Theory of probability and mathematical statistics

Tomáš Mrkvička

Chapter 1

Probability theory

1.1 Definition of probability

Definition 1.1 Classical definition: *Let $A_1, \ldots, A_n$ be random events, such that*

- *every time one and only one random event happen,*
- *all the event are equally probable.*

And let the event A happen if happen one of the event $A_{i_1}, \ldots, A_{i_k}$. Then the probability of A is $P(A) = k/n$.

1.1.1 Kolmogorov definition of probability

Probability space $(\Omega, \mathcal{A}, P)$.
Ω is nonempty set of all results of the random experiment, results are ω.
$\mathcal{A}$ is σ-algebra on Ω. (The set of all "nice" subsets of Ω)
$P : \mathcal{A} \to \mathbb{R}$ is function giving every set $A \in \mathcal{A}$ its probability ($0 \leq P(A) \leq 1$, $P(\Omega) = 1$). This is probability measure.

Rules

$$P(A^C) = 1 - P(A)$$

$$P(A \cup B) = P(A) + P(B), \text{if A, B are disjoint}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Examples

1. A phone company found that 75% of customers want text messaging, 80% photo capabilities and 65% both. What is the probability that customer will want at least one of these?
2. What is the probability, that there exists two students, in a class with n students, who have the birth dates in a same day.
3. Two boats, which uses same harbor, can enter whenever during 24 hours. They enter independently. The first boat stays 1 hour, second 2 hours. What is the probability, that no boat will has to wait to enter the harbor.
4. Student choose 3 questions from 30 during an exam. There are 10 questions from algebra, 15 from analysis and 5 from 5 geometry. What is the probability, that he choose at least two questions from the same area.

1.2 Conditional probability

Definition 1.2 *Let $(\Omega, \mathcal{A}, P)$ is probability space and A, B are random events, where $P(B) > 0$. We define the conditional probability of A under the condition B by relation*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1.1}$$

1.3 Independence

Consider now two random events A and B. If the following holds

$$P(A|B) = P(A) \quad a \quad P(B|A) = P(B), \tag{1.2}$$

then we speak about its independence. From (**??**) wee see, that the probability of A under the condition B do not depend on B and vica versa. From (**??**) and definition of conditional probability we have the following definition of the independence.

Definition 1.3 *Random events A and B are independent, if*

$$P(A \cap B) = P(A) \cdot P(B). \tag{1.3}$$

Theorem 1.1 *Let $A_1, \ldots, A_n$ are independent. Then*

$$P(\cup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} [1 - P(A_i)]. \tag{1.4}$$

Rule: When A and B are not independent, one can use:

$$P(A \cap B) = P(A|B)P(B).$$

Examples

1. A phone company found that 75% of customers want text messaging, 80% photo capabilities and 65% both. What are the probabilities that a person who wants text messaging also wants photo capabilities and that a person who wants photo capabilities also wants text messaging?

2. The players A and B throw a coin and they alternate. A starts, then B, then A, ... The game ends when the first one obtain head on the coin. What is the probability of winning of A and B.

3. The probability that one seed grow up is 0.2. You have 10 seeds. What is the probability that exactly 1 seed grow up. (2 seeds, ...) What is the most probable result?

4. The probability that one seed grow up is 0.2. How many seeds you have to use, to have 99% probability, that at least 1 seed will grow up.

1.4 Bayes theorem

Theorem 1.2 (Total probability) *Let A_1, A_2, ... are random events which partition Ω, i.e.*

$$A_i \cap A_j = \emptyset, \ \forall i \neq j \ \ a \ \ \cup_{i=1}^{\infty} A_i = \Omega.$$

These random events have the probabilities $P(A_1), P(A_2), \ldots$, where $P(A_i) > 0, \ \forall i = 1, 2, \ldots$ Assume an event B, where we know the conditional probabilities

$$P(B|A_i), \ \forall i = 1, 2, \ldots$$

Then

$$P(B) = \sum_{i=1}^{\infty} P(A_i) \cdot P(B|A_i). \tag{1.5}$$

Theorem 1.3 (Bayes theorem) *Assume the same assumption as in Theorem* **??**. *Then*

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(A_j) \cdot P(B|A_j)}, \quad i = 1, 2, \ldots \tag{1.6}$$

Example 1.1 *During examination of a patient there is suspicion on 3 different ilnesses. Probability of accurence of the first ilness is 0,3, second 0,5 and third 0,2. Laboratory gives positive result for 15% of people having first ilness, for 30% of people having second ilness and for 30% of people having third ilness. What is the probability that a patient has second ilness under the condition of the positive test?*

Remark 1.1 *Probabilities* $P(A_1), P(A_2), \ldots$ *are called apriory and events* A_1, A_2 *are called hypothesis. Probabilities* $P(A_i|B)$ *are called aposteriory.*

Example 1.2 (AIDS). *Two posible errors in the blood test*

1. *Test falsly indicate positivity,*
2. *Test falsly indicate negativity.*

$P(Pos|Inf) = 0,995$, *thus, probability of the error of the first kind is* $P(Neg|Inf$
$0,005$.

$P(Neg|NeInf) = 0,995$ *thus, probability of the error of the second kind is* $P(Poz|NeInf) = 0,005$.

$P(Inf|Poz)$?

$$P(Inf|Poz) = \frac{P(Poz|Inf) \cdot P(Inf)}{P(Poz|Inf) \cdot P(Inf) + P(Poz|NeInf) \cdot P(NeInf)}.$$

Apriory probabilities:

$$P(Inf) = 0,001 \ a \ P(NeInf) = 0,999.$$

Aposteriory probabilities:

$$P(Inf|Poz) = \frac{0,995 \cdot 0,001}{0,995 \cdot 0,001 + 0,005 \cdot 0,999} = 0,16.$$

Examples

1. The test is made in four classes. 75% of the students pass in the first class (50 students). 50% of the students pass in the second class (35 students). 65% of the students pass in the third class (40 students). 60% of the students pass in the first class (30 students). What is a probability that a random student pass the test?

2. There are 36 students in the class. 15 boys, 21 girls. 5 boys and 7 girls are interested in economics. What is a probability that random child is interested in economics?

3. Compute the probability in exercise 1) that a student is from the first class under the condition that he/she passes the test.

4. The messages emited by Morse alphabet have the following statistics: If dot is emited then in 1/10 cases comma is observed. If comma is emited then in 1/15 cases dot is observed. The dots and commas are emited with ratio 5:3. What is a probability that emited was dot under the condition that observed

is comma? What is a probability that emited was comma under the condition that observed is dot?

1.5 Random variables

Definition 1.4 Random variable *is every measurable mapping* X *from* $(\Omega, \mathcal{A}, P)$ *to* $\mathbb{R}$*.*

Definition 1.5 Distribution function F *of a random variable* X *is given by*

$$F(x) = P(\omega : X(\omega) \leq x).$$

Shortly

$$F(x) = P(X \leq x).$$

Discrete random variables

Random variable X can have maximally countably many values $x_1, x_2, \ldots$.
$P(X = x_i) = p_i$, $i = 1, 2, \ldots$. $\sum p_i = 1$.
Formulas:

$$P(X \in B) = \sum_{i: x_i \in B} p_i.$$

Expectation X

$$\mathbb{E}X = \sum_i x_i p_i. \tag{1.7}$$

Expectation of a function of X

$$\mathbb{E}g(X) = \sum_i g(x_i) p_i. \tag{1.8}$$

Basic distributions of discrete random variables

Alternativ distribution A(p) represents success/unsuccess of the experiment $0 < p < 1$.

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

$$\mathbb{E}X = p, \quad \mathrm{Var}(X) = p(1 - p).$$

Binomial distribution Bi(n, p) represents number of successes in n independent experiments. The probability of success is $0 < p < 1$. In other words, binomial distribution is sum of n independent alternative distributions.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

$$\mathbb{E}X = np, \quad \mathrm{Var}(X) = np(1 - p).$$

Hypergeometric distribution HGeom(n, M, N) is used instead of Binomial in experiments, where n represents number of draws without returning (Binomial - n is with returning.) from box which has N elements, and M elements represent success. (Binomial $M/N = p$.) Hypergeometrical distribution then represents number of success in this experiment.

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, n.$$

$$\mathbb{E}X = n\frac{M}{N}, \quad \mathrm{Var}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

Poisson distribution Po(λ) $\lambda > 0$ represents number of events which appear in time of length t.

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

$$\mathbb{E}X = \lambda, \quad \mathrm{Var}(X) = \lambda.$$

Geometrical distribution Geom(p) represents number of experiment until first success appears. The probability of success is $0 < p < 1$.

$$P(X = k) = p(1-p)^k.$$

$$\mathbb{E}X = \frac{1-p}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Examples

1. The phone central connects 15 talks during one hour in average. What is the probability that during 4 minutes it connects: a) exactly one talk, b) at least one talk, c) at least two talks and in maximum 5 talks.
2. The man phone to phone central during the maximal load, when there is probability 0.25 to be connected. He repeat the experiments until he is successful. Compute the probability of being connected during 5 try. Compute the expectation of unsuccessful tries.
3. Compute the probability that among 136 products are at least 3 broken. We know that there is 2.6% of broken products.
4. The restaurant gives to each meal a picture of the local basketball team (basic set 5 players). Every time, you go in the restaurant, you obtain one picture in random. How many times in average you have to go in the restaurant to have whole set of pictures?

Continuous random variables

Random variable X can have uncountably many values.
Density $f(x)$ gives relative probability of occurrence x as a result of X.
Distribution function

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

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

Formulas:

$$P(X \in B) = \int_B f(x)dx.$$

Expectation X

$$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx. \tag{1.9}$$

Expectation of a function of X

$$\mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x)f(x)dx. \tag{1.10}$$

Variance X (basic characteristic of dispersion of random variable) is given by

$$\mathrm{Var}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Variance is denoted also by σ^2 and σ is then called standard deviation.

Theorem 1.4 *Let $Y = a + bX$. If $\mathbb{E}X$ exists, then $\mathbb{E}Y = a + b\mathbb{E}X$. Furthermore if $\mathbb{E}X^2 < \infty$, then $\mathrm{Var}Y = b^2\mathrm{Var}X$.*

Basic distributions of continuous random variables

Uniform distribution on $[A, B]$, U$[A, B]$. All points in the interval $[A, B]$ have the same probability of occurrence.

$$f(x) = \frac{1}{B - A}, \quad \text{pro } x \in [A, B], \quad f(x) = 0, \quad \text{jinak}.$$

$$\mathbb{E}X = \frac{A + B}{2}, \quad \operatorname{Var}(X) = \frac{1}{12}(B - A)^2.$$

Exponential distribution Exp(λ) represents waiting time until certain event appear, for example time until engine breaks down.

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda}, \quad \text{pro } x > 0, \quad f(x) = 0, \quad \text{jinak}.$$

$$\mathbb{E}X = \lambda, \quad \operatorname{Var}(X) = \lambda^2.$$

1.5.1 Normal distribution

Normal distribution with expectation μ and variance σ^2 has density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

Figure 1.1: The plot of normal density - solid N(0,1), dashed N(0,2), dotted N(0,1/2).

The most important is normalised normal distribution N(0,1). Its density is

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad x \in \mathbb{R}$$

and its distribution function is denoted by

$$\Phi(x) = \int_{-\infty}^{x} \phi(u)du.$$

Function ϕ is even, thus $\Phi(-x) = 1 - \Phi(x)$.

Examples

1. The durability of a product follow exponential distribution with expected value 3 years. How long should be the guarantee to have 10% of the product returned during the guarantee time.
2. What is the probability, that a random variable U with distribution N(0,1) is a) less than 1.64, b) more than -1.64, c) between -1.96 and 1.96, d) more than 2.33, e) less than -2.33.
3. The error of measurement follows N(0,σ^2). How big is σ^2, if we know, that the probability, that the absolute value of the error is smaller than 1, is equal to 0.95.
4. Compute the expectation and the variance for the result of a classical dice.

Theorem 1.5 Central limit theorem *Let $X_1, \ldots, X_n$ is progeression of independent identically distributed random variables with expectation μ and with finite variance σ^2. Then*

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$$

has for $n \to \infty$ asymptotically distribution N(0,1).

Examples

1. The probability that the device breaks down during the test of device reliability is 0.05. What is the probability that during testing of 1000 devices there will be more than 75 devices broken down. 1) Use CLT. 2) Use Binomial distribution.

2. We throw 100 times by a die. Denote S_{100} the sum of these 100 results. Calculate the probability
$$P(320 < S_{100} < 380).$$

3. The insurance company insures 1000 people of the same age. The probability of death in this age is 0.01. In the case of death the company pay out 80000 CZK. What price of the insurance should be set in order to have 99% that the company do not loose any money.

4. The time to repair one electricity issue on a electric suply has exponential distribution with mean 4 hours. There was a big storm and 30 issue appeared on the supply. What is the time which guarantee with 95% that all the issue will be repaired?

1.6 Random vectors

If random variables $X_1, \dots, X_n$ are defined on the same probability space $(\Omega, \mathcal{A}, P)$, then

$$\mathbf{X} = (X_1, \dots, X_n)^T$$

is called random vector.
Distribution function of random vector is

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Expectation is $\mathbb{E}\mathbf{X} = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)^T$.

We will work now only with two random variables X, Y. For more random variables it is analogous.

Discrete case: Common probability distribution is given by

$$P(\mathbf{X} = (x_i, y_j)) = p_{ij}, \quad i = 1, 2, \dots, j = 1, 2, \dots.$$

Marginal distribution is distribution of the part of the vector. Let's

$$p_i = \sum_j p_{ij}, \quad p_j = \sum_i p_{ij}.$$

Thus the marginal distr. are:

$$P(X = x_i) = p_i, \quad P(Y = y_j) = p_j, \quad i = 1, 2, \ldots, j = 1, 2, \ldots.$$

Expected value of a function of the random vector is given by

$$\mathbb{E}g(\mathbf{X}) = \sum_{i,j} g(x_i, y_j) p_{ij}.$$

Continuous case: Common probability distribution is given by density

$$f_{\mathbf{X}}(x, y), \quad x, y \in \mathbb{R}.$$

Distribution function

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv.$$

Densities of marginal distributions are

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy, \quad x \in \mathbb{R}, \quad f_Y(y) = \int_{\mathbb{R}} f(x,y)dx, \quad y \in \mathbb{R}.$$

Expected value of a function of the random vector is given by

$$\mathbb{E}g(\mathbf{X}) = \int_{\mathbb{R}} \int_{\mathbb{R}} g(x,y)f(x,y)dxdy.$$

Covariance of random variables X and Y is

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y.$$

It is clear, that $\mathrm{Var}X = \mathrm{Cov}(X,X)$. Covariance of random variables X and Y is denoted by σ_{XY}.

Theorem 1.6

$$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y,$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}X + 2\mathrm{Cov}(X,Y) + \mathrm{Var}Y,$$

if all expressions on the right hand side exist.

We say that two random variables X and Y are **independent**, if

$$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$

In continuous case it is equivalent to

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \forall\ x, y.$$

In discrete case it is equivalent to

$$p_{ij} = p_i p_j \quad \forall\ i, j.$$

This is a mathematical definition of the term independence which is commonly used in speech.

Theorem 1.7 *Let X and Y are independent random variables with finite expectations. Then*

$$\mathbb{E}(XY) = (\mathbb{E}X)(\mathbb{E}Y).$$

Theorem 1.8 *Let X and Y are independent random variables with finite variances. Then*

$$\mathrm{Cov}(X,Y) = 0.$$

If $\mathrm{Cov}(X, Y) = 0$, then we say that random variables are **uncorrelated**. The uncorrelateness does not imply independence! But the previous theorem is often used in tests of independence. Instead of covariance, there is used its normalized version **correlation coefficient:**

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Theorem 1.9 *It holds $-1 \leq \rho \leq 1$. Furthermore $\rho = 1$, iff $Y = a + bX$, $b > 0$. A $\rho = -1$, iff $Y = a + bX$, $b < 0$.*

Examples

1. The bank company runs two products both with expected profit μ and variance of the profit σ^2. Compute the expected profit of the company and variance of the profit of the company if a) the products are independent, b) the products are positively correlated with $\rho = 1/2$ and c) the products are negatively correlated with $\rho = -1/2$.
2. Compute the expectation and the variance of the sum of results on 10 independent dice.
3. The box contains 2 white balls and 2 black balls, we choose 2 balls without returning. The random variable X is 1 if the first ball is white and 0 otherwise. The random variable Y is 1 if the second ball is white and 0 otherwise. Determine common distribution of (X, Y), compute marginal distributions and determine if the random variables are independent.
4. The same experiment as above. Determine common distribution of (X, Y) and compute $\mathbb{E}X, \mathbb{E}Y, \mathrm{Var}X, \mathrm{Var}Y, \rho$.

1.7 Distributions derived from normal distribution

Gama and Beta function.

$$\Gamma(a) = \int_0^\infty x^{a-1} \cdot e^{-x} dx, \quad a > 0$$

Properties: $\Gamma(a+1) = a \cdot \Gamma(a)$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$B(a,b) = \frac{\Gamma(a) \cdot \Gamma(b)}{\Gamma(a+b)}$$

1.7.1 Pearson distribution

Let U_1, U_2,..., U_k are i.i.d. with N(0,1). Then

$$\chi_k^2 = \sum_{i=1}^{k} U_i^2$$

has χ^2 with k degrees of freedom. The density (for $u > 0$) is

$$f_k(u) = \frac{1}{\Gamma(k/2) \cdot 2^{k/2}} \cdot u^{(k/2)-1} \cdot e^{-u/2}, \quad u > 0.$$

$$\mathbb{E}\chi_k^2 = k, \qquad \text{Var } \chi_k^2 = 2k.$$

Figure 1.2: The plot of the density of Pearson distribution - solid χ^2_{10}, dashed χ^2_{20}, dotted χ^2_5.

1.7.2 Student distribution

U is N(0,1) random variable
V is χ^2 r. v. with k degree of freedom.
U, V are independent, then

$$T_k = \frac{U}{\sqrt{V}} \cdot \sqrt{k}$$

has Student distribution t with density

$$f_k(t) = \frac{1}{B(\frac{1}{2}, \frac{k}{2}) \cdot \sqrt{k}} \cdot (1 + \frac{t^2}{k})^{-(k+1)/2}, \quad t \in \mathbb{R}$$

with k degree of freedom.

$$\mathbb{E}T_k = 0, \qquad \text{Var } T_k = \frac{k}{k-2}, \qquad t_k \to_{k\to\infty} \Phi.$$

Figure 1.3: The density plot of Student distribution - solid N(0,1), dashed t_{10}, dotted t_5.

1.7.3 Fisher-Snedecor distribution.

U is χ^2 with k degree of freedom,
V is χ^2 with n degree of freedom.
U, V are independent, then

$$F_{k,n} = \frac{U/k}{V/n}$$

has Fisher-Snedecor distribution with k and n degreeof freedom with density

$$f_{k,n}(z) = \frac{1}{B(\frac{k}{2}, \frac{n}{2})} \cdot \left(\frac{k}{n}\right)^{k/2} \cdot \frac{z^{(k-2)/2}}{(1 + z \cdot \frac{k}{n})^{(k+n)/2}}, \quad z > 0.$$

$$\mathbb{E}F_{k,n} = \frac{n}{n-2}, \qquad \text{Var } F_{k,n} = \frac{2n^2(n+k-2)}{(n-2)^2(n-4)k}.$$

Figure 1.4: The density plot of Fisher-Snedecor distribution - solid $F_{10,10}$, dashed $F_{20,10}$, dotted $F_{5,10}$.

1.8 Critical values

Critical value is the border which random variable exceed with a given probability α. The Excel functions are NORM.INV, CHI.INV, T.INV, F.INV.

Critical value of normal distribution $u(\alpha)$

$$X \sim \mathrm{N}(0,1), \qquad P[X \leq u(\alpha)] = \alpha.$$

Critical value of Pearson distribution $\chi^2_k(\alpha)$

$$X \sim \chi^2_k, \qquad P[X \leq \chi^2_k(\alpha)] = \alpha.$$

Critical value of Student distribution $t_k(\alpha)$

$$X \sim t_k, \qquad P[X \leq t_k(\alpha)] = \alpha.$$

Critical value of Fisher-Snedecor distribution $F_{k,n}(\alpha)$

$$X \sim F_{k,n}, \qquad P[X \leq F_{k,n}(\alpha)] = \alpha.$$