Definitions:

L2P1

- A finite set $\Omega \neq \varnothing$ called ***space*** (***sample space***)/**пространство элементарных событий (исходов)** and it's elements are ***elementary events***, (*samples*, *choices*, *options*, etc).

- All subsets of $\boldsymbol{\Omega}$ are called ***events***.

- The empty set $\varnothing$ is called *the* ***impossible event***;

- The space Ω is called *the* ***certain/sure event***;

- Disjoint events (i.e. $A \cap B = \varnothing$) are called (*mutually*) ***exclusive***.

- **P($\boldsymbol{\Omega}$)** or $\mathbf{2}^{\Omega}$ is the power-set (the set of all events (in finite case!))

L2P2

- **P**: $2^{\Omega} \to [0,1]$ – **probability/вероятность** (probability function) is a non-negative additive function P (measure) to all events and satisfying normalization condition $P(\Omega)=1$.

- A (finite) **probability space/вероятностное пространство** is a triple (Ω, F, P), where
  Ω is a finite event/sample space,
  $\mathrm{F} = 2^{\Omega}$ is the set of events,
  P: $\mathrm{F} \to [0,1]$ a (total) probability function.

- A ***probability space*** is a sample space Ω together with a set of events $\mathrm{F} \subseteq 2^{\Omega}$ (that must be a σ-algebra) and a non-negative additive probability function P to all events and satisfying normalization condition $P(\Omega)=1$. (week 10)

L2P3

- **!n** is the number of derangements (also known as the ***subfactorial***);
  !n=(n-1)(!(n-1)+!(n-2))
  with initial values !0= 1 and !1 = 0.

L3P2

- $\mathbf{A_n^k}$ – ***arrangement/*****размещение *(число размещений)*** of k elements from a set of n elements is any sequence (vector) of k different elements from the set;

  $\mathbf{A_n^k} = n*(n-1)*\ldots*(n-k) = n!/(n-k)!$

- $\mathbf{P_n}$ – ***permutation/перестановка (число перестановок)*** is an "arrangement of n form n",

  $\mathbf{P_n} = n!, \quad 0!=1$

- $\mathbf{C_n^k}$ – ***k-combination/сочетание (число сочетаний)*** of a set that has n elements is a subset of k distinct elements of the set; $\mathbf{C_n^k}$ also denoted as $(_n^k)$;

  $\mathbf{C_n^k} = n!/(k!(n-k)!)$

L4P1

- Given two events A and B of a probability space with P(B)>0. **P(A|B)** called ***conditional probability/условная вероятность*** of A given/assuming B is defined as the quotient of the probability of the product $A \cap B$, and the probability of B: $\mathbf{P(A|B)} = P(A \cap B)/P(B)$

L5P1

- ***Probability Tree*** (decision tree) is the obvious scheme (method) for calculations of the conditional probabilities.

L5P2

- Events A and B are ***independent*** if P(A) = P(A|B) and/or P(B) = P(B|A); otherwise the events are ***dependent***.

- A set of events is ***mutually independent/независимы в совокупности*** if each event in the set is independent with every product of other events in the set.
  Attention! Mutually independent events are **pair-wise/попарно** independent (by definition), but **not vice versa**.

L5P3

- ***Partition/разбиение*** of a sample space Ω is presentation of the space as a sum of mutually exclusive events that are called ***hypothesizes***:

  $\Omega = \cup_{n\in[0..m]}H_n = \Sigma_{n\in[0..m]}H_n$
  where $H_i \cap H_j = \varnothing$ for all $0 \leq i<j \leq m$.

L6P1

- ***Discrete random variable/дискретная случайная величина*** is any (total) real function on finite domain $X:\Omega\to R$.
- In general: a ***random variable*** is any (total) function $X: \Omega\to R$ that range (co-domain) is an interval of real numbers (finite or infinite).

  Notation convention: If X, Y. Z are random variables then x, y, and z are reserved for values of these functions.

- ***(Frequency) distribution/частотное распределение (распределение частот)*** is a table/function that assigns number (i.e. *non-normalized frequency*) $|X^{-}(x)|$ of the corresponding outcomes to each value x of the variable X.

  Here

  X^{-} is the inverse of X as a function,

|…| is the number of elements of a set.

- ***(Discrete) probability distribution/ распределение вероятностей/ probability mass function/функция вероятности*** is a table/function that assigns probability (i.e. *the normalized frequency*)
  $\mathbf{P_X(x)} = P(X=x) = |X^{-}(x)|/|\Omega|$
  of the corresponding outcomes to each value x in the range of the variable X.

L6P2

- Let $X: \Omega \to R$ be a (discrete) random variable is any (total) real function on finite domain.

  It defines probability mass function P_X and ***cumulative distribution function/(интегральная)функция распределения:***
  $\mathbf{F_X(x)} = P_X(X \le x) = \Sigma_{y \le x} P_X(y)$.

- Let $X:\Omega \to R$ be a random variable and $g:R \to R$ be function (that is defined on the range of X at least).
  This function ***g*** defines a ***function of random variable X*** with the following probability distribution $\mathbf{P_{Y=g(X)}(y)} = \Sigma_{g(x)=y} P_X(x)$.

L7P1

- Let $X:\Omega \to R$ be a random variable and P_X be its probability distribution. ***Mean/главное (значение), average/среднее (значение), expectation/(математическое) ожидание***, the ***first moment/первый момент*** of X is defined as
  $$\mathbf{M(X) = E(X)} = \Sigma_{x \in R} x * P_X(x).$$

- *Random variables $X,Y:\Omega \to R$ are said to be* ***independent*** if (X=x) and (Y=y) are independent events for all $x,y \in R$.

- if $a,b \in R$ are a constants, $X,Y:\Omega \to R$ are random variables, then:
  ✓ $\mathbf{M(a) = a}$;
  ✓ $\mathbf{M(a*X) = a*M(X)}$;

- ✓ **M(a*X + b*Y)= a*M(X) + b*M(Y)**.
- ✓ Let g:R→R be function (that is defined on the range of X at least);
- ✓ if Y=g(X) then **M(Y) = $\Sigma_{x\in R}$ g(x)*P_X(x).**
- ✓ if random variables X,Y:Ω→R are independent then
- ✓ **M(X*Y) = M(X)*M(Y)**.

L7P2

- ***Discrete uniform distribution /равномерное распределение*** corresponds to a random variable X that get exactly n values $\{x_1, \ldots, x_n\}$ with a flat probability:
  - ✓ **$P_X(x_k)$**=1/n for all $k\in[1..n]$;
  - ✓ $M(X) = (x_1 + \ldots + x_n)/n$.

- ***Bernoulli distribution/распределение Бернулли***: X = Bernoulli(p), where $p\in[0, 1]$, gets just 2 conventional values 0 (fail) and 1 (success).
  - ✓ **P(Bernoulli(p)=1) = p** and **P(Bernoulli(p)=0) = 1-p=q**;
  - ✓ **M(Bernoulli(p))** = 0*(1-p) + 1*p = **p**;
  - ✓ **D(Bernoulli(p))= p*q**

- ***Bernoulli trials (or binomial experiment)*** consist of some fixed number n of independent Bernoulli trials, each with a probability of success p, and counts the number of successes.
  A corresponding random variable is denoted by ***B(n,p)*** or by ***binomial(n,p)***, and is said to have a ***Binomial distribution)/биномиальное распределение***:
  - ✓ **$P_{B(n,p)}$(k)** = P(B(n,p)=k) = **$C_n^k p^k (1-p)^{(n-k)}$** ;
  - ✓ $B(n,p) = X_1 + \ldots + X_n$ where all $X_1,\ldots X_n$ =Bernoulli(p);
  - ✓ **M(B(n,p))** = **n*p**;
  - ✓ **D(B(n,p))= n*p*(1-p)**

- ***Geometric Distribution/геометрическое распределение:*** the probability distribution of the number k in { 1, 2, 3, ...} of Bernoulli(p) trials needed to get the first success. A corresponding random variable is denoted by **Geom(p)**:
  - ✓ **$P_{Geom(p)}$ (k)** = P(Geom(p)=k) = **$(1-p)^{(k-1)}$*p**;
  - ✓ **M(Geom(p))** = $\Sigma_{k>0}$ $k*p*(1-p)^{(k-1)}$ = **1/p**

- ✓ **D(Geom(p)) = (1-p)/p^2**

Attention! the set of outcomes Ω must be **infinite**!

L7P3

- Let $A \subseteq \Omega$ is an "event" in 2Ω. Then ***conditional (discrete) random variable* X|A** is the restriction of the function X on the domain A.
  It defines ***conditional distribution/условное распределение***
  $P_{X|A}$(x) = P((X|A)=x) = |(X=x) $\cap$ A|/|A| =|X^{-}(x) $\cap$ A|/|A| = P((X=x)|A).
  Corollary: P((X=x) $\cap$ A) = $P_{X|A}$(x)*P(A)

- ***Conditional expectation/условное математическое ожидание*** – expectation of conditional random variable:

$$M(X|A) = \Sigma_{x \in R} x * P_{X|A}(x)$$

L7P4

- ***Memoryless Property/свойство отсутствия памяти.***
  Memorylessness refers to the cases when the distribution of a waiting until a certain event does not depend on time: P(X>m+n|X>m) = P(X>n), m,n ∈ ℕ.
  Only two kinds of distributions are memoryless: ***exponential*** and ***geometric*** distributions.

L8P3

- Let k>0 be an integer, and X be a discrete random variable.
  k-th (initial) moment of X is expectation of X_k:
  M(X_k) = $\Sigma_{x \in R}\, x^k * P_X(x)$;
  k-th central moment of X is expectation of the random variable $[X - E(X)]^k$:
  M($[X - E(X)]^k$) =$\Sigma_{x \in R}\, (x-E(X))^k * P_X(x)$.

- ***Variance/дисперсия*** of a (discrete) random variable X is *its* ***2-nd central moment:***
  D(X) = var(X) = $M([X – E(X)]^2) = M(X^2) – M^2(X)$.
  Properties:
  if a∈R is a constant then

- ✓ D(a)= 0;
- ✓ D(X + a)= D(X);
- ✓ D(a*X)= a^2 *D(X);
- ✓ if X and Y have the same outcomes then
  D(X + Y) = D(X) + D(Y) + 2[M(X*Y) – M(X)*M(Y)];
- ✓ If X and Y are independent random variables than
  D(X + Y) = D(X) + D(Y).

- For any discrete random variable X its ***(standard) deviation/девиация is σ(X) = $D^{1/2}$(X).***

- For any discrete random variable X and Y (with same set of outcomes) their ***covariance/ковариация*** is
  cov(X,Y) = M[(X-E(X))*(Y-E(X))] = **M(X*Y) – M(X)*M(Y)**;
  if X and Y are independent then cov(X,Y)=0.

- For any discrete random variable X and Y (with same set of outcomes) their ***correlation coefficient/коэффициент корреляции*** is
  r(X,Y) = **ρ(X,Y)** = **corr(X,Y)** =cov(X,Y)/σ(X)* σ(Y).
  - ✓ -1≤cor(X,Y) ≤1 for all random variables X;
  - ✓ if X and Y are independent then cor(X,Y)=0 (they are ***uncorrelated***);
  - ✓ if Y= a*X + b (a,b = const) then
    corr(X,Y) = $\begin{cases} 1, \text{if a} > 0 \\ -1, \text{if a} < 0 \end{cases}$

L9P2

- ***Функция Лапласа/ /Gauss error function/функция ошибок (erf)***

$$\Phi_0(x) = \int_0^x \phi_0(t)dt = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt, \quad \text{where}$$

$$\phi_0(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

.

L11P1

- ***Cumulative probability distribution/cumulative distribution function/(интегральная)функция распределения $F_X(x)$ (or $\Phi(x)$)***
  Assume that
  –the domain of a random variable X is the sample space of some probability space with probability function P,
  –and for each x∈R the pre-image of (-∞,x] is an event in this space;
  then **(*cumulative probability*) *distribution function (CDF)*** is the following function $F:R^+\to[0,1]$ defined as $F_X(x) = P(-\infty<X<x)$;
  ✓ $F_X(a)\leq F_X(b)$ for all $a\leq b$ (monotonicity);
  ✓ $F_X(-\infty)=0$ and $F_X(\infty)=1$;
  ✓ $P(a\leq X<b) = F_X(b) - F_X(a)$ for all $a\leq b$.

- ***Continuous random variables/непрерывные случайные величины*** and their distributions.
  If distribution $F_X(x)$ of a random variable X is a continuous function then the distribution $F_X(x)$ is called a ***continuous distribution*** and the variable X is called a ***continuous random variable***.
  If $F_X(x)$ is a continuous distribution than $P(a)= P(x=a)=0$ for all a∈R.

- ***Probability density function/плотность распределения (плотность вероятности)***
  If $F_X(x) = \int_{-\infty}^{x}\varphi(t)dt$ then
  $\varphi(t) = F_X'(x)$ and $\varphi(x)$ is called ***probability density function (PDF)***;
  $P(a\leq X<b) = F_X(b) - F_X(a) = \int_a^b \varphi(t)dt$ for all $a\leq b$.

- Let k>0 be an integer and X be a continuous random variable with distribution $F_X(x)$ and density $\varphi(x)$.
  Then ***k-th (initial) moment*** of X must be defined as

$$M(X^k) = \int_{-\infty}^{\infty} x^k\varphi(x)dx$$

  and ***k-th central moment*** – as

$$M([X - E(X)]^k) = \int_{-\infty}^{\infty} [x - E(X)]^k \varphi(x) dx$$

L11P3

- ***The standard normal distribution/стандартное нормальное распределение*** $\Phi_0(x)$ with expectation 0 and deviation 1 has the probability density function

$$\varphi_0(x) = \phi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

- ***Normal (Gaussian) distribution/*нормальное распределение** with a finite expectation μ and finite non-zero deviation σ has the probability density function

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-{(x-\mu)^2}/{2\sigma^2}}$$

L12P1

- A ***(random) sample/(случайная) выборка*** of length/size n for a (fixed) distribution F is a set of n IID random variables with distribution F.

- A ***(sample) statistic/(выборочная) статистика*** is a (real) function of a sample where the function is independent of the sample's distribution (e.g., its arithmetic mean value). The term statistic used both for the function and for the value of the function on a given sample.

- An ***estimator/оценка*** is a rule to calculate an estimate a parameter (e.g. mean, variance) based on observed data. Estimator example: estimate the second moment as the mean squared in a random point $[(X_1(r))^2 + \ldots + (X_n(r))^2]/n$.

- An estimator of parameter is ***consistent*** if it *converges in probability* (due to use of a random population point) the true value.

- The ***bias/смещение*** of an estimator $T(X_1,...X_n)$ of a parameter θ is the difference $E[T(X_1,...X_n)] - \theta$.

- An estimator is ***unbiased/несмещённая*** if it has 0 bias.

L12P2

- A ***confidence interval/доверительный интервал*** of a ***confidence level* p** ($0 \leq p \leq 1$) for a parameter θ is an interval [L,U] such that $P(L \leq \theta \leq U) = p$. The value $q=1-p$ called

L12P3

- ***Less than by probability***
  Let P be the probability of joint distribution of random variables X and Y; let X<Y by probability, if $P(X<Y) > P(X \leq Y)$ (i.e. $P(X<Y) > ½$).

- ***Stochastically/стохастически less than***
  Let X and Y be random variables; let X<Y stochastically, if $P_X(X<z) \leq P_Y(Y<z)$ for all $z \in R$ and $P_X(X<z) > P_Y(Y<z)$ for some $z \in R$.

L13P1

- ***Pearson's chi-squared test/критерий согласия Пирсона (χ^2)***
  Chi-squared tests a hypothesis that the observed frequency distribution of events is consistent with a particular distribution. It used chi-squared test statistic $\chi^2 = \sum_{i \epsilon [1 \ldots k]} (n_i - n * p_i)^2 / (n * p_i)$ and Pearson's distribution chi-squared with k-1 freedom degrees χ^2_{k-1}; hypothesis holds with some ***significance level (critical value) q/ уровень значимости, вероятность ошибки***.

L13P3

- ***(Statistical) hypothesis***: a statement about the discribution.

Simple hypothesis: any hypothesis which specifies the distribution exactly.

- ✓ *Composite hypothesis*: any hypothesis which does not specify the distribution completely

- ✓ *Null hypothesis* (H_0): usually a simple hypothesis one would like to prove.
- ✓ *Alternative hypothesis* (H_1): a hypothesis (often composite) opposite to the null hypothesis.

- ***Statistical test***: a procedure whose inputs are (accidental) samples and hypothesis and whose result is hypothesis acceptance or refutation of hypothesis.

- ***Region of acceptance***: the set of values of the test statistic for which we fail to reject the null hypothesis.

- ***Region of rejection / Critical region***: the set of values of the test statistic for which the null hypothesis is rejected.

- ***Critical value***: the threshold value delimiting the regions of acceptance and rejection for the test statistic.