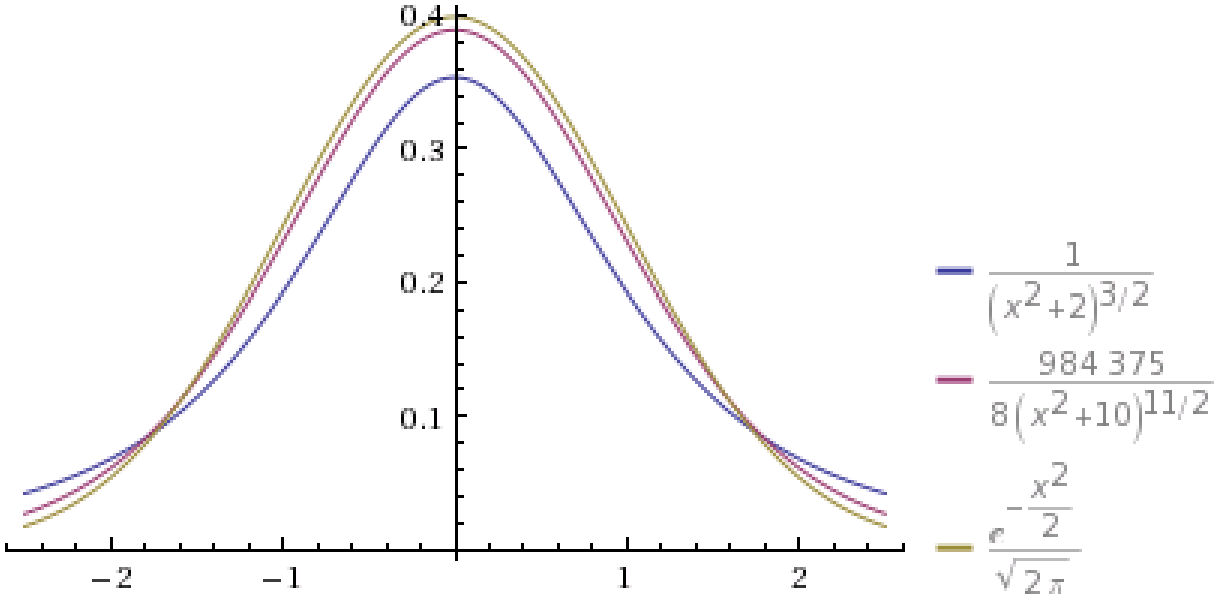
Bardzo czesto wykorzystywany w statystyce:

pojawia się w modelach o „standardowych” założeniach przy testowaniu hipotez, konstrukcji przedziałów ufności oraz jako „składnik” wielu innych rozkładów

Rozkład *t*-Studenta *t*(*n*), *n*=1,2,…

$n^{1/2}X/Y^{1/2}$ dla X i Y niezależnych $X \sim N(0,1)$, $Y \sim \chi^2(n)$

$$g_n(x) = \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma(\frac{1}{2}(n+1))}{\Gamma(\frac{1}{2}n)} \left(1 + \frac{x^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad n = 1, 2, \ldots.$$



$\mathbb{E}X = 0$ dla $n > 1$ $\quad \mathrm{Var}X = n/(n-2)$ dla $n > 2$

Wykorzystanie rozkładu *t*-Studenta

Statystyka: regresja liniowa, testowanie hipotez, budowa przedziałów ufności

Ekonometria: jako alternatywa dla rozkładu normalnego (ma „grubsze” ogony)

Dla dużych n: rozkład prawie normalny

Rozkład *F*-Snedecora $F(d_1,d_2)$, d_1,d_2 =1, 2,…

X ma rozkład $F(d_1,d_2)$, jeśli $X = (Y_1/d_1)/(Y_2/d_2)$, gdzie Y_i są niezależne o rozkładzie $\chi^2(d_i)$

$$g_{d_1,d_2}(x) = \frac{\left(\frac{d_1\,x}{d_1\,x+d_2}\right)^{d_1/2}\left(1-\frac{d_1\,x}{d_1\,x+d_2}\right)^{d_2/2}}{x\,\mathrm{B}(d_1/2,d_2/2)}1_{(0,\infty)}(x)$$

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},\quad a,\,b>0$$

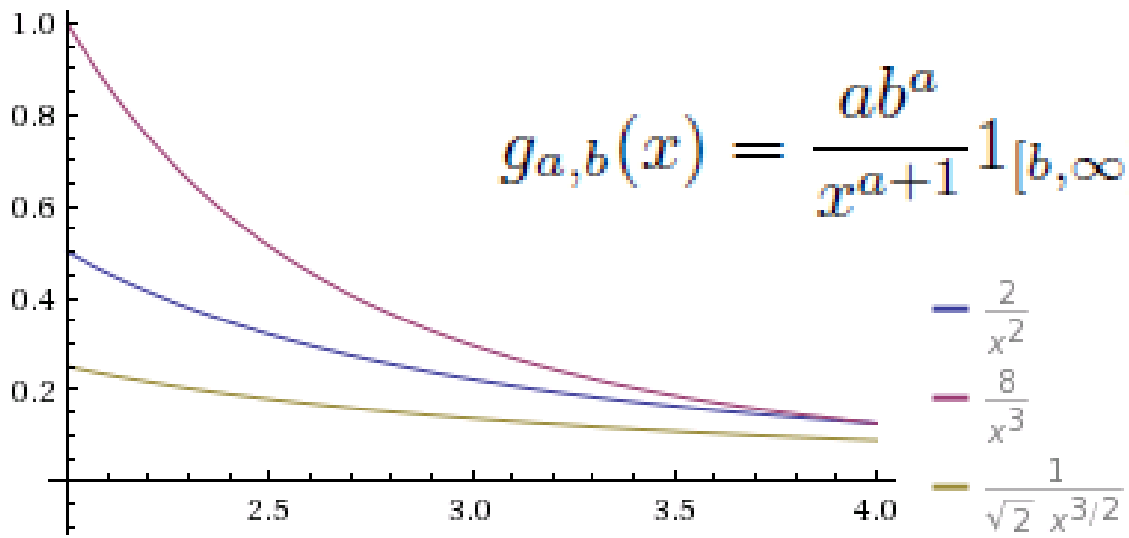
$$\mathbb{E}X = d_2/(d_2-2) \text{ dla } d_2>2$$

$$\mathrm{Var}X = \frac{2\,d_2^2\,(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)} \text{ dla } d_2>4$$

Wykorzystanie rozkładu *F*-Snedecora

Statystyka: testowanie hipotez statystycznych

Rozkład Pareto



$$g_{a,b}(x) = \frac{ab^a}{x^{a+1}} 1_{[b,\infty)}(x), \quad a,\, b > 0$$

$$\mathbb{E}X = ab/(a-1) \text{ dla } a > 1$$

$$\mathrm{Var}X = ab^2/[(a-1)^2(a-2)] \text{ dla } a > 2$$

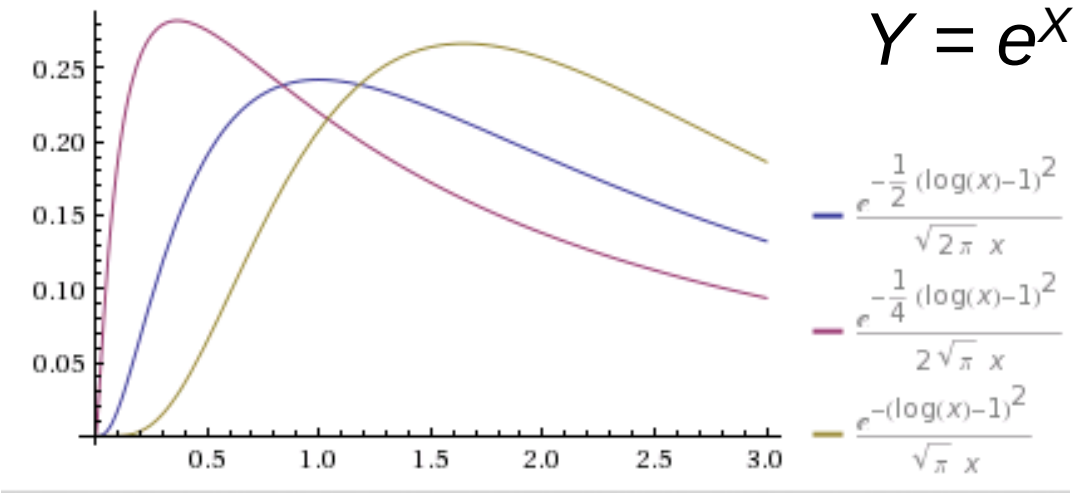
Wykorzystanie rozkładu Pareto

Pareto opisywał rozkłady dochodów (własność: większa część bogactwa w danej społeczności jest w rękach mniejszości → reguła Pareto 80-20, odpowiadająca pewnemu $\alpha > 1$)

Stosowane nie tylko do dochodów i bogactwa, ale również np. finanse, ubezpieczenia, nauki aktuarialne

Rozkład lognormalny (logarytmicznie normalny) $L(m,\sigma^2)$, $m\in\Re$, $\sigma>0$

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp\left(-\frac{(\ln x - m)^2}{2\sigma^2}\right)1_{(0,\infty)}(x)$$



$Y = e^X$, gdzie $X \sim N(m, \sigma^2)$

$$\mathbb{E}Y = \exp(m + \tfrac{1}{2}\sigma^2)$$

$$\mathrm{Var}Y = (\exp(\sigma^2) - 1)\cdot\exp(2m + \sigma^2)$$

Twierdzenie: Rozkład iloczynu zmiennych z rozkładów lognormalnych ma rozkład lognormalny

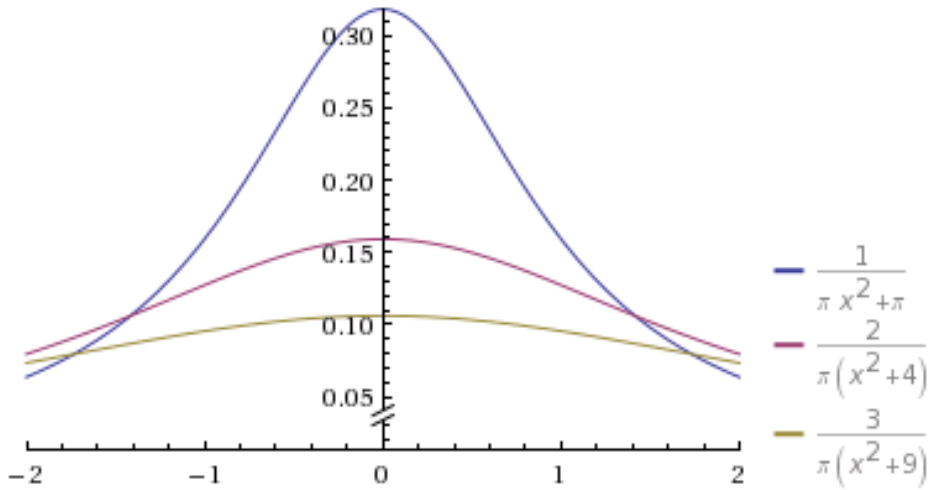
Wykorzystanie rozkładu lognormalnego

Do modelowania wszędzie tam, gdzie zmienne są nieujemne i mają charakter multiplikatywny (można stosować CTG dla logarytmów)

Do modelowania dochodów (poza procentem najbogatszych). W finansach (np. model Blacka-Scholesa)

Rozkład Cauchy'ego Cau(*a*, *m*), gdzie *a*>0, *m*∈ℜ

$$g(x) = \frac{a}{\pi((x-m)^2 + a^2)}$$



Nie ma wartości oczekiwanej ani innych momentów.

Nie można do niego stosować PWL, CTG!

Własności rozkładu Cauchy'ego

Twierdzenie: Średnia identycznych niezależnych zmiennych z rozkładu Cauchy'ego ma rozkład Cauchy'ego

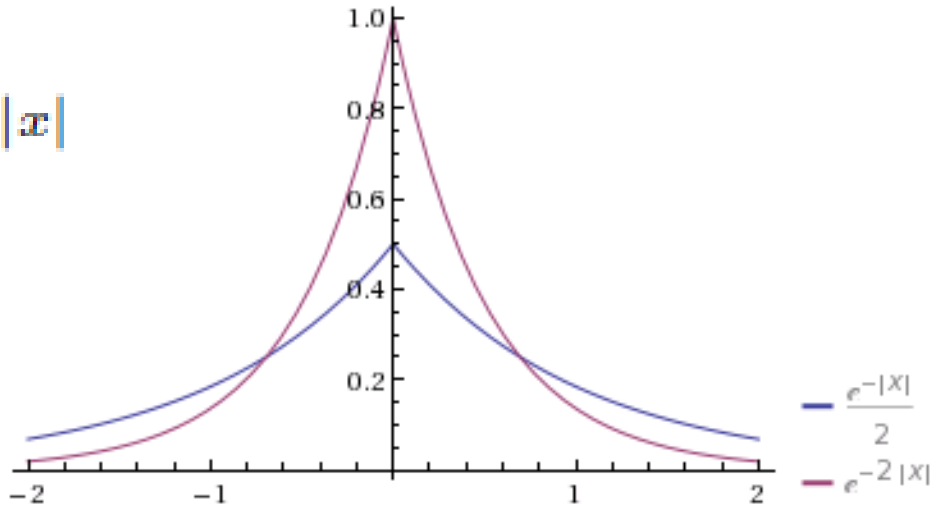
→ Wnioskowanie na podstawie średniej jest bez sensu

Iloraz dwóch zmiennych o standardowych rozkładach normalnych ma rozkład Cauchy.

Rozkład Cau(1,0) to rozkład t-Studenta (1).

Rozkład wykładniczy dwustronny (rozkład Laplace'a) z parametrem $\lambda>0$

$$g(x) = \frac{1}{2}\lambda e^{-\lambda|x|}$$



$$\mathbb{E}X = 0,\ \mathrm{Var}X = 2/\lambda^2$$

różnica dwóch niezależnych zmiennych z rozkładu wykladniczego ma rozkład Laplace'a

Rozkład Weibulla

Inne uogólnienie rozkładu wykładniczego

$$g(x) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-(x/\beta)^{\alpha}}1_{(0,\infty)}(x), \quad \alpha, \beta > 0.$$

$$\mathbb{E}X = \beta\Gamma(1+\frac{1}{\alpha})$$

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

Do modelowania czasów trwania, w zależności od wartości α: malejąca, stała, rosnąca zapadalność. Nauki aktuarialne.

Ostatnia ankieta….

https://forms.gle/3VMYtLYTvXDUG1Fp6