Probabilities and Statistics - Lecture 6

Olariu E. Florentin

March 23, 2026

Table of contents

Markov inequality

Theorem 1.1

(Markov inequality.) *Let $X \geqslant 0$ be a discrete random variable with $\mathbb{E}[X] = \mu$. Then*

$$P\{X \geqslant t\} \leqslant \frac{\mu}{t}, \forall t > 0.$$

proof: When X is a discrete random variable having the following distribution

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix},$$

where $x_1 < x_2 < \dots < x_n < \dots$. Let us suppose that $t \in (x_{k-1}, x_k]$ ($x_0 = -\infty$), then $\mu = \mathbb{E}[X] = \sum_i p_i x_i \geqslant \sum_{i \geqslant k} p_i x_i \geqslant t \sum_{i \geqslant k} p_i = t \cdot P\{X \geqslant t\}$. ■

Observation. The probability that a random variable $X \geqslant 0$ (with finite expectation), has values greater than a given $t > 0$, becomes smaller as t increases.

Markov inequality

Proposition 1

The Markov inequality becomes equality if and only if

$$P\{X=0\}+P\{X=t\}=1.$$

Example. Let $X \geqslant 0$ be a random variable with $\mathbb{E}[X]=1$. Find upper bounds for the following probabilities.

$$P\{X \geqslant 2\}, P\{X \geqslant 4\} \text{ and } P\{X \geqslant 2^k\}.$$

Solution: Using Markov inequality

$$P\{X \geqslant 2\} \leqslant \frac{\mathbb{E}[X]}{2}=\frac{1}{2}, P\{X \geqslant 4\} \leqslant \frac{\mathbb{E}[X]}{4}=\frac{1}{4},$$

$$P\{X \geqslant 2^k\} \leqslant \frac{\mathbb{E}[X]}{2^k}=\frac{1}{2^k} \clubsuit$$

Chebyshev inequality

Theorem 1.2

(Chebyshev inequality.) Let X be a discrete random variable with $\mathbb{E}[X] = \mu$ and $Var[X] = \sigma^2$. Then

$$P\{|X - \mu| \geqslant t\} \leqslant \frac{\sigma^2}{t^2}, \forall t > 0.$$

proof: Consider the variable $Y = (X - \mu)^2$ for which $\mathbb{E}[Y] = Var[X]$, then using Markov inequality,

$$P\{|X - \mu| \geqslant t\} = P\left\{(X - \mu)^2 \geqslant t^2\right\} \leqslant \frac{\mathbb{E}[Y]}{t^2} = \frac{\sigma^2}{t^2}, \forall t > 0.$$

■

Chebyshev inequality

- A possible interpretation of this inequality: if a variable has small variance, then the probability that this variable takes values far away from its expectation is small.
- The following Chebyshev inequality consequence says that the probability that a variable takes values at least k standard deviations far from its expectation is at most $\frac{1}{k^2}$.

Corollary 1.1

Let X be a variable with $\mathbb{E}[X] = \mu$ and $Var[X] = \sigma^2 > 0$.

$$P\left\{|X - \mu| \geqslant k\sigma\right\} \leqslant \frac{1}{k^2}, \forall k > 0.$$

- In this way we can say that the standard deviation measures the spreading of the values of a variable around its expectation.

Chebyshev inequality

Proposition 2

The Chebyshev inequality becomes equality if and only if

$$P\{X = \mu - t\} + P\{X = \mu\} + P\{X = \mu + t\} = 1.$$

Example. Let X be a random variable with $\mathbb{E}[X] = 1$ and $Var[X] = 4$. Find (lower or upper) bounds for the following probabilities

$$P\{X \geqslant 3\}, P\{|X - 1| \geqslant 6\} \text{ and } P\{X \leqslant -9\}.$$

Solution: Using Chebyshev inequality

$$P\{X \geqslant 3\} = P\{X - 1 \geqslant 2\} \leqslant P\{|X - 1| \geqslant 2\} \leqslant \frac{Var[X]}{2^2} = 1,$$

$$P\{|X - 1| \geqslant 6\} \leqslant \frac{4}{36} = \frac{1}{9},$$

$$P\{X \leqslant -9\} = P\{X - 1 \leqslant -10\} \leqslant P\{|X - 1| \geqslant 10\} \leqslant \frac{1}{25} \clubsuit$$

Chernoff bounds

Theorem 2.1

Let $(X_i)_{1 \leqslant i \leqslant n}$ independent variables, each being Bernoulli distributed with a parameter p_i, respectively. If $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$, then

$$P\{X > (1+\delta)\mu\} < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu} \text{ (upper tail), } \forall \delta > 0 \text{ and}$$

$$P\{X < (1-\delta)\mu\} < \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu} \text{ (lower tail), } \forall \delta \in [0,1).$$

Chernoff bounds

- The first inequality says that the sum of Bernoulli independent variables exponentially decays as we move to the right of its expectation:

$$\lim_{\delta \to +\infty} \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right]^{\mu} = 0.$$

- Both inequalities has simpler forms:

Corollary 2.1

Let $(X_i)_{1 \leqslant i \leqslant n}$ *be independent variables, having Bernoulli distribution with parameter* p_i. *If* $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$, *then*

$$P\{X > (1 \pm \delta)\mu\} < \exp\left(\frac{-\delta^2 \mu}{2+\delta}\right), \forall \delta \geqslant 0.$$

Chernoff bounds - an application

Application. We flip a coin n times; let X_i be a variable equal with 1 if we get the head at the i-th flip and 0 otherwise. $X = \sum_{i=1} X_i$ is the number of heads from all flips.

$$\mathbb{E}[X_i] = p_i = \frac{1}{2}, Var[X_i] = p_i(1 - p_i) = \frac{1}{4},$$

$$\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2} \text{ and } \sigma^2 = Var[X] = \sum_{i=1}^{n} Var[X_i] = \frac{n}{4}.$$

(This is a different way to compute the characteristics of a binomial variable.) Using Chernoff bounds we get

Chernoff bounds - an application

$$P\{X > \mu + \lambda\} = P\left\{X > \left(1 + \frac{\lambda}{\mu}\right)\mu\right\} < exp\left(\frac{-\lambda^2}{\lambda + 2\mu}\right) =$$
$$= \exp\left(\frac{-\lambda^2}{\lambda + n}\right).$$

We compare this result with those from Markov and Chebyshev inequalities:

$$P\{X \geqslant \mu + \lambda\} \leqslant \frac{\mu}{\mu + \lambda} = \frac{n}{n + 2\lambda} \text{ (Markov)},$$

$$P\{X \geqslant \mu + \lambda\} \leqslant P\{|X - \mu| \geqslant \lambda\} \leqslant \frac{\sigma^2}{\lambda^2} = \frac{n}{4\lambda^2} \text{ (Chebyshev)}.$$

Note that Markov inequality is weaker than Chebyshev inequality which is weaker than Chernoff bounds. On the other hand Markov inequality (and Chebyshev's) don't need the independence of the n variables.

Hoeffding bounds

Theorem 2.2

Let $X_1, X_2, \ldots, X_n$ be independent bounded random variables: $a_i \leqslant X_i \leqslant b_i$, $a_i \neq b_i \in \mathbb{R}$, $i = \overline{1, n}$ and $X = \sum_{i=1}^{n} X_i$. Then

$$P\{X - \mathbb{E}[X] \geqslant \delta\} \leqslant \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \forall \delta \geqslant 0.$$

Corollary 2.2

In the above conditions we have

$$P\{|X - \mathbb{E}[X]| \geqslant \delta\} \leqslant 2\exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \forall \delta \geqslant 0.$$

Continuous random variables - Random events

- When $|\Omega| \geqslant |\mathbb{R}|$ (i.e., Ω has, at least, a continuous cardinal), random events are defined in a different manner.
- The most notably difference in definition is that it is possible to have subsets $A \subseteq \Omega$ that are not random events: for technical reasons the random events family forms a σ-algeba $\mathcal{A} \subseteq \mathcal{P}(\Omega)$:
  - $\varnothing, \Omega \in \mathcal{A}$;
  - if $A_1, A_2 \in \mathcal{A}$, then $A_1 \cap A_2 \in \mathcal{A}$;
  - if $(A_n)_{n\geqslant 1} \subseteq \mathcal{A}$, then $\bigcup_{n\geqslant 1} A_n \in \mathcal{A}$.
- The probability function is defined only on $\mathcal{A}$ (with known axioms):

$$P : \mathcal{A} \to [0, 1].$$

Continuous random variables

- A function $X : \Omega \to \mathbb{R}$ is called random variable if

$$\text{for every interval } J \subseteq \overline{\mathbb{R}},\ X^{-1}(J) \in \mathcal{A}.$$

- A random variable $X : \Omega \to \mathbb{R}$ is called continuous if its distribution function is a continuous one (*Sometimes, this definition addresses the situations when $X(\Omega)$ has a continuous cardinality*).
- The distribution of such a variable is given by its cumulative distribution function (or for short distribution function):

$$F : \mathbb{R} \to [0,1],\ F(a) = P(X \leqslant a),$$

- or by its probability density function, $f : \mathbb{R} \to [0, +\infty)$, such that F can be described like follows

$$F(a) = P(X \leqslant a) = \int_{-\infty}^{a} f(t)\, dt.$$

Continuous random variables

- Any function $f : \mathbb{R} \to [0, +\infty)$, such that $\int\limits_{-\infty}^{\infty} f(t)\, dt = 1$, is the density function for a certain continuous random variable (or simply a continuous distribution).
- Using the probability density function we can compute (if the integrals exist) the expectation and the variance:

$$\mathbb{E}[X] = \int\limits_{-\infty}^{+\infty} tf(t)\, dt \text{ and } Var[X] = \int\limits_{-\infty}^{+\infty} (t - \mathbb{E}[X])^2 f(t)\, dt.$$

- If $h : \mathbb{R} \to \mathbb{R}$ is a real (say, continuous) function, and X is a random variable with the density f, then $h(X)$ is a random variable having the following expected value

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(t)f(t)\, dt.$$

Continuous random variables

- The associated probabilities are computed like this

$$P(a < X \leqslant b) = F(b) - F(a) = \int_a^b f(t)\, dt$$

which is the area under the graph of f between $t = a$ and $t = b$.

- If F is continuous, $P(X = a) = 0$ and $P(a \leqslant X < b) = P(a < X \leqslant b) = P(a < X < b)$.
- For a given random variable $X : \Omega \to \mathbb{R}$, standardization consists in the following transformation of X:

$$Y = \frac{X - \mathbb{E}[X]}{StDev[X]}.$$

- The new variable is "standard", that is,

$$\mathbb{E}[Y] = 0 \text{ and } Var[Y] = 1.$$

Continuous random variables - Example

Example The life time, in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \frac{k}{x^4}, & x \geqslant 1 \\ 0, & x < 1 \end{cases}$$

Find k, its distribution function, and the probability for life-time to exceed 2 years.

Solution. We must have $f(t) \geqslant 0, \forall t \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f(t)\, dt = 1$, therefore

$k \geqslant 0$ and $1 = \int_1^{\infty} \frac{k}{t^4}\, dt = \left[-\frac{k}{3t^3} \right]_1^{\infty} = \frac{k}{3}$ which gives $k = 3$.

Continuous random variables - Example

The distribution function is $F(x) = \int_{-\infty}^{x} f(t)\, dt$, therefore

$$F(x) = \begin{cases} 1 - \dfrac{1}{x^3}, & x \geqslant 1 \\ 0, & x < 1 \end{cases}$$

Let X be the life time of this electronic component, the probability that the life-time exceeds 2 years is $P(X \geqslant 2) = 1 - P(X < 2) = 1 - F(2) = 1/8$ (because F is continuous).

Remarkable continuous distributions

Uniform distribution. It is denoted by $U(a,b)$ and have the density function

$$f(t) = \begin{cases} 0, & x < a \\ \dfrac{1}{b-a}, & x \in [a,b] \\ 0, & x > b \end{cases}$$

If $X : U(a,b)$, then $\mathbb{E}[X] = \dfrac{a+b}{2}$ and $Var[X] = \dfrac{(b-a)^2}{12}$.

$U(0,1)$ is called *the standard uniform distribution.*

Remarkable continuous distributions

Exponential distribution. It is abbreviated by $Exp(\lambda)$ and have the density function ($\lambda > 0$ is the rate parameter)

$$f(t) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geqslant 0 \end{cases}$$

$X : Exp(\lambda)$, $\mathbb{E}[X] = \dfrac{1}{\lambda}$, $Var[X] = \dfrac{1}{\lambda^2}$.

Exponential distribution is used to model waiting time, interarrival time, hardware lifetime, failure time; in a sequence of rare events the time between events is exponentially distributed.

The Exponential distribution is memoryless (having waited for x_0 minutes get forgotten): regardless of the event $X > x$, when the total waiting time exceeds x, the remaining waiting time still has Exponential distribution: $P(X > x + \Delta x | X > x) = P(X > \Delta x)$ (why?).

Remarkable continuous distributions

Gaussian (normal) distribution. It is denoted by $N(\mu, \sigma^2)$ with the density function

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\dfrac{(t-\mu)^2}{2\sigma^2}}.$$

If $X : N(\mu, \sigma^2)$, then $\mathbb{E}[X] = \mu$ and $Var[X] = \sigma^2$.
The distribution $N(0, 1)$ is called *the standard normal distribution*.
The values of a normal distributed variable have the following spreading (symmetrical around the mean): %68 belongs to a interval $[\mu - \sigma, \mu + \sigma]$, %95 belongs to $[\mu - 2\sigma, \mu + 2\sigma]$, and %99.7 belongs to $[\mu - 3\sigma, \mu + 3\sigma]$.

Remarkable continuous distributions

- Normal distribution has a proeminent role in Probability and Statistics for at least two reasons.
- As a consequence of the Central Limit Theorem (CLT - see below) sums and/or averages of identical distributed independent random variables have approximatively a Normal distribution.
- Normal distribution was found to be a good model for variables like temperature, weight, height or even student grades.
- The Normal distribution was tacitly used by de Moivre for approximating to the binomial distribution and was later used by Laplace and Gauss.

Remarkable continuous distributions

Student (or t) distribution. It is denoted by $t(r)$ with the density function

$$f(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)}\left(1+\frac{x^2}{r}\right)^{-\frac{r+1}{2}},$$

where $\Gamma(y) = \int\limits_0^{+\infty} x^{y-1} e^{-x}\, dx$. For a random variable $X : t(r)$, we have

$\mathbb{E}[X] = 0$ and $Var[X] = \dfrac{r}{r-2}$.

The larger the number of degrees of freedom, the more the distribution looks like the standard normal distribution.

Remarkable continuous distributions

Gamma distribution. It is denoted by $\Gamma(\alpha, \lambda)$ with the density function

$$f(x) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases},$$

where $\Gamma(t) = \int\limits_0^{+\infty} x^{t-1} e^{-x} \; dx$. α is the shape parameter and λ is the rate (or the frequency) parameter. For a random variable $X : \Gamma(\alpha, \lambda)$, we have $\mathbb{E}[X] = \dfrac{\alpha}{\lambda}$ and $Var[X] = \dfrac{\alpha}{\lambda^2}$.

Suppose that we have a process that consists of α independent steps, and each step takes $Exp(\lambda)$ amount of time, then the total time follows a Gamma distribution.

That is, *Gamma* distribution is a sum of α independent Exponential variables.

Remarkable continuous distributions

Chi-squared distribution. It is denoted by $\chi^2(r)$, where $r \in \mathbb{N}^*$ is the number of degrees of freedom. Its density function is

$$f(x) = \begin{cases} \dfrac{x^{r/2-1}e^{-x/2}}{2^{r/2}\Gamma(r/2)} x^{\alpha-1}e^{-\lambda x}, & x > 0 \\ 0, & x \leqslant 0 \end{cases},$$

where $\Gamma(t) = \int\limits_0^{+\infty} x^{t-1}e^{-x}\ dx$.

The chi-squared distribution is a particular case of Gamma distribution: when $X : \chi^2(r)$, then $X : Gamma(r/2, 1/2)$.

This distribution is widely used in inferential statistic (we will see later in which way) for goodness of fit of a given sample distribution or for independence testing of two categorical variables.

Remarkable continuous distributions

Fisher distribution. It is denoted by $F(r_1, r_2)$, where $r_1, r_2 \in \mathbb{N}^*$ are the two numbers of degrees of freedom. Its density function is

$$f(x) = \begin{cases} \dfrac{\sqrt{\dfrac{(r_1 x)^{r_1} r_2^{r_2}}{(r_1 x + r_2)^{r_1 + r_2}}}}{x B(r_1/2, r_2/2)}, & x > 0 \\ 0, & x \leqslant 0 \end{cases},$$

where $B(t_1, t_2) = \int\limits_0^1 x^{t_1 - 1}(1 - x)^{t_2 - 1}\, dx$.

If $X_1 : \chi^2(r_1)$ and $X_2 : \chi^2(r_2)$, then $X = \dfrac{X_1 r_2}{X_2 r_1} : F(r_1, r_2)$.

Like the $Exp(\lambda)$, $Gamma(\alpha, \lambda)$, and $\chi^2(r)$, the Fisher distribution is nonnegative and nonsymmetrical. It is used in inferential statistic and play a proeminent role in the *analysis of variance - ANOVA*.

Markov and Tchebychev inequalities revisited

Proposition 3

Let $X \geqslant 0$ be a non-negative continuous random variable which has a probability density function. If $a > 0$, then

$$P(X \geqslant a) \leqslant \frac{\mathbb{E}[X]}{a}.$$

proof:

$$\mathbb{E}[X] = \int_0^{+\infty} tf(t)dt = \int_0^a tf(t)dt + \int_a^{+\infty} tf(t)dt \geqslant$$

$$\int_a^{+\infty} tf(t)dt \geqslant a \int_a^{+\infty} f(t)dt = aP(X \geqslant a).$$

■

In a similar we can get the Tchebychev's inequality for continuous random variables having a probability density function.

Tchebychev's theorem

Theorem 1.1

Let $(X_n)_{n \geqslant 1}$ be a sequence of independent random variables having finite variances, uniformly bounded, that is $Var[X_n] \leqslant c$, for every $n \geqslant 1$. Then

$$\lim_{n \to \infty} P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] \right| < \epsilon \right) = 1.$$

proof: We know that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] \text{ and}$$

$$Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} Var[X_i] < \frac{c}{n}.$$

Tchebychev's theorem

Applying the Tchebychev's inequality for the variable $\frac{1}{n}\sum_{i=1}^{n} X_i$ we get

$$1 \geqslant P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right| < \epsilon\right) \geqslant 1 - \frac{Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}{\epsilon^2} \geqslant 1 - \frac{c}{n\epsilon^2}.$$

Taking to the limit we obtain

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right| < \epsilon\right) = 1. \blacksquare$$

The weak law of large numbers

- The laws of large numbers say that as the number of identically distributed and independent random variables increases, their sample mean approaches their theoretical common mean (expectation).

Theorem 2.1

(The weak law of large numbers, Khintchine's law) Let $(X_n)_{n\geqslant 1}$ be a sequence of identically distributed independent random variables having mean μ and variance σ^2. Then

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1 \text{ or}$$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geqslant \epsilon\right) = 0.$$

The strong law of large numbers

proof: It is a consequnce of the previous theorem, since

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mu.$$

∎

Theorem 2.2

(The strong law of large numbers) Let $(X_n)_{n\geqslant 1}$ be a sequence of identically distributed and independent random variables having mean μ and variance σ^2. Then

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = \mu\right) = 1.$$

proof: It is too complex and will be omitted.

An example - frequencies

- Bernoulli is credited with the first proof of the weak law of large numbers, but only for Bernoulli distributions.
- Suppose that we have a random experience and a related random event A with $P(A) = p$. We independently perform the experience, and consider the following sequence of random variables: $X_i = 1$ if A occurs at the ith performing, and 0 otherwise.
- The variables are independent and Bernoulli distributed with parameter p. The law of large numbers says that, with probability 1,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to p.$$

- $\sum_{i=1}^{n} X_i$ is the number of occurrences of A after n performings of the experience. In other words the law of large numbers says that the A occurs with frequency p.

Bernoulli's theorem

Proposition 4

Let α_n be a the number of occurrences of an event A in n independent performings of a random experience. If $f_n = \dfrac{\alpha_n}{n}$ is the relative frequency of occurrence of A, then the sequence $(f_n)_{n\geqslant 1}$ converges in probability to p, the probability of A.

proof: $\alpha_n = nf_n$ is a binomial distributed variable, hence $\mathbb{E}[\alpha_n] = np$ and $Var[\alpha_n] = np(1-p)$. Moreover

$$P(|f_n - p| < \epsilon) = P(|\alpha_n - np| < n\epsilon) = P(|\alpha_n - \mathbb{E}[\alpha_n]| < n\epsilon) \geqslant$$

$$\geqslant 1 - \frac{p(1-p)}{n\epsilon^2}.$$

Obviously, $\lim\limits_{n\to\infty} P(|f_n - p| < \epsilon) = 1$, for every $\epsilon > 0$. ∎

Some History

- James Bernoulli proved the weak law of large numbers in 1700; Poisson generalized his result around 1800.
- Tchebychev discovered his inequality in 1866, and Markov extended Bernoulli's theorem to dependent random variables.
- In 1909 Émile Borel proved what today is known as the strong law of large numbers that further generalizes Bernoulli's theorem.
- In 1926 Kolmogorov derived a more general condition that was sufficient for a set of mutually independent random variables to obey the law of large numbers. This condition is

$$\sum_{n \geqslant 1} \frac{Var[X_n]}{n^2} < +\infty.$$

Exercises for Seminar

- Markov and Chebyshev's Inequalities: I.1, I.3, I. 5, I.6, I.7, I.8, I.9.
- Continuous random variables: II.1, II. 4, II.5, II.6 (f_1, f_3).
- Reserve: I.10, I.11, II.3.

End

Exercises - Markov and Chebyshev's Inequalities

I.1. A random variable $X \geqslant 0$ has its expectation and variance both equal with 20. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for $P\{X \geqslant 40\}$ and $P\{-60 \leqslant X \leqslant 100\}$.

I.2. Let $X \geqslant 0$ be random variable with $\mathbb{E}[X] = Var[X] = 1$. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for

$$P\{X \geqslant 2\}, P\{|X-1| \geqslant 2\}, P\{X \leqslant -3\}.$$

I.3. Let $X \geqslant 0$ be random variable with $\mathbb{E}[X] = Var[X] = 2$. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for $P\{X \geqslant 8\}$ and $P\{|X-2| \geqslant 8\}$.

I.4. Let $X \geqslant 0$ be random variable with $\mathbb{E}[X] = 2$ and $Var[X] = 1$. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for $P\{X \geqslant 6\}$ and $P\{|X-1| \geqslant 5\}$.

Exercises - Markov and Chebyshev's Inequalities

I.5. A random variable $X \geqslant 0$ has $\mathbb{E}[X] = 2$ and $Var[X] = 3$. Using Markov and/or Chebyshev's inequalities what can you say about $P\{X \geqslant 8\}$ and $P\{|X - 2| \geqslant 4\}$?

I.6. The probability to get the head on a biased coin is 0.3. We toss this coin 300 times. Find an upper bound for the probability that we get the head at least 100 times.

I.7. The probability to get the head on a biased coin is 0.2. We toss this coin n times. Find an upper bound for the probability that we get the head at least 50% times.

I.8. Two biased coins have the probabilities of the tail occurrence 0.25 and 0.8, respectively. The coins are 25 times flipped. Using Markov and Chebyshev inequalities find upper bounds for the probability that we get two tails at least 10 times.

Exercises - Markov and Chebyshev's Inequalities

I.9. We roll two dice 36 times. Using Markov and Chebyshev inequalities find upper bounds for the probability that we get two a product which is a prime number at least 10 times.

I.10. Two biased coins are flipped 32 times. The probability of getting a tail is 1/3 for the first coin and 3/4 for the second. Using Markov and Chebyshev inequalities find upper bounds for the probability that we get tail on both coins at least 12 times.

I.11. We toss a fair coin n times. Let X be the number of tails we get. Find upper bounds for

(a) $P\{|X - n/2| > \sqrt{n}\}$ and $P\{X > n/2 + \sqrt{n}\}$;

(b) $P\{|X - n/2| > 5\sqrt{n}\}$ and $P\{X > n/2 + 5\sqrt{n}\}$.

I.12. Let X be a Poisson distributed variable with cu parameter λ. Estimate the probability that X deviates from $\mathbb{E}[X]$ with at least $2\sqrt{\lambda}$.

Exercises - Markov and Chebyshev's Inequalities

I.13*. (Borel-Cantelli lemma) Let $(A_n)_{n\geqslant 1}$ be a sequence of random events such that $\sum\limits_{n\geqslant 1} P(A_n) < +\infty$. Prove that the probability of occurrence of at most k of these events is at least

$$1 - \frac{\sum\limits_{n\geqslant 1} P(A_n)}{k}.$$

(*Hint: Use Markov's inequality for a variable which numbers the occurring events.*)

Exercises - Continuous random variables

II.1. Let X be a continuous random variable with the following probability density function

$$f(t) = \begin{cases} \alpha t, & 0 \leqslant t \leqslant 2 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find α and the (cumulative) distribution function of X.

(b) Compute the expectation and the variance of X.

II.2. The time in minutes, it takes to reboot a certain system is a continuous variable with the probability density function

$$f(t) = \begin{cases} C(10-t)^2, & 0 < t < 10 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find C and the (cumulative) distribution function of X.

(b) Compute the probability that it takes between 1 and 2 minutes to reboot that system and $P(X > 1 | X < 3)$.

Exercises - Continuous random variables

II.3. Lifetime in years, of a certain HD is a continuous random variable with the probability density function

$$f(t) = \begin{cases} K - \dfrac{t}{50}, & 0 < t < 10 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find K and the (cumulative) distribution function of X.

(b) Compute the probability of a failure within first 5 years, and the expectation of the lifetime.

II.4. Consider a continuous random variable having the following probability density function

$$f(t) = \begin{cases} \alpha\sqrt{t}, & 0 < t < 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) Find α and the (cumulative) distribution function of X.

(b) Compute the probability $P(0.3 < X < 0.6)$ and $\mathbb{E}[X]$.

Exercises - Continuous random variables

II.5. Consider a continuous random variable having the following (cumulative) distribution function

$$f(t) = \begin{cases} 0, & t < 0 \\ \dfrac{t}{t+1}, & t \geqslant 0 \end{cases}$$

(a) Compute $P(0 < X < 3)$, $P(X > 0 | X \leqslant 2)$, and $\mathbb{E}[X]$.

(b) Find the probability density function of X.

II.6. Which of the following functions can be probability density functions? (Determine the constant α in the affirmative case.)

$$f_1(t) = \begin{cases} \alpha(t^2 - t), & 0 \leqslant t \leqslant 1 \\ 0, & \text{otherwise} \end{cases} ; f_2(t) = \begin{cases} \alpha t^2, & 0 \leqslant t \leqslant 2 \\ 0, & \text{otherwise} \end{cases} ;$$

$$f_3(t) = \begin{cases} \alpha(t^3 - 2t), & -1 \leqslant t \leqslant 1 \\ 0, & \text{otherwise} \end{cases} ; f_4(t) = \begin{cases} \dfrac{\alpha}{t}, & 0 < t \leqslant 2 \\ 0, & \text{otherwise} \end{cases} .$$

Bibliography I

- Bertsekas, D. P., J. N. Tsitsiklis, *Introduction to Probability*, Athena Scietific, 2002.
- Gordon, H., *Discrete Probability*, Springer Verlag, New York, 1997.
- Johnosn, J. L., *Probability and Statistics for Computer Science*, Wiley Interscience, 2008.
- Lipschutz, S., *Theory and Problems of Probability*, Scahaum's Outline Series, McGraw-Hill, 1965.
- Mitzenmacher, M., E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.
- Ross, S. M., *A First Course in Probability* , Prentice Hall, 5th edition, 1998.

Bibliography II

Stone, C. J., *A Course in Probability and Statistics*, Duxbury Press, 1996.