NMAI059 Probability and statistics 1
Class 7

Robert Šámal

Overview

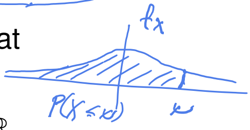
Continuous random variables

Particular continuous distributions and their parameters

General and continuous random variable – what we have learned

- R.v. is a mapping $X : \Omega \to \mathbb{R}$, that for every $x \in \mathbb{R}$ satisfies $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$.
- Discrete r.v. is a r.v.
- CDF of a r.v. X is a function $F_X(x) := P(X \leq x)$.
- CDF F_X is nondecreasing right-continuous function with limits in ± 1 equal to 0/1.

- A continuous r.v. has a PDF $f_X \geq 0$ such that $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$.



- $P(a \leq X \leq b) = \int_a^b f_X(t)dt$ for every $a, b \in \mathbb{R}$.
- $P(X \in A) = \int_A f_X(t)dt$ for a “reasonable set A”.

Expectation of a continuous r.v.

$\circledast = \int \left(\left\lfloor \frac{x}{\Delta} \right\rfloor \cdot \Delta\right) \cdot f_X(x) =: EX$

Definice

Consider a continuous r.v. X with PDF f_X. Then its expectation (expected value, mean) is denoted by $\mathbb{E}(X)$ and defined by

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

$y = n\Delta$

for discrete r.v. Y: $E(Y) = \sum_{y \in \mathrm{Im}(Y)} y \cdot P(Y=y)$

whenever the integral is defined; that is unless it is of type $\infty - \infty$.

$Y = \left\lfloor \frac{X}{\Delta} \right\rfloor \cdot \Delta$ ... Y is X rounded down to a multiple of Δ

integer

- An analogy with computing a center of mass of a pole from a formula for its density.
- Discretization.

$\Delta > 0$

$\Delta \doteq 0$

$EY = \sum_{y \in \mathbb{Z}\cdot\Delta} y\, P(Y=y) = \sum y\, P\left(\frac{y}{\Delta} \le \frac{X}{\Delta} < \frac{y}{\Delta}+1\right)$

$= \sum y\, P(y \le X < y+\Delta)$

$= \sum_{n \in \mathbb{Z}} n\Delta \cdot \int_{n\Delta}^{(n+1)\Delta} f_X(t)\,dt =: \circledast$

$P(x \le X \le x+\Delta) = \int_x^{x+\Delta} f_X(t)\,dt$

f_X is density of probability

$\boxed{\frac{P(x \le X \le x+\Delta)}{\Delta}}$ } average prob. around x

Properties of expectation

PDF

Thm ~~Věta~~ (LOTUS)

Consider a continuous r.v. X with density f_X and a real function g. Then we have

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx,$$

whenever the integral is defined.

by def. we would have to compute

$$\int_{-\infty}^{\infty} y \cdot f_Y(y)\, dy$$

$Y = g(X)$

(We skip the proof.)

Theorem ~~Věta~~ (Linearity of expectation)

For $X_1, \ldots, X_n$ discrete or continuous random variables we have

$$\boxed{\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n).}$$

(Proof later.)

Variance of a continuous r.v.

$\mathbb{E}((X-\mathbb{E}(X))^2)$

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

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2\, f_X(x) dx$$

Writing $\mu = \mathbb{E}(X)$, we have

LOTUS

$$var(X) := \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2\, f_X(x) dx.$$

~~Nota~~ Theorem

For continuous random variables we have the same formula as for discrete ones, $\boxed{var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2}$.

(Proof is the same as for discrete r.v.)

Variance of a sum

Thm

~~Věta~~ (Variance of a sum)

For $X_1, \ldots, X_n$ independent discrete or continuous r.v. we have

$$\boxed{var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n).}$$

(Proof later.)

$\boxed{1}$ indep. is important: $X_2 = -X_1$, X_1 arbitr.

$var(X_1 + X_2) = 0$, $var X_2 = var X_1 \neq 0$

$\boxed{2}$ $X \sim Bin(n,p)$ $\quad var(X) = var X_1 + \ldots + var X_n$

$\quad = n \cdot var X_1 = n \cdot p(1-p)$

"

$X_1 + \ldots + X_n$

i.i.d. $\quad X_i \sim Bern(p)$

indep. id. distr.

Overview

Continuous random variables

Particular continuous distributions and their parameters

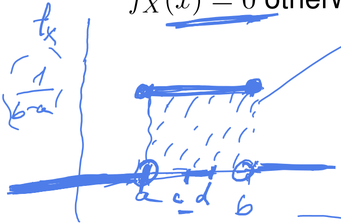
Uniform distribution

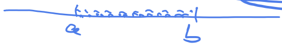
$E(X) = \int_{-\infty}^{\infty} x f_X(x) = \int_a^b x \cdot \frac{1}{b-a} = \left[\frac{x^2}{2}\frac{1}{b-a}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \boxed{\frac{a+b}{2}}$

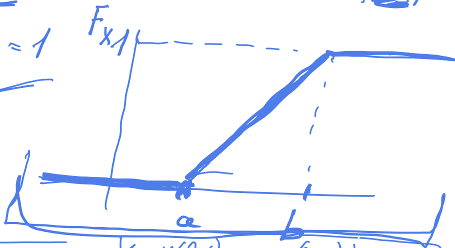
def. of E def of f_X

- R.v. X has a uniform distribution on $[a, b]$, we write $X \sim U(a, b)$, if $f_X(x) = 1/(b - a)$ for $x \in [a, b]$ and $f_X(x) = 0$ otherwise.

$\int_{-\infty}^{\infty} f_X = 1$







$X \sim U(0,1) \to a+(b-a)X \sim U(a,b)$

$P(c \le X \le d) =$ what our depends $\propto |d-c|$

$= \int_c^d f_X = \int_c^d \frac{1}{b-a} = \frac{d-c}{b-a}$

$c=a \quad d=x$

$P(a \le X \le x)$

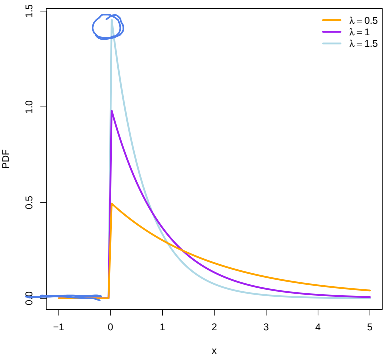
$= \frac{x-a}{b-a}$

Exponencial distribution $Exp(\lambda)$ with rate $\boxed{\lambda > 0}$

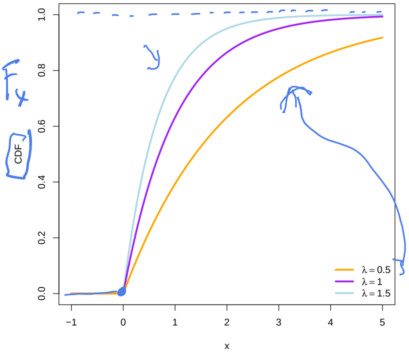
$$F_X(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 - e^{-\lambda x} & \text{for } x \ge 0 \end{cases}$$

$$f_X(x) = \begin{cases} 0 & x \le 0 \\ 0 - e^{-\lambda x} \cdot (-\lambda) = \lambda e^{-\lambda x} & \text{for } x > 0 \end{cases}$$

f_X



F_X



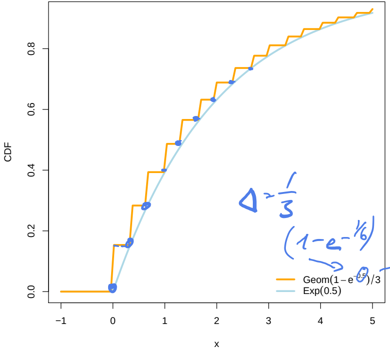
- X models time before next phone call in a call-center / web-server response / time till another lightning in a storm / ...

Relating $Exp(\lambda)$ and $Geom(p)$

$X \sim$ $\qquad Y \sim$

- $P(X > x) = e^{-\lambda x}$ for $x > 0$
- $P(Y > n) = (1-p)^n$ for $n \in \mathbb{N}$

$\Delta \doteq 0$



want $X \doteq \Delta \cdot Y$

$P(X > x) \doteq P(\Delta Y > x)$

$\boxed{x = n \cdot \Delta}$

$e^{-\lambda \Delta n} \qquad P(Y > n) = (1-p)^n$

$e^{-\lambda\Delta} = 1 - p$ (Taylor aprox.)

$p = 1 - e^{-\lambda\Delta} \doteq \lambda\Delta$

EX, $\mathrm{var}\,X \longrightarrow$ exercise

Standard normal distribution

$\int_{-\infty}^{\infty} e^{-x^2/2}$ $\boxed{\int_{-\infty}^{\infty} \varphi(x) = 1}$

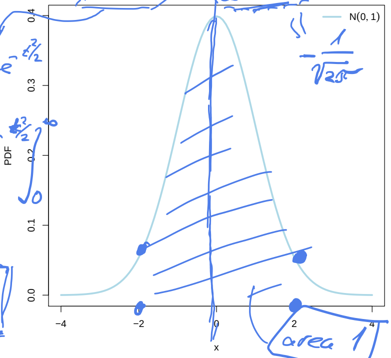
- $\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}}} e^{-x^2/2}$ PDF
- $\Phi(x)$ – antiderivative of φ … doesn't have closed form, ~ Error function
- *Standard normal distribution* $N(0,1)$ has PDF φ and CDF Φ.

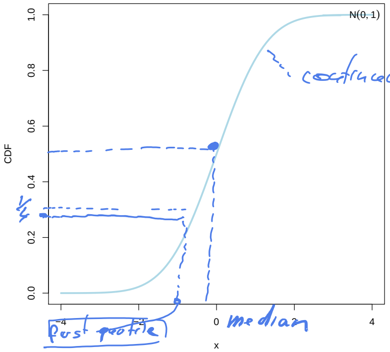
$var(Z) = \mathbb{E}(Z^2) = \int_{-\infty}^{\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-t^2/2} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} t \cdot (t e^{-t^2/2}) = \left[\frac{1}{\sqrt{2\pi}} t e^{-t^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot 1 \cdot e^{-t^2/2} = \int \varphi(t) = 1$

- If $Z \sim N(0,1)$, then $\mathbb{E}(Z) = 0$, $var(Z) = 1$.

$\mathbb{E}(Z) = \int_{-\infty}^{\infty} t\varphi(t) = \int_{-\infty}^{0} t\varphi(t) + \int_{0}^{\infty} t\varphi(t) = 0$ $\quad \sigma_Z = 1$

$\int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} t e^{-t^2/2} = \left[\frac{-1}{\sqrt{2\pi}} e^{-t^2/2}\right]_{0}^{\infty} = 0 - \frac{-1}{\sqrt{2\pi}} = \boxed{\frac{1}{\sqrt{2\pi}}}$



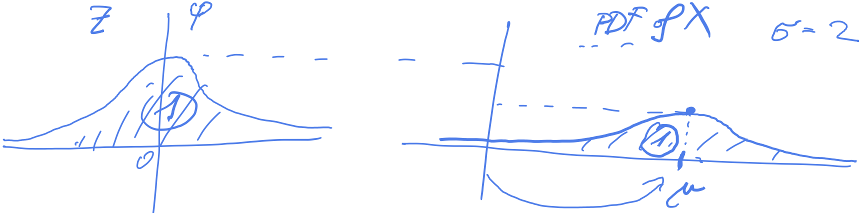


General normal distribution

$EX = \mu + \sigma \cdot EZ = \mu$ / $\mathrm{var}(X) = 0 + \sigma^2 \mathrm{var}\, Z = \sigma^2$

- For $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ we put $\boxed{X = \mu + \sigma \cdot Z}$, where $Z \sim N(0,1)$.

  $EX = \mu$, $\mathrm{var}(X) = \sigma^2$
- We write $X \sim N(\mu, \sigma^2)$ – general normal distribution
- Normal distribution $N(\mu, \sigma^2)$ has density $\frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma})$.



$\Phi(z) = P(Z \le z) = P(X \le \mu + \sigma z) = F_X(\mu + \sigma z)$ / $'$

$\varphi(z) = \Phi'(z) = [F_X(\mu + \sigma z)]' = f_X(\mu + \sigma z) \cdot \sigma$

$\Rightarrow f_X(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$

$z = \frac{x-\mu}{\sigma}$

Resistance to a sum

- Suppose $X_1, \ldots, X_k$ are independent r.v., where $X_i \sim N(\mu_i, \sigma_i^2)$. Then

$$X_1 + \cdots + X_k \sim N(\mu, \sigma^2),$$

where $\mu =$ $E(X_1 + \cdots + X_k) = EX_1 + \cdots + EX_k = \mu_1 + \cdots + \mu_k$

$\sigma^2 =$ $\mathrm{var}(X_1 + \cdots + X_k) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_k)$

$= \sigma_1^2 + \cdots + \sigma_k^2$

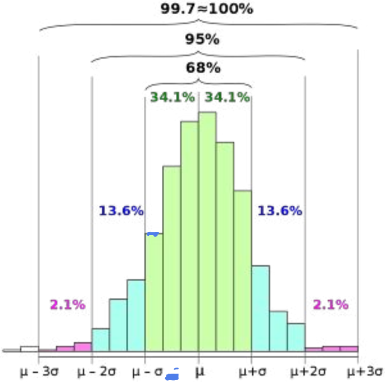
Normal distribution – key properties

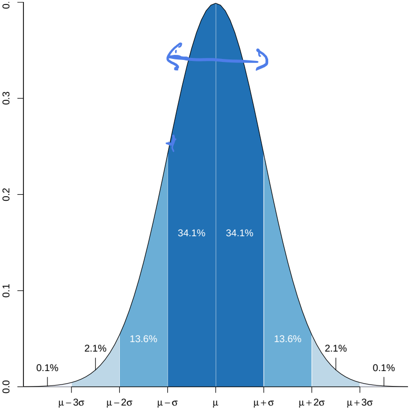
$P(\mu-\sigma < X < \mu+\sigma) = \Phi(1) - \Phi(-1)$

$\doteq 0.68$

$P(\mu-2\sigma < X < \mu+2\sigma) \doteq 0.95$

- 68–95–99.7 rule (3σ rule)
- Central limit theorem





(Image on the left is from Wikipedia, author Melikamp.)

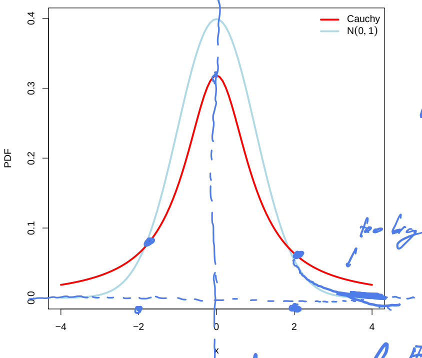
sample $N(\mu, \sigma^2)$, draw histogram

$\varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$

Cauchy distribution

- *Cauchy distribution:* PDF $f(x) = \frac{1}{\pi(1+x^2)}$
- does not have an expectation!

$$\int_{-\infty}^{\infty} f(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1}{1+x^2} = \frac{1}{\pi}\Big[\arctan x\Big]_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \frac{-\pi}{2}\right) = \underline{1}$$



$$EX = \int_{-\infty}^{\infty} x f(x) = \int_{-\infty}^{0} x f(x) + \int_{0}^{\infty} x f(x)$$

$$-\infty$$

$$\frac{1}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2} = \frac{1}{2\pi}\Big[\log(1+x^2)\Big]_{0}^{\infty} = \infty$$

$$= \infty - \infty = ?$$

undefined

if we sample EX

we sample 1000 # out of Cauchy distr, take average

Gamma distribution

- $Gamma(w, \lambda)$, *gamma distribution with parameters* $w > 0$ *and* $\lambda > 0$ has PDF

$$f(x) = \begin{cases} 0 & \text{pro } x \leq 0 \\ \frac{1}{\Gamma(w)} \lambda^w x^{w-1} e^{-\lambda x} & \text{pro } x \geq 0 \end{cases}$$

  where $\Gamma(w) = (w-1)! = \int_0^\infty x^{w-1} e^{-x} dx$.
- For $w = 1$ we get exponencial distribution again.
- If $X_1, \ldots, X_n$ are i.i.d with distribution $Exp(\lambda)$, then $X_1 + \cdots + X_n \sim Gamma(n, \lambda)$.
- Models lifetime of an electronic component, total of rainfall in a year, web-server latency.

A many others

- $Beta(s,t)$ – beta distribution
- χ^2 distribution with k degrees of freedom = chi-square (χ^2_k) is an alternative name for $Gamma(\frac{1}{2}k, \frac{1}{2})$. It is the distribution $Z_1^2 + \cdots + Z_k^2$, where $Z_i \sim N(0,1)$ are i.i.d.
- Student t-distribution
- etc. etc.

Uniform distribution

- R.v. X has a uniform distribution on $[a, b]$, we write $X \sim U(a, b)$, if $f_X(x) = 1/(b - a)$ for $x \in [a, b]$ and $f_X(x) = 0$ otherwise.

Universality of uniform

Věta

Let X be a r.v. with CDF $F_X = F$, let F be continuous and increasing. Then $F(X) \sim U(0,1)$.

Věta

Let F be a function "of CDF-type": nondecreasing right-continuous function with $\lim_{x\to-\infty} F(x) = 0$ a $\lim_{x\to+\infty} F(x) = 1$. Let Q be the corresponding quantile function.

Let $U \sim U(0,1)$ and $X = Q(U)$. Then X has CDF F.