NMAI059 Probability and statistics 1
Class 8

Robert Šámal

Overview

Continuous distributions

Random vectors

Back to the basics

Which distributions we have seen

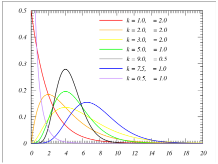
- $U(a, b)$ – uniform on interval $[a, b]$
- $Exp(\lambda)$ – exponential – how long till something happens
- $N(\mu, \sigma^2)$ – normal – how much does a bread weigh

Gamma distribution

- *$Gamma(w, \lambda)$, 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.
- 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_k^2) 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

Theorem

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

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.

Overview

Continuous distributions

Random vectors

Back to the basics

Joint cdf

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

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

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

Joint pdf

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

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

- 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".

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

- $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$

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

LOTUS

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

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

- And as in the discrete case we conclude:

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

Independence of continuous random variables

Definition

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

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$
$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Theorem

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

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

Multidimensional normal distribution

- $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$
- $f(t_1, \ldots, t_n) = \varphi(t_1)\varphi(t_2)\cdots\varphi(t_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{t_1^2 + \cdots + t_n^2}{2}}$
- $f(t_1, \ldots, t_n) = (2\pi)^{-n/2} e^{-r^2/2}$, where $r^2 = t_1^2 + \cdots + t_n^2$
  radially symmetric function

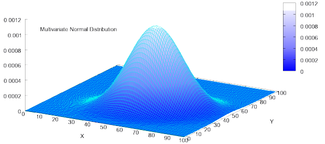


Image by Wikipedia editor Piotrg.

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

- $Z/\|Z\|$ 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 $N(0, 1)$

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!

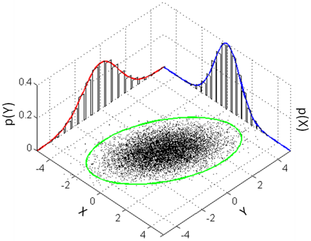


Image by Wikipedia editor Bscan.

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 $(\Omega, \mathcal{F}, P)$, $B \in \mathcal{F}$.

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

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

Theorem

Let B_1, B_2, $\dots$ be a partition of Ω. 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*

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