Definition 1 *Suppose that (X, Y) is a 2-dimensional absolutely continuous random variable with cumulative density function $f_{X,Y}(x, y)$. The* ***conditional density function of** X **given** Y* *is $f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$, where $f_Y(y)$ is the marginal density, i.e., the density function of random variable Y.*

Remark 1 *Multiplying both sides by $f_Y(y)$ it gives that $f_{X,Y}(x, y) = f_{X|Y}(x|y) \cdot f_Y(y)$, and the density of X is $f_X(x) = \int\limits_{-\infty}^{\infty} f_{X|Y}(x|y) \cdot f_Y(y)dy$, as we have already seen. This is the law of total probability applied to density functions.*

Proposition 1 (Markov type inequalities) *Let $X \geq 0$ be a positive random variable, $g : \mathbb{R}^+ \to \mathbb{R}^+$ monotonically increasing function and $\varepsilon > 0$ a real number. Then $P(X \geq \varepsilon) \leq \frac{E(g(X))}{g(\varepsilon)}$. In special case, if $g(x) = x$, we get* ***Markov's inequality***

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}. \tag{1}$$

Corollary 1 (Chebyshev's inequality) *Suppose that $g(x) = x$, and apply Statement to $(X - E(X))^2$, instead of X. This follows that $P\left((X - EX)^2 \geq \varepsilon^2\right) \leq \frac{E((X-EX)^2)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$.*

Remark 2 *The inequalities above are sharp. It means that for any $\varepsilon > 0$ exists a random variable X, which satisfies the inequality with equality.*

Definition 2 (convergence in probability) *A sequence $X_1, X_2, \ldots$ of random variables* ***converges in probability*** *towards random variable X, if $\forall \varepsilon > 0$ $\lim\limits_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$.*

Theorem 1 (Law of large numbers) *Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables, $\sigma^2 = Var(X_i) < \infty$, and let $m := E(X_i)$ denote their expected values. Then for any ε, δ positive numbers there is an $n_0 \in \mathbb{N}$ that for $n > n_0$:*

$$P\left(\left|\frac{X_1 + \ldots + X_n)}{n} - m\right| \geq \varepsilon\right) \leq \delta.$$

Proof: $\frac{E(X_1+\ldots+X_n)}{n} = m$ and $Var\left(\frac{X_1+\ldots+X_n}{n}\right) = \frac{\sigma^2}{n}$. Using Chebyshev's inequality the following holds: $P\left(\left|\frac{X_1+\ldots+X_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n}$. At last, we choose $n_0 := \lceil \frac{\sigma^2}{\varepsilon^2 \delta} \rceil$. □

Remark 3

1. *Theorem 1 says that the mean sequence $\frac{X_1+\ldots+X_n}{n}$ converges in probability to the expected value m.*
2. *The statement of Theorem 1 remains also for weakened conditions. It is enough to require the expected value to be finite. $(E(X_i) < \infty)$*
3. *if $X_1 = X_2 = \ldots = X_n$, then $\frac{X_1+\ldots+X_n}{n} = X_1$ does not converge to a constant*

Example 1 *Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables. Then the limit of relative frequency is equal to p, where $p = P(X = 1)$ the probability of one successful trial.*

Definition 3 (weak convergence or convergence in distribution) *A sequence $X_1, X_2, \ldots, X_n, \ldots$ of random variables* ***converges to** X **weakly**, if $\lim\limits_{n\to\infty} F_n(x) = F(x)$ (distribution functions) for every real number x at which F is continuous.*

Definition 4 (almost surely convergence or convergence with probability 1)
$X_n \xrightarrow{a.s.} X$, if $P(\lim\limits_{n\to\infty} X_n = X) = 1$.

Theorem 2 (Central limit theorem) *Suppose that $X_1, \ldots, X_n$ are independent, identically distributed random variables. Let $\sigma^2 := Var(X_1)$ be finite and let $m := E(X_1)$ denote their expected value. Let the standardized sum of the variables be $Z_n := \frac{X_1+\ldots+X_n-n\cdot m}{\sqrt{n}\sigma}$. Then Z_n converges weakly to the standard normal distribution, i.e.,*

$$P\left(\frac{X_1 + \ldots + X_n - n \cdot m}{\sqrt{n}\sigma} < z\right) \xrightarrow{d} \Phi(z),$$

where $\Phi(z)$ denotes the standard normal distribution function.

Remark 4 *It is not necessary to require the identical distribution of variables. The conditions can be weakened, but here we do not discuss it.*

Theorem 3 (Moivre-Laplace) *Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p) random variables. This follows that $Y = X_1 + \ldots + X_n$ is Binomial(n, p). The theorem states that*

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}.$$

Statistical hypothesis testing

H_0 (null hypothesis) says that the unknown parameter is in the subset $\Theta_0 \subset \Theta$, where Θ is the parameter space. In contrast, H_1 (alternative hypothesis) says that $\vartheta \in \Theta_1 := \Theta \backslash \Theta_0$.

Definition 5 ***Type I error****: when the null hypothesis is incorrectly rejected.* ***Type II error****: when one fails to reject an incorrect null hypothesis.*

Definition 6 *The* ***critical region*** *of a hypothesis test is the set of all outcomes which cause the null hypothesis to be rejected in favor of the alternative hypothesis. This region will be denoted by $\mathfrak{X}_c$. In other words, it contains all the $\underline{X}$ observations for which H_0 is rejected.*

Remark 5 *Usually we define it using $T(\underline{X})$ statistics.*

Definition 7 *The* ***probability of Type I error*** *is $\alpha := \sup\limits_{\vartheta \in \Theta_0} P_\vartheta(\mathfrak{X}_c)$.*
The ***probability of Type II error*** *is $\beta(\vartheta) := 1 - P_\vartheta(\mathfrak{X}_c) = P_\vartheta(\mathfrak{X}_a)$, where $\vartheta \in \Theta_1$ and $\mathfrak{X}_a$ denotes the* ***region of acceptance****.*

	H_0 is true	H_1 is true
accept null hypothesis	right decision	wrong decision (Type II error)
reject null hypothesis	wrong decision (Type I error)	right decision