# NMAI059 Probability and statistics 1
# Class 8

Robert Šámal

# Overview

Continuous distributions

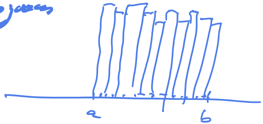
Random vectors

Back to the basics

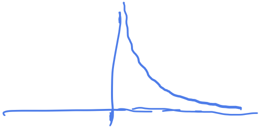
# Which distributions we have seen

- $U(a, b)$ – uniform on interval $[a, b]$

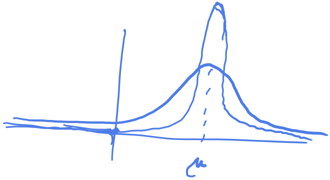

histograms




- $Exp(\lambda)$ – exponential – how long till something happens



- $N(\mu, \sigma^2)$ – normal – how much does a bread weigh




→ Cauchy distr.

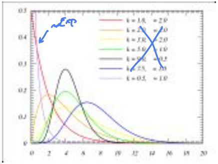
# Gamma distribution

$\int_{-\infty}^{\infty} f(x) = 1$

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

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

  where $\Gamma(w) = (w-1)! = \int_0^\infty x^{w-1} e^{-x} dx$.
- For $w = 1$ we get exponential distribution again. $\sim \frac{1}{0!} \lambda^1 e^{-\lambda x}$
- If $X_1, \ldots, X_n$ are i.i.d with distribution $Exp(\lambda)$, then $X_1 + \cdots + X_n \sim Gamma(n, \lambda)$. (exercise)
- Models lifetime of an electronic component, total of rainfall in a year, web-server latency.

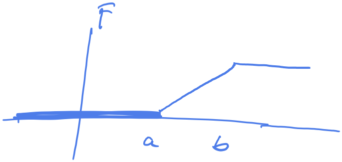


# A many others

- $Beta(s,t)$ – beta distribution
- $\chi^2$ distribution with $k$ degrees of freedom = chi-square ($\chi^2_k$) 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.

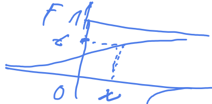
# Uniform distribution

- 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

$Q = F^{-1}$
$x = Q(t)$




### Theorem (universality of verifying the uniform)

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

$\exists$ unique $x$ : $F(x) = t$

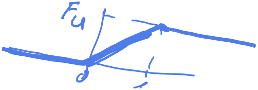
$\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{F} \mathbb{R}\ [0,1]$

Proof $P(F(X) \le t) = P(X \le x) = F(x) = t$

$t \in (0,1)$ $\quad$ F is increasing

CDF of $F(X)$ $\quad$ $F(X) = F \circ X$

$P(F(X) \le 0) = 0$
$P(F(X) \le 1) = 1$




$\Rightarrow F(X) \sim U(0,1)$

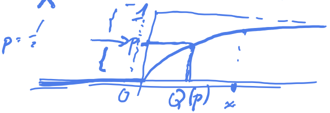
### Theorem (*)

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

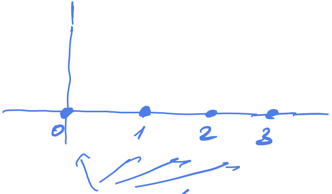
*Let $U \sim U(0,1)$ and $X = Q(U)$. Then $X$ has CDF $F$.*

(universality of uniform generator)

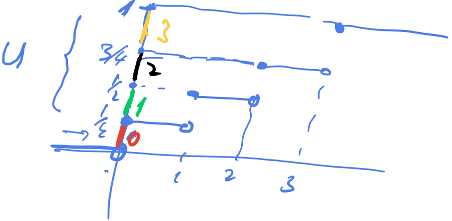
Proof $F_X(x) = P(Q(U) \le x) = P(U \le F(x)) = F_U(F(x)) = F(x)$




$Q(\frac{1}{2})$ = median




X prob. 1/4




$$Q(p) := \min\{x : F(x) \geq p\}$$

$p \in (\frac{1}{4}, \frac{1}{2}]$ --- $[1, \infty)$ $\quad Q(p) = 1$

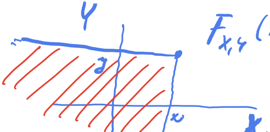
$$Q(p) \leq x \iff p \leq F(x)$$

$F(x) = 1 - e^{-\lambda x} = p$ $\qquad x = Q(p) = \dfrac{\log(1-p)}{-\lambda}$

$U \sim U(0,1)$ $\qquad Q(U) = \dfrac{\log(1-U)}{-\lambda} \sim Exp(\lambda)$

# Overview

# Joint cdf




$F_{X,Y}(x,y) = P($


## Definition

*For r.v. $X, Y$ on probability space $(\Omega, \mathcal{F}, P)$ we define their joint cdf $F_{X,Y} : \mathbb{R}^2 \to [0,1]$ by*

$$F_{X,Y}(x,y) = P(\underbrace{\{\omega \in \Omega : X(\omega) \le x \,\&\, Y(\omega) \le y\}}_{\parallel}).$$

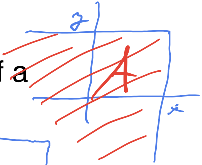
- Formal condition: we need $\{X \le x \,\&\, Y \le y\} \in \mathcal{F}$, otherwise $(X, Y)$ is not a random vector.
- We can define this also for more than two r.v.: $F_{X_1,\dots,X_n}(x_1,\dots,x_n) = P(X_1 \le x_1 \,\&\, X_2 \le x_2 \,\&\dots\&\, X_n \le x_n)$
- From here we can derive the probability of a rectangle: $P(X \in (a,b] \,\&\, Y \in (c,d]) = \boxed{F(b,d) - F(a,d) - F(b,c) + F(a,c)}$

$\boxed{\text{Proof}}$

$= P(A \cup B \cup C \cup D) - P(B \cup D) - P(A \cup D) + P(D)$

$= \cancel{P(A)} + \cancel{P(B)} + \boxed{P(C)} + \cancel{P(D)} - \cancel{P(D)} + \cancel{P(D)}$

# Joint pdf




- Often we can write a joint cdf as an integral of a nonnegative function $f_{X,Y}$

$P(X,Y \in A) =$

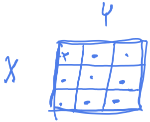
$$\boxed{F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\cancel{dsdt}.}$$

$dt\,ds$

- Then we call r.v. $X$, $Y$ *jointly continuous*. Function $f_{X,Y}$ is their *joint pdf*.
- As in the one-dimension case we can have $f_{X,Y} > 1$.
- As in the one-dimension case we can use joint pdf to find other probabilities for a "reasonable set $A$".

$$P((X,Y) \in A) = \int_A f_{X,Y}(x,y)dxdy.$$

$$\int_{X\times Y} f(x,y) = \int_X\int_Y f(x,y)\,dy\,dx = \int_Y\int_X f(x,y)\,dx\,dy$$





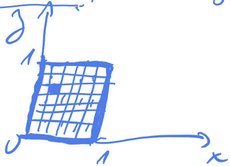
$$\int_0^1\left(\int_0^1 xy^2\,dx\right)dy = \int_0^1 \left[\frac{x^2}{2}y^2\right]_0^1 dy$$

$$\int_0^1 \frac{1}{2}y^2\,dy = \left[\frac{y^3}{2\cdot 3}\right]_0^1 = \frac{1}{6}$$

$$\int_0^1\int_0^1 xy^2\,dy\,dx = \int_0^1 \left[x\,\frac{y^3}{3}\right]_0^1 dx$$

$$= \int_0^1 \frac{x}{3} = \left[\frac{x^2}{2\cdot 3}\right]_0^1 = \frac{1}{6}$$

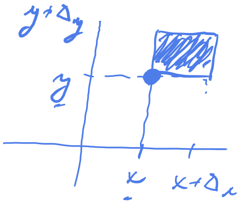
$X = Y = [0,1]$



- $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$ → $F_{X,Y}(x,y)$ differ. by x then by y

partial derivative

- $f_{X,Y}(x,y) \doteq \frac{P(x \leq X \leq x+\Delta_x \;\&\; y \leq Y \leq y+\Delta_y)}{\Delta_x \Delta_y}$




$$P(\blacksquare) = \int_x^{x+\Delta_x} \int_y^{y+\Delta_y} f_{X,Y}(s,t)\, dt\, ds \doteq f_{X,Y}(x,y) \cdot \Delta_x \Delta_y$$

IF $f_{X,Y}$ is nice (continuous) then should be true for small $\Delta_x, \Delta_y$

$\Delta_x, \Delta_y \doteq 0$

# LOTUS

$\sum_{x,y} g(x,y) P(X=x, Y=y)$

$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)$ ? LATER

- We have a similar formula as for the discrete case:

Thm

$$\mathbb{E}(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy.$$

- And as in the discrete case we conclude: LINEARITY OF EXP.

$g(x,y) = ax+by+c$

$$\boxed{\mathbb{E}(aX+bY+c) = a\cdot\mathbb{E}(X) + b\cdot\mathbb{E}(Y) + c.}$$

$\mathbb{E}(g(X,Y)) = \iint g(x,y) f(x,y) = \iint (ax+by+c) f(x,y)$

$= \iint ax\, f(x,y) + \iint by\, f(x,y) + c \underbrace{\iint f(x,y)}_{=1}$

$a \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y)\,dy\right) dx \overset{?}{=} a \int_{-\infty}^{\infty} x\, f_X(x)\,dx = a\,\mathbb{E}(X)$

# Independence of continuous random variables

$(\{X=x\}$ & $\{Y=y\}$ for discrete r.v.$)$

### Definition

*We call random variables $X$, $Y$ independent, if the events $\{X \le x\}$ and $\{Y \le y\}$ are independent for any $x, y \in \mathbb{R}$. Equivalently,*

$$\boxed{\begin{aligned} P(X \le x, Y \le y) &= P(X \le x)P(Y \le y), \\ F_{X,Y}(x,y) &= F_X(x)F_Y(y) \end{aligned}}$$

$\boxed{\Uparrow}$ $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \underbrace{f_{X,Y}(s,t)}_{f_X(s)\cdot f_Y(t)}\, dt\, ds = \int_{-\infty}^{x} f_X(s) \cdot \int_{-\infty}^{y} f_Y(t) = F_X(x) F_Y(y)$

### Theorem

*Let $X$, $Y$ have joint pdf $f_{X,Y}$ (and pdf's $f_X$, $f_Y$). The following are equivalent:*

$\Uparrow \Downarrow$

- $X$, $Y$ are independent
- $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

$\boxed{\Downarrow}$ $F_{X,Y}(x,y) = F_X(x)\,F_Y(y)$

diff/by $x$ & $y$

$f_{X,Y}(x,y)$ $\qquad$ $f_X(x) \cdot f_Y(y)$

# Multidimensional normal distribution

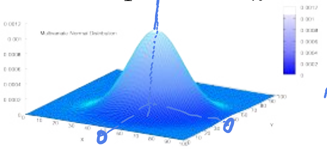
independence

- $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ pdf of $N(0,1)$
- $f(t_1, \dots, t_n) \overset{(*)}{=} \varphi(t_1)\varphi(t_2)\cdots\varphi(t_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{t_1^2+\cdots+t_n^2}{2}}$
- $f(t_1, \dots, t_n) = (2\pi)^{-n/2} e^{-r^2/2}$, where $r^2 = t_1^2 + \cdots + t_n^2$

  $\vec{0}$ *radially symmetric function*

  $r =$ dist. from $\vec{0}$ to $(t_1, \dots, t_n)$

$\int_{\mathbb{R}^n} f = 1$



$n=2$

Image by Wikipedia editor Piotrg.

- Let $Z = (Z_1, \dots, Z_n)$ have a pdf $f$.
- $(*) \Rightarrow$ $Z_1, \dots, Z_n$ are i.i.d., $Z_i \sim N(0,1)$ (pdf of $Z_i$)

  independent

  (if $Z \neq \vec{0}$)
- $\boxed{Z/\|Z\|}$ is a uniformly random point on a unit sphere in $\mathbb{R}^n$.
- Thus the inner product of $Z$ with any unit vector is $N(0,1)$.
- $\langle u, Z\rangle = \sum_{i=1}^n u_i Z_i$ follows $N(0,1)$

  robustness of NORMAL FOR SUM

  we can take basis $\|u\|=1$ starting with $u$

# General multidimensional normal distribution

- In general we can take a random vector with joint pdf $c \cdot e^{-Q(t)}$, where $c > 0$ is an appropriate constant and $Q(t)$ is a positive definite quadratic function.
- Is used in machine learning.
- Coordinates are not independent!

$$Q(t_1 \dots t_n) = \frac{[t_1^2 + \dots + t_n^2]}{2}$$

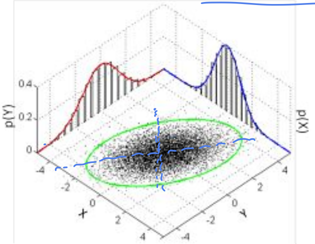



Image by Wikipedia editor Bscan.

# Sum of continuous random variables

### Theorem

*Suppose* $X, Y$ *are independent continuous variables. Then* $Z = X + Y$ *is a continuous random variable and its pdf is obtained by a convolution of* $f_X$ *and* $f_Y$*. Explicitly,*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx.$$

# Conditioning

## Definition

*$X$ is a r.v. on $(\Omega, \mathcal{F}, P)$, $B \in \mathcal{F}$.*

$$F_{X|B}(x) := P(X \le x \mid B)$$

*The corresponding pdf is denoted by $f_{X|B}$.*

## Theorem

*Let $B_1$, $B_2$, $\ldots$ be a partition of $\Omega$. Then*

$$F_X(x) = \sum_i F_{X|B_i} P(B_i) \quad \textit{and}$$

$$f_X(x) = \sum_i f_{X|B_i} P(B_i).$$

Proof: Theorem on total probability.

# Overview

Continuous distributions

Random vectors

Back to the basics

# Covariance

## Definition

*For r.v.'s $X$, $Y$ we define their covariance by formula*

$$cov(X,Y) = \mathbb{E}\big((X - \mathbb{E}X)(Y - \mathbb{E}Y)\big).$$

## Theorem

$$cov(X,Y) = \mathbb{E}\big(XY\big) - \mathbb{E}\big(X\big)\mathbb{E}\big(Y\big)$$

- $var(X) = cov(X,X)$
- $cov(X, aY + bZ + c) = a\,cov(X,Y) + b\,cov(X,Z)$
- $cov(X,Y) = 0$ if $X$,$Y$ are independent
- but not only then

# Correlation

## Definition

*Correlation of random variables $X$, $Y$ is defined by*

$$\varrho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)\,var(Y)}}.$$

- "scaled covariance"
- $-1 \leq \varrho(X,Y) \leq 1$ (exercise)
- Correlation does not imply causation! (In particular, correlation is symmetric.)
- OTOH, uncorrelation does not imply independence. (Extreme case: $X$ any r.v., $Y = +X$ or $Y = -X$, both with the same probability.)

# Variance of a sum

## Theorem

*Let* $X = \sum_{i=1}^n X_i$. *Then*

$$var(X) = \sum_{i=1}^n \sum_{j=1}^n cov(X_i, X_j) = \sum_{i=1}^n var(X_i) + \sum_{i \neq j} cov(X_i, X_j).$$

*In particular, if* $X_1, \ldots, X_n$ *are independent, then*

$$var(X) = \sum_{i=1}^n var(X_i).$$