Probability Calculus

Anna Janicka

lecture XIII, 27.01.2022

Central Limit Theorem – examples

Some interesting distributions

Plan for Today

- Central Limit Theorem – examples of applications
- Some interesting distributions

Central Limit Theorem – reminder

1. Classical version:

Let $X_1, X_2, \ldots$ *be identically distributed independent random variables, such that* $\mathbb{E}X_1^2 < \infty$. *If by* $m = \mathbb{E}X_1$ *we denote the mean, and by* $\sigma^2 = \mathrm{Var}X_1$ *the variance of this distribution, then for any* $t \in \mathbb{R}$, *we have that*

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

where $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx$

is the CDF of the standard normal distribution.

also:

$$\mathbb{P}\left(s \leqslant \frac{X_1 + X_2 + \ldots + X_n - nm}{\sigma\sqrt{n}}\right) \xrightarrow{n\to\infty} 1 - \Phi(s)$$

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

De Moivre-Laplace Theorem – reminder

2. Theorem:

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables, such that $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0)$.
Then, we have that for any $s < t$,

$$\mathbb{P}\left(s \leqslant \frac{X_1+X_2+\ldots+X_n-np}{\sqrt{np(1-p)}} \leqslant t\right) \xrightarrow{n\to\infty} \Phi(t) - \Phi(s).$$

each inequality (both in the CLT and in dML) may be changed to strict without consequences

Central Limit Theorem

3. Examples

- *boys and girls*
- how many students should be accepted?
- aggregate errors
- confidence intervals

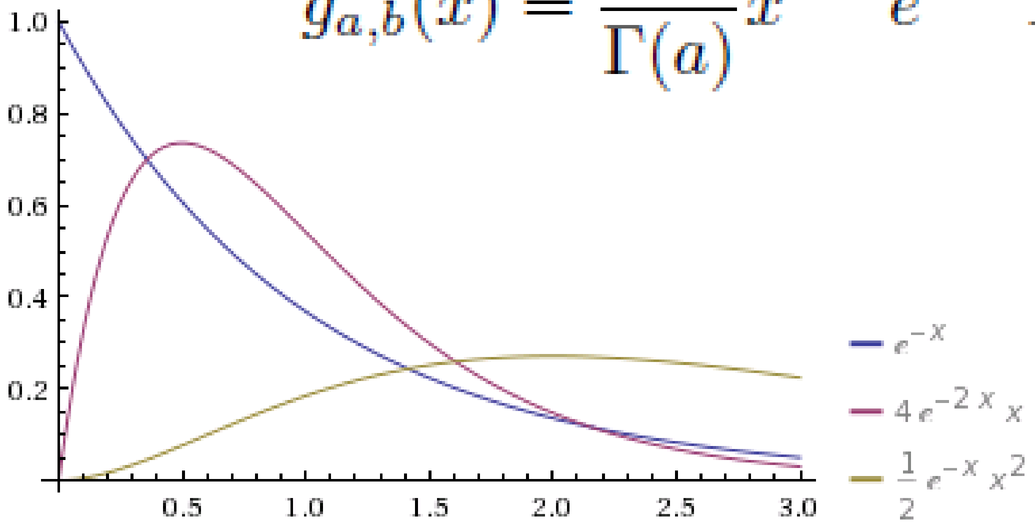
Some useful distributions

used in statistics and economics when modeling or as a result of operations performed on other distributions

Gamma Distribution Γ(*a*,*b*), *a*,*b*>0

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

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



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

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

Gamma Distribution properties

- for a=1 – exponential distribution exp(b)
- for integer a – Erlang Distribution
- Γ(n/2, 1/2) – **chi-squared distibution** $\chi^2(n)$

Theorem:

A sum of independent random variables from gamma distributions $\Gamma(a_i, b)$ has a $\Gamma(\Sigma a_i, b)$ distribution

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

How the Gamma Distribution is used

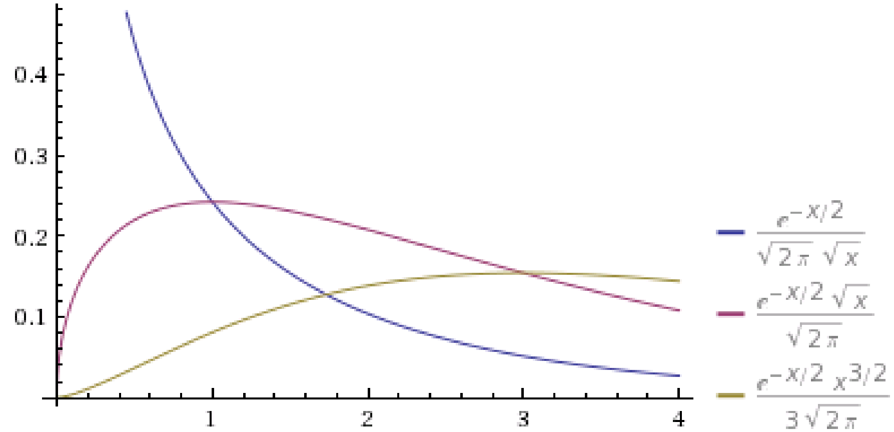
In econometrics: duration of events

In statistics: bayesian statistics.

....the chi-squared distribution

Describes the time until the *a*th event in a Poisson proces in which there is on average 1/*b* events per unit of time

Chi-squared Distribution $\chi^2(n)$



Theorem: sum of squares of *n* IIN random variables (independent identically N(0,1) distributed) has a $\chi^2(n)$ distribution

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

for large n: like the normal distr.

The distribution of the sample mean and variance

Theorem:

Let $X_1, X_2, \ldots, X_n$ *be independent identically distributed random variables from a standard normal distribution and let* $\bar{X}, s^2$ *be the sample mean and variance, respectively:*

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

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

Then $\sqrt{n}\bar{X}$ *has a* $\mathcal{N}(0,1)$, *distribution,*
ns^2 *has a* χ^2_{n-1} *distribution*
and $\bar{X}, s^2$ *are independent.*

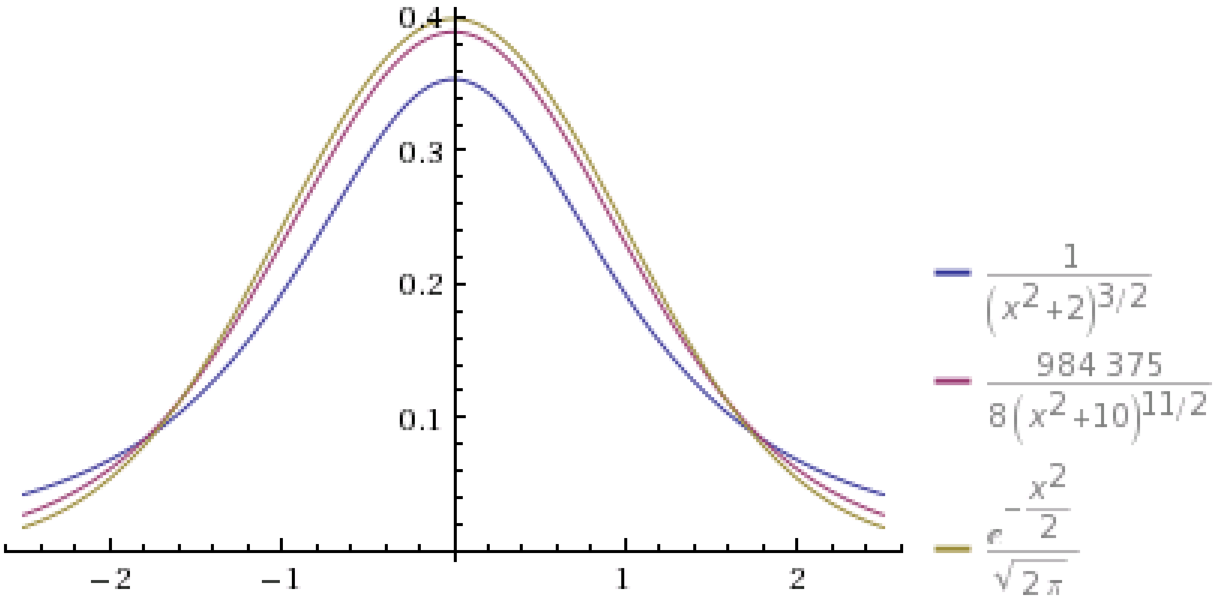
Chi-squared distribution uses

Used very frequently in statistics: appears in „standard" models when testing hypotheses, cosntructing confidence intervals, as a „component" of other distributions.

t-Student Distribution $t(n)$, n=1,2,…

$n^{1/2}X/Y^{1/2}$ for independent X and Y, $X \sim N(0,1)$, $Y \sim \chi^2(n)$

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



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

t-Student distribution uses

Statistics: regression analysis, hypothesis testing, construction of confidence intervals

Econometrics: as an alternative for the normal distribution in cases where fat tails are needed

for large n: almost N(0,1).

F-Snedecor $F(d_1,d_2)$, $d_1,d_2 = 1, 2, \ldots$

X has a $F(d_1,d_2)$ distribution, if $X = (Y_1/d_1)/(Y_2/d_2)$, where Y_i are independent, $\chi^2(d_i)$

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

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

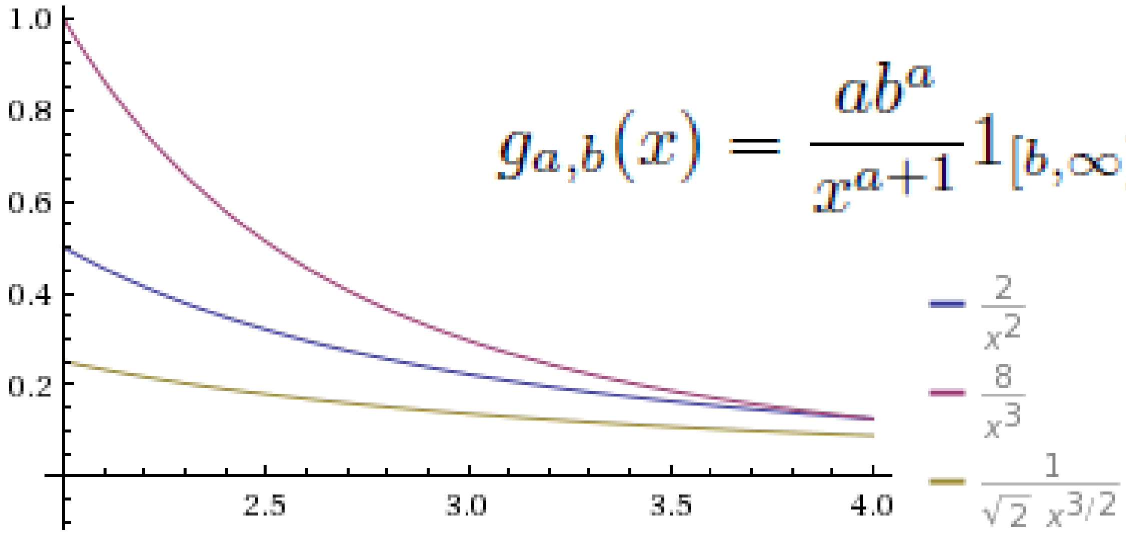
$$\mathbb{E}X = d_2/(d_2-2) \qquad d_2 > 2$$

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

F distribution uses

Statistics: testing hypotheses

Pareto distribution



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

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

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

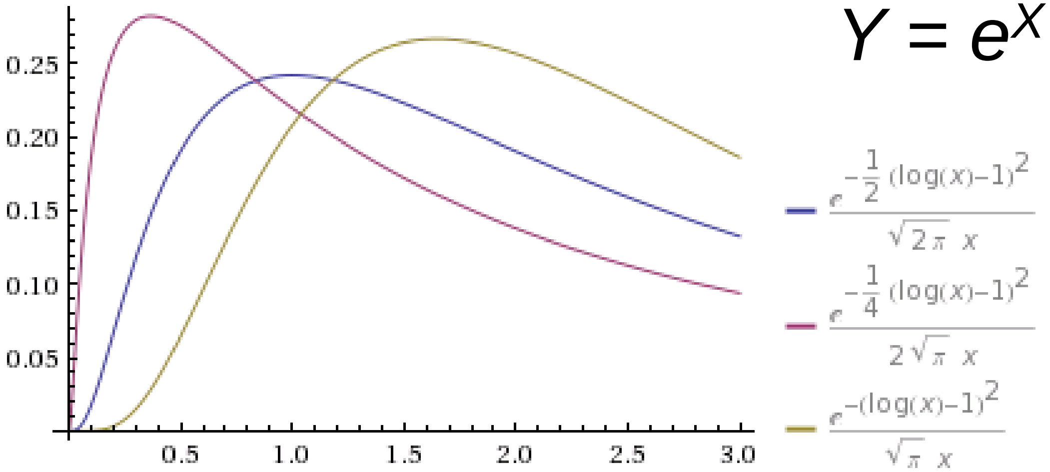
Pareto distribution uses

Pareto described the distribution of wealth/income among various groups in a population
(property: larger part of wealth accumulated by a smaller part of the population
→ Pareto 80-20 rule, which corresponds to a value $\alpha > 1$)

Used not only for income and wealth, but also in finance, insurance, actuarial

Lognormal distribution $L(m,\sigma^2)$, where $m \in \Re$, $\sigma > 0$

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



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

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

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

Theorem: A product of lognormal variables has a lognormal distribution

Lognormal distribution uses

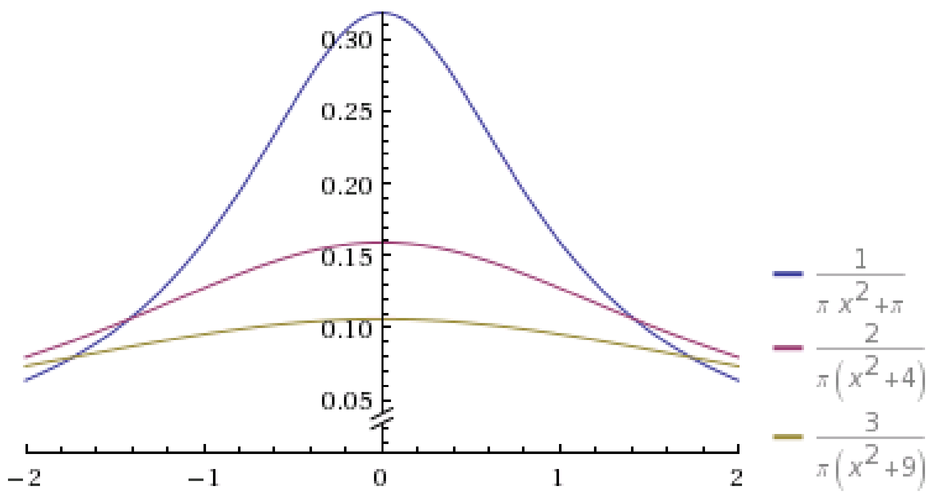
In most cases where we model positive random variables of multiplicative character (the CLT may be used for logarithms)

For modeling income (apart from the most rich).

In finance (eg. Black-Scholes model)

Cauchy distribution Cau(*a*, *m*), where *a*>0, *m*∈ℜ

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



Does not have an expected value nor higher moments.

The LLN nor the CLT may not be applied!

Cauchy distribution properties

Theorem: an average of identical independent Cauchy random variables is a Cauchy variable

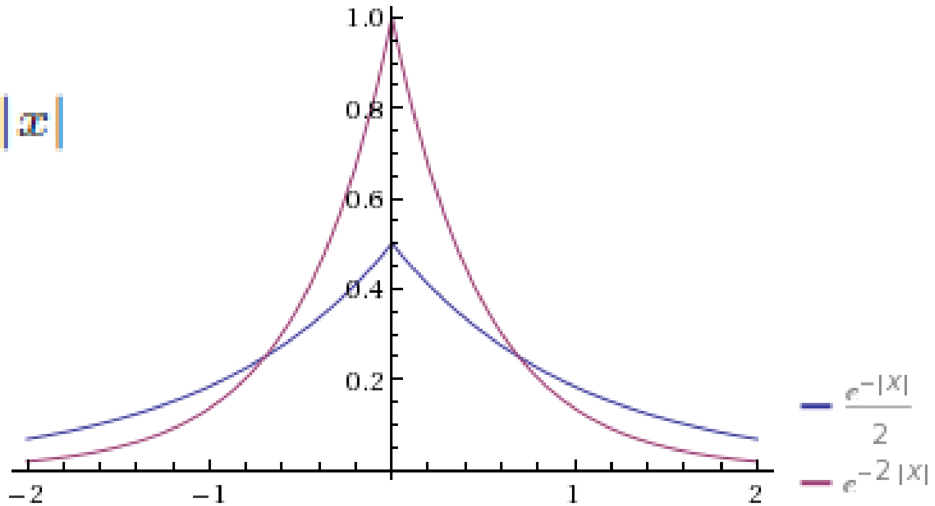
→ conclusions based on the mean are worthless

A ratio of two IIN random variables has a Cauchy distribution

Cau(1,0) is equivalent to the t-Student distribution with one degree of freedom (1).

Two-sided exponential distribution (Laplace distribution) with parameter λ>0

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



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

a difference of two independent exponential random variables has a Laplace distribution

Weibull distribution

A different generalization of the exp. distr.

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

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

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

For modeling duration; depending on the value of α: decreasing, constant, increasing incidence. Actuarial sciences.

Survey

The last quiz…

https://forms.gle/c4dC773k5uDTWzMD7