Lecture Notes on Probability Theory

E. Y. Emelyanov

Department of Mathematics
Middle East Technical University, 06800 Ankara, Turkey.
Email: eduard@metu.edu.tr

N. Erkurşun-Özcan

Department of Mathematics,
Hacettepe University, 06800 Ankara, Turkey.
Email: erkursun.ozcan@hacettepe.edu.tr

S. G. Gorokhova

Sobolev Institute of Mathematics, 630090 Novosibirsk, Russia.
Email: lanagor71@gmail.com

April 18, 2019

The main concepts and principles of basic probability theory are introduced: random variables and random processes, laws of large numbers, central limit theorem, Markov chains, and limit theorems for them. They are illustrated by supplied examples and exercises. The course does not require knowledge of measure theory, and we refer to standard facts as they are needed.

Content

1 Historical remarks (optional)

1.1 Mathematical gambling

The probability theory is a branch of math which studies *random phenomena*. The main concept of the probability theory is an *event* (e.g., "coin toss is head", "roll a dice by side 3", etc.). Now days, there are plenty amazing illustrations of usefulness of this concept almost everywhere. To begin with our introduction to the probability theory, let us chose the following famous historical problem which had led to developing basic ideas of the probability theory three and half centuries ago.

In 1654, Chevalier de Mere has met some paradox in his gambling practice. He wrote an angry letter to his friend Blaise Pascal (a mathematician with whom de Mere discussed mathematical aspects of gambling), accusing mathematicians in creating incorrect theories contradicting the common practice. Now this "paradox" is known as the **Mere problem** or the **problem of points**: *3 fair dice are thrown, which sum of numbers "appears more often", 11 or 12?*

De Mere was thinking that the expectation of 11 and of 12 must be equal (accordingly to what he understand from "mathematical gambling"), indeed

$$11 = 6 + 4 + 1 = 6 + 3 + 2 = 5 + 5 + 1 = 5 + 4 + 2 = 5 + 3 + 3 = 4 + 4 + 3,$$

the six equally likely possibilities, and

$$12 = 6 + 5 + 1 = 6 + 4 + 2 = 6 + 3 + 3 = 5 + 5 + 2 = 5 + 4 + 3 = 4 + 4 + 4,$$

the six equally likely possibilities! But, after a long series of experiments, de Mere got different "practical" answer. What is wrong? His friend Pascal (after a quite long mail conversation with Piere de Fermat) solve the problem as follows.

Consider the six combinations of numbers (from 1 to 6) a,b, and c (with $a \neq b \neq c \neq a$):

$$(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a),$$

as *different* (with the equal probability of appearance, of course).

The unordered triple $\{6, 4, 1\}$ appears 6 times by ordered triples

$$(6, 4, 1), (1, 6, 4), (4, 1, 6), (6, 4, 1), (1, 6, 4), \textit{and } (4, 1, 6);$$

the triple $\{6, 3, 2\}$ – 6 times; $\{5, 5, 1\}$ – 3 times; $\{5, 4, 2\}$ – 6 times; $\{5, 3, 3\}$ – 3 times; and $\{4, 4, 3\}$ – 3 times. Thus, the sum 11 appears

$$6 + 6 + 3 + 6 + 3 + 3 = 27$$

times.

The triple $\{6, 5, 1\}$ appears 6 times; the triple $\{6, 4, 2\}$ – 6 times; $\{6, 3, 3\}$ – 3 times; $\{5, 5, 2\}$ – 3 times; $\{5, 4, 3\}$ – 6 times; and $\{4, 4, 4\}$ – 1 time. Thus, the sum 12 appears

$$6 + 6 + 3 + 3 + 6 + 1 = 25$$

times ($25 \neq 27$) out of $6 \times 6 \times 6 = 216$ possible results.

Hence, in a long series of dice throwing, the sum 11 appears more often than 12.

1.2 Modeling of events

We need a math model for describing random events that are result of performing an experiment. We cannot use frequency of occurrence as a model, because it does not have the power of prediction. First, we define a *state space* (or *sample space*) that we will denote by S. We consider elements of S as outcomes of the experiment.

Then, we specify a collection $\mathcal{A}$ of subsets of S. Each of these subsets is called an *event*. These events are sets we can talk about the probability of them. When S is finite, $\mathcal{A}$ can be taken to be the collection of all subsets of S.

The followings are the examples for the finite S.

Example 1.1 *Roll a six-sided die; what is the probability of rolling the six or five? First, write a sample space. Here is a natural one:*

$$S = \{1, 2, 3, 4, 5, 6\}$$

In this case, S is finite and we want $\mathcal{A}$ to be the collection of all subsets of S. Clearly (outcomes are equally likely, since the dice thrown suppose to be fair)

$$P(\{6, 5\}) = P(\{6\}) + P(\{5\}) = 1/6 + 1/6 = 1/3. \quad \square$$

Example 1.2 *For the de Mere problem, the sample space is*

$$S = \{(a, b, c)| \ a, b, c \in \{1, 2, 3, 4, 5, 6\}\}.$$

All $(a, b, c) \in S$ are equi-likely. The cardinality *(the* power, *or the* dimension*) of S is $6 \times 6 \times 6 = 216$. So, we have 216 equi-likely outcomes. In this case $P((a, b, c)) = 1/216$ for every $a, b, c \in \{1, 2, 3, 4, 5, 6\}$,*

$$P(\{(a, b, c)| \ a + b + c = 11\}) = 27/216,$$

and

$$P(\{(a, b, c)| \ a + b + c = 12\}) = 25/216. \quad \square$$

Let $\mathcal{A}$ be the class of all events for the sample set S. Under the operations $\cup$, $\cap$, and ^{c} is a Boolean algebra.

Definition 1.1 *A Boolean algebra is a non-empty set $\mathcal{A}$ on which operations union $\cup$, intersection $\cap$ and complement ^{c} are defined such that*

$$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A} \ and A \cap B \in \mathcal{A}$$

$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$

and satisfying the following axioms

(i) Commutativity $A \cup B = B \cup A, \quad A \cap B = B \cap A$

(ii) Associativity $(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)$

(iii) Distributivity $(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

(iv) Absorption $(A \cap B) \cup B = B, \quad (A \cup B) \cap B = B$

(v) *Associativity* $(A \cap A^c) \cup B = B, \quad (A \cup A^c) \cap B = B$

Also from (iii) and (iv), idempotent laws are satisfied, i.e. $A \cup A = A$ and $A \cap A = A$, $\forall A \in \mathcal{A}$.

If $A = A \cap B$ or $B = A \cup B$ then the event B is said to include an event A. The inclusion is denoted by $\subset$ and the inclusion relation is a partial order relation on $\mathcal{A}$.

Moreover $\emptyset = A \cap A^c$ and $S = A \cup A^c$ are called the imposible event and the sure event of $\mathcal{A}$. The definition does not depend on $A \in \mathcal{A}$. Two elements A, B are called dsjoint if $A \cap B = \emptyset$.

Definition 1.2 *An element $A \in \mathcal{A}$ is said to be an atom if $A \neq \emptyset$ and $B \subset A$ implies $B = \emptyset$ or $B = A$.*

Lemma 1.1 *If A is an atom then for any element $D \in \mathcal{A}$ either $A \subset D$ or $A \subset D^c$.*

Proof: Let A be an atom. Then for any element $D \in \mathcal{A}$, $D \cap A \subset A$. Hence either $A \cap D = \emptyset$ or $A \cap D = A$. Therefore if $A \cap D = \emptyset$ then $A \subset D^c$ or if $A \cap D = A$ then $A \subset D$. ■

Remark that if A and B are distinct atoms of $\mathcal{A}$ then A and B are disjoint.

We want more than algebra for events, because we want to deal with sequences of events. Therefore we define σ-algebra.

Definition 1.3 *Let S be a sample space. A non-empty class $\mathcal{A}$ is called a σ-algebra if*

(i) $S \in \mathcal{A}$,

(ii) *if* $A \in \mathcal{A}$ *then* $A^c \in \mathcal{A}$,

(iii) *given a sequence* $(A_n)_n \subseteq \mathcal{A}$, *we have* $\cup_n A_n \in \mathcal{A}$.

v

1.3 Function of probability

Once we have a sample space and an event-space, we need to assign a probability to every event. This assignment has to satisfy some properties.

Rule 1. $0 \leq P(A) \leq 1$ for every event A.

Rule 2. $P(S) \leq 1$. Something will happen with probability one.

Rule 3. (Addition rule) If A and B are disjoint events (i.e. $A \cap B = \emptyset$), then the probability, that is at least one of the two occurs, is the sum of the individual probabilities. More precisely, put $P(A \cup B) = P(A) + P(B)$.

In what follows, we use freely the common set theoretical symbols and the terminology. For instance, $\mathcal{P}(S)$ is the power set of S (= the set of all subsets of S), $A \setminus B$ the set of elements of A which are not in B, $B^c = S \setminus B$ is the complement of $B \subseteq S$ in S, etc.

1.4 Countable additivity

Recall that a set X is said to be *countable infinite* (or, just *countable*) if there is a bijective (1-to-1 and onto) function from $\mathbb{N} = \{1, 2, 3, ...\}$ onto X, in other words, if we can count X. Examples of countable sets are $\mathbb{N}, \mathbb{Z}, \mathbb{Z}^2$, and $\mathbb{Q}$. An example of an uncountable set is any nonempty interval in the real line $\mathbb{R}$.

One can form countable union of sets by defining $\bigcup_{i \geq 1} A_i$ to be the set of elements that are in at least one of the sets A_i. Similarly, $\bigcap_{i \geq 1} A_i$ is the set of elements that are in all of the A_is simultaneously (if there are no such elements, then the intersection is empty).

Rules 1 – 3 suffice if we want to study only finite sample spaces. But infinite sample spaces are also interesting and much more useful. This happens, for example, if we want to write a model that answers, what is the probability that we toss a coin 12 times before we toss heads? This leads us to the next, and final, rule of probability.

Rule 4. (Extended addition rule) If $A_1, A_2, \ldots$ are (countably-many) pairwise disjoint events, then

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

This rule will be extremely important to us. It needs to be assumed as a part of the model of the probability theory, which we going to study in this course. In the next lecture, we continue with exact definitions of this model and consider further methods for defining (and calculating) of the probability.

2 Axiomatic approach

2.1 Probability axioms

Given a sample space S, an algebra $\mathcal{A} \subseteq \mathcal{P}(\mathcal{S})$, and a function $P : \mathcal{A} \to \mathbb{R}$. The triple $(S, \mathcal{A}, P)$ is called a **probability space** (and $P(A)$ is called the **probability** of an event A) if $\mathcal{A}$ is a σ-algebra (**algebra of events**) and

1. $0 \leq P(A) \leq 1$ for each $A \in \mathcal{A}$.

2. $P(S) = 1$.

3. $P(\bigcup_n A_n) = \sum_n P(A_n)$ for every pairwise disjoint sequence A_n of events.

It should be emphasized that, in many cases, a choice of a proper probability space is the subject of an experiment.

Example 2.1 *If S is either finite or countable, then $(S, \mathcal{A}, P)$ is called a* **discrete probability space**. *Clearly, in this case, P is uniquely determined by values $P(\{a\})$ for all $a \in S$ such that $\{a\}$ is an event, namely*

$$P(A) = \sum_{a \in A, \{a\} \in \mathcal{A}} P(\{a\}) \quad (A \in \mathcal{A}). \ \square$$

Consider the following example of a "non-discrete" probability space.

Example 2.2 *Suppose that we choose randomly a point in the interval $[0, 1]$. In this case, events are exactly the Lebesgue measurable subsets of $[0, 1]$, and the probability is the Lebesgue measure μ of these subsets. This simple probability function $P(A) = \mu(A)$ of a dropped randomly point to fall into $A \subseteq [0, 1]$ is called the* **geometric probability**.

A slight modification can be obtained if we replace $[0, 1]$ by any Lebesgue measurable subset S of $\mathbb{R}^n$ of a finite Lebesgue measure $\lambda(S)$ and define P as follows

$$P(A) = \lambda(A \cap S) \quad (A \text{ is a Lebesgue measurable subset of } \mathbb{R}^n). \ \square$$

Example 2.2 motivates the following definition which is fairly important in the probability theory. Later on we will extend this definition to an infinite sequence of probability spaces.

Definition 2.1 *Let* $(S_1, \mathcal{A}_1, P_1)$, $(S_2, \mathcal{A}_2, P_2)$, ... $(S_n, \mathcal{A}_n, P_n)$ *be probability spaces. Define a probability function* P *on rectangles* $A_1 \times A_2 \times ... \times A_n$, *where* $A_k \in \mathcal{A}_k$, *by*

$$P(A_1 \times A_2 \times ... \times A_n) := P_1(A_1) \cdot P_2(A_2) \cdot ... \cdot P_n(A_n);$$

and "extend" P *to arbitrary "measurable" subsets of the Cartesian product* $S_1 \times S_2 \times ... \times S_n$ *(at this point some measure theory is needed, but it is not included in our course, so, we just accept such possibility of this extension as well as its uniqueness).*

The above defined probability P on Lebesgue measurable subsets of the Cartesian product $S = S_1 \times S_2 \times ... \times S_n$ gives a rise to the probability space $(S, \mathcal{A}, P)$, where the algebra of events is the Lebesgue algebra of S. This probability space is called the product of probability spaces $(S_1, \mathcal{A}_1, P_1)$, $(S_2, \mathcal{A}_2, P_2)$, ... , and $(S_n, \mathcal{A}_n, P_n)$.

Example 2.3 *Let* $(S, \mathcal{P}(S), P)$ *be a probability space. In order to define* P *it suffices to assign a probability* $p(s)$ *to each outcome* $s \in S$ *so that:*

$$0 \le p(s) \le 1 \quad (\forall s \in S) \qquad \& \qquad \sum_{s \in S} p(s) = 1.$$

Indeed, in this case the probability function P *is given by* $P(A) = \sum_{s \in A} p(s)$ *for all* $A \subseteq S$. □

2.2 Independence

Definition 2.2 *Two events* A *and* B *are* **independent** *whenever*

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

Example 2.4 *Suppose you draw a card from a a standard deck. Let* H *be you drew a heart and* K *be that you drew a king. Since*

$$P(H \cap K) = P(king\ of\ heart) = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = P(H)P(K),$$

the events H *and* K *are independent.* □

Lemma 2.1 *For any two events A and B the following conditions are equivalent:*

(i) A and B are independent;

(ii) A and B^c are independent;

(iii) A^c and B^c are independent.

Proof: Since $A^{cc} = A$, it is sufficient to prove $(i) \Rightarrow (ii)$:

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A) \cdot P(B) = P(A) \cdot (1 - P(B)) = P(A)P(B^c). \quad \blacksquare$$

Definition 2.3 *A* family $\{A_\gamma\}_{\gamma \in \Gamma}$ *of* events *is called:*

(i) **independent** *if, for each* finite subfamily $\{A_{\gamma_i}\}_{i=1}^{i=k}$, *there holds*

$$P(A_{\gamma_1} \cap A_{\gamma_2} ... \cap A_{\gamma_k}) = P(A_{\gamma_1}) \cdot P(A_{\gamma_2}) \cdot ... \cdot P(A_{\gamma_k});$$

(ii) **pairwise independent** *if, for every* pair $(A_{\gamma_1}, A_{\gamma_2})$ *of events with $A_{\gamma_1} \neq A_{\gamma_2}$, there holds*

$$P(A_{\gamma_1} \cap A_{\gamma_2}) = P(A_{\gamma_1}) \cdot P(A_{\gamma_2}).$$

If events A and B are independent then the family $\{A, B\}$ is independent. Moreover, the family $\{\emptyset, A, S\}$ is independent for any event A such that $A \neq \emptyset$ and $A \neq S$.

Proposition 2.1 *For any* family $\mathcal{A} = \{A_\gamma\}_{\gamma \in \Gamma}$ *of* events:

(i) if $\mathcal{A}$ is pairwise independent then the family $\mathcal{A}' = \{A_\gamma^c\}_{\gamma \in \Gamma}$ is pairwise independent;

(ii) if $\mathcal{A}$ is independent then the family $\mathcal{A}' = \{A_\gamma^c\}_{\gamma \in \Gamma}$ is independent.

Proof: (i) It follows directly from Lemma 2.1.

(ii) It is enough to show that if $P(\bigcap\limits_{j=1}^{j=m} A_{\gamma_{k_j}}) = \prod\limits_{k=j}^{j=m} P(A_{\gamma_{k_j}})$ for any subset $\{A_{\gamma_{k_j}}\}_{j=1}^{j=m}$ of arbitrary finite subset $\{A_{\gamma_k}\}_{k=1}^{k=n} \subseteq \mathcal{A}$ then $P(\bigcap\limits_{k=1}^{k=n} A_{\gamma_k}^c) = \prod\limits_{k=1}^{k=n} P(A_{\gamma_k}^c))$. The case of $n = 1$ is trivial. The case of $n = 2$

follows from (i). We prove the case $n = 3$ only and leave the case of arbitrary n to the reader as an exercise.

Let a family of events $\{A, B, C\}$ be independent. Then

$$P(A^c \cap B^c \cap C^c) = P((A \cup B \cup C)^c) = 1 - P(A \cup B \cup C) =$$
$$1 - [P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)] =$$
$$1 - P(A) - P(B) - P(C) + P(A)P(B) + P(B)P(C) + P(C)P(A) - P(A)P(B)P(C) =$$
$$(1 - P(A))(1 - P(B))(1 - P(C)) = P(A^c)P(B^c)P(C^c). \quad \blacksquare$$

Example 2.5 *The collection of all events which are independent with a given event A need not to be an algebra. To see this, take the unit square $S = [0,1]^2$ with the geometric probability and consider the following events:*

$$S_{kj} = \big\{(x,y) \in S : \frac{k-1}{3} \le x \le \frac{k}{3}; \frac{j-1}{3} \le x \le \frac{j}{3}\big\} \quad (k, j = 1, 2, 3),$$

and let

$$A := S_{12} \cup S_{22} \cup S_{32}, \quad B := S_{13} \cup S_{12} \cup S_{11} \cup S_{22} \cup S_{21} \cup S_{33}, \quad C = S_{33} \cup S_{32} \cup S_{12}.$$

Notice that $P(S_{kj}) = \frac{1}{9}$ for all $k, j = 1, 2, 3$. Since

$$P(A \cap B) = P(S_{12} \cup S_{22}) = \frac{2}{9} = \frac{1}{3} \cdot \frac{2}{3} = P(A) \cdot P(B) \quad \textit{and}$$

$$P(A \cap C) = P(S_{33}) = \frac{1}{9} = \frac{1}{3} \cdot \frac{1}{3} = P(A) \cdot P(C),$$

the events B and C are both independent with A. However

$$P(A \cap [B \cap C]) = P(A \cap S_{33}) = P(\emptyset) = 0 \neq \frac{1}{27} = \frac{1}{3} \cdot \frac{1}{9} = P(A) \cdot P(B \cap C),$$

which shows that $B \cap C$ is not independent with A. $\square$

Clearly, every independent family is pairwise independent. However, as the following example shows, it could happened that a family of events is pairwise independent but not independent.

Example 2.6 (*S.N. Bernstein, 1927*) *There are four tickets in a box with numbers abc which are* (112), (121), (211), *and* (222). *The probability of taking any ticket is* 1/4. *Let $A_1 = \{abc|\ c = 1)$, $A_2 = \{abc|\ b = 1)$, and $A_3 = \{abc|\ a = 1)$. Then events A_1, A_2, and A_3 are pairwise independent. Indeed,*

$$P(A_1) = P(A_2) = P(A_3) = 2/4 = 1/2,$$

$$P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = 1/4 = (1/2) \cdot (1/2).$$

However,

$$P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq 1/8 = P(A_1) \cdot P(A_2) \cdot P(A_3). \quad \square$$

The Definitions 2.1, 2.3 are both generalized by the following definition.

Definition 2.4 *Let $(S, \mathcal{A}, P)$ be a probability space and let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two σ-algebras of $\mathcal{A}$.* $\mathcal{A}_1$ **is said to be independent with** $\mathcal{A}_2$ *if, whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$,*

$$\mathrm{P}(A_1 \cap A_2) = \mathrm{P}(A_1) \cdot \mathrm{P}(A_2).$$

An arbitrary family of σ-algebras $(\mathcal{A}_i)_{i \in I}$ of events *of $\mathcal{A}$ is said to be* **pairwise independent** *if any pair (A_1, A_2) of events is independent, whenever $A_1 \in \mathcal{A}_{i_1}$, $A_2 \in \mathcal{A}_{i_2}$, and $i_1 \neq i_2$.*

Likewise, a finite family *of σ-algebras $(\mathcal{A}_i)_{i \in I}$ is said to be* **independent** *iff*

$$\forall \left(A_i\right)_{i \in I} \in \prod_{i \in I} \mathcal{A}_i \ : \ \mathrm{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathrm{P}\left(A_i\right),$$

and an arbitrary family of σ-algebras *is said to be* **independent** *iff all its finite subfamilies are independent.*

Example 2.7 *The* sequence of n Bernoulli trials (n-binomial trials) *is a series of n independent random experiments; each of them has two possible outcomes: "success" (the success is usually denoted by 1) with the probability p, and "failure" (the failure is usually denoted by 0), with the probability $q = 1 - p$, where $p = P(1)$ is the same every time the experiment is conducted.*

Consider $A_k = \{(a_1, a_2, ..., a_n) | a_k = 1\}$, where $(a_1, a_2, ..., a_n)$ is a sequence of n-Bernoulli trials. Clearly, $P(A_k) = p$ where we consider as a sample space the Cartesian product of n copies of elementary probability spaces

$$(\{0, 1\}, \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}, P(\{0\}) = 1 - p, P(\{1\}) = p).$$

Moreover, the sequence $A_1, A_2, ..., A_n$ is independent. For example, if k_1, k_2, and k_3 are distinct, then

$$P(A_{k_1} \cap A_{k_2} \cap A_{k_3}) = p^3 = P(A_{k_1}) \cdot P(A_{k_2}) \cdot P(A_{k_3}).$$

Generalized Bernoulli trails *corresponds to the following probability space*

$$(\{0, 1, 2, ...k\}, \ \mathcal{P}(\{0, 1, 2, ...k\}), \ P(\{k\}) = p_k, \sum_{i=1}^{k} p_k = 1).$$

Clearly, the above sequence $A_1, A_2, ..., A_n$ is independent. □

2.3 Conditional probability

Definition 2.5 *Let* $(S, \mathcal{A}, P)$ *be a probability space, and* $A, B \in \mathcal{A}$ *be two events with* $P(B) > 0$*. The* **conditional probability of** A **given** B *is defined by*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

If A and B are independent with $P(B) > 0$, then $P(A|B) = P(A)$.

The formula $P(A \cap B) = P(A|B) \cdot P(B)$ may be extended to more than two events, for example

$$P(A_3 \cap A_2 \cap A_1) = P(A_3 \cap (A_2 \cap A_1)) = P(A_3|(A_2 \cap A_1)) \cdot P(A_2 \cap A_1) =$$

$$P(A_3|(A_2 \cap A_1)) \cdot P(A_2|A_1) \cdot P(A_1)$$

assuming $P(A_1 \cap A_2) > 0$. More generally:

Theorem 2.1 **(Multiplication Theorem)** *Let* A_1*,* A_2*, ...,* A_n *be a sequence of events with* $P(\cap_{k=1}^{n-1} A_k) > 0$*. Then*

$$P(\cap_{k=1}^{n} A_k) = P(A_1)P(A_2|A_1) \cdot ... \cdot P(A_n| \cap_{k=1}^{n-1} A_k). \quad \blacksquare$$

The following theorem although trivial is very useful.

Theorem 2.2 **(Theorem of Total Probability)** *Let* B_1*,* B_2*, ... be a finite or countable sequence of disjoint events with all* $P(B_k) > 0$ *and* $\cup_{k=1}^{\infty} B_k = S$*. Then, for any event* A*,*

$$P(A) = \sum_{k=1}^{\infty} P(B_k) \cdot P(A|B_k).$$

Proof:

$$P(A) = P(A \cap S) = P(A \cap (\cup_{k=1}^{\infty} B_k)) = P(\cup_{k=1}^{\infty} (A \cap B_k)) = \sum_{k=1}^{\infty} P(A \cap B_k) = \sum_{k=1}^{\infty} P(B_k) \cdot P(A|B_k). \quad \blacksquare$$

Formula

$$P(A) = \sum_{k=1}^{\infty} P(B_k) \cdot P(A|B_k). \tag{2}$$

is called *Formula of Total Probability*. As a consequence of this formula we obtain so-called Bayes's formula.

Theorem 2.3 (Bayes's formula) *Let B_1, B_2, ... be a finite or countable sequence of disjoint events with all $P(B_k) > 0$ and $\cup_{k=1}^{\infty} B_k = S$. Then, for any event A with $P(A) > 0$, we have*

$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{k=1}^{\infty} P(B_k) \cdot P(A|B_k)}.$$

Proof: Indeed, by (1) and (2), we obtain

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A \cap B_k)}{P(A)} = \frac{P(B_k) \cdot P(A|B_k)}{P(A)} = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{k=1}^{\infty} P(B_k) \cdot P(A|B_k)}. \quad \blacksquare$$

Example 2.8 *From a box containing M white balls and $N-M$ black balls, one ball of unknown color is lost. Find the probability of taking a white one.*

Solution: *Consider two events $B_1 = \{white\ ball\ is\ lost\}$ and $B_2 = \{black\ ball\ is\ lost\}$. Clearly, $P(B_1) = \frac{M}{N}$ and $P(B_2) = \frac{N-M}{N}$. Denote also*

$$A = \{a\ ball\ which\ is\ taken\ from\ the\ box\ is\ white\}.$$

By the formula of total probability,

$$P(A) = P(B_1) \cdot P(A|B_1) + P(B_2) \cdot P(A|B_2) = \frac{M}{N}\frac{M-1}{N-1} + \frac{N-M}{N}\frac{M}{N-1} = \frac{M}{N}$$

The probability of taking a white ball from the box, before *one ball was lost, is the same.* $\square$

2.4 Exercises

Exercise 2.1 *Coefficients p and q of the quadratic equation $x^2 + px + q = 0$ are taken randomly from the interval $[0, 1]$. Find the probability of the event R that the roots of the equation are real.*

Solution: Roots $x_{1,2} = \frac{-p \pm \sqrt{p^2-4q}}{2}$ are real iff $p^2 - 4q \geq 0$ or $q \leq \frac{1}{4}p^2$.

The sample space

$$S = \{(p,q)|0 \leq p \leq 1,\ 0 \leq q \leq 1\}$$

with the uniform probability $P(A) = \text{area}(A)$.

$$R = \{(p,q)|q \leq \frac{1}{4}p^2 \ \&\ (p,q) \in S\}.$$

$$P(R) = \text{area}(R) = \int_0^1 \frac{p^2}{4} dp = \frac{p^3}{12}\Big|_0^1 = \frac{1}{12}. \ \square$$

Exercise 2.2 *Let $A_1, A_2, ..., A_n$ be independent events with $P(A_k) = 1/2$ for all $k = 1, 2, ..., n$. Find $P(\bigcup_{k=1}^{k=n} A_k)$.*

Solution: By Proposition 2.1, the events $A_1^c, A_2^c, ..., A_n^c$ are independent. Therefore,

$$P(\bigcup_{k=1}^{k=n} A_k) = 1 - P([\bigcup_{k=1}^{k=n} A_k]^c) = 1 - P(\bigcap_{k=1}^{k=n} A_k^c) = 1 - \prod_{k=1}^{k=n} P(A_k^c) = 1 - \frac{1}{2^n} = \frac{2^n - 1}{2^n}. \ \square$$

Exercise 2.3 *Three fair dice are rolled. Denote by S_k the score on k-th die.* 1) *Given A=$\{S_1 = 2\}$, and B=$\{S_1 + S_2 \leq 5\}$.*
a) Find the conditional probability $P(A|B)$.
b) Are the events A and B independent?

2) *Given C=$\{S_1 = S_2\}$, D=$\{S_2 = S_3\}$, and G=$\{S_3 = S_1\}$.*
a) Is the set of events $\{C, D\}$ independent?
b) Is the set of events $\{C, D, G\}$ independent?

Solution: 1)a)

$P(B) = P(\{S_1 = 1, S_2 = 1\} \cup \{S_1 = 1, S_2 = 2\} \cup \{S_1 = 1, S_2 = 3\} \cup \{S_1 = 1, S_2 = 4\} \cup \{S_1 = 2, S_2 = 1\} \cup$

$\{S_1 = 2, S_2 = 2\} \cup \{S_1 = 2, S_2 = 3\} \cup \{S_1 = 3, S_2 = 1\} \cup \{S_1 = 3, S_2 = 2\} \cup \{S_1 = 4, S_2 = 1\}) = \frac{10}{6 \cdot 6} = \frac{10}{36}.$

$$P(A \cap B) = P(\{S_1 = 2, S_2 = 1\} \cup \{S_1 = 2, S_2 = 2\} \cup \{S_1 = 2, S_2 = 3\}) = \frac{3}{6 \cdot 6} = \frac{3}{36} = \frac{1}{12}.$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3}{10}.$$

1)b) No. Since

$$P(A \cap B) = \frac{3}{36} \neq P(A)P(B) = \frac{1}{6} \cdot \frac{10}{36} = \frac{5}{3} \cdot \frac{1}{36}.$$

2)a) Yes. Since

$$P(C \cap D) = P(S_1 = S_2 = S_3) = \frac{6}{6 \cdot 6 \cdot 6} = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \frac{6}{36} \cdot \frac{6}{36} = P(C) \cdot P(D).$$

2)b) No. Since

$$P(C \cap D \cap G) = P(S_1 = S_2 = S_3) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = P(C) \cdot P(D) \cdot P(G). \ \square$$

Exercise 2.4 *Three fair dice are rolled. Denote by S_k, the score on k-th die. Given A=$\{S_1 = 2\}$, and B=$\{S_1 + S_2 + S_3 \le 5\}$.*

a) Find the conditional probability $P(A|B)$.

b) Are the events A and B independent?

c) Given C=$\{S_1 = S_2\}$, D=$\{S_2 = S_3\}$, and G=$\{S_3 = S_1\}$. Is the set of events $\{C, D, G\}$ independent?

Solution: *a)*

$$P(B) = P((111), (112), (113), (121), (122), (131), (211), (221), (212), (311)) = \frac{10}{6^3}.$$

$$P(A \cap B) = P((121), (122), (221)) = \frac{3}{6^3}.$$

Hence

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3}{10}.$$

b)

$$P(A) = \frac{1}{6} \neq \frac{3}{10} = P(A|B).$$

Hence A is not independent with B.

c) Since $P(C) = P(D) = P(G) = \frac{6}{6 \cdot 6} = \frac{1}{6}$ then

$$P(C \cap D \cap G) = P(S_1 = S_2 = S_3) = \frac{6}{6 \cdot 6 \cdot 6} = \frac{1}{36} \neq P(C) \cdot P(D) \cdot P(G) = \frac{1}{216}.$$

Thus the set of events $\{C, D, G\}$ is not independent. □

Exercise 2.5 *There are 10 white and 5 green balls in the box H_1. In the box H_2 there are 3 white, 5 green, and 7 red balls. One of balls was transferred randomly from H_2 into H_1. Then we took randomly a ball from H_1. Find the probability $P(G)$ of the event that we got green one.*

Solution: Let $G_T = \{$a green ball was transferred$\}$ and $NG_T = \{$not a green ball was transferred$\}$.

Then $P(G) = P(G|G_T) \cdot P(G_T) + P(G|NG_T) \cdot P(NG_T) = \frac{6}{16} \cdot \frac{5}{15} + \frac{5}{16} \cdot \frac{10}{15} = \frac{6}{48} + \frac{10}{48} = \frac{1}{3}$. □

Exercise 2.6 *Let $\{A, B, C\}$ be an independent family of events with $P(A) = 1/2$, $P(B) = 1/3$, and $P(C) = 1/4$. Find $P(A \cup B \cup C)$.*

Solution: The set $\{A^c, B^c, C^c\}$ is independent by Proposition 2.1. Then

$$P(A \cup B \cup C) = 1 - P([A \cup B \cup C]^c) = 1 - P(A^c \cap B^c \cap C^c) = 1 - P(A^c) \cdot P(B^c) \cdot P(C^c) = 1 - (1 - 1/2) \cdot (1 - 1/3) \cdot (1 - 1/4) = \frac{3}{4}.$$

Exercise 2.7 *There are 5 white and 10 black balls in the box H_1. In the box H_2, there are 3 white and 7 black balls. One of balls was transferred randomly from H_2 into H_1. Then, we took randomly a ball from H_1. Find the probability of the event that this ball is white.*

Solution:

$A = \{\text{the transferred ball is white}\};\; B = \{\text{the transferred ball is black}\};\; C = \{\text{the taken ball is white}\}.$

$$P(A) = \frac{3}{10};\; P(B) = \frac{7}{10},\; P(C|A) = \frac{5+1}{16} = \frac{6}{16};\; P(C|B) = \frac{5}{16}.$$

$$P(C) = P(C \cap A) + P(C \cap B) = P(A) \cdot P(C|A) + P(B) \cdot P(C|B) = \frac{3}{10} \cdot \frac{6}{16} + \frac{7}{10} \cdot \frac{5}{16} = \frac{53}{160}. \quad \square$$

Exercise 2.8 *Initially, there are 3 red balls, 4 blue, and 2 white ones in a box. In a trail one of ball is taken randomly out of box without returning back. Consider a series of three trails.*

a) Find the probability of the event A that the ball in the first trail is red, it the second one is blue, and in the last one is white.

b) Find probability of the event B that all three balls taken out form the box are of different colors.

c) Find probability of the event C that all three balls taken out form the box are of the same color.

d) Find the conditional probability of A given B.

Solution:

a) $P(A) = P(RBW) = \frac{3}{9} \cdot \frac{4}{8} \cdot \frac{2}{7} = \frac{1}{21}$.

b) $P(B) = P(RBW) + P(RWB) + P(BRW) + P(BWR) + P(WRB) + P(WBR) = 6 \cdot \frac{1}{21} = \frac{2}{7}$.

c)

$$P(C) = P(RRR) + P(BBB) + P(WWW) = \frac{3}{9} \cdot \frac{2}{8} \cdot \frac{1}{7} + \frac{4}{9} \cdot \frac{3}{8} \cdot \frac{2}{7} + \frac{2}{9} \cdot \frac{1}{8} \cdot \frac{0}{7} = \frac{1}{4} \cdot \frac{1}{21} + \frac{1}{21} = \frac{5}{4} \cdot \frac{1}{21} = \frac{5}{84}.$$

d) $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/21}{2/7} = \frac{1}{6}$. $\square$

Exercise 2.9 *Given two independent* Uniform (0,2) *distributed random variables X and Y. Calculate the probability that $XY \le 1$.*

Solution: Let D be a subset of $\mathbb{R}^2$: $D = \{(x,y) \in \mathbb{R}^2 | 0 \le x \le 2 \;\&\; 0 \le y \le 2 \;\&\; xy \le 1\}$. This is region bounded by the lines $xy = 1$, $x = 0$, $x = 2$, $y = 0$, and $y = 2$.

$$P(XY \le 1) = \frac{1}{4} \iint\limits_D dxdy = \frac{1}{4} \cdot 2 \cdot \frac{1}{2} + \frac{1}{4} \int_{1/2}^{2} \frac{1}{x} dx = \frac{1}{4} + \frac{1}{4}\left(\ln 2 - \ln(2^{-1})\right) = \frac{1}{4} + \frac{1}{4} \cdot 2 \ln 2 = \frac{1}{4} + \frac{\ln 2}{2}. \quad \square$$

Exercise 2.10 *Pit and Bob agree to meet at a certain place some time between 23 and 24 o'clock. Pit will stay 20 min. Bob will stay 10 min. Assuming that the arrival times are independent and uniformly distributed, find the probability that they will meet.*

Solution:

$$P(\text{Bob meets Pit}) = \iint\limits_D dxdy = 1 - \frac{1}{2}\left(\left(\frac{2}{3}\right)^2 + \left(\frac{5}{6}\right)^2\right) = 1 - \frac{4}{18} - \frac{25}{72} = \frac{72 - 16 - 25}{72} = \frac{31}{72}. \quad \square$$

Exercise 2.11 *Pit and Bull take turns throwing a biased (=unfair) coin that produces a head with probability α. Pit is starting first, and he will win the game, if Pit gets a head. Bull will win, if Bull gets a tail.*
a) Find the probability $P(P^c \cap B^c) = P(TH, TH, \ldots, TH, \ldots)$ that none will win. Justify you answer.

b) Find the value(s) of α for which the game is fair: that is $P(P) = P(B)$, where $P(P)$ is the probability that Pit wins the game and $P(B)$ is the probability that Bull wins.

Solution: *a*)

$$0 \le P(P^c \cap B^c) \le P(\underbrace{TH, TH, \ldots, TH}_{n \text{ times}}) = [(1-\alpha)\alpha]^n \to 0 \;\Rightarrow P(P^c \cap B^c) = 0.$$

b) By a),

$$P(P) = P(B) \Rightarrow P(P) = P(B) = \frac{1}{2}.$$

Note that $P = \bigcup_{n=0}^{\infty} P_n$, where $P_0 = H$, $P_1 = THH$, $\ldots$, $P_n = (TH)^n H$.

$$P(P) = \sum_{n=0}^{\infty} P(P_n) = P(H) \sum_{n=0}^{\infty} P\left((TH)^n\right) = \alpha \cdot \sum_{n=0}^{\infty} (\alpha(1-\alpha))^n = \alpha \cdot \frac{1}{1-\alpha(1-\alpha)} = \frac{1}{2}$$

$$\Rightarrow 2\alpha = 1 - \alpha(1-\alpha) \Rightarrow \alpha^2 - 3\alpha + 1 = 0 \Rightarrow \alpha = \frac{3 \pm \sqrt{9-4}}{2} = \frac{3}{2} \pm \frac{\sqrt{5}}{2}.$$

Since $0 \le \alpha \le 1$ then

$$\alpha = \frac{3-\sqrt{5}}{2}. \quad \square$$

3 Random variables

In what follows, we denote a probability space by $(S, \mathcal{A}, P)$.

3.1 Definition of a random variable

Definition 3.1 *A function $f : S \to \mathbb{R}$ is called a* **random variable** *on S, if $f^{-1}((-\infty, a]) \in \mathcal{A}$ for every $a \in \mathbb{R}$.*
Two random variables f and g on S are said to be **equal almost surely** *$(f = g(a.s.))$ if $P(\{f = g\}) = 1$.*

By convention, *random variables are represented with capital roman letters*, and the event like $\{s \in S : X(s) \in B\}$, where B is a Borel subset of $\mathbb{R}$ and X is an RV, with the simplified notations $\{X \in B\}$. In the case when $B = \{b\}$, $B = \{r : r \leq b\}$, ..., we write $\{X = b\}$, $\{X \leq b\}$, etc.

Remind that the **Borel algebra** $\mathcal{B}(\mathbb{R})$ of subsets of the real line is defined as the σ-algebra generated by all intervals in $\mathbb{R}$. Clearly, it is enough to consider only intervals $(-\infty, a]$ with $a \in \mathbb{Q}$. Given a random variable X. Notice that $X^{-1}(B) \in \mathcal{A}$ for any $B \in \mathcal{B}(\mathbb{R})$.

Among random variables on S, the simplest are **indicator functions** of events (it can be shown that the indicator function $\mathbb{I}_A(s)$ of a subset A of the sample space S is a random variable iff A is an event).

Notice that not every function $f : S \to \mathbb{R}$ is a random variable. *For example, consider $S = \mathbb{Z}$, and the algebra of events $\mathcal{A}$ consisting of those subsets A of $\mathbb{Z}$ for which $a \in A$ iff $a^2 \in A$. It is left to the reader to check that, for the identity function $f(t) = t$, $f^{-1}((-\infty, a]) \notin \mathcal{A}$ for every $a \in \mathbb{R}$.*
□

It can be shown that the collection of all random variables on $(S, \mathcal{A}, P)$ is a vector space w.r. to usual operations on real-valued functions. A composition $f \circ X : S \to \mathbb{R}$ of any random variable X with a *Borel measurable* function $f : \mathbb{R} \to \mathbb{R}$ is a random variable.

Sometimes, the term *random vector* is used for a function $F : S \to \mathbb{R}^n$, all whose components are random variables.

Definition 3.2 *An RV X is* **discrete** *if there is at most countable set $B \subseteq \mathbb{R}$ such that $P(X \in B) = 1$. An RV X is* **continuous** *if $P(X = r) = 0$ for every $r \in \mathbb{R}$. An RV X which is neither discrete nor continuous is said to be* **mixed**.

The **probability function** of an RV X is a function $pf_X : \mathbb{R} \to [0, 1]$ such that $pf_X(r) = P(X = r)$ for all $r \in \mathbb{R}$. RV X is *continuous* iff $\sum_{r \in \mathbb{R}} pf_X(r) = 0$, and X is *discrete* iff $\sum_{r \in \mathbb{R}} pf_X(r) = 1$.

Here we include several examples of discrete RVs.

X is **discrete uniform** *if there is a finite set of reals $\{r_1, ..., r_n\}$ such that $pf_X(r_k) = \frac{1}{n}$ for all k, $1 \le k \le n$.*

X is p-**Bernoulli**, *where $0 \le p \le 1$, if $pf_X(1) = p$ and $pf_X(0) = 1 - p$.*

X is np-**Binomial**, *where $0 \le p \le 1$ and $n \in \mathbb{N}$, if*

$$pf_X(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \qquad (0 \le k \le n).$$

X is λ-**Poisson**, *where $\lambda > 0$, if $pf_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for every integer $k \ge 0$.* $\square$

3.2 Distribution function of a random variable

Definition 3.3 *Let $X : S \to \mathbb{R}$ be a random variable on S. Let*

$$F_X(r) \equiv P(X \le r), \; r \in \mathbb{R}\,.$$

Then F_X is called the **distribution function (DF)** *of X. The probability measure P_X on the Borel algebra $\mathcal{B}(\mathbb{R})$, given by*

$$P_X(B) \equiv P(X \in B), \quad B \in \mathcal{B}(\mathbb{R}),$$

is called the **probability distribution (PD)** *of X.*

One may also define the joint distribution function and the **joint probability distribution (JPD)** *of a* random vector *by similar formulas. We shall turn to this later. Notice that, in some books the distribution function might be defined by the slightly different formula $\mathcal{F}_X(r) \equiv P(\{s : X(s) < r\})$. Obviously,*

$$\mathcal{F}_X(r) = P(-r < -X < +\infty) = 1 - P(-\infty < -X \le -r) = 1 - F_{-X}(-r)$$

and $F_X(r) = 1 - \mathcal{F}_{-X}(-r)$. □

Any distribution function is *non-decreasing and continuous from the right*, that is, for any $a \in \mathbb{R}$,

$$F_X(a) = F_X(a^+) \equiv \lim_{0<h\to 0} F_X(a+h)$$

Notice also that $\mathcal{F}_X$ is *continuous from the left*.

Example 3.1

$$F_{\mathbb{I}_A}(r) = \begin{cases} 0, & r < 0; \\ 1-p, & 0 \le r < 1; \quad (p = P(A)) \\ 1, & 1 \le r. \end{cases}$$

Clearly, the distribution function of $\mathbb{I}_A$ depends only on $P(A)$, but not on A. PD of $\mathbb{I}_A$ is given by the formula $P_{\mathbb{I}_A}((-\infty, r]) = F_{\mathbb{I}_A}(r)$, hence

$$P_{\mathbb{I}_A}(B) = \begin{cases} 0, & \{0,1\} \cap B = \emptyset; \\ 1-p, & 0 \in B,\ 1 \notin B; \quad B \in \mathcal{B}(\mathbb{R}) \\ 1, & \{0,1\} \subseteq B. \end{cases} \quad \square$$

For any random variable X, the distribution function F_X of X is given by the formula

$$F_X(x) = \int_{-\infty}^{x} dF_X = \int_{-\infty}^{x} p_X(t)dt,$$

where

$$p_X(t) = F'_X(t) = \lim_{\Delta t\to 0} \frac{\Delta F_X(t)}{\Delta t} = \lim_{\Delta t\to 0} \frac{P(t < X \le t + \Delta t)}{\Delta t}$$

is the *density of the probability distribution* or the *distribution density*.

The density "function" of the distribution of the indicator function of A with $p = P(A)$ is given by using the **Dirac "function"** δ, which is formally defined by:

$$\int_A \delta(t)dt = \begin{cases} 0, & 0 \notin A, \\ 1, & 0 \in A. \end{cases}$$

Namely, $p_{\mathbb{I}_A}(t) = (1-p)\delta(t) + p\delta(t-1)$. Indeed, this "function" is the *generalized derivative* of $P_{\mathbb{I}_A}(A)$.

More generally, given a discrete RV X which takes values $t_1 < t_2 < \ldots < t_n$ with probabilities $P(\{t_k\}) = p_k$, $p_1 + p_2 + \ldots + p_n = 1$, then density of the distribution function of X is

$$p(t) = p_1\delta(t-t_1) + p_2\delta(t-t_2) + \ldots + p_k\delta(t-t_k).$$

By using the **Heaviside function**

$$H(t) = \begin{cases} 0, & t < 0; \\ 1, & t \geq 0, \end{cases}$$

we may write the distribution function of X as follows

$$F_X(t) = \sum_{i=1}^{n} p_i H(t-t_i).$$

Notice that the *generalized derivative* of $H(t)$ is the "function" δ.

Example 3.2 *Find $a \in \mathbb{R}$ such that the function*

$$p(t) = \begin{cases} 0, & t \leq 0; \\ \frac{a}{\sqrt{t}}, & 0 < t \leq 1; \\ 0, & t > 1 \end{cases}$$

is the distribution density of some random variable X. Find its distribution function and the probability of $\{-\frac{1}{3} \leq X < \frac{4}{5}\}$.

Solution. *Using $\int_{-\infty}^{+\infty} p(t)dt = 1$, obtain $a = \frac{1}{2}$. Thus,*

$$F_X(t) = \int_{-\infty}^{t} p(s)ds = \int_{-\infty}^{t} \frac{ds}{2\sqrt{s}} = \begin{cases} 0, & t \leq 0; \\ \sqrt{t}, & 0 \leq t \leq 1; \\ 1, & t > 1. \end{cases}$$

Since $F_X(t)$ is continuous,

$$P\Big\{-\frac{1}{3} \leq X < \frac{4}{5}\Big\} = F_X\Big(\frac{4}{5}\Big) - F_X\Big(-\frac{1}{3}\Big) = F_X\Big(\frac{4}{5}\Big) = \frac{2}{\sqrt{5}}. \quad \square$$

Example 3.3 *Given the distribution function*

$$F_X(t) = \begin{cases} 0, & t < 2; \\ (t-2)^2, & 2 \leq t \leq 3; \\ 1, & t > 3 \end{cases}$$

of a random variable X. Find the probability of $\{\frac{5}{2} < X < \frac{7}{2}\}$. Find the distribution density of X.

Solution. *Since $F_X(t)$ is continuous,*

$$P\left\{\frac{5}{2} < X < \frac{7}{2}\right\} = F_X\left(\frac{7}{2}\right) - F_X\left(\frac{2}{5}\right) = 1 - \left(\frac{5}{2} - 2\right)^2 = \frac{3}{4}.$$

Using $p(t) = F'_X(t)$, obtain

$$p(t) = \begin{cases} 0, & t < 2, \\ 2(t-2), & 2 \le t \le 3, \\ 0, & t > 3. \end{cases} \quad \square$$

Example 3.4 *Given a sequence of n Bernoulli trials with the probability of success p (it is denoted by (n, p)). There are two typical problems related to (n, p). The first one is to find the probability $P(m)$ that the success in (n, p) appears m times. The second one is to find the most probable number k of successes in (n, p). The first problem being solved by using the classical Bernoulli formula*

$$P_n(m) = C_n^m p^m (1-p)^{n-m}, \quad C_n^m = \frac{n!}{m!(n-m)!}. \tag{3}$$

For instance, the probability of $\{\ge m\}$ successes in (n, p) is

$$P_n(\{\ge m\}) = \sum_{i=m}^{n} C_n^i p^i (1-p)^{n-i}.$$

To solve the second problem, we begin with the following two inequalities:

$$P_n(k-1) \le P_n(k), \quad P_n(k+1) \le P_n(k).$$

From the first inequality, by using (3), one gets $\frac{C_n^{k-1}}{C_n^k} \le \frac{1-p}{p}$ or just $k \le (n+1)p$. From the second one, we obtain $\frac{C_n^{k+1}}{C_n^k} \le \frac{1-p}{p}$ or $(n+1)p \le k+1$. Thus

$$k \le (n+1)p \le k+1. \tag{4}$$

It follows from (4) that if $(n+1)p \notin \mathbb{N}$, then there is only one solution k. If $(n+1)p \in \mathbb{N}$, then there are two most probable numbers, $k = (n+1)p$ and $k = (n+1)p - 1$. The situation becomes more clear when we consider the graph of the following "density"

$$p(t) = \begin{cases} 0, & t < 0; \\ P_n(i), & i \le t < i+1; \\ 0, & t \ge n. \end{cases} \quad \square$$

3.3 Distribution function of a random vector

Definition 3.4 *Given a system $\vec{X} = \langle X_1, X_2, \ldots X_n \rangle$ of random variables on S. Sometimes we shall call $\vec{X}$ a* **random vector** *on S. The probability measure $P_{\vec{X}}$ on the Borel algebra $\mathcal{B}(\mathbb{R}^n)$, given by*

$$P_{\vec{X}}(A) \equiv P(\vec{X}^{-1}(A)), \ A \in \mathcal{B}(\mathbb{R}^n),$$

is called the **probability distribution** *of $\vec{X}$.*

The **probability function** of an RV $\vec{X} = \langle X_1, \ldots, X_n \rangle$ is a function $pf_{\vec{X}} : \mathbb{R}^n \to [0, 1]$, such that,, for all $\vec{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n$,

$$pf_{\vec{X}}(\vec{r}) = P(X_1 = r_1 \,\&\ldots\&\, X_n = r_n).$$

The **probability density** of a continuous RV $\vec{X}$ is a function $p_{\vec{X}} : \mathbb{R}^n \to \mathbb{R}_+$ such that, for all $\vec{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n$ and for all $\vec{\Delta r} = (\Delta r_1, \ldots, \Delta r_n) \in \mathbb{R}^n$,

$$P(\vec{X} \in [\vec{r}, \vec{r} + \vec{\Delta r}]) = \int_{r_1}^{r_1+\Delta r_1} \ldots \int_{r_n}^{r_n+\Delta r_n} p_{\vec{X}}(\vec{r}) dr_1 \ldots dr_n.$$

The **distribution function** of an RV $\vec{X} = \langle X_1, \ldots, X_n \rangle$ is a function $F_{\vec{X}} : \mathbb{R}^n \to [0, 1]$ such that

$$F_{\vec{X}}(\vec{r}) = P(X_1 \leq r_1 \,\&\ldots\&\, X_n \leq r_n).$$

If a random vector $\vec{X}$ is discrete (*that is: all* X_k *are discrete for* $1 \leq k \leq n$) then

$$F_{\vec{X}}(\vec{r}) = \sum_{\{t_k \leq r_k,\ 1 \leq k \leq n\}} pf_{\vec{X}}(\vec{t}\,),$$

where $\vec{r} = (r_1, \ldots, r_n)$ and $\vec{t} = (t_1, \ldots, t_n)$.

If a random vector $\vec{X}$ is continuous (*that is: all* X_k *are continuous for* $1 \leq k \leq n$) then

$$F_{\vec{X}}(\vec{r}) = \int_{-\infty}^{r_1} \ldots \int_{-\infty}^{r_n} p_{\vec{X}}(\vec{t}\,) dt_1 \ldots dt_n.$$

3.4 Independent families of random variables

Definition 3.5 *A collection* $\{X_i\}_{i \in I}$ *of random variables on the same probability space is called* **independent** *if the collection of events* $\{X_i^{-1}(-\infty, r_i)\}_{i \in I}$ *is independent for arbitrary* $r_i \in \mathbb{R}$.

It follows immediately that, for an independent collection $\{X_i\}_{i\in I}$ of RVs, any collection $\{B_i\}_{i\in I}$ of Borel subsets of $\mathbb{R}$, and any distinct indices $i_1, i_2, \ldots i_k$ in I, we have

$$P(X_{i_1} \in B_{i_1}, \ldots X_{i_k} \in B_{i_k}) = P(X_{i_1} \in B_{i_1}) \cdot \ldots \cdot P(X_{i_k} \in B_{i_k}).$$

From this remark we obtain

Proposition 3.1 *For any* independent *collection* $\{X_i\}_{i\in I}$ *of random variables on the same probability space and for* Borel measurable *functions* $\{f_i\}_{i\in I}$*, the collection* $\{f_i \circ X_i\}_{i\in I}$ *of random variables is independent.*

Proof: For any distinct $i_1, i_2, \ldots i_k$ in I and any $r_1, r_2, \ldots r_k$ in $\mathbb{R}$, we have (for Borel sets $B_{i_j} = f_{i_j}^{-1}(-\infty, r_j)$):

$$P\Big(\bigcap_{j=1}^{j=k} (f_{i_j} \circ X_{i_j})^{-1}(-\infty, r_j)\Big) = P\Big(\bigcap_{j=1}^{j=k} X_{i_j}^{-1}(f_{i_j}^{-1}(-\infty, r_j))\Big) = \prod_{j=1}^{j=k} P(X_{i_j}^{-1}(f_{i_j}^{-1}(-\infty, r_j))) = \prod_{j=1}^{j=k} P((f_{i_j} \circ X_{i_j})^{-1}(-\infty, r$$

Proposition 3.2 *Given a random vector* $\vec{X} = \langle X_1, X_2, \ldots X_n\rangle$ *(on the same probability space, of course). Then* $\vec{X}$ *is independent iff*

$$F_{\vec{X}}(x_1, x_2, \ldots x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \ldots \cdot F_{X_n}(x_n).$$

Proof: Consider the case $n = 2$ only (the general case is left to the reader as an exercise).

Sufficiency:

$$F_X(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2) = P(\{X_1 \le x_1\} \cap \{X_2 \le x_2\}) = P(X_1 \le x_1) \cdot P(X_2 \le x_2) = F_{X_1}(x_1) \cdot F_{X_2}(x_2). \quad (5)$$

Necessity: By (5), the necessity follows from Definition 3.5. ■

Definition 3.6 *If a random vector* $\vec{X} = \langle X_1, X_2, \ldots, X_n\rangle$ *has the DF* $F_{\vec{X}}$ *which is* n *times continuously differentiable in some domain* $D \subseteq \mathbb{R}^n$*, then the restriction of* **distribution density** *of* $\vec{X}$ *on* D *is*

$$p_{\vec{X}}(\vec{x}) = \frac{\partial^n F_{\vec{X}}(\vec{x})}{\partial x_1 \partial x_2 \ldots \partial x_n}(\vec{x}) \qquad (\vec{x} \in D).$$

Indeed, we may define DD of an RV X without any assumption of differentiability of DF F_X by using generalized derivatives as in the one-dimensional case.

One of the main reasons for introducing DD is the following formula:

$$P(X \in D) = \int_D p_{\vec{X}}(\vec{x}) d\vec{x} \quad (D \in \mathcal{B}(\mathbb{R}^n)).$$

Proposition 3.3 *For any independent RV $\vec{X} = \langle X_1, \ldots, X_n \rangle$,*

$$p_{\vec{X}}(\vec{x}) \equiv p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n). \quad (\vec{x} \in \mathbb{R}^n)$$

Proof:

$$\int\limits_{\{x_k \le r_k:\ k=1,2,\ldots n\}} p_{\vec{X}}(\vec{x}) d\vec{x} = F_{\vec{X}}(\vec{r}) = F_{X_1}(r_1) \cdot \ldots \cdot F_{X_n}(r_n) =$$

$$\int_{-\infty}^{r_1} p_{X_1}(x_1) dx_1 \cdot \ldots \cdot \int_{-\infty}^{r_n} p_{X_n}(x_n) dx_n = \int\limits_{\{x_k \le r_k:\ k=1,2,\ldots n\}} p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n) d\vec{x}.$$

Therefore, $p_{\vec{X}}(\vec{x}) \equiv p_{X_1}(x_1) \cdot \ldots \cdot p_{X_n}(x_n)$. ■

3.5 Convolution theorem

Theorem 3.1 (Convolution theorem) *Let X and Y be independent RVs with distribution densities p_X and p_Y. Then $X + Y$ has a distribution density given by*

$$p_{X+Y}(r) = \int\limits_{-\infty}^{\infty} p_X(x) p_Y(r - x) dx = \int\limits_{-\infty}^{\infty} p_Y(y) p_X(r - y) dy. \quad (6)$$

Proof: Since

$$F_{X+Y}(t) = \int\limits_{-\infty}^{t} p_{X+Y}(r)\, dr = P(X + Y \le t) = \iint\limits_{x+y \le t} p_{\langle X,Y \rangle}(x, y)\, dx\, dy = \iint\limits_{x+y \le t} p_X(x) p_Y(y)\, dx\, dy =$$

$$\int\limits_{-\infty}^{\infty} p_X(x) \Bigg[\int\limits_{-\infty}^{t-x} p_Y(y)\, dy \Bigg] dx = \int\limits_{-\infty}^{\infty} p_X(x) \Bigg[\int\limits_{-\infty}^{t} p_Y(r - x)\, dr \Bigg] dx = \int\limits_{-\infty}^{t} \Bigg[\int\limits_{-\infty}^{\infty} p_X(x) p_Y(r - x)\, dx \Bigg] dr,$$

then $p_{X+Y}(r) = \int\limits_{-\infty}^{\infty} p_X(x) p_Y(r - x)\, dx$. The second equality in (6) is due to symmetry. ■

Example 3.5 *Let* X *and* Y *be* independent and identically distributed *RVs* **(iid RVs)** *with* $p_X(t) = p_Y(t) = \frac{t}{2}\mathbb{I}_{(0,2)}(t)$. *Then*

$$p_{X+Y}(s) = \int_{-\infty}^{\infty} p_X(s-x)p_Y(x)\,dx = \frac{1}{4}\int_{-\infty}^{\infty} (s-x)x\mathbb{I}_{(s-2,s)}(x)\mathbb{I}_{(0,2)}(x)\,dx =$$

$$\frac{1}{4}\int_0^2 (s-x)x\mathbb{I}_{(s-2,s)}(x)\,dx = \frac{1}{4}\begin{cases} 0, & s \notin [0,4]; \\ \int_0^s (s-x)x\,dx, & 0 \le s < 2; \\ \int_{s-2}^s (s-x)x\,dx, & 2 \le s \le 4. \end{cases} =$$

$$\frac{1}{4}\left(\mathbb{I}_{[0,2)}(s)\int_0^s (s-x)x\,dx + \mathbb{I}_{[2,4]}(s)\int_{s-2}^s (s-x)x\,dx\right) = \mathbb{I}_{[0,2)}(s)\frac{s^3}{24} + \mathbb{I}_{[2,4]}(s)\left[-\frac{2}{3} + \frac{s}{2} - \frac{s(s-2)^2}{8} + \frac{(s-2)^3}{12}\right]. \ \square$$

3.6 Conditional DF and Bayes' Rule

Let us consider shortly a system of two RVs on the same sample space S, say $\langle X, Y\rangle$.

Definition 3.7 *The* **marginal probability density** *of* $\langle X, Y\rangle$ *are defined as follows*

$$p_X^m(x) = \int_{-\infty}^{+\infty} p_{\langle X,Y\rangle}(x,y)dy, \quad p_Y^m(y) = \int_{-\infty}^{+\infty} p_{\langle X,Y\rangle}(x,y)dx\,.$$

The marginal PD $p_X^m(x)$ is the usual probability density $p_X(x)$ of X, if Y is not considered.

Definition 3.8 *Assume that* $p_X(x) > 0$. *Then the* **conditional distribution** *of* Y *given* $X = x$ *is defined by*

$$F_Y(y|x) = \frac{\int_{-\infty}^{y} p_{\langle X,Y\rangle}(x,t)dt}{p_X(x)},$$

and the **conditional probability density** *of* Y *given* $X = x$ *is*

$$p_Y(y|x) = \frac{p_{\langle X,Y\rangle}(x,y)}{p_X(x)}.$$

[*The last formula can be considered as a version of* Bayes's theorem *for densities.*]

Notice that if the system $\langle X, Y \rangle$ is independent, then

$$p_Y(y|x) = \frac{p_{\langle X,Y \rangle}(x,y)}{p_X(x)} = \frac{p_X(x) \cdot p_Y(y)}{p_X(x)} = p_Y(y) \, .$$

One of most typical cases of Bayes rule occurs when X is a continuous RV and Y is a discrete RV. Then:

$$p_{Y|X}(y|x) = \frac{p_{X|Y}(x|y) \cdot p_Y(y)}{p_X(x)},$$

where the **conditional probability mass function of Y given X** is

$$p_{Y|X}(y|x) = \lim_{\Delta x \to 0} P(Y = y | X \in [x, x + \Delta x]).$$

In this case:

$$P(X \leq a | Y = y) = \int_{-\infty}^{a} p_{X|Y}(x|y) dx.$$

Indeed, applying Bayes rule to the events $Y = y$ and $x \leq X \leq x + \Delta x$

$$\lim_{\Delta x \to 0} P(Y = y | x \leq X \leq x + \Delta x) = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)|Y = y)P(Y = y}{P(x \leq X \leq x + \Delta x)}$$

$$\lim_{\Delta x \to 0} \frac{\left(\int_x^{x+\Delta x} p_{Y|X}(s|y) ds \right) \cdot P(Y = y)}{\int_x^{x+\Delta x} p_Y(s) ds} = \lim_{\Delta x \to 0} \frac{\Delta x p_{Y|X}(x|y) p_Y(y)}{\Delta x p_Y(y)} = p_{Y|X}(x|y$$

Then $P(X \leq a | Y = y) = \int\limits_{-\infty}^{a} p_{X|Y}(x|y) dx.$ ■

3.7 Exercises

Exercise 3.1 *Given the following function* $p(t) = \begin{cases} \alpha(4|t| - t^2) & if \quad |t| \leq 2, \\ 0 & if \quad |t| > 2. \end{cases}$

a) Find α for which $p(t) = p_X(t)$ is the distribution density of a random variable X.

b) For the value of α which is found in a) and for the correspondent RV X:
(i) Find the distribution function $F_X(r)$ for $r \geq 0$.
(ii) Find the probability of event $-1 \leq X \leq 1$.

c) Find the values of $F_{X^2}(1)$ and $F_{X^3}(1)$.

Solution: *a)*

$$\int_{-\infty}^{\infty} p(t)dt = \int_{-2}^{2} \alpha(4|t| - t^2)dt = 2\alpha \int_{0}^{2} (4t - t^2)dt = 2\alpha \left(2t^2 - \frac{1}{3}t^3\right)|_0^2 = 2\alpha\left(8 - \frac{8}{3}\right) = 16\alpha \cdot \frac{2}{3} = \frac{32}{3}\alpha = 1.$$

So we have $\alpha = 3/32$.

b)(i)

$$F_X(r) = \int_{-\infty}^{r} p(t)dt = \frac{1}{2} + \int_{0}^{r} \frac{3}{32}(4t - t^2)dt = \frac{1}{2} + \frac{3}{32}\left[2t^2 - \frac{1}{3}t^3\right]|_0^r = \frac{1}{2} + \frac{3}{32}\left[2r^2 - \frac{r^3}{3}\right] = \frac{1}{2} + \frac{3}{16}r^2 - \frac{1}{32}r^3.$$

b)(ii)

$$P(-1 \le X \le 1) = \int_{-1}^{1} p(t)dt = 2\int_{0}^{1} \frac{3}{32}(4t - t^2)dt = \frac{6}{32}\left[2t^2 - \frac{1}{3}t^3\right]|_0^1 = \frac{3}{16}\left(2 - \frac{1}{3}\right) = \frac{5}{16}.$$

c) $F_{X^2}(1) = P(X^2 \le 1) = P(-1 \le X \le 1) = \frac{5}{16}$, by b(ii).

$F_{X^3}(1) = P(X^3 \le 1) = P(X \le 1) = [\text{see b(i)}] = \frac{1}{2} + \frac{3}{16} - \frac{1}{32} = \frac{16+6-1}{32} = \frac{21}{32}$. $\square$

Exercise 3.2 *Random variables X and Y are independent and (1)-exponentially distributed.*

a) Calculate the probability that $1 \le Y \le X \le 2$.

b) Calculate the probability density of $X + Y$.

Solution: a) Let D be a subset of $\mathbb{R}^2$: $D = \{(x, y) \in \mathbb{R}^2 | 1 \le y \le x \le 2\}$. This is the triangle bounded by the lines $y = x$, $x = 1$, and $x = 2$.

$$P(1 \le Y \le X \le 2) = \int\int_D p_{\langle X,Y\rangle}(x, y)dxdy = [\text{X and Y are indep.}] = \int\int_D p_X(x)p_Y(y)dxdy = \int_1^2 e^{-x}\left(\int_1^x e^{-y}dy\right) dx$$

$$\int_1^2 e^{-x}(e - e^{-x})dx = e\int_1^2 e^{-x}dx - \int_1^2 e^{-2x}dx = e(e^{-1} - e)^{-2} + \frac{1}{2}\int_1^2 (e^{-2x})'dx = 1 - \frac{1}{e} + \frac{1}{2}e^{-2x}\Big|_1^2 = 1 - \frac{1}{e} + \frac{1}{2e^4} - \frac{1}{2e^2}.$$

b)

$$p_{X+Y}(s) = \int_{-\infty}^{\infty} p_X(s - t)p_Y(t)dt = \int_{-\infty}^{\infty} e^{-(s-t)} \cdot \mathbb{I}_{[0,\infty)}(s - t) \cdot e^{-t} \cdot \mathbb{I}_{[0,\infty)}(t)dt = \int_0^{\infty} e^{-s} \cdot \mathbb{I}_{[0,\infty)}(s - t)dt =$$

$$\int_0^{\infty} e^{-s} \cdot \mathbb{I}_{(-\infty,s]}(t)dt = \begin{cases} 0 & \text{if } s < 0; \\ \int_0^s e^{-s}dt = se^{-s} & \text{if } s \ge 0. \end{cases} \quad \square$$

Exercise 3.3 *Let X and Y be independent (1)-exponentially distributed random variables, i.e.*

$$F_X(t) = F_Y(t) = \begin{cases} 0 & if \quad t < 0, \\ 1 - e^{-t} & if \quad t \geq 0. \end{cases}$$

Find the probability of $0 \leq Y \leq X \leq 1$.

Solution: Let $D = \{(x, y) \in \mathbb{R}^2 | 0 \leq y \leq x \leq 1\}$. This is the triangle bounded by the lines $x = y$, $x = 1$, and $y = 0$.

$$P(0 \leq Y \leq X \leq 1) = \int\int_D P_{\langle X,Y\rangle}(x, y)dxdy = [\text{X and Y are independent}] = \int\int_D P_X(x)P_Y(y)dxdy =$$

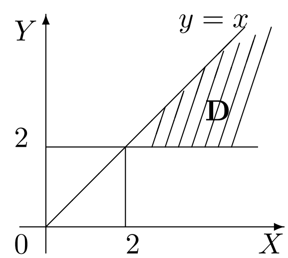
$$\int_0^1 e^{-x}\left[\int_0^x e^{-y}dy\right]dx = \int_0^1 e^{-x}(1-e^{-x})dx = -e^{-x}|_0^1 - (-\frac{1}{2}e^{-2x})|_0^1 = -e^{-1}+1+\frac{1}{2}e^{-2}-\frac{1}{2} = \frac{1}{2}-\frac{1}{e}-\frac{1}{2e^2}. \square$$

Exercise 3.4 *Random variables X and Y are independent and* **(1)-exponentially distributed** *(that is $p_X(t) = p_Y(t) = \mathbb{I}_{\mathbb{R}_+}(t)e^{-t}$). Calculate the probability that $X \geq Y \geq 2$.*

Solution. Let D be a subset of $\mathbb{R}^2$:

$$D = \{(x, y) \in \mathbb{R}^2 | x \geq y \geq 2\}.$$

This is the open angle bounded by the lines $y = x$ and $y = 2$.



$$P(X \geq Y \geq 2) = \int\int_D p_{\langle X,Y\rangle}(x, y)dxdy \underset{\text{because of the independence}}{=}$$

$$\int\int_D p_X(x)p_Y(y)dxdy = \int_2^\infty e^{-x}\left(\int_2^x e^{-y}dy\right)dx = \int_2^\infty e^{-x}(e^{-2} - e^{-x})dx =$$

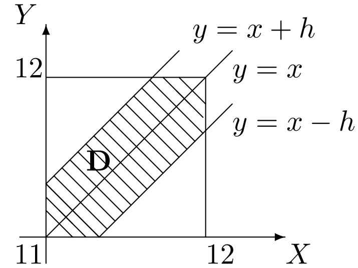
$$e^{-2}\int_2^\infty e^{-x}dx - \int_2^\infty e^{-2x}dx = e^{-4} - \frac{1}{2}e^{-4} = \frac{1}{2}e^{-4}. \square$$

Exercise 3.5 *Two friends agree to meet at a certain place some time between 11 and 12 o'clock. They agree that the one arriving first will wait h hours, $0 \leq h \leq 1$, for other to arrive. Assuming that the arrival times are independent and* **uniformly distributed**, *find the probability that they will meet.*

Solution.

Denote by X the arrival time of the first friend and by Y of the second one.
Denote by D the strip $|x - y| \leq h$.

Then



$$P(|X-Y| \le h) = \iint\limits_D p_{\langle X,Y\rangle}(x,y)dxdy \underset{\text{because of the independence}}{=} \iint\limits_D p_X(x)p_Y(y)dxdy =$$

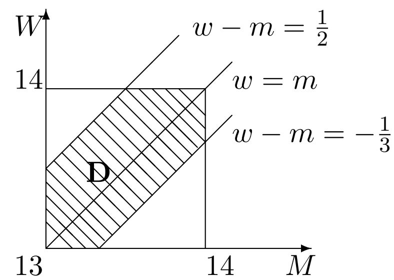
$$\iint\limits_D \mathbb{I}_{[11,12]}(x) \cdot \mathbb{I}_{[11,12]}(y)dxdy = \iint\limits_{D\cap[11,12]^2} dxdy \underset{\text{the shadowed area}}{=} 1-(1-h)^2 = 2h-h^2. \square$$

Exercise 3.6 *A man and a woman agree to meet at a cafe some time between 13 and 14 o'clock. A man will stay 30 min. A woman will stay 20 min. Assuming that the arrival times are independent and uniformly distributed, find the probability that they will meet.*

Solution:

Denote by D the strip $\frac{1}{2} \ge w-m \ge -\frac{1}{3}$.

Then



$$P(D) = \iint\limits_D p_{\langle M,W\rangle}(m,w)dmdw \underset{\text{because of the independence}}{=} \iint\limits_D p_M(m)p_W(w)dmdw =$$

$$\iint\limits_D \mathbb{I}_{[13,14]}(m) \cdot \mathbb{I}_{[13,14]}(w)dmdw = \iint\limits_{D\cap[13,14]\times[13,14]} dmdw \underset{\text{the shadowed area}}{=}$$

$$1 - \frac{1}{2}\cdot\left(\frac{1}{2}\right)^2 - \frac{1}{2}\cdot\left(\frac{2}{3}\right)^2 = \frac{72-9-16}{72} = \frac{47}{72}. \square$$

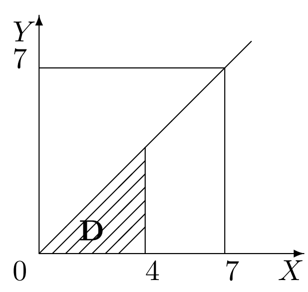
Exercise 3.7 *Given two independent* $(0,7)$*-uniform RVs* X *and* Y*. Calculate the probability* $P(0 < Y < X < 4)$*.*

Solution:

Let $D = \{0 < Y < X < 4\}$.

$$P(0<Y<X<4) =$$

$$\left(\frac{1}{7}\right)^2 \cdot \frac{1}{2}\cdot 4^2 = \frac{1}{2}\cdot\frac{1}{2}\cdot 16 = \frac{8}{49}. \square$$



Exercise 3.8 *Let* X *and* Y *be independent* (1)*-exponentially distributed random variables Find* $F_{X+Y}(3)$ *[hint: use the convolution theorem].*

Solution: Remark that $F_X(t) = F_Y(t) = (1 - e^{-t}) \cdot \mathbb{I}_{[0,\infty)}(t)$.

$P_X(t) = \frac{d}{dt}F_X(t) = \begin{cases} 0 & if \quad t < 0 \\ e^{-t} & if \quad t \geq 0 \end{cases}$ and $P_Y(t) = P_X(t)$. Then, by Theorem 3.1,

$$p_{X+Y}(s) = \int\limits_{-\infty}^{\infty} p_X(s-t)p_Y(t)dt = \int\limits_{-\infty}^{\infty} e^{-(s-t)}\mathbb{I}_{[0,\infty)}(s-t) \cdot e^{-t}\mathbb{I}_{[0,\infty)}(t)dt =$$

$$\int\limits_{0}^{\infty} e^{-s}\mathbb{I}_{[0,\infty)}(s-t)dt = \int\limits_{0}^{\infty} e^{-s}\mathbb{I}_{(-\infty,s]}(t)dt = \begin{cases} 0 & if \quad s < 0 \\ se^{-s} & if \quad s \geq 0 \end{cases} = se^{-s}\mathbb{I}_{[0,\infty)}(s).$$

$F_{X+Y}(r) = \int\limits_{-\infty}^{r} p_{X+Y}(s)ds = \int\limits_{0}^{r} se^{-s}ds = s(-e^{-s})|_0^r - \int\limits_{0}^{r}(-e^{-s})ds = -re^{-r} - e^{-s}|_0^r = -re^{-r} - e^{-r} + 1 = 1 - (r+1)e^{-r}.$

$F_{X+Y}(3) = 1 - (3+1)e^{-3} = 1 - 4e^{-3}$. □

Exercise 3.9 *Random variables X and Y are independent and* Exponential (β) *distributed with $EX = EY = 2$.*
a) *Find the value of β.*
b) *Calculate the probability density $p_{X+Y}(t)$ of $X + Y$.*

Solution: a)

$$2 = EX = \int\limits_{0}^{\infty} t \cdot \frac{1}{\beta} \cdot e^{-t/\beta}dt = \left(-t \cdot e^{-t/\beta}|_0^{\infty} - \int\limits_{0}^{\infty}(-e^{-t/\beta})dt \right) = \int\limits_{0}^{\infty} e^{-t/\beta}dt = -\beta \cdot e^{-t/\beta}|_0^{\infty} = \beta.$$

b) By using of the convolution theorem,

$$p_{X+Y}(s) = \int\limits_{-\infty}^{\infty} p_X(s-t) \cdot p_Y(t) \cdot \mathbb{I}_{(0,\infty)}(s-t) \cdot \mathbb{I}_{(0,\infty)}(t)dt = \int\limits_{0}^{s} \frac{1}{2} \cdot e^{-(s-t)/2} \cdot \frac{1}{2} \cdot e^{-t/2}dt = \frac{1}{4}\int\limits_{0}^{s} e^{-s/2}dt =$$

$$\frac{1}{4} \cdot s \cdot e^{-s/2} \text{ for } s \geq 0; \text{ and } p_{X+Y}(s) \equiv 0 \text{ for } s < 0. \ \square$$

Exercise 3.10 *Given two independent random variables X and Y with densities*

$$p_X(t) = p_Y(t) = 3t^2 \cdot \mathbb{I}_{[0,1]}(t).$$

a) *Find the distribution function $F_{-X}(t)$ of the random variable $-X$.*

b) *Find the density $p_{X+Y}(t)$ of the random variable $X + Y$.*

Solution:
a)

$F_{-X}(t) = P(-X \leq t) = P(X \geq -t) = 1 - P(X \leq -t) = 1 - \int\limits_{-\infty}^{-t} 3s^2 \cdot \mathbb{I}_{[0,1]}(s)ds = 1 - \int\limits_{0}^{-t} 3s^2 ds = 1 + t^3$ for $t \in [-1, 0]$;

and

$$F_{-X}(t) = 1 \text{ for } t \geq 0$$
$$F_{-X}(t) = 0 \text{ for } t \leq -1.$$

b)

$$p_{X+Y}(a) = \int_{-\infty}^{\infty} p_X(a-t)p_Y(t)dt = \int_{-\infty}^{\infty} 9(a-t)^2t^2 \cdot \mathbb{I}_{\{0\leq a-t\leq 1\}}(t) \cdot \mathbb{I}_{\{0\leq t\leq 1\}}(t)dt = 9\int_{-\infty}^{\infty} (a^4-2at^3+a^2t^2) \cdot \mathbb{I}_{\{a-1\leq t\leq a\}}(t) \cdot \mathbb{I}_{\{0\leq t}$$

So, for $0 \leq a \leq 1$, we have

$$p_{X+Y}(a) = 9\int_0^a (a^4 - 2at^3 + a^2t^2)dt = 9\left(\frac{t^5}{5} - \frac{2at^4}{4} + \frac{a^2t^3}{3}\right)\Bigg|_0^a = \frac{3}{10}a^5.$$

And, for $1 \leq a \leq 2$, we have

$$p_{X+Y}(a) = 9\int_{a-1}^{1} (a^4-2at^3+a^2t^2)dt = 9\left(\frac{t^5}{5} - \frac{2at^4}{4} + \frac{a^2t^3}{3}\right)\Bigg|_{a-1}^{1} = 9\left(\frac{1}{5} - \frac{2a}{4} + \frac{a^2}{3}\right) - 9\left(\frac{(a-1)^5}{5} - \frac{2a(a-1)^4}{4} + \frac{a^2}{}\right.$$

For $a \notin [0,2]$, we have $p_{X+Y}(a) = 0$. $\square$

4 Mathematical expectation and variance

4.1 The expectation

Definition 4.1 *Given a random variable X on $(S, \mathcal{A}, P)$, then the* **expected value** (**mathematical expectation**, **average value**, *or* **mean**) *of X, denoted by $E(X) = EX$, is defined as the* Lebesgue integral

$$EX \equiv \int_S X dP = \int_S X(\omega)P(d\omega), \tag{7}$$

provided that the integral is well defined. That is, at least one of two values $\int_S X_+ dP$ and $\int_S X_- dP$ is finite. If X has a distribution density $p(t)$, then

$$EX = \int_{-\infty}^{+\infty} sp(s)ds. \tag{8}$$

Notice that in most cases which are considered in this course, the definition of the expectation by formula (8) is applicable (although, the distribution density $p_X(t)$ may involve the Dirac "function"). The proof of the following proposition is based on the Lebesgue integration and therefore is omitted. [*Try to understand the formula (8) for $g = \mathbb{I}_{(-\infty,a]}$, where $r \in \mathbb{R}$. The rest of the proof is based on an approximation.*]

Proposition 4.1 (The change of variables) *Let X be a random variable on $(S, \mathcal{A}, P)$. For any Borel measurable $g(t) : \mathbb{R} \to \mathbb{R}$,*

$$\int_S |g \circ X| dP = \int_{-\infty}^{+\infty} |g(t)| P_X(dt) = \int_{-\infty}^{+\infty} |s| P_{g \circ X}(ds).$$

In particular, if $\int_S |g \circ X| dP < \infty$, then

$$E(g \circ X) = \int_{-\infty}^{+\infty} g(t) P_X(dt) = \int_{-\infty}^{+\infty} s P_{g \circ X}(ds)\,.$$

When g is non-negative, then $E(g \circ X) = \int_{-\infty}^{+\infty} g(t) P_X(dt)$ even if $E(g \circ X) = \infty$. ■

If $EX = 0$, then X is said to be **central** random variable. In what follows, CRV stands for a *central RV*. Now we study some basic properties of the mathematical expectation.

(a) $E(\alpha) = \alpha$ and $E(\alpha X) = \alpha EX$ for any RV X and any $\alpha \in \mathbb{R}$. It follows immediately from the definition.

(b) $E(X - EX) = 0$. Indeed, denote by $m = EX$ and apply (8) (by using of (7), it is obvious):

$$E(X - m) = \int_{-\infty}^{+\infty} (s - m)p(s)ds = m - m = 0.$$

(c) Given two RVs X and Y on $(S, \mathcal{A}, P)$, then

$$E(X + Y) = EX + EY, \tag{9}$$

assuming that both EX and EY exist and the sum $EX + EY$ is well defined. In particular, $X \le Y$ implies $EX \le EY$. Formula (9) is obvious as we use the definition of the expected value by the Lebesgue integral. Using of (8) is more complicated here and is omitted.

(d) Given two *independent* RVs X and Y on $(S, \mathcal{A}, P)$, then $E(XY) = EX \cdot EY$ (assuming that both EX and EY exist and the product $EX \cdot EY$ is well defined). [*Remind that a collection* $\{X_i\}_{i\in I}$ *of RVs on the same probability space is called independent if the collection of the following sets* $\{X_i^{-1}(-\infty, r_i)\}_{i\in I}$ *is independent for arbitrary* $r_i \in \mathbb{R}$.] The proof is obvious in the case if all X_i are indicator functions of events (of course, $E(\mathbb{I}_A) = P(A)$). For proving (d) in the general case, some approximation and the theory of Lebesgue integral is needed, and hence the proof in the general case is omitted. Notice that the equality $E(XY) = EX \cdot EY$ does not guaranties that X and Y are independent (see the proof of Proposition 4.10).

4.2 Conditional expectation

Definition 4.2 *Let X be a RV and $P(A) > 0$ then the* **conditional expectation of X given A** *is*

$$E(X|A) := \frac{E(X \cdot \mathbb{I}_A)}{P(A)}.$$

Example 4.1 *Let $p_{\langle X,Y\rangle}(x,y) = \frac{3x^2}{1-x} \cdot \mathbb{I}_{\{0<x<y<1\}}(x,y)$. Find $E(Y|\frac{1}{4})$.*

$$p_X(x) = \int_{\mathbb{R}} p_{\langle X,Y\rangle}(x,y)dxdy = \mathbb{I}_{(0,1)}(x) \cdot \int_{y=x}^{1} \frac{3x^2}{1-x}dy = 3x^2 \cdot \mathbb{I}_{(0,1)}(x),$$

$$p(y|x) = \frac{p_{\langle X,Y\rangle}(x,y)}{p_X(x)} = \frac{\frac{3x^2}{1-x} \cdot \mathbb{I}_{\{0<x<y<1\}}(x,y)(x,y)}{3x^2 \cdot \mathbb{I}_{[0,1]}(x)} = \frac{\mathbb{I}_{\{x<y<1\}}(x,y)}{1-x},$$

$$E(Y|X) = \int_{\mathbb{R}} yp(y|x)dy = \int_{y=x}^{1} \frac{y}{1-x}dy = \frac{1-x^2}{2(1-x)} = \frac{1+x}{2}.$$

Thus, $E(Y|\frac{1}{4}) = \frac{1+1/4}{2} = 5/8$. □

4.3 The moments and the variance

Definition 4.3 *Given an RV X and $n \in \mathbb{N}$. Then the n'th* **moment** *of X is defined by*

$$\mu_n(X) = \mu_n \equiv EX^n,$$

provided that EX^n is well defined. The **variance** *of X is defined by*

$$\mathrm{Var}X \equiv E(X - EX)^2 = EX^2 - EX \cdot EX,$$

provided $EX^2 < \infty$.

Proposition 4.2 *If EX^n if finite then EX^k is also finite for all $0 \le k \le n$.*

Proof: To see this, notice that

$$|X(s)|^k \le 1 + |X(s)|^n \quad (\forall s \in S),$$

and hence $E|X|^k \le 1 + E|X|^n$. ■

Notice that the first moment of X is EX. Sometimes the term **central n^{th} moment** of an RV X is used for $E(X-EX)^n$. Clearly the variance of X is nothing else than the central second moment of X. Sometimes, the so-called **standard deviation** $\sigma_X \equiv +\sqrt{\text{Var}X}$ is more useful than $\text{Var}X$.

Let us study some elementary properties of the variance.

(a) $\text{Var}(\alpha) = 0$ and $\text{Var}(\alpha X) = \alpha^2 \text{Var}X$ for any RV X and for any $\alpha \in \mathbb{R}$. In particular, $\text{Var}(-X) = \text{Var}X$. Property (a) follows immediately from properties of the expected value.

(b) If X and Y are two *independent* RVs, then

$$\text{Var}(X \pm Y) = \text{Var}X + \text{Var}Y,$$

Provided that $\text{Var}X$ and $\text{Var}Y$ exist.

Proof: In view of the fact that X and Y are independent iff X and $-Y$ are independent, and the fact that $\text{Var}(-Y) = \text{Var}Y$, it is enough to prove $\text{Var}(X+Y) = \text{Var}X + \text{Var}Y$. By properties (b) and (c) of the expectation, obtain

$$\text{Var}(X+Y) = E(X+Y-E(X+Y))^2 = E(X^2+2XY+Y^2-2(X+Y)E(X+Y)+(E(X+Y))^2) =$$
$$EX^2 + 2E(XY) + EY^2 - 2E(X+Y)E(X+Y) + E(X+Y)E(X+Y) =$$
$$EX^2+2E(XY)+EY^2-E(X+Y)\cdot E(X+Y) = EX^2+2EX\cdot EY+EY^2-(EX)^2-2EX\cdot EY-(EY)^2 =$$
$$\{EX^2-(EX)^2\} + \{EY^2-(EY)^2\} = \text{Var}X + \text{Var}Y. \ \blacksquare$$

Notice that the equality $\text{Var}(X \pm Y) = \text{Var}X + \text{Var}Y$ does not imply the independence of X and Y (see the proof of Proposition 4.10).

(c) If X and Y are two *independent* RVs, then

$$\text{Var}XY = \text{Var}X \cdot \text{Var}Y + \text{Var}X \cdot (EY)^2 + \text{Var}Y \cdot (EX)^2.$$

In particular, for any two independent CRV X and Y we have $\text{Var}XY = \text{Var}X \cdot \text{Var}Y$.

Proof: As X and Y are independent, then X^2 and Y^2 are independent by Proposition 3.1. Thus

$$\text{Var}XY = E(XY-E(XY))^2 = E(X^2Y^2 - 2XYE(XY) + E(XY)E(XY)) =$$
$$EX^2 \cdot EY^2 - 2EX \cdot EY \cdot EX \cdot EY + EX \cdot EY \cdot EX \cdot EY = EX^2 \cdot EY^2 - (EX)^2 \cdot (EY)^2 =$$
$$(\text{Var}X + (EX)^2) \cdot (\text{Var}Y + (EY)^2) - (EX)^2 \cdot (EY)^2 = \text{Var}X \cdot \text{Var}Y + \text{Var}X \cdot (EY)^2 + \text{Var}Y \cdot (EY)^2. \ \blacksquare$$

Definition 4.4 *The* **moment generating function** (**MGF**) *of an RV X is defined by*

$$M_X(t) \equiv E(\exp(tX)), \quad t \in \mathbb{R}.$$

Notice that $M_X(t)$ is well defined, since $\exp(tX)$ is non-negative. Certainly, an MGF could be infinity for some t. Proposition 4.1 allows sometimes to compute moments without computing distributions of X^k. We shall consider this later. Now we state connections between the MGF and moments. Unfortunately, complete proofs of the following two propositions are based on Lebesgue integration, and we omit them.

Proposition 4.3 *Let $X \geq 0$ be an RV and let $t \geq 0$. Then*

$$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n \mu_n}{n!},$$

where $\mu_n = EX^n$. ■

The proof is actually nothing but the possibility of term by term integration (in sense of the Lebesgue integral) of the formula $\exp(tX) = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}$.

Proposition 4.4 *Let X be an RV with $M_X(t) < \infty$ for $|t| < \varepsilon$. Then $E(|X|^n) < \infty$ for all $n \in \mathbb{N}$. Moreover, $M_X(t) = \sum_{n=0}^{\infty} \frac{t^n \mu_n}{n!}$ for $|t| < \varepsilon$ and $M_X(t)$ has all derivatives for $|t| < \varepsilon$, namely:*

$$M_X^{(k)}(t) = \sum_{n=0}^{\infty} \frac{t^n \mu_{n+k}}{n!} = E(\exp(tX) \cdot X^k), \quad |t| < \varepsilon. \qquad (10)$$

In particular, $M_X^{(k)}(0) = EX^k = \mu_k$. ■

[It is important to notice that $\frac{|t|^n \cdot |X|^n}{n!} \leq \exp(|tX|) \leq \exp(tX) + \exp(-tX)$. Integrating this formula for some $0 < |t_0| < \varepsilon$, one gets $E|X|^n \leq \frac{n!}{t_0^n}[M_X(t) + M_X(-t)] < \infty$. The formula $M_X(t) = \sum_{n=0}^{\infty} \frac{t^n \mu_n}{n!}$ for $|t| < \varepsilon$ requires more knowledge of the Lebesgue integration. After getting the formula, differentiating and obtaining (10) becomes an easy exercise.]

Proposition 4.5 *Let* $M_X(t) = M_Y(t) < \infty$ *for* $|t| < \varepsilon \neq 0$. *Then* $F_X(t) \equiv F_Y(t)$.

Proof: To be included later. ■

Example 4.2 *Let* $\langle X, Y\rangle$ *be independent with* $p_X(t) = p_Y(t) = \frac{\mathbb{I}_{[0,2]}(t)}{2}$. *(i) Find* $P(Y(X+1) < 1)$.
(ii) Find $M_X(t)$.
(iii) Find $M_{X+Y}(t)$.

$$(i): \ P(Y(X+1) < 1) = P(Y < (X+1)^{-1}) = \int_0^2 \int_0^{(x+1)^{-1}} \frac{dydx}{4} =$$

$$\frac{1}{4}\int_0^2 \frac{dx}{x+1} = \frac{1}{4}\cdot\Big[\log(x+1)\Big]_{x=0}^{x=2} = \frac{\log 3 - \log 1}{4} = \frac{\log 3}{4}.$$

$$(ii): \ M_X(t) = \frac{1}{2}\int_0^2 e^{tx}dx = \frac{1}{2t}\int_0^2 te^{tx}dx = \frac{1}{2t}\Big[e^{tx}\Big]_{x=0}^{x=2} = \frac{e^{2t}-1}{2t}.$$

$$(iii): \ M_{X+Y}(t) =^{[\langle X,Y\rangle \ is \ independent]} = M_X(t)M_Y(t) = \frac{(e^{2t}-1)^2}{4t^2}. \ \square$$

4.4 Several Inequalities and the Covariance

Proposition 4.6 (**Jensen's Inequality**) *Let* $P(X \in (a,b)) = 1$, $\theta : \mathbb{R} \to \mathbb{R}$ *be convex on* (a,b), *and* $E|X| + E|\theta(X)| < \infty$. *Then*

$$\theta(EX) \leq E\theta(X). \tag{11}$$

Proof: Notice that both

$$\theta'_+(t) = \lim_{s\downarrow t}\frac{\theta(s)-\theta(t)}{s-t}, \quad \theta'_-(t) = \lim_{s\uparrow t}\frac{\theta(s)-\theta(t)}{s-t}$$

exist and are finite whenever $a < t < b$. Moreover, $\theta'(t) = \lim_{s\to t}\frac{\theta(s)-\theta(t)}{s-t}$ exists except on the at most countable set of discontinuity points of θ'_+ and θ'_-. Furthermore,

$$\theta(t) - \theta(c) \geq \max[\theta'_+(c)(t-c), \theta'_-(c)(t-c)] \quad (t, c \in (a,b)). \tag{12}$$

Applying (12) for $t = X(\omega)$ and $c = EX$, obtain that $P(Y \geq 0) = 1$, where the RV Y is defined by

$$Y(\omega) := \theta(X(\omega)) - \theta(c) - \theta'_+(c)(X(\omega) - c) \quad (\omega \in S).$$

Since $\int_S [X(\omega) - c]dP = 0$, then

$$E\theta(X) - \theta(EX) = \int_S [\theta(X(\omega)) - \theta(c)]dP = \int_S [\theta(X(\omega)) - \theta(c) - \theta'_+(c)(X(\omega)-c)]dP = \int_S Y(\omega)dP \geq 0. \ ■$$

Corollary 4.1 *For any RV X, it holds that*
a) $(EX)^{2n} \le EX^{2n}$ *for* $n \in \mathbb{N}$.
b) $e^{EX} \le M_X(1)$.
If $X \ge 0$ *then*
c) $(EX)^{-(2n+1)} \le E(X^{2n+1})$ *for* $n \in \mathbb{N}$.
d) $E(\ln X) \le \ln EX$.

Proof: *a)* There is nothing to prove when $EX^{2n} = \infty$. If $EX^{2n} < \infty$, then by Proposition 4.2, $E|X| < \infty$. Now apply Proposition 4.6 to convex function $\theta(t) = t^{2n}$.

b) WlOG, suppose that $M_X(1) \le \infty$. Then also $EM_X(0) \le \infty$. Now apply Proposition 4.6 to convex function $\theta(t) = \exp(t)$.

Proof of *c)* and *d)* are similar. ∎

Proposition 4.7 (**Markov's inequality**) *Let $X \ge 0$ be an RV, then*

$$P(X \ge r) \le \frac{EX}{r} \tag{13}$$

for each $r > 0$.

Proof:

$$r \cdot P(X \ge r) = r \cdot \int_r^\infty p_X(t)dt \le \int_r^\infty tP_X(dt) \le \int_{-\infty}^\infty tP_X(dt) = EX. \quad \blacksquare$$

Corollary 4.2 (**Chebyshev's inequality**) *If* $\mathrm{Var}Y < \infty$, *then, for every* $r > 0$,

$$P(|Y - EY| \ge r) \le \frac{\mathrm{Var}Y}{r^2}.$$

Proof: Applying Proposition 4.7 to the RV $X = (Y - EY)^2 \ge 0$, one gets

$$P(|Y - EY| \ge r) = P(X \ge r^2) \le \frac{EX}{r^2} = \frac{\mathrm{Var}Y}{r^2}. \quad \blacksquare$$

For the proof of the following proposition, we refer the reader to any advanced textbook in analysis.

Proposition 4.8 (**Cauchy-Schwarz Inequality**) *If $EX^2 < \infty$ and $EY^2 < \infty$, then*

$$(EXY)^2 \le EX^2 \cdot EY^2 .$$

If $(EXY)^2 = EX^2 \cdot EY^2$, then $Y = \alpha X$ almost surely *for some $\alpha \in \mathbb{R}$.*
■

Definition 4.5 *The* **covariance** *of the RVs X and Y is*

$$\mathrm{Cov}(X, Y) \equiv E[(X - EX)(Y - EY)] = EXY - EXEY .$$

In particular, $\mathrm{Var}X = \mathrm{Cov}(X, X)$.

If $\mathrm{Var}X \cdot \mathrm{Var}Y \neq 0$, then the number

$$\rho(X, Y) \equiv \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}X \cdot \mathrm{Var}Y}}$$

is called the **correlation coefficient** *of X and Y.*

Notice that the property $\mathrm{Var}X = \mathrm{Cov}(X, X)$ has the following immediate extension

$$\mathrm{Var}\sum_{i=1}^{n} \alpha_i X_i = E\Big(\sum_{i=1}^{n} \alpha_i X_i - \sum_{i=1}^{n} \alpha_i EX_i\Big)^2 =$$

$$E\Big(\sum_{i=1}^{n} \alpha_i (X_i - EX_i)\Big)^2 = E\sum_{i=1}^{n} \alpha_i (X_i - EX_i) \cdot \sum_{j=1}^{n} \alpha_j (X_j - EX_j) =$$

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j E[(X_i - EX_i)(X_j - EX_j)] = \sum_{i,j=1}^{n} \alpha_i \alpha_j \mathrm{Cov}(X_j, X_j) .$$

Proposition 4.9 *Let X, Y be two RVs with $\mathrm{Var}X \cdot \mathrm{Var}Y \neq 0$, then $|\rho(X, Y)| \le 1$. If $|\rho(X, Y)| = 1$, then $Y = aX + b$ almost surely for some $a, b \in \mathbb{R}$.*

Proof: By the *Cauchy-Schwarz Inequality*, the first property follows from

$$\mathrm{Cov}(X,Y)^2 = (E(X-EX)(Y-EY))^2 \le E(X-EX)^2 \cdot E(Y-EY)^2 = \mathrm{Var}X \cdot \mathrm{Var}Y\,.$$

For the second one, assume $|\rho(X,Y)| = 1$. Then

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

and hence

$$(E(X-EX)(Y-EY))^2 = E(X-EX)^2 \cdot E(Y-EY)^2\,. \tag{14}$$

By Proposition 4.8, the condition (14) implies that

$$Y - EY = \alpha(X - EX)$$

almost surely for some $\alpha \in \mathbb{R}$. Then

$$Y = \alpha X + (EY - \alpha EX) \ \ (a.s.)\,.$$

Take $a = \alpha$ and $b = EY - \alpha EX$. ■

Proposition 4.10 *If random variables X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$, but not conversely in general.*

Proof:

$$\mathrm{Cov}(X,Y) = E((X-EX)(Y-EY)) = EXY - E(XEY) - E(YEX) + EXEY = EXY - EXEY = 0\,.$$

For the rest of the proof, consider the unit circle $S = \Gamma$ in the complex plane with the standard probability measure P on the σ-algebra of Lebesgue measurable subsets of Γ. Take RVs

$$X(\exp(it)) = \cos(t), \quad Y(\exp(it)) = \sin(t); \quad \exp(it) \in \Gamma\,.$$

RV's X and Y *are not independent*, since, for example,

$$P\Big(X \le \frac{1}{\sqrt{2}}\Big)P\Big(Y \le \frac{1}{\sqrt{2}}\Big) = \frac{3}{4}\frac{3}{4} = \frac{9}{16} \ne \frac{1}{2} = P\Big(X \le \frac{1}{\sqrt{2}}, Y \le \frac{1}{\sqrt{2}}\Big)\,.$$

However,

$$EXY = \frac{1}{2\pi}\int_0^{2\pi} \cos(t)\sin(t)dt = \frac{1}{4\pi}\int_0^{2\pi} \sin(2t)dt = 0 = \frac{1}{2\pi}\int_0^{2\pi} \cos(t)dt \cdot \frac{1}{2\pi}\int_0^{2\pi} \cos(t)dt = EXEY\,.$$

In the proof of $\mathrm{Cov}(X,Y) = 0$, we have used only that $EXY = EXEY$. Thus, in our example, we have $\mathrm{Cov}(X,Y) = 0$ *without having independence of X and Y*. ■

4.5 Random sequences

The sample space will be denoted by S, as usual. The space of all real valued sequences will be denoted by m.
Given a countable system $X = \langle X_1, X_2, \dots X_n, \dots \rangle$ of random variables on our sample space S. We call X a **random sequence** on S. The **JDF** of X is the function

$$F_X(\text{seq}(x)) \equiv P(\{X_1 \le x_1, X_2 \le x_2, \dots X_n \le x_n, \dots\}),$$

where $\text{seq}(x) = (x_1, x_2, \dots x_n, \dots) \in m$.
Given a random vector $\vec{X} = \langle X_1, X_2, \dots X_n \rangle$, then the **JDD** of $\vec{X}$ is the following generalized function $\mathbf{p}_{\vec{X}} : FS(\mathbb{R}^n) \to \mathbb{R}_+$, defined by:

$$(\mathbf{p}_{\vec{X}}, \mathbb{I}_B) = P(F_{\vec{X}} \in B) \quad (B \in \mathcal{B}(\mathbb{R}^n)).$$

It can be shown that $\mathbf{p}_{\vec{X}}(\vec{x}) = D_{x_1} D_{x_2} \dots D_{x_n} F_{\vec{X}}(\vec{x})$, where $\vec{x} = (x_1, x_2, \dots x_n)$.
Given an independent random vector $X = \langle X_1, X_2, \dots X_n \rangle$, then the **marginal distribution density** of X_k is the generalized function $\mathbf{p_k} \in FS'(\mathbb{R}^n)$ defined by $\mathbf{p_k} = D_{x_k} F_X$. Clearly, the marginal distribution density $\mathbf{p_k}$ of X is the usual DD $\mathbf{p_{X_k}} \in FS'(\mathbb{R})$ of X_k, if X_j are not considered for $j \ne k$. Denoting $Q_k = \{\vec{x} | x_k \le r_k\} \subset \mathbb{R}^n$ for a $\vec{r} = (r_1, r_2, \dots r_n) \in \mathbb{R}^n$, then

$$(\mathbf{p}, \mathbb{I}_{\cap_{k=1}^n Q_k}) = F_X(\vec{r}) = F_{X_1}(r_1) \cdot F_{X_2}(r_2) \cdot \dots F_{X_n}(r_n) =$$
$$(\mathbf{p_{X_1}}, \mathbb{I}_{\{x_1 \le r_1\}}) \cdot (\mathbf{p_{X_2}}, \mathbb{I}_{\{x_2 \le r_2\}}) \cdot \dots \cdot (\mathbf{p_{X_n}}, \mathbb{I}_{\{x_n \le r_n\}}) = (\mathbf{p_1}, \mathbb{I}_{Q_1}) \cdot (\mathbf{p_2}, \mathbb{I}_{Q_2}) \cdot \dots \cdot (\mathbf{p_n}, \mathbb{I}_{Q_n}).$$

The approximation of functions in $FS(\mathbb{R}^n)$ by the step functions gives us the following formula

$$(\mathbf{p}, \phi(\vec{x})) = (\mathbf{p_1}, \phi_1(x_1)) \cdot \dots \cdot (\mathbf{p_n}, \phi_n(x_n)) = (\mathbf{p_1} \cdot \dots \cdot \mathbf{p_n}, \phi(\vec{x})) \tag{15}$$

for all $\phi_k(x_k) \in FS(\mathbb{R})$, where $\phi(\vec{x}) = \phi_1(x_1) \cdot \phi_2(x_2) \cdot \dots \cdot \phi_n(x_n)$. Taking the formula (15) as the definition of the product of generalized functions $\mathbf{p_k}$ depending only on x_k and using the density of linear span of all functions $\phi(\vec{x}) = \phi_1(x_1) \cdot \phi_2(x_2) \cdot \dots \cdot \phi_n(x_n)$ in $FS(\mathbb{R}^n)$, obtain $\mathbf{p} = \mathbf{p_1} \cdot \dots \cdot \mathbf{p_n}$.
Let $X = \langle X_1, X_2, \dots X_n \rangle$ be an *independent random vector*. Then:
I. Given $k \in \mathbb{N}$, then

$$E(X_1^k \cdot \dots \cdot X_n^k) = EX_1^k \cdot \dots \cdot EX_n^k, \tag{16}$$

provided that EX_j^k are all finite. We know already this fact for two independent RVs and $k = 2$. Let us prove the formula (16) for arbitrary n and k.
Proof: Denote by $\mathbf{p_{X_k}} \in FS'(\mathbb{R})$ the DD of X_k, by $\mathbf{p} \in FS'(\mathbb{R}^n)$ the DD of X, and by $\mathbf{p_k} \in FS'(\mathbb{R}^n)$ the marginal distribution density of X_k. By (15), we obtain

$$\mu_k(X_1^k \cdot \dots \cdot X_n^k) = (\mathbf{p}, x_1^k \cdot \dots \cdot x_n^k) = (\mathbf{p_1}, x_1^k) \cdot \dots \cdot (\mathbf{p_n}, x_n^k) = \mu_k(X_1) \cdot \dots \cdot \mu_k(X_n). \quad \blacksquare$$

It follows immediately from (13) that
II.

$$E(X_1 \cdot X_2 \cdot \dots \cdot X_n) = EX_1 \cdot EX_2 \cdot \dots \cdot EX_n,$$

provided that EX_j are all finite.
III.

$$M_{X_1 + \dots + X_n}(t) = Ee^{tX_1 + \dots + tX_n} = Ee^{tX_1} \cdot \dots \cdot Ee^{tX_n} = M_{X_1}(t) \cdot \dots \cdot M_{X_n}(t), \tag{17}$$

provided that $M_{X_j}(t)$ are all well defined.
IV. If all our RVs X_j are central (that is $EX_k = 0$), then

$$\text{Var}(X_1 \cdot \dots \cdot X_n) = \mu_2(X_1) \cdot \dots \cdot \mu_2(X_n) - \mu_1^2(X_1) \cdot \dots \cdot \mu_1^2(X_n) = \mu_2(X_1) \cdot \dots \cdot \mu_2(X_n) =$$
$$(\mu_2(X_1) - \mu_1^2(X_1)) \cdot \dots \cdot (\mu_2(X_2) - \mu_1^2(X_2)) = \text{Var}X_1 \cdot \dots \cdot \text{Var}X_n,$$

provided that $\text{Var}X_j$ are all well defined.

4.6 Exercises

Exercise 4.1 *Given a random variable X with density*

$$p_X(t) = \frac{t^3+1}{2} \cdot \mathbb{I}_{[-1,1]}(t)$$

a) *Find* $\mathrm{Var}X$.

b) *Find the density $p_Y(t)$ of $Y = X^2$.*

Solution:
a)

$$EX = \int_{-1}^{1} \frac{1}{2}(s^4+s)ds = \left(\frac{s^5}{10} - \frac{s^2}{4}\right)\Bigg|_{-1}^{1} = \frac{1}{5}; \quad EX^2 = \int_{-1}^{1} \frac{1}{2}(s^5+s^2)ds = \frac{1}{3}.$$

$$\mathrm{Var}X = EX^2 - (EX)^2 = \frac{1}{3} - \left(\frac{1}{5}\right)^2 = \frac{22}{75}.$$

b)

$$F_Y(t) = P(Y \le t) = P(X^2 \le t) = P(-\sqrt{t} \le X \le \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} p_X(s)ds = \begin{cases} \int_{-\sqrt{t}}^{\sqrt{t}} \frac{s^3+1}{2} \cdot \mathbb{I}_{[-1,1]}(s)ds & \text{for} \quad t \ge 0, \\ 0 & \text{for} \quad t < 0 \\ 1 & \text{for} \quad t \ge 1 \end{cases}$$

Hence $p_Y(t) = \mathbb{I}_{[0,1]}(t) \cdot \frac{\partial}{\partial t} F_Y(t) =$

$$\mathbb{I}_{[0,1]}(t) \cdot \frac{d}{dt} \int_{-\sqrt{t}}^{\sqrt{t}} \frac{s^3+1}{2} ds = \mathbb{I}_{[0,1]}(t) \cdot \left(\frac{t^{3/2}+1}{2} + \frac{-t^{3/2}+1}{2}\right) \cdot \frac{1}{2}t^{-1/2} = \mathbb{I}_{[0,1]}(t) \cdot \frac{1}{2}t^{-1/2}. \quad \square$$

Exercise 4.2 *Seven fair dice are rolled. Denote by S_k, the score on k-th die. Given the following random variables: $X = S_1 - S_2$, and $Y = S_5 + S_6 + S_7$.*

a) Find the expected value EX.

b) Find the expected value EY.

c) Find the variance $\mathrm{Var}X$.

d) Find the covariance $\mathrm{Cov}(X, Y)$.

Solution:
a)

$$ES_k = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}; \quad EX = E(S_1 - S_2) = ES_1 - ES_2 = 0.$$

b)

$$EY = ES_5 + ES_6 + ES_7 = 3 \cdot \frac{7}{2} = \frac{21}{2}.$$

c) Note that S_k is independent with S_j for $k \neq j$. Then we can write

$$EX^2 = E(S_1 - S_2)^2 = ES_1^2 - 2ES_1ES_2 + ES_2^2 = 2 \cdot \frac{91}{6} - 2 \cdot \frac{7}{2} \cdot \frac{7}{2} = \frac{35}{6}.$$

We have used here the following:

$$ES_k^2 = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}.$$

So

$$\mathrm{Var}X = EX^2 - (EX)^2 = \frac{35}{6} - 0^2 = \frac{35}{6}.$$

d) $X = S_1 - S_2$ depends only on 1-st and 2-nd dice; $Y = S_5 + S_6 + S_7$ depends only on 5-th, 6-th, and 7-th dice. Consequently X and Y are independent, hence $\mathrm{Cov}(X, Y) = 0$. □

Exercise 4.3 *Given the following* **continuous** *distribution function*
$F_X(t) = \begin{cases} 0 & \text{if} \quad t < 3, \\ C(t-3)^2 & \text{if} \quad 3 \le t \le 5, \\ 1 & \text{if} \quad 5 < t \end{cases}$ *of a random variable* X.

a) Find the coefficient C.

b) Find the distribution density of the random variable X.

c) Find the probability of the event $\{X \in [3, 4)\}$.

d) Find the expected value EX.

e) Find the variance $\mathrm{Var}X$.

Solution:

a)

$$\lim_{t \to 5+} F_X(t) = 1 = \lim_{t \to 5-} C(t-3)^2 = C \cdot 4 \Rightarrow C = \frac{1}{4}.$$

b)

$$p(t) = \frac{\partial}{\partial t} F_X(t) = \begin{cases} 0 & \text{if} \quad t \notin [3, 5], \\ \frac{1}{4} \cdot 2(t-3) = \frac{t-3}{2} & \text{if} \quad t \in [3, 5], \end{cases}$$

c)

$$P(X) = F_X(4) - F_X(3) = \frac{1}{4}(4-3)^2 - 0 = \frac{1}{4}.$$

d)

$$EX = \int_{-\infty}^{\infty} tp(t)dt = \int_3^5 \frac{t(t-3)}{2} dt = \left(\frac{t^3}{6} - \frac{3}{4}t^2 \right) \Bigg|_3^5 = \frac{125 - 27}{6} - \frac{3}{4}(25 - 9) = \frac{49}{3} - 12 = \frac{13}{3}.$$

e)

$$EX^2 = \int_3^5 t^2 \frac{t-3}{2} dt = \left(\frac{t^4}{8} - \frac{t^3}{2} \right) \Bigg|_3^5 = \frac{625 - 81}{8} - \frac{125 - 27}{2} = \frac{544}{8} - \frac{98}{2} = 68 - 49 = 19.$$

$$\mathrm{Var}X = EX^2 - (EX)^2 = 19 - \left(\frac{13}{3} \right)^2 = \frac{171}{9} - \frac{169}{9} = \frac{2}{9}. \quad \square$$

Exercise 4.4 *Three fair dice are rolled. Denote by S_k, the score on k-th die. Given the following random variables: $X = S_1 - S_2$, and $Y = S_1 - S_2 + S_3$.*

Find EX, EY, $\mathrm{Var}X$, and $\mathrm{Var}Y$

Solution:

$$ES_k = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{3 \cdot 7}{6} = \frac{7}{2}; \quad ES_k^2 = \frac{1 + 4 + 9 + \cdots + 36}{6} = \frac{91}{6}; \quad \mathrm{Var}S_k = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

$$EX = ES_1 - ES_2 = 0 - 0 = 0; \quad EY = ES_1 - ES_2 + ES_3 = \frac{7}{2}.$$

S_k are independent for different k. Therefore

$$\mathrm{Var}X = \mathrm{Var}S_1 + \mathrm{Var}S_2 = \frac{70}{12} = \frac{35}{6}; \quad \mathrm{Var}Y = \mathrm{Var}S_1 + \mathrm{Var}S_2 + \mathrm{Var}S_3 = \frac{35}{4}. \quad \square$$

Exercise 4.5 *Let $r > 1$ be a real number. Consider a random variable X which takes values r with probability p and $1/r$ with probability $1 - p$. Compute $\mathrm{Var}(2X + 1)$ for $r = 2$ and $p = 1/2$.*

Solution:

$$EX = r \cdot p + \frac{1}{r}(1 - p); \; EX^2 = r^2 \cdot p + \frac{1}{r^2}(1 - p).$$

$$\mathrm{Var}X = EX^2 - (EX)^2 = r^2 \cdot p + \frac{1}{r^2}(1 - p) - \left(r \cdot p + \frac{1}{r}(1 - p)\right)^2.$$

$$\mathrm{Var}(2X + 1) = 4\mathrm{Var}X = 4\left[4 \cdot (1/2) + \frac{1}{4}(1/2) - \left(2 \cdot (1/2) + \frac{1}{2}(1/2)\right)^2\right] = \frac{39}{4}. \quad \square$$

Exercise 4.6 *Consider two random variables X and Y with joint probability distribution given in the table:*

$X \setminus Y$	0	1	2
0	1/6	1/4	1/8
1	1/8	1/6	1/6

(i) Find $P(X = 0 | Y > 0)$.

(ii) Are the random variables X and Y independent?

Solution:
(i)

$$P(X = 0 | Y > 0) = \frac{P(X = 0 \;\&\; Y > 0)}{P(Y > 0)} = \frac{1/4 + 1/8}{1/4 + 1/8 + 1/6 + 1/6} = \frac{3/8}{17/24} = \frac{9}{17}.$$

(ii)

$$P(X = 0 | Y > 0) = \frac{9}{17} \neq P(X = 0) = 1/6 + 1/4 + 1/8 = \frac{13}{24}.$$

Thus, X and Y are not independent. $\square$

Exercise 4.7 *A fair coin is rolled 3 times. Let X be the number of heads in first 2 trials and Y be the number of heads in first 3 trials. The corresponding joint probability distribution of $< X, Y >$ is given in the table*

$X \setminus Y$	0	1	2	3	
0	1/8	1/8	0	0	$P(X=0)=1/4$
1	0	1/4	1/4	0	$P(X=1)=1/2$
2	0	0	1/8	1/8	$P(X=2)=1/4$
	$P(Y=0)=1/8$	$P(Y=1)=3/8$	$P(Y=2)=3/8$	$P(Y=3)=1/8$	

a) Find $E(XY)$.

b) Find $\mathrm{Var}(X)$.

c) Find $\mathrm{Var}(X+Y)$.

d) Find $\mathrm{Var}(XY)$.

e) Find $\mathrm{Cov}(X,Y)$.

f) Find the conditional expectation $E(X|Y \geq \sqrt{2})$.

Solution:

a)

$$EXY = 1 \cdot 1 \cdot \frac{1}{4} + 1 \cdot 2 \cdot \frac{1}{4} + 2 \cdot 2 \cdot \frac{1}{8} + 2 \cdot 3 \cdot \frac{1}{8} = 2.$$

b)

$$EX = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1. \quad EX^2 = 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{3}{2}. \quad \mathrm{Var}X = EX^2 - (EX)^2 = \frac{3}{2} - 1 = \frac{1}{2}.$$

c)

$$\mathrm{Var}(X+Y) = E(X+Y)^2 - (E(X+Y))^2 = EX^2 + 2EXY + EY^2 - (EX)^2 - 2EXEY - (EY)^2 =$$

$$\mathrm{Var}X + \mathrm{Var}Y + 2(EXY - EX \cdot EY).$$

$$EY = 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = \frac{3}{2}. \quad EY^2 = 1 \cdot \frac{3}{8} + 4 \cdot \frac{3}{8} + 9 \cdot \frac{1}{8} = \frac{24}{8} = 3. \quad \mathrm{Var}Y = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}.$$

$$\mathrm{Var}(X+Y) = \frac{1}{2} + \frac{3}{4} + 2\left(2 - 1 \cdot \frac{3}{2}\right) = \frac{5}{4} + 2 \cdot \frac{1}{2} = \frac{9}{4}.$$

d)

$$E(XY)^2 = 1^2 \cdot 1^2 \cdot \frac{1}{4} + 1^2 \cdot 2^2 \cdot \frac{1}{4} + 2^2 \cdot 2^2 \cdot \frac{1}{8} + 2^2 \cdot 3^2 \cdot \frac{1}{8} = \frac{1}{4} + 1 + 2 + \frac{9}{2} = 3 + \frac{19}{4} = \frac{31}{4}.$$

$$\mathrm{Var}XY = E(XY)^2 - (EXY)^2 = \frac{31}{4} - 2^2 = \frac{15}{4}.$$

e)

$$\mathrm{Cov}(X,Y) = E((X-EX)(Y-EY)) = E(XY) - EX \cdot EY = 2 - 1 \cdot \frac{3}{2} = \frac{1}{2}. \ \square$$

f)

$$E(X|Y \geq \sqrt{2}) = \frac{E\left(X \cdot \mathbb{I}_{\{Y \geq \sqrt{2}\}}\right)}{P(Y \geq \sqrt{2})} = \frac{1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8}}{P(Y=2) + P(Y=3)} = \frac{\frac{3}{4}}{\frac{3}{8} + \frac{1}{8}} = \frac{3}{2}. \quad \square$$

Exercise 4.8 *Let X and Y be independent random variables, $EX = 0$, $EY = 1$, and $\text{Var}X = \text{Var}Y = 1$. Find the variance $\text{Var}[(X-1)(Y+1)]$.*

Solution:
Note that $X - 1$ is independent with $Y + 1$. Hence we have

$$\text{Var}[(X-1)(Y+1)] = \text{Var}(X-1)\text{Var}(Y+1) + \text{Var}(X-1)(E(Y+1))^2 + \text{Var}(Y+1)(E(X-1))^2 =$$

$$\text{Var}X \cdot \text{Var}Y + \text{Var}X(1+1)^2 + \text{Var}Y(-1)^2 = 1 + 4 + 1 = 6. \quad \square$$

Exercise 4.9 *Given the following function*

$$f(x) = \begin{cases} 0 & \text{if} \quad x < 0, \\ \lambda(4x - x^2) & \text{if} \quad 0 \le x \le 2, \\ 0 & \text{if} \quad 2 < x. \end{cases}$$

a) Find a value of λ for which f is the distribution density of the distribution function F_X of some random variable X.

b) the correspondent λ and X find:
(i) F_X;
(ii) the probability of the event $\{X \in [1,3)\}$;
(iii) the expected value EX;
(iv) the variance $\text{Var}X$.

Solution:
a) $f(x) \ge 0$,

$$1 = \int_{-\infty}^{\infty} f(x)dx = \lambda \int_0^2 (4x - x^2)dx = \lambda \left(2x^2 - \frac{x^3}{3}\right)\Big|_0^2 = \lambda(8 - 8/3) = \lambda \cdot \frac{16}{3}.$$

So, $\lambda = \frac{3}{16}$.
b)(i) $F_X(t) = 0$ for $t \le 0$ and $F_X(t) = 1$ for $t \ge 2$. For $t \in (0,2)$:

$$F_X(t) = \int_0^t f(x)dx = \frac{3}{16}\int_0^t (4x - x^2)dx = \frac{3}{16}\left(2x^2 - \frac{x^3}{3}\right)\Big|_0^t = \frac{3}{16} \cdot \left(2t^2 - \frac{t^3}{3}\right).$$

b)(ii)

$$P(0 \le X < 3) = F_X(3) - F_X(1) = 1 - \frac{3}{16}(2 \cdot 1 - 1/3) = 1 - \frac{3}{16} \cdot \frac{5}{3} = 1 - \frac{5}{16} = \frac{11}{16}.$$

b)(iii)

$$EX = \int_{-\infty}^{\infty} tf(t)dt = \frac{3}{16}\int_0^2 (4t^2 - t^3)dt = \frac{3}{16}\left(\frac{4}{3}t^3 - \frac{1}{4}t^4\right)\Big|_0^2 = \frac{3}{16} \cdot \left(\frac{32}{3} - \frac{16}{4}\right) = 2 - \frac{3}{4} = \frac{5}{4}.$$

b)(iv)

$$EX^2 = \frac{3}{16}\int_0^2 (4t^3 - t^4)dt = \frac{3}{16}\left(t^4 - \frac{1}{5}t^5\right)\Big|_0^2 = \frac{3}{16} \cdot \left(16 - \frac{32}{5}\right) = 3 - \frac{6}{5} = \frac{9}{5}.$$

Hence

$$\text{Var}X = EX^2 - (EX)^2 = \frac{9}{5} - \frac{25}{16} = \frac{144 - 125}{80} = \frac{19}{80}. \quad \square$$

Exercise 4.10 *Given two independent random variables X and Y with the moment generating functions*

$$M_X(t) = e^{t^2/2} \text{ and } M_Y(t) = (1+e^t)/2.$$

a) *Find EXY and $E(X-Y)$.*

b) *Find* $\text{Var}(X-Y)$.

c) *Find* $\text{Var}(X \cdot (2Y-1))$.

Solution:
a) $EX = M_X'(0) = te^{t^2/2}\Big|_{t=0} = 0, \quad EY = M_Y'(0) = \frac{1}{2}e^t\big|_{t=0} = 1/2$. Hence $EXY = 0$ and $E(X-Y) = EX - EY = -1/2$..

b) $\mu_2(X) = (e^{t^2/2})''\Big|_{t=0} = (t \cdot e^{t^2/2})'\Big|_{t=0} = (e^{t^2/2} + t^2e^{t^2/2})\Big|_{t=0} = 1.$

$\mu_2(Y) = \left(\frac{1+e^t}{2}\right)''\Big|_{t=0} = \left(\frac{e^t}{2}\right)'\Big|_{t=0} = \frac{e^t}{2}\Big|_{t=0} = \frac{1}{2}.$

$\text{Var}X = \mu_2(X) - (EX)^2 = 1 - 0 = 1.$

$\text{Var}Y = \mu_2(Y) - (EY)^2 = 1/2 - (1/2)^2 = 1/2 - 1/4 = 1/4.$

$\text{Var}(X-Y) = \text{Var}X + \text{Var}Y = 1 + 1/4 = 5/4.$

c) $EX = 0$ and $E(2Y-1) = 2EY - 1 = 2 \cdot 1/2 - 1 = 0.$

Thus, both the factors a central random variables. Since they are independent, one gets

$\text{Var}(X \cdot (2Y-1)) = \text{Var}X \cdot \text{Var}(2Y-1) = \text{Var}X \cdot \text{Var}2Y = \text{Var}X \cdot 4 \cdot \text{Var}Y = 1 \cdot 4 \cdot 1/4 = 1.$ □

Exercise 4.11 *Given two independent random variables X and Y with the expected values $EX = 1$, $EY = 2$, and the variances $VarX = VarY = 1$.*

a) Find the second moment $\mu_2(X) = EX^2$.

b) Find the variance $\text{Var}(X+1)(Y+2)$.

Solution:
a) $\text{Var}X = EX^2 - (EX)^2 = 1$, so

$$EX^2 = \text{Var}X + (EX)^2 = 1 + 1^2 = 2.$$

b) $X+1$ and $Y+2$ are independent too. Hence

$$\text{Var}(X+1)(Y+2) = \text{Var}(X+1)\text{Var}(Y+2) + \text{Var}(X+1)(E(Y+2))^2 + \text{Var}(Y+2)(E(X+1))^2 =$$

$$\text{Var}X \cdot \text{Var}Y + \text{Var}X(E(Y+2))^2 + \text{Var}Y(E(X+1))^2 = 1 + (2+2)^2 + (1+1)^2 = 1 + 16 + 4 = 21. \quad \square$$

Exercise 4.12 *Moment generating functions of independent RVs X and Y are:*

$$M_X(t) = \frac{1}{1-t} \quad \& \quad M_Y(t) = e^{\frac{t^2}{2}}.$$

a) Find $\text{Var}(XY)$.

b) Find $\text{Cov}(X,Y)$.

Solution:
a) Since X and Y are independent, then $\mathrm{Var}XY = \mathrm{Var}X \cdot \mathrm{Var}Y + \mathrm{Var}X(EY)^2 + \mathrm{Var}Y(EX)^2$.

$$EX = M'_X(t)\big|_{t=0} = \frac{1}{(1-t)^2}\bigg|_{t=0} = 1. \quad EX^2 = M''_X(t)\big|_{t=0} = \frac{2}{(1-t)^3}\bigg|_{t=0} = 2. \quad \mathrm{Var}X = 2 - 1^2 = 1.$$

$$EY = \left(e^{\frac{t^2}{2}}\right)'\bigg|_{t=0} = t \cdot e^{\frac{t^2}{2}}\bigg|_{t=0} = 0. \quad EY^2 = \left(t \cdot e^{\frac{t^2}{2}}\right)'\bigg|_{t=0} = \left(e^{\frac{t^2}{2}} + t^2 \cdot e^{\frac{t^2}{2}}\right)\bigg|_{t=0} = 1. \quad \mathrm{Var}Y = 1 - 0^2 = 1.$$

$$\mathrm{Var}XY = 1 \cdot 1 + 1 \cdot 0^2 + 1 \cdot 1^2 = 2.$$

b) Since X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$. □

Exercise 4.13 *Given a random variable X with the expected value: $EX = 1$; the variance: $\mathrm{Var}(X) = 2$; and the third moment $\mu_3(X) = \mu_3 = 2$.*

a) Find the third centered moment $\mu_3^0(X)$. [Hint: use the formula $\mu_3^0 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3$].

b) Show that the random variables $X + 1$ and X^2 are not independent.

c) Let a random variable Y be independent with X, and $EY = -2$, and $\mathrm{Var}Y = 1$. Find the variance $\mathrm{Var}(X - 1)(Y + 2)$.

Solution:
a)

$$\mathrm{Var}X = \mu_2 - \mu_1^2 \Rightarrow \mu_2 = \mathrm{Var}X + \mu_1^2 = 2 + 1^2 = 3.$$

$$\mu_3^0 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 = 2 - 3 \cdot 1 \cdot 3 + 2 \cdot 1^2 = 4 - 9 = -5.$$

b)

$$E[(X+1)X^2] = EX^2 + EX^3 = \mu_2 + \mu_3 = 3 + 2 = 5.$$

$$E(X+1) \cdot EX^2 = (EX + 1) \cdot \mu_2 = (1+1) \cdot 3 = 6.$$

If $X + 1$ and X^2 are independent, then

$$5 = E[(X+1)X^2] = E(X+1)EX^2 = 6, \quad \text{a contradiction.}$$

Hence, $X + 1$ and X^2 are not independent.

c) As X and Y are independent, then $X - 1$ and $Y + 2$ are independent too. Since $E(X-1) = EX - 1 = 0$ and $E(Y+2) = EY + 2 = 0$, then $X - 1$ and $Y + 2$ are both centered RV's. Hence

$$\mathrm{Var}(X-1)(Y+2) = \mathrm{Var}(X-1)\mathrm{Var}(Y+2) = \mathrm{Var}X \cdot \mathrm{Var}Y = 2 \cdot 1 = 2. \ \square$$

Exercise 4.14 *Given two independent random variables X and Y with the same distribution functions* $F_X(t) = F_Y(t) = \begin{cases} 1 - \exp(-t) & (t \geq 0), \\ 0 & (t < 0). \end{cases}$

a) Find the distribution density $p_X(t)$ of the random variable X.

b) Find the expected value EX.

c) Find the variance $\mathrm{Var}X$.

d) Find the variance $\mathrm{Var}(X - 2Y)$.

Solution:
a)

$$p_X(t) = \frac{d}{dt}F_X(t) = \begin{cases} e^t & (t \geq 0), \\ 0 & (t < 0). \end{cases}$$

b)

$$EX = \int\limits_{-\infty}^{\infty} tp_X(t)dt = \int\limits_{0}^{\infty} te^{-t}dt = [v := t, u' := e^{-t}] = vu|_0^\infty - \int\limits_{0}^{\infty} v'udt =$$

$$t(-e^{-t})\Big|_0^\infty - \int\limits_{0}^{\infty} 1 \cdot (-e^{-t})dt = 0 + \int\limits_{0}^{\infty} e^{-t}dt = -e^{-t}\Big|_0^\infty = 0 - (-1) = 1.$$

c)

$$EX^2 = \int\limits_{0}^{\infty} t^2e^{-t}dt = t^2(-e^{-t})\Big|_0^\infty - 2\int\limits_{0}^{\infty} t \cdot (-e^{-t})dt = 0 + 2 \cdot 1 = 2.$$

$$\text{Var}X = EX^2 - (EX)^2 = 2 - 1^2 = 1.$$

d) Since X and Y are independent, then X and $-2Y$ are independent too. So we have

$$\text{Var}(X - 2Y) = \text{Var}X + \text{Var}(-2Y) = \text{Var}X + 4\text{Var}Y = 1 + 4 \cdot 1 = 5. \ \square$$

Exercise 4.15 *Given a (1)-exponentially distributed random variable X.*

a) Find the moment generating function $M_X(t)$ and its domain.

b) Find the variance of X.

c) Find the variance of X^2.

Solution:
a)

$$p_X(s) = \begin{cases} e^{-s} & (s \geq 0); \\ 0 & (s < 0). \end{cases}$$

$$M_X(t) = Ee^{tX} = \int\limits_{0}^{\infty} e^{ts} \cdot e^{-s}ds = \int\limits_{0}^{\infty} e^{s(t-1)}ds = \frac{1}{t-1} \cdot e^{s(t-1)}\Big|_0^\infty = 0 - \frac{1}{t-1} = \frac{1}{1-t}. \quad (-\infty < t < 1)$$

b)

$$EX = M_X'(0) = \left(\frac{1}{1-t}\right)'\Bigg|_{t=0} = \frac{1}{(1-t)^2}\Bigg|_{t=0} = 1.$$

$$EX^2 = M_X''(0) = \left(\frac{1}{1-t}\right)''\Bigg|_{t=0} = \frac{2}{(1-t)^3}\Bigg|_{t=0} = 2.$$

Hence $\text{Var}X = EX^2 - (EX)^2 = 2 - 1 = 1$.

c)

$$\left(\frac{1}{1-t}\right)''' = \frac{6}{(1-t)^4}, \quad \left(\frac{1}{1-t}\right)^{IV} = \frac{24}{(1-t)^5} \ \Rightarrow \ EX^4 = 24.$$

$$\text{Var}X^2 = E(X^2)^2 - (EX^2)^2 = EX^4 - (EX^2)^2 = 24 - 2^2 = 20. \ \square$$

Exercise 4.16 *The distribution of the system $\{X, Y\}$ of random variables is given by:*

$X \setminus Y$	-2	0
0	1/4	1/6
1	1/3	1/4

(i) *Find the covariance* $\mathrm{Cov}(X, Y)$

(ii) *Find the conditional expectation* $E(X|X + Y \geq 0)$

Solution:
(i)

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

$E(XY) = 1 \cdot (-2) \cdot 1/3 = -2/3$.

$EX = 0 \cdot (1/4 + 1/6) + 1 \cdot (1/3 + 1/4) = 7/12$.

$EY = (-2) \cdot (1/4 + 1/3) + 0 \cdot (1/6 + 1/4) = -7/6$.

$$\mathrm{Cov}(X, Y) = -\frac{2}{3} - \frac{7}{12} \cdot \left(-\frac{7}{6}\right) = -\frac{2}{3} + \frac{49}{72} = -\frac{48}{72} + \frac{49}{72} = \frac{1}{72}.$$

(ii)

$$E(X|X + Y \geq 0) = E(X|Y = 0) = 0 \cdot P(X = 0|Y = 0) + 1 \cdot P(X = 1|Y = 0) =$$

$$P(X = 1|Y = 0) = \frac{P(X = 1 \;\&\; Y = 0)}{P(Y = 0)} = \frac{1/4}{1/6 + 1/4} = \frac{1/4}{5/12} = \frac{3}{5}. \quad \square$$

Exercise 4.17 *A point (R_1, R_2) is taken randomly in the parallelogram with vertices $(0, 0)$, $(2, 0)$, $(3, 1)$, and $(1, 1)$. Find the function $E(R_2|R_1 = x)$ of the conditional expectation of R_2, given $R_1 = x$.*

Solution:
For $0 \leq x \leq 1$:

$$E(R_2|R_1 = x) = \int_0^x y \cdot p(y)dy = \frac{1}{x} \cdot y\Big|_0^x = \frac{1}{x} \cdot \frac{x^2}{2} = \frac{x}{2}.$$

For $1 \leq x \leq 2$:

$$E(R_2|R_1 = x) = \int_0^1 y \cdot p(y)dy = \int_0^1 ydy = \frac{1}{2}.$$

For $2 \leq x \leq 3$:

$$E(R_2|R_1 = x) = \int_{x-2}^1 y \cdot p(y)dy = \frac{1}{1 - (x - 2)} \int_{x-2}^1 y \; dy = \frac{1}{3 - x} \cdot \frac{y^2}{2}\Big|_{x-2}^1 =$$

$$\frac{1}{2(3 - x)}(1 - (x - 2)^2) = \frac{1}{2(3 - x)}(-x^2 + 4x - 3) = \frac{(3 - x)(x - 1)}{2(3 - x)} = \frac{x - 1}{2}. \quad \square$$

5 Standard distributions

Here we consider several standard distributions.

5.1 Uniform, Exponential, and Normal distributions

I. A RV X is called (a, b)**-Uniform** (or $U_{a,b}$), where $-\infty < a < b < \infty$, if

$$p_X(t) = \frac{\mathbb{I}_{(a,b)(t)}}{b-a}.$$

$$EX = \int_a^b x(b-a)^{-1}dx = \frac{x^2}{2(b-a)}\Big|_{t=a}^{t=b} = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}.$$

$$EX^2 = \int_a^b x^2(b-a)^{-1}dx = \frac{x^3}{3(b-a)}\Big|_{t=a}^{t=b} = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ba+a^2}{3}.$$

$$VarX = EX^2 - (EX)^2 = \frac{b^2+ba+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}. \quad \square$$

The MGF is given by

$$M_X(t) = E(e^{tX}) = \int_a^b e^{tx}(b-a)^{-1}dx = \frac{e^{bt}-e^{at}}{(b-a)t} \qquad (t \neq 0).$$

II. A RV X is called λ**-Exponential** (or E_λ), where $\lambda > 0$, if

$$p_X(t) = \lambda \cdot e^{-\lambda t} \cdot \mathbb{I}_{\mathbb{R}_+}(t).$$

The MGF is given by

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{(t-\lambda)x}dx = -\frac{\lambda}{t-\lambda} \quad = \frac{1}{1-t/\lambda} \qquad (-\infty < t < \lambda).$$

$$EX = M'_X(t)|_{t=0} = \frac{1}{\lambda(1-t/\lambda)^2}\Big|_{t=0} = \frac{1}{\lambda}.$$

$$EX^2 = M''_X(t)|_{t=0} = \frac{2}{\lambda^2(1-t/\lambda)^3}\Big|_{t=0} = \frac{2}{\lambda^2}.$$

$$VarX = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}. \quad \square$$

III. A RV X is called (a, σ^2)**-Normal** or (a, σ^2)**-Gaussian** or just N_{a,σ^2}, where $a, \sigma \in \mathbb{R}$, if

$$p_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(t-a)^2}{\sigma^2}}.$$

The MGF is given by

$$M_X(t) = E(e^{tX}) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int\limits_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-a)^2}{\sigma^2}} dx = \exp(at + \frac{1}{2}\sigma^2 t^2) \quad (-\infty < t < \infty).$$

$$EX = M'_X(t)\Big|_{t=0} = (a + \sigma^2 t) \cdot \exp(at + \frac{1}{2}\sigma^2 t^2)\Big|_{t=0} = a.$$

$$EX^2 = M''_X(t)\Big|_{t=0} = [(a + \sigma^2 t)^2 + \sigma^2] \cdot \exp(at + \frac{1}{2}\sigma^2 t^2)\Big|_{t=0} = a^2 + \sigma^2.$$

$$VarX = EX^2 - (EX)^2 = a^2 + \sigma^2 - a^2 = \sigma^2.$$

The next important property follows directly from Theorem 3.1.

Proposition 5.1 *If RVs X and Y are independent, with $X \in N_{a_1,\sigma_1^2}$ and $Y \in N_{a_2,\sigma_2^2}$ then $X + Y \in N_{a_1+a_2,\sigma_1^2+\sigma_2^2}$.* ■

Proof: It follows from (17) and Proposition 4.5. ■

5.2 More on Dirac δ-function

Consider the space $FS(\mathbb{R}^n)$ of infinitely many times differentiable finitely supported real-valued functions on $\mathbb{R}^n$. Denote by $FS'(\mathbb{R}^n)$ the space of all [*continuous w.r.t. to a "certain natural topology" in* $FS(\mathbb{R}^n)$]. The following functional $\delta = \delta_0 \in FS'(\mathbb{R})$:

$$\delta(\phi) = (\delta, \phi) \equiv \phi(0) \qquad (\phi \in FS(\mathbb{R}))$$

is called the **Dirac δ-function**. Respectively,

$$\delta_q(\phi) = (\delta_q, \phi) \equiv \phi(q) \qquad (\phi \in FS(\mathbb{R}))$$

Notice that any locally integrable function g (in symbols: $g \in \mathrm{Loc}(\mathbb{R}^n)$) is in $FS'(\mathbb{R}^n)$, indeed:

$$g(\phi) = (g, \phi) \equiv \int_{\mathbb{R}^n} g(t)\phi(t)dt \qquad (\phi \in FS(\mathbb{R}^n)).$$

Definition 5.1 *Given $u \in FS'(\mathbb{R})$, the following functional $u' \in FS'(\mathbb{R})$:*

$$u'(\phi) = (u', \phi) \equiv -(u, \phi') \qquad (\phi \in FS(\mathbb{R}))$$

is said to be the **generalized derivative** *of u and is denoted by $D_t(u)$. Similarly, we define* **generalized partial derivatives** *of $v \in FS'(\mathbb{R}^n)$.*

By induction,

$$\left(u^{(m)}, \phi\right) = (-1)^m (u, \phi^{(m)}) \qquad (\phi \in FS(\mathbb{R})),$$

for all $u' \in FS'(\mathbb{R})$, $m \in \mathbb{N}$.

Remark that for a differentiable function f on a domain in $\mathbb{R}^n$, the generalized derivative coincides with the *usual derivative*. Consider only the one dimensional case, $n = 1$.

$$\left(\frac{\partial f}{\partial t}, \phi\right) = \int_{-\infty}^{\infty} \frac{\partial f}{\partial t}\phi(t)dt = \int_{-\infty}^{\infty} f\phi(t)dt - \int_{-\infty}^{\infty} f\frac{\partial \phi}{\partial t}dt = -\int_{-\infty}^{\infty} f\frac{\partial \phi}{\partial t}dt = -\left(f, \frac{\partial \phi}{\partial t}dt\right) = \left(D_t f, \phi\right)$$

for all **test functions** $\phi \in FS(\mathbb{R})$, and therefore $\frac{\partial f}{\partial t} = D_t f$.

For example, the Heaviside function $H \in \text{Loc}(\mathbb{R}) \subset FS'(\mathbb{R})$ has the following derivative:

$$(H', \phi) = -(u, \phi') = -\int_{-\infty}^{\infty} H(t)\frac{d\phi}{dt}(t)dt = -\int_{0}^{\infty} H\phi' dt = \phi|_0^{\infty} = \phi(0) = (\delta, \phi) \qquad (\phi \in FS(\mathbb{R})).$$

In other words, $H' = \delta$.

One can do many of usual analytic manipulations in *generalized sense* with functional from $FS'(\mathbb{R})$. For instance, if $f : \mathbb{R} \to \mathbb{R}$ has countably many discontinuities with left and right limits at them, say

$$f(t) = f_0(t) + \sum_{n=1}^{\infty} h_n H(t - t_n),$$

where f_0 is continuous. Then

$$D_t f = [f_0'(t)] + \sum_{n=1}^{\infty} h_n \delta(t - t_n) = \sum_{n=1}^{\infty} h_n \delta_{t_n}(t),$$

where $f_0'(t)$ is the usual derivative of $f(t)$ at $t \neq t_n$.

5.3 Discrete distributions

Discrete RVs are given by distribution functions of the following form.

$$F_X(x) = \sum_{n=0}^{\infty} p_n H(x - x_n),$$

where $\sum_{n=0}^{\infty} p_n = 1$, and $(x_n)_{n=0}^{\infty}$ is a sequence in $\mathbb{R}$. In most of cases, $x_n = n$. Consider several standard discrete distributions.

IV. **(p)-Bernoulli** (or B_p): $0 \leq p \leq 1$, and the DF is given by

$$F_X(x) = \sum_{i=0}^{1} p^i (1-p)^{1-i} H(x - i) = (1-p)H(x) + pH(x-1).$$

The parameter p is interpreted as the probability of success in a Bernoulli trial. The "distribution density" is given by

$$\mathbf{p_X}(x) = D_x\Big((1-p)H(x) + pH(x-1)\Big) = (1-p)\delta_0(x) + p\delta_1(x).$$

Given a (p)-Bernoulli distributed RV X, then

$$EX = \int_{-\infty}^{\infty} x\mathbf{p_X}(x)dx = (1-p)\int_{-\infty}^{\infty} x\delta_0(x)dx + p\int_{-\infty}^{\infty} x\delta_1 dx =$$

$$(1-p)(\delta_0, x) + p(\delta_1, x) = (1-p)x(0) + px(1) = (1-p)\cdot 0 + p \cdot 1 = p.$$

The k-th moment is:

$$EX^k = (1-p)(\delta_0, x^k) + p(\delta_1, x^k) = (1-p)x(0) + px^k(1) = p.$$

Thus,

$$\text{Var}X = \mu_2 - \mu_1^2 = p - p^2 = p(1-p)$$

The MGF is given by

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\mathbf{p_X}(x)dx = (1-p)(\delta_0, e^{tx}) + p(\delta_1, e^{tx}) = (1-p)e^{tx}|_{x=0} + pe^{tx}|_{x=1} = (1-p) + pe^t \quad (t \in \mathbb{R}). \ \square$$

V. **(n,p)-Binomial** (or $B_{n,p}$): $n \in \mathbb{N}$, $0 \le p \le 1$, and the DF is given by

$$F_X(x) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} H(x-i).$$

The "distribution density" is given by

$$\mathbf{p_X}(x) = D_x\bigg(\sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} H(x-i)\bigg) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \delta_i(x).$$

Given a (n,p)-Binomial distributed RV X, then

$$EX = (\mathbf{p_X}, x) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} (\delta_i, x) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} i = np.$$

The k-th moment is:

$$EX^k = (\mathbf{p_X}, x^k) = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} i^k, \quad \text{Var}X = np(1-p), \quad \textit{and} \quad M_X(t) = ((1-p) + pe^t)^n. \ \square$$

VI. **(p)-Geometric** (or G_p): $0 < p < 1$, and the DF is given by

$$F_X(x) = \sum_{i=1}^{\infty} p(1-p)^{i-1} H(x-i).$$

The "distribution density" is given by

$$\mathbf{p_X}(x) = D_x\bigg(\sum_{i=1}^{\infty} p(1-p)^{i-1} H(x-i)\bigg) = \sum_{i=1}^{\infty} p(1-p)^{i-1} \delta_i(x).$$

Given an (p)-Geometric distributed RV X, then

$$EX = (\mathbf{p_X}, x) = \sum_{i=1}^{\infty} p(1-p)^{i-1} (\delta_i, x) = \sum_{i=1}^{\infty} p(1-p)^{i-1} x(i) = \sum_{i=1}^{\infty} p(1-p)^{i-1} i = \frac{1}{p}.$$

$$EX^k = (\mathbf{p_X}, x^k) = \sum_{i=1}^{\infty} p(1-p)^i i^k, \quad \text{Var}X = \frac{1-p}{p^2},$$

and $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$ for $t < -\log(1-p)$. $\square$

Consider a *sequence of (p)-Bernoulli trials*. Then $P(F.S. = k) = p(1-p)^k$ is the probability of the event that *first success occurs in the k-th trial*.

Proposition 5.2 $P(F.S. > n+k | F.S. > n) = P(F.S. > k)$.

Proof:

$$P(F.S. > n+k | F.S. > n) = \frac{P(F.S. > n+k \ \& \ F.S. > n)}{P(F.S. > n)} = \frac{P(F.S. > n+k)}{P(F.S. > n)} = \frac{(1-p)^{n+k}}{(1-p)^n} = (1-p)^k = P(F.S. > k).$$

Meaning of Proposition 5.2 is the following. Assume that we have a collection of devices of different age, whose life-time are (p)-Geometric distributed. Then there is no reason to replace some of them unless they are broken.

VII. **(λ)-Poisson** (or P_λ): $0 < \lambda < \infty$, and the DF is given by

$$F_X(x) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} H(x-i).$$

The "distribution density" is given by $\mathbf{p_X}(x) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} \delta_i(x)$. Given an (λ)-Poisson distributed RV X, then

$$EX = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} (\delta_i, x) = \sum_{i=0}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!} i = e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{(m+1)}}{m!} = e^{-\lambda} \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda,$$

$$EX(X-1) = \sum_{i=0}^{\infty} i(i-1) \frac{\lambda^i}{i!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{i=2}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2,$$

$$\text{Var} X = EX^2 - (EX)^2 = [EX(X-1) + EX] - (EX)^2 = [\lambda^2 + \lambda] - \lambda^2 = \lambda,$$

and $M_X(t) = \exp\Big(\lambda(e^t - 1)\Big)$ for $t \in \mathbb{R}$. $\square$

5.4 Exercises

Exercise 5.1 *Let X and Y be two independent random variables. Find* $\text{Var}XY$ *in each of the following two cases.*

a) *X is $B_{1/2}$ and Y is $G_{1/2}$.*

b) *X is $N_{-2,1}$ and Y is $U_{-2,1}$.*

Solution: a) $\text{Var}X = 1/2(1-1/2) = 1/4$. $\text{Var}Y = \frac{1-1/2}{(1/2)^2} = 2$. $EX = \frac{1}{2}$. $EY = \frac{1}{1/2} = 2$.

Since X and Y are independent, then

$$\text{Var}XY = \text{Var}X \cdot \text{Var}Y + \text{Var}X \cdot (EY)^2 + \text{Var}Y \cdot (EX)^2 = \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 2^2 + 2 \cdot (1/2)^2 = 1/2 + 1 + 1/2 = 2.$$

b)

$$\text{Var}X = 1; \quad \text{Var}Y = \frac{(1-(-2))^2}{12} = 9/12 = 3/4; \quad EX = -2; \quad EY = \frac{1+(-2)}{2} = -1/2;$$

$$\text{Var}XY = 1 \cdot \frac{3}{4} + 1 \cdot \left(-\frac{1}{2}\right)^2 + \frac{3}{4} \cdot (-2)^2 = 3/4 + 1/4 + 12/4 = 4. \ \square$$

6 Laws of Large Numbers

Here we consider the limit behavior of sequences of independent random variables.

6.1 Chebyshev's Law of Large Numbers

We begin with the following proposition.

Proposition 6.1 *Assume that RVs X_j are only pairwise independent. Then*

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathrm{Var}X_1 + \mathrm{Var}X_2 + \ldots + \mathrm{Var}X_n, \tag{18}$$

provided that $\mathrm{Var}X_j$ *are all well defined for all* $1 \le j \le n$.

Proof: Since $EX_jX_k = EX_jEX_k$ for $j \ne k$, we obtain

$$\mathrm{Var}\sum_{j=1}^n X_j = E\Big(\sum_{j=1}^n X_j - \sum_{j=1}^n EX_j\Big)^2 = E\Big(\sum_{j=1}^n (X_j - EX_j)\Big)^2 = \sum_{j=1}^n E(X_j - EX_j)^2 - \sum_{j\ne k} E(X_j - EX_j)(X_k - EX_k) =$$

$$\sum_{j=1}^n \mathrm{Var}X_j - \sum_{j\ne k}(EX_jX_k - EX_jEX_k - EX_jEX_k + EX_jEX_k) = \sum_{j=1}^n \mathrm{Var}X_j - \sum_{j\ne k}(EX_jX_k - EX_jEX_k) = \sum_{j=1}^n \mathrm{Var}X_j. \quad \blacksquare$$

Now, let us remind the **Markov's inequality** (cf., Proposition 4.7). Given a non-negative RV R, then

$$P(R \ge \varepsilon) \le \frac{ER}{\varepsilon} \qquad (\forall \varepsilon > 0). \tag{19}$$

It follows from (19) that for any random variable Y and for any number $c \in \mathbb{R}$:

$$P(|Y - c| \ge \varepsilon) = P(|Y - c|^m \ge \varepsilon^m) \le \frac{E|Y - c|^m}{\varepsilon^m} \qquad (\forall \varepsilon > 0, m \in \mathbb{N}). \tag{20}$$

In particular if $m = 2$, $c = EY$ then

$$P(|Y - EY| \ge \varepsilon) \le \frac{E(Y - EY)^2}{\varepsilon^2} = \frac{\mathrm{Var}Y}{\varepsilon^2} \qquad (\forall \varepsilon > 0). \tag{21}$$

Theorem 6.1 (Chebyshev's law of large numbers) *Let* $(X_j)_{j=1}^{\infty}$ *be a sequence of pairwise independent RVs having finite second moment. Assume also that* $\mathrm{Var}X_j \leq M$ *for all* j *and denote* $S_n = \sum_{j=1}^{n} X_j$. *Then, for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\frac{1}{n}\left|S_n - ES_n\right| \geq \varepsilon\right\} = 0.$$

Proof: Letting $Y = \frac{1}{n}S_n$ in (21) and using independence, we get

$$P\left\{\frac{1}{n}|S_n - ES_n| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2}\mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2\varepsilon^2}\sum_{j=1}^{n}\mathrm{Var}X_j \leq \frac{1}{n^2\varepsilon^2}nM = \frac{M}{n\varepsilon^2} \to 0. \quad (22) \ \blacksquare$$

Corollary 6.1 *Let* $X_1,X_2,\ldots,X_j,\ldots$ *be the results of independent trails in an experiment with* $m = EX_j < \infty$ *and* $\sigma^2 = \mathrm{Var}X_j < \infty$. *Then, for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - m\right| < \varepsilon\right\} = 1. \ \blacksquare$$

Corollary 6.2 *Consider a sequence of* (p)*-Bernoulli trails, and let* X_j *be the number which appears on trail* j. *Then, for any* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|\frac{S_n}{n} - p\right| < \varepsilon\right\} = 1. \ \blacksquare$$

6.2 Central limit theorem

A sequence X_k of RVs is said to **converge in distribution** to an RV X if

$$\lim_{n\to\infty} F_{X_k}(t) = F_X(t),$$

for every $t \in \mathbb{R}$, at which $F_X(t)$ is continuous. In this case, we write $X_k \xrightarrow{d} X$.

Theorem 6.2 (**Central limit theorem**) *Let $(X_j)_j$ be a sequence of i.i.d. (independent identically distributed) RVs with $EX_j = m$ and $\mathrm{Var}X_j = \sigma^2$. Then, for every $a < b$*

$$P\left(a < \frac{S_n - nm}{\sigma\sqrt{n}} < b\right) \to F_{N_{0,\sigma^2}}(b) - F_{N_{0,\sigma^2}}(a) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt \ ,$$

where $S_n = \sum_{j=1}^{n} X_j$. In other words,

$$\frac{S_n - ES_n}{\sqrt{\mathrm{Var}S_n}} \xrightarrow{d} Z \in N_{0,1} \qquad (n \to \infty).$$

Proof: To be included later. ∎

A a special case of Theorem 6.2 is:

Theorem 6.3 (**Moivre-Laplace's theorem**) *Let $(X_j)_j$ be a sequence of independent (p)-Bernoulli distributed RVs. Then, for every $0 \le \alpha < \beta$*

$$P(\alpha < S_n < \beta) \approx F_{N_{0,1}}\left(\frac{\beta - np}{\sqrt{np(1-p)}}\right) - F_{N_{0,1}}\left(\frac{\alpha - np}{\sqrt{np(1-p)}}\right), \quad (23)$$

where $F_{N_{0,1}}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$.

Proof:

$$P(\alpha < S_n < \beta) = P(\alpha - np < S_n - np < \beta - np) = P\left(\frac{\alpha - np}{\sigma\sqrt{n}} < \frac{S_n - np}{\sigma\sqrt{n}} < \frac{\beta - np}{\sigma\sqrt{n}}\right) =$$

$$P\left(\frac{\alpha - np}{\sqrt{np(1-p)}} < \frac{S_n - np}{\sigma\sqrt{n}} < \frac{\beta - np}{\sqrt{np(1-p)}}\right) \approx F_{N_{0,1}}\left(\frac{\beta - np}{\sqrt{np(1-p)}}\right) - F_{N_{0,1}}\left(\frac{\alpha - np}{\sqrt{np(1-p)}}\right). \quad ∎$$

6.3 Exercises

Exercise 6.1 *Given a sequence X_n of independent identically distributed random variables with moment generating functions $M_{X_n}(t) = \frac{n-1+e^t}{n}$. Are the conditions of Law of Large Numbers satisfied for the following sequences.*

a) For the sequence X_n.

b) For the sequence $Y_n = \sum_{i=1+2^{n^2}}^{2^{(n+1)^2}} X_i$.

c) For the sequence $Z_n = X_1 + (-1)^n X_2$.

Solution:
a) The sequence X_n is independent.

$$EX_n = M'_{X_n}(t)\big|_{t=0} = \frac{1}{n}e^t\bigg|_0 = \frac{1}{n}; \quad EX_n^2 = \left(\frac{1}{n}e^t\right)'\bigg|_0 = \frac{1}{n}; \quad \mathrm{Var}X_n = EX_n^2 - (EX_n)^2 = \frac{n-1}{n^2} \le 1.$$

So $\mathrm{Var}X_n$ is uniformly bounded. Hence the conditions of LLN are satisfied for X_n.

b) The sequence Y_n is independent, since the families $\{X_{1+2^{n^2}}, \dots, X_{2^{(n+1)^2}}\}$ are disjoint. But

$$\mathrm{Var}Y_n = \sum_{i=1+2^{n^2}}^{2^{(n+1)^2}} \mathrm{Var}X_i = \sum_{i=1+2^{n^2}}^{2^{(n+1)^2}} \frac{i-1}{i^2} \ge \sum_{i=1+2^{n^2}}^{2^{(n+1)^2}} \frac{1}{i} \sim \int_{2^{n^2}}^{2^{(n+1)^2}} \frac{1}{t}dt = \ln t\big|_{2^{n^2}}^{2^{(n+1)^2}} = \ln 2^{(n+1)^2} - \ln 2^{n^2} = \ln 2^{2n+1} \to \infty.$$

So $\mathrm{Var}Y_n$ are not uniformly bounded, and the conditions of LLN are not satisfied.

c) The sequence $(Z_n)_n$ contains the same RV on odd places, so $(Z_n)_n$ is not pairwise independent, and hence the conditions of LLN are not satisfied. □

Exercise 6.2 *1000 fair dice are rolled. Consider the following event:*

$A = \{290 < \textit{sum of scores on first 100 dice} < 410\}$; $B = \{3400 < \textit{sum of scores on all 1000 dice} < 3600\}$.

a) *Show that* $P(A) > 91\%$.

b) *Show that* $P(B) > 2/3$.

Solution:
a) Consider random variables X_j that are the scores of j-th die for $j = 1, \dots, 1000$.

$$EX_j = \frac{1}{6}(1+2+\dots+6) = \frac{7}{2}; \quad EX_j^2 = \frac{1}{6}(1+4+\dots+36) = \frac{91}{6}; \quad \mathrm{Var}X_j = EX_j^2 - (EX_j)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

Denote $S_{100} = \sum_{j=1}^{100} X_j$. Then

$$P(A) = P\left(\left|\sum_{j=1}^{100} X_j - 350\right| < 60\right) = P\left(\frac{1}{100}|S_{100} - ES_{100}| < 0.6\right) = 1 - P\left(\frac{1}{100}|S_{100} - ES_{100}| \ge 0.6\right) \ge$$

$$1 - \frac{\mathrm{Var}S_{100}}{(100)^2 \cdot (0.6)^2} = 1 - \frac{100\mathrm{Var}X_j}{3600} = 1 - \frac{1}{36}\cdot\frac{35}{12} > 1 - \frac{1}{12} = \frac{11}{12} > \frac{91}{100} = 91\%.$$

b)

$$P(B) = P\left(\left|\sum_{j=1}^{1000} X_j - 3500\right| < 100\right) = P\left(\frac{1}{1000}|S_{1000} - ES_{1000}| < 0.1\right) = 1 - P\left(\frac{1}{1000}|S_{1000} - ES_{1000}| \ge 0.1\right) \ge$$

$$1 - \frac{\mathrm{Var}S_{1000}}{(1000)^2 \cdot (0.1)^2} = 1 - \frac{1000 \cdot 35/12}{(1000)^2 \cdot 0.01} = 1 - \frac{1}{10}\cdot\frac{35}{12} = 1 - \frac{35}{120} = \frac{85}{120} = \frac{17}{24} > \frac{16}{24} = \frac{2}{3}. \quad \square$$

Exercise 6.3 *Given a sequence* $(X_n)_{n=1}^{\infty}$ *of independently distributed random variables with the moment generating functions* $M_{X_n}(t) = \frac{e^t}{2-e^t}$, $-\infty < t < \ln 2$.

a) *What is the expected value* EX_n *and the variance* $\mathrm{Var}X_n$*?*

b) *Show that* $P(\{1900 < \sum_{n=1}^{n=1000} X_n < 2100\}) \ge 80\%$.

Solution: a)

$$M'_{X_n}(t) = \frac{2e^t}{(2-e^t)^2} \Rightarrow EX_n = M'_{X_n}(0) = 2.$$

$$M''_{X_n}(t) = \left[\frac{2e^t}{(2-e^t)^2}\right]' = \frac{2e^t(2-e^t)^2 + 4e^{2t}(2-e^t)}{(2-e^t)^4}.$$

So $EX_n^2 = M''_{X_n}(0) = 6$ and $\mathrm{Var}X_n = EX_n^2 - (EX_n)^2 = 6 - 2^2 = 2$.

b)

$$P = P\left(\left|\sum_{n=1}^{1000} X_n - \sum_{n=1}^{1000} EX_n\right| < 100\right) = [\text{since} = \ EX_n = 2] = 1 - P\left(\frac{1}{1000}\left|\sum_{n=1}^{1000} X_n - 2000\right| \geq 0.1\right) \geq$$

$$1 - \frac{\mathrm{Var}\sum_{n=1}^{1000} X_n}{(1000)^2 \cdot (0.1)^2} = 1 - \frac{2000}{10000} = 1 - 0.2 = 80\%. \quad \square$$

Exercise 6.4 *Given a sequence $(X_n)_{n=1}^{\infty}$ of independent equally distributed random variables with the moment generating function: $M_{X_n}(t) = \frac{1}{1-\frac{t}{2}}$.*
a) Find the expected value EX_n and the variance $\mathrm{Var}X_n$.
b) Show that $P(\{100 < \sum_{n=1}^{n=400} X_n < 300\}) \geq 99\%$.

Solution: *a)*

$$M'_{X_n}(t) = -\frac{1}{(1-\frac{t}{2})^2} \cdot \left(1 - \frac{t}{2}\right)' = \frac{1}{2} \cdot \frac{1}{(1-\frac{t}{2})^2} \Rightarrow EX_n = M'_{X_n}(0) = \frac{1}{2}.$$

$$M''_{X_n}(t) = \frac{1}{2}(-2) \cdot \frac{1}{(1-\frac{t}{2})^3} \cdot \left(-\frac{1}{2}\right) = \frac{1}{2} \cdot \frac{1}{(1-\frac{t}{2})^3} \Rightarrow EX_n^2 = M''_{X_n}(0) = \frac{1}{2}.$$

$$\mathrm{Var}X_n = EX_n^2 - (EX_n)^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

b) $EX_n = \frac{1}{2}$, $\mathrm{Var}X_n = \frac{1}{4}$.

$$P\left(100 < \sum_{n=1}^{400} X_n < 300\right) = P\left(\left|\sum_{n=1}^{400} X_n - 200\right| < 100\right) = P\left(\left|\frac{\sum_{n=1}^{400} X_n}{400} - \frac{E(\sum_{n=1}^{400} X_n)}{400}\right| < \frac{1}{4}\right) =$$

$$1 - P\left(\frac{1}{400}\left|\sum_{n=1}^{400} X_n - 200\right| \geq \frac{1}{4}\right) \geq 1 - \frac{\mathrm{Var}\left(\sum_{n=1}^{400} X_n\right)}{400^2 \cdot \left(\frac{1}{4}\right)^2} = 1 - \frac{400 \cdot \frac{1}{4}}{400^2 \cdot \left(\frac{1}{4}\right)^2} = 1 - \frac{1}{100} = 99\%. \quad \square$$

Exercise 6.5 *Given a sequence $(X_n)_{n=1}^{\infty}$ of independent (p)-geometrically distributed random variables for $0 < p < 1$.*

a) *What is the expected value EX_n and the variance $\mathrm{Var}X_n$? [Hint: You may use the fact that $M'_{X_n}(t) = \frac{pe^t}{(1-e^t+pe^t)^2}$ for $t < -\ln(1-p)$, where $M_{X_n}(t)$ is the moment generating function of X_n.]*

b) *Assume that $p = 1/2$. Estimate the probability $P(B)$ of the following event: $B = \{800 < \sum_{n=1}^{n=500} X_n < 1200\}$.*

Solution:

a)

$$M'_{X_n}(0) = \frac{p \cdot 1}{(1-1+p)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

$$M''_{X_n}(t) = \frac{pe^t(1-e^t+pe^t)^2 - pe^t \cdot 2(1-e^t+pe^t)(-e^t+pe^t)}{(1-e^t+pe^t)^4}.$$

$$M''_{X_n}(0) = \frac{p \cdot 1(1-1+p)^2 - p \cdot 1 \cdot 2(1-1+p)(-1+p)}{(1-1+p)^4} = \frac{p^3 - 2p^2(p-1)}{p^4} = \frac{2-p}{p^2}.$$

$$EX_n = M'_{X_n}(0) = \frac{1}{p}.$$

$$\text{Var}X_n = M''_{X_n}(0) - (M'_{X_n}(0))^2 = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}.$$

b)

$$p = 1/2 \Rightarrow EX_n = \frac{1}{1/2} = 2 \text{ and } \text{Var}X_n = \frac{1-1/2}{(1/2)^2} = \frac{1}{1/2} = 2.$$

$$P(B) = P\left(\left|\sum_{n=1}^{500} X_n - 1000\right| < 200\right) = P\left(\left|\frac{\sum_{n=1}^{500} X_n}{500} - 2\right| < \frac{2}{5}\right) = 1 - P\left(\left|\frac{1}{500}\sum_{n=1}^{500} X_n - 2\right| \geq \frac{2}{5}\right) \geq$$

$$1 - \frac{1}{500^2 \cdot \left(\frac{2}{5}\right)^2} \cdot \text{Var}\sum_{n=1}^{500} X_n = 1 - \frac{25}{4 \cdot 250000} \cdot 500 \cdot 2 = 1 - \frac{25}{2 \cdot 500} = 1 - 2.5\% = 97.5\%. \quad \square$$

Exercise 6.6 *Given a sequence $(X_k)_{k=1}^{\infty}$ of independent random variables with the moment generating functions $M_{X_k}(t) = \frac{1}{1-t}$, $|t| < 1$.*

a) What is the expected value $E(X_k^2)$ and the variance $\text{Var}(X_k^2)$?

b) Show that $P(1500 < \sum_{k=1}^{k=1000} X_k^2 < 2500) > 91\%$.

Solution:

a)

$$M_{X_k}(t) = \sum_{j=0}^{\infty} t^j, \quad EX_k^2 = M''_{X_k}(t)|_{t=0} = 2! = 2.$$

$$EX_k^4 = M^{(4)}_{X_k}(t)\Big|_{t=0} = 4! = 24.$$

$$\text{Var}(X_k^2) = E(X_k^2)^2 - (EX_k^2)^2 = 24 - 2^2 = 20. \ \square$$

b)

$$P(1500 < \sum_{k=1}^{1000} X_k^2 < 2500) = 1 - P(|S_{1000} - 2000| \geq 500) =$$

$$1 - P\left(\frac{|S_{1000} - ES_{1000}|}{1000} \geq 0.5\right) \overset{1000=n, 0.5=\varepsilon}{\geq} 1 - \frac{1}{\varepsilon^2}\text{Var}\left(\frac{S_n - ES_n}{n}\right) =$$

$$1 - \frac{1}{0.25} \cdot \frac{1}{n^2}\text{Var}S_n = 1 - \frac{1}{250\,000} \cdot 1000 \cdot 20 = 1 - \frac{2}{25} = 0.92 > 91\%. \ \square$$

Exercise 6.7 *A market every day gets from a farm chicken eggs in boxes. Each box contains 300 eggs. The standard percentage of the crashing is* 15%*. How many (in* %*) boxes of usual day purchase contain less than 40 crashed eggs?*

Solution:
Clear, we need only to find the probability of the event {there are $0 \le S_{300} \le 40$ crashed eggs in a box}. We apply the Moivre-Laplace for independent (p)-Bernoulli distributed RVs $X_j, j = 1, \ldots, 300$, where $p = 0.15$.

$$EX_j = p = 0.15, \quad \text{Var}X_j = p(1-p) = 0.15 \cdot 0.85 = 0.1275, \quad \alpha = 0, \beta = 40, \quad n = 300.$$

By (23),

$$P(\alpha < S_n < \beta) \approx F_{N_{0,1}}\left(\frac{\beta - np}{\sqrt{np(1-p)}}\right) - F_{N_{0,1}}\left(\frac{\alpha - np}{\sqrt{np(1-p)}}\right).$$

Therefore,

$$P(0 < S_{300} < 40) \approx F_{N_{0,1}}\left(\frac{40 - 300 \cdot 0.15}{\sqrt{300 \cdot 0.1275}}\right) - F_{N_{0,1}}\left(\frac{0 - 300 \cdot 0.15}{\sqrt{300 \cdot 0.1275}}\right) = F_{N_{0,1}}\left(\frac{40-45}{\sqrt{38.25}}\right) - F_{N_{0,1}}\left(\frac{0-45}{\sqrt{38.25}}\right) =$$

$$F_{N_{0,1}}\left(\frac{-5}{6.18}\right) - F_{N_{0,1}}\left(\frac{-45}{6.18}\right) = F_{N_{0,1}}(-0.81) - F_{N_{0,1}}(-7.28) = 1 - F_{N_{0,1}}(0.81) - 1 + F_{N_{0,1}}(7.28) =$$

$$[\text{use the table 8.1 from Appendix}] = 1 - 0.7910 - 1 + 1 = 0.209.$$

Thus, $\approx 21\%$ of boxes contain less than 40 crashed eggs. □

Exercise 6.8 *The probability* p *of the event* A *is* 0.8 *for each trial. How many times we have to repeat the trial, if we want to expect with the probability* 90% *that the frequency of occurrence of the event will deviant from* p *less than by* 0.05*?*

Solution:

$$\left|\frac{S_n}{n} - 0.8\right| < 0.05 \quad \Rightarrow \quad 0.75n < S_n < 0.85n.$$

Put in Theorem 6.3 $\alpha = 0.75n$ and $\beta = 0.85n$. So, by (23),

$$P(0.75n < S_n < 0.85n) \approx F_{N_{0,1}}\left(\frac{0.85n - 0.8n}{\sqrt{n \cdot 0.8 \cdot 0.2}}\right) - F_{N_{0,1}}\left(\frac{0.75n - 0.8n}{\sqrt{n \cdot 0.8 \cdot 0.2}}\right) = F_{N_{0,1}}\left(\frac{0.05n}{\sqrt{n \cdot 0.16}}\right) - F_{N_{0,1}}\left(\frac{-0.05n}{\sqrt{n \cdot 0.16}}\right) =$$

$$F_{N_{0,1}}\left(0.125\sqrt{n}\right) - F_{N_{0,1}}\left(-0.125\sqrt{n}\right) = F_{N_{0,1}}\left(0.125\sqrt{n}\right) - 1 + F_{N_{0,1}}\left(0.125\sqrt{n}\right) =$$

$$2F_{N_{0,1}}\left(0.125\sqrt{n}\right) - 1 = 0.9 \Rightarrow F_{N_{0,1}}\left(0.125\sqrt{n}\right) = 0.95.$$

Using the table 8.1 from Appendix, we get

$$0.125\sqrt{n} = 1.64 \quad \Rightarrow \quad n = \left(\frac{1.64}{0.125}\right)^2 = 13.12^2 = 172.1344$$

Thus it will be enough 172 trials. □

Exercise 6.9 *A fair coin is tossed 400 times. By using of the Moivre – Laplace theorem, calculate.*

a) The probability of getting the tails strictly between 190 and 210 times?

b) The probability of getting the heads strictly between 180 and 200 times?

Solution:
a)

$$EX_j = p = 0.5, \quad EX_j^2 = p = 0.5, \quad \mathrm{Var}X_j = p - p^2 = 0.25.$$

$$P(190 < S_{400} < 210) \approx F_{N_{0,1}}\left(\frac{210 - 400 \cdot 0.5}{\sqrt{400 \cdot 0.25}}\right) - F_{N_{0,1}}\left(\frac{190 - 400 \cdot 0.5}{\sqrt{400 \cdot 0.25}}\right) =$$

$$F_{N_{0,1}}(1) - F_{N_{0,1}}(-1) = F_{N_{0,1}}(1) - 1 + F_{N_{0,1}}(1) \overset{\text{use the table}}{=} 2 \cdot 0.8413 - 1 = 0.6826 \approx 68\%. \ \square$$

b)

$$P(180 < S_{400} < 200) \approx F_{N_{0,1}}\left(\frac{200 - 400 \cdot 0.5}{\sqrt{100}}\right) - F_{N_{0,1}}\left(\frac{180 - 400 \cdot 0.5}{\sqrt{100}}\right) =$$

$$F_{N_{0,1}}(0) - F_{N_{0,1}}(-2) = 0.5 - 1 + F_{N_{0,1}}(2) \overset{\text{use the table}}{=} 0.9772 - 0.5 = 0.4772 \approx 48\%. \ \square$$

Exercise 6.10 *To get the result of an experiment, one needs to add 2500 numbers. The rounding precision of each number is* 10^{-4}*. Suppose that occurred (by rounding) errors are independent and uniformly distributed in the interval* $(-0.5 \cdot 10^{-4}, 0.5 \cdot 10^{-4})$*. Find the interval* $(-r, r)$ *that contains the total (sum) error of this counting with the probability* 0.99*.*

Solution:
We need to find $r > 0$ such that $P(-r < S_n < r) \approx 0.99$, where $n = 2500$, $S_n = X_1 + \cdots + X_n$, and X_j are independent RVs with the uniform distribution in the interval $(-0.5 \cdot 10^{-4}, 0.5 \cdot 10^{-4})$. [For uniform distribution, we have math expectation is the middle of the interval and variance is equal to $d^2/12$, where d is a length of the interval.] $m = EX_j = 0$ and $\sigma^2 = \mathrm{Var}X_j = \frac{10^{-8}}{12}$, $\sigma = \frac{1}{2\sqrt{3} \cdot 10^4}$.

By Theorem 6.2,

$$P(-r < S_n < r) \approx F_{N_{0,1}}\left(\frac{r}{\sigma\sqrt{n}}\right) - F_{N_{0,1}}\left(-\frac{r}{\sigma\sqrt{n}}\right) = F_{N_{0,1}}\left(\frac{2\sqrt{3} \cdot 10^4 r}{500}\right) - F_{N_{0,1}}\left(-\frac{2\sqrt{3} \cdot 10^4 r}{500}\right) =$$

$$F_{N_{0,1}}\left(40\sqrt{3}r\right) - F_{N_{0,1}}\left(-40\sqrt{3}r\right) = 2F_{N_{0,1}}\left(\sqrt{3}r\right) - 1.$$

Using the table from Appendix, we find u such that that

$$2F_{N_{0,1}}(u) - 1 = 0.99 \quad \Leftrightarrow \quad F_{N_{0,1}}(u) = 0.995.$$

That is $u = 2.58$. After solving the following equation

$$40\sqrt{3}r = 0.995 \quad \Rightarrow \quad r = \frac{0.995}{40\sqrt{3}} = \frac{0.995}{40 \cdot 1.732} = \frac{0.995}{69.28} = 0.014,$$

we get that the the total error will be in the interval $(-1.4 \cdot 10^{-4}, 1.4 \cdot 10^{-4})$ $\square$

7 Markov chains with finite number of states

A **stochastic process** is an ordered set $\mathcal{X} = \{\vec{X}_i\}_{i\in I}$ of random vectors. When the indexing set I is countable, $\mathcal{X}$ is called a **discrete time stochastic process**. When the indexing set I is a nontrivial interval in $\mathbb{R}$, $\mathcal{X}$ is called a **continuous time stochastic process**.

In this section, $\mathcal{X} = (X_n)_{n=0}^{\infty}$ is a sequence of $\mathbb{N}$-valued RVs on a probability space $(S, \mathcal{A}, P)$. Instead of saying $P(X_n = k) = \alpha$ we also say that **X_n is in state k with probability α**. The collection $E = \{k\}_{k=1}^{\infty}$ is called the **state space of the random process (or, stochastic process) $(X_n)_{n=0}^{\infty}$**.

Example 7.1 *Let X_n be the number of students who planning to be graduated from ODTU in $(2018+n)$-th year. A probability space $(S, \mathcal{A}, P)$ can be arbitrary. Notice that, in general*

$$P(X_{n+1} = n+1 | X_0 = k_0, \ldots, X_{n-1} = k_{n-1}, X_n = k_n) \neq P(X_{n+1} = n+1 | X_n = k_n). \quad \square$$

7.1 Markov chains

If the process , given the present, the future is independent of the past, the process is called a Markov chain.

Definition 7.1 *$\mathcal{X} = (X_n)_{n=0}^{\infty}$ is said to be a* **Markov chain** *whenever*

$$P(X_{n+1} = n+1 | X_0 = k_0, \ldots, X_n = k_n) = P(X_{n+1} = n+1 | X_n = k_n) \tag{24}$$

for every $n \geq 0$ and every states $k_0, \ldots, k+1$.
A Markov chain $\mathcal{X}$ is said to be **homogeneous** *whenever*

$$P(X_{n+1} = k_2 | X_n = k_1) = P(X_{n+l+1} = k_2 | X_{n+l} = k_1) \tag{25}$$

for every $n, l \geq 0$ and every states k_1, k_2.

Example 7.2 *Any independent $\mathbb{N}$-valued stochastic process $(X_n)_{n=0}^{\infty}$ is a homogeneous Markov chain.* $\square$

In what follows $\mathcal{X}$ is a **homogeneous Markov chain with the finite state space $E = \{1, 2, \ldots, m\}$**, and we shall say that $\mathcal{X}$ is just a Markov chain.

7.2 Transition matrices

Let $n \in \mathbb{N}$, $1 \leq i, j \leq m$. Denote by

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i) \tag{26}$$

the **n-step transition probability from state i to state j**. The 1-step transition probabilities will be simple denoted by p_{ij} and called by the **transition probabilities** of the Markov chain $(X_n)_{n=0}^{\infty}$. The $m \times m$-matrix

$$P^{(n)} = \begin{bmatrix} p_{11}^{(n)} & p_{12}^{(n)} & \dots & p_{1m}^{(n)} \\ p_{21}^{(n)} & p_{22}^{(n)} & \dots & p_{2m}^{(n)} \\ \vdots & \vdots & & \vdots \\ p_{m1}^{(n)} & p_{m2}^{(n)} & \dots & p_{mm}^{(n)} \end{bmatrix}$$

is called the **n-step transition matrix**. The 1-step transition matrix will be simple denoted by P and called the **transition matrix of the Markov chain**.

Clearly, every matrix $P^{(n)}$ satisfies

$$P^{(n)} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Therefore, $P^{(n)}$ is **stochastic** (that is: $p_{ij}^{(n)} \geq 0$ for all $1 \leq i, j \leq m$, and $\sum_{j=1}^{m} p_{ij}^{(n)} = 1$ for every $1 \leq i \leq m$).

Since

$$p_{ij}^{(n)} = \sum_{k=1}^{m} P(X_n = j | X_{n-1} = k) \cdot P(X_{n-1} = k | X_0 = j) = \sum_{k=1}^{m} p_{ik}^{(n-1)} \cdot p_{kj}^{1},$$

we obtain

$$P^{(n)} = P^{(n-1)} \cdot P = P^{(n-2)} \cdot P^2 = \ldots = P^n. \tag{27}$$

7.3 Probability distribution of a Markov chain

The number

$$p_k(n) := P(X_n = k)$$

is called the **probability of state k at time $n \geq 0$**. The vector

$$\vec{p}(n) := (p_1(n), p_2(n), \ldots, p_m(n))$$

is called the **probability distribution of the Markov chain $(X_j)_{j=0}^{\infty}$ at time $n \geq 0$**. Clearly,

$$\vec{p}(n+l) := \vec{p}(n) \cdot P^l \qquad (\forall n, l \geq 0). \tag{28}$$

The vector

$$\vec{p} := (p_1(1), p_2(1), \ldots, p_m(1))$$

is called the **probability distribution of $(X_j)_{j=0}^{\infty}$**.

7.4 Classification of states

We say that **the state j is accessible from i** if $p_{ij}^{(n)} > 0$ for some $n \geq 1$. Then we write $i \to j$. If $p_{ij}^{(n)} = 0$ for all $n \geq 1$, we write $i \not\to j$. We say that i **communicates** with j if $i \to j$ and $j \to i$. In this case we write $i \leftrightarrow j$. Otherwise, we write $i \not\leftrightarrow j$. Clearly, $\leftrightarrow$ is an equivalence relation on the state space E. A state i is said to be **absorbing** if $p_{ii} = 1$.

Definition 7.2 *A Markov chain is said to be* **irreducible** *if every two of its states communicate.*

Notice that a Markov chain is irreducible iff the state space E is the only one equivalence class of the relation $\leftrightarrow$.

Definition 7.3 *The* **period** $\mathrm{Per}(i)$ *of a state i is the greatest common divisor of all $n \geq 1$ such that $p_{ii}^{(n)} > 0$. If $p_{ii}^{(n)} = 0$ for all $n \geq 1$, we say that* $\mathrm{Per}(i)$ *is undefined. The state i is called* **aperiodic** *if* $\mathrm{Per}(i) = 1$.

Notice that if $i \leftrightarrow j$ then $\text{Per}(i) = \text{Per}(j)$.

Example 7.3 *Given a transition matrix* $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$ *The correspondent Markov chain is not irreducible, since* $2 \nrightarrow 1$. *Indeed,* $P^{(n)} = P^n = \begin{bmatrix} \frac{1}{2^n} & g_n \\ 0 & 1 \end{bmatrix}$ *for every* n, *and hence* $p_{21}^{(n)} = [P^{(n)}]_{2,1} = 0$ *for all* n. *Both states* 1 *and* 2 *are aperiodic, but only state* 2 *is absorbing.* □

Example 7.4 *Given a transition matrix* $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ *The correspondent Markov chain is irreducible, since* $i \leftrightarrow j$ *for every* i *and* j. *Moreover,* $\text{Per}(i) = 3$ *for* $i = 1, 2, 3$. □

Denote

$$f_{ii}^{(n)} := P(X_n = i | X_k \neq i[k = 1, 2, \ldots, n-1] | X_0 = i)$$

and

$$f_{ii}^* := \sum_{n=1}^{\infty} f_{ii}^{(n)}.$$

Thus, the number f_{ii}^* is the probability of the event that *the process starting from state i would return to i in a finite time*.

Definition 7.4 *A state* i *is called* **recurrent** *if* $f_{ii}^* = 1$ *(in other words, if starting from* i*, eventual return to* i *is certain). A state* i *is called* **transient** *if* $f_{ii}^* < 1$.

In Example 7.4 all states are recurrent.

Theorem 7.1 *A state* i *is recurrent if and only if* $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

Proof: To be included later. ■

As an application of Theorem 7.1, one may easily obtain that in Example 7.3 the state 1 is transient and the state 2 is recurrent.

7.5 Limiting behavior

Theorem 7.2 *If a Markov chain with the transition matrix P is irreducible and every state is aperiodic, then there exists $N \in \mathbb{N}$ such that for every $n \geq N$, the matrix P^n has no nonzero elements.*

Proof: To be included later. ■

Theorem 7.3 (A.A. Markov) *Suppose that for some k all entries of P^k are nonzero. Then there are strictly positive numbers $p_1, p_2, \ldots, p_m$ such that*

$$\lim_{n\to\infty} p_{ij}^{(n)} = p_j, \quad \sum_{j=1}^{m} p_j = 1.$$

in other words the sequence $(P^{(n)})_n$ of n-step transition matrices entries-wise converges to a matrix with constant nonzero columns.

Proof: To be included later. ■

Definition 7.5 *A probability distribution $\vec{p}$ of $(X_j)_{j=0}^{\infty}$ is called* **stationary** *if $\vec{p} = \vec{p} \cdot P$.*

It follows from Theorem 7.3 that if for some k all entries of P^k are nonzero then $\vec{p} := (p_1, p_2, \ldots, p_m)$ is a stationary probability distribution of $(X_j)_{j=0}^{\infty}$.

Definition 7.6 *If for some k all entries of P^k are nonzero, then the Markov chain is said to be* **ergodic**.

Example 7.5 *If a student did not took a course last semester then he would take the course this semester with probability 30%. If he took the course last semester then he would not take the course this semester with probability 60%. Find the probability that the student take the course.*

Consider a Markov chain with states: 1 that the student fails to pass a course this year and 2 that he passes the course this year. The corresponding transition matrix $P = \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$ *is ergodic. Hence,*

$$p_1 + p_2 = 1, \quad 0.7p_1 + 0.6p_2 = p_1; \;\Rightarrow\; \quad p_1 = 2p_2 = 1/3.$$

Therefore, the probability that the student the course is 1/3. □

Theorem 7.2 and Theorem 7.3 imply that *every irreducible Markov chain, in which all states are aperiodic, has a strictly positive stationary distribution*. Indead¡ it is true even more, namely.

Theorem 7.4 *A Markov chain is ergodic if and only if it is irreducible and all its states are aperiodic.*

Proof: To be included later. ■

7.6 Exercises

Exercise 7.1 *Given a Markov chain with the state space $E = \{1, 2, 3, 4, 5, 6, 7\}$ and the transition matrix:*

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Draw the corresponding directed graph and

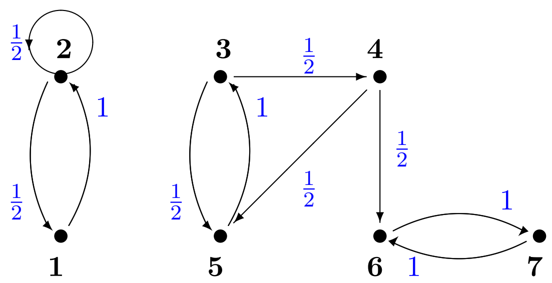
a) *find an equivalence class of E consisting of intercommunicating recurrent states;*

b) *find an equivalence class of E consisting of intercommunicating transient states;*

c) *Find the period of each state in E;*

d) *Find a stationary probability distribution $\vec{p} = (p_1, p_2, \ldots, p_7)$ for the Markov chain.*

Solution:



a) $E_1 = \{6, 7\}$ is a class consisting of intercommunicating recurrent states.

b) $E_2 = \{3, 4, 5\}$ is a class consisting of intercommunicating transient states.

c) $\text{Per}(1) = \text{Per}(2) = 1$.

$\text{Per}(3) = \text{Per}(4) = \text{Per}(5) = 1$.

$\text{Per}(6) = \text{Per}(7) = 2.$

d) It can be done inside of any of our 3 classes. Let us do this in $E_1 = \{6,7\}$, so $p_1 = p_2 = p_3 = p_4 = p_5 = 0$.

$$\vec{p} = \vec{p} \cdot p \Rightarrow (p_6, p_7) = (p_6, p_7) \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow (p_6, p_7) = (p_7, p_6) \Rightarrow p_6 = p_7.$$

Since $\sum_{k=1}^{7} p_k = 1$, we have $\vec{p} = (0,0,0,0,0,\frac{1}{2},\frac{1}{2})$. □

Exercise 7.2 *Given* $P = \begin{bmatrix} \alpha & \alpha & 0 \\ 0 & \alpha & \alpha \\ \alpha & 0 & \alpha \end{bmatrix}$.

a) *Find the value of* α *such that* P *is a transition matrix of a Markov chain* $(X_n)_{n=0}^{\infty}$.

b) *Denote by* $\{1,2,3\}$ *the state space of the Markov chain* $(X_n)_{n=0}^{\infty}$ *from* **a)**.

(i) *Is the Markov chain* $(X_n)_{n=0}^{\infty}$ *ergodic?*

(ii) *Find the conditional probability of* $X_2 = 3$ *given* $X_0 = 1$.

Solution:
a)

$$\alpha + \alpha = 1 \Rightarrow \alpha = \frac{1}{2}.$$

b) (i)

$$P^2 = \begin{bmatrix} \alpha & \alpha & 0 \\ 0 & \alpha & \alpha \\ \alpha & 0 & \alpha \end{bmatrix} \cdot \begin{bmatrix} \alpha & \alpha & 0 \\ 0 & \alpha & \alpha \\ \alpha & 0 & \alpha \end{bmatrix} = \begin{bmatrix} \alpha^2 & 2\alpha^2 & \alpha^2 \\ \alpha^2 & \alpha^2 & 2\alpha^2 \\ 2\alpha^2 & \alpha^2 & \alpha^2 \end{bmatrix}.$$

$P^{(2)} = P^2$ has all entries nonzero, hence $(X_n)_n$ is ergodic.

b) (ii)

$$P(X_2 = 3 | X_0 = 1) = P_{13}^{(2)} = [P^2]_{13} = \alpha^2 = \frac{1}{4}. \quad \square$$

Exercise 7.3 *Given a Markov chain with the state space* $E = \{1,2,3,4,5,6,7\}$ *and the transition matrix:*

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Draw the corresponding directed graph and then:

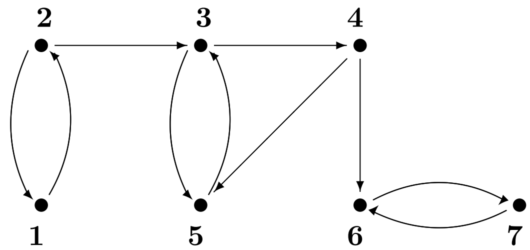
a) *Find an equivalence class of* E *consisting of intercommunicating recurrent states.*

b) *Find two equivalence classes of* E *consisting of intercommunicating transient states.*

c) *Find the period of every state in* E.

d) *Find a stationary probability distribution* $\vec{p} = (p_1, p_2, \ldots, p_7)$ *for the Markov chain.*

Solution:



a) For example, $C_1 = \{6, 7\}$.

b) For example, $C_2 = \{1, 2\}$, $C_3 = \{3, 4, 5\}$.

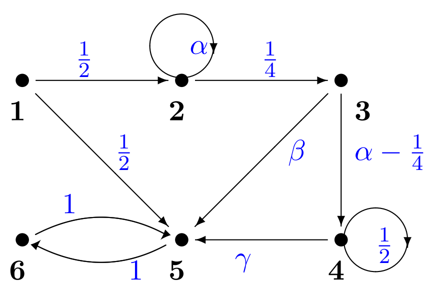
c) $\text{Per}(1) = \text{Per}(2) = 2$.

$\text{Per}(3) = \text{Per}(4) = \text{Per}(5) = 1$.

$\text{Per}(6) = \text{Per}(7) = 2$.

d) $C_1 = \{6, 7\}$ is recurrent, so we may take a stationary state (p_6, p_7) of C_1 which is $(\frac{1}{2}, \frac{1}{2})$ and complement it by $p_1 = p_2 = p_3 = p_4 = p_5 = 0$, that is $\vec{p} = (0, 0, 0, 0, 0, \frac{1}{2}, \frac{1}{2})$. $\square$

Exercise 7.4 *A homogeneous Markov chain is given by the following directed graph:*



a) Find values of α, β, *and* γ.

b) Write down the transition matrix of the Markov chain.

c) Find $P(X_{2018} = 3 | X_{2015} = 1)$.

d) Find periods of all 6 states.

e) Find a stationary probability distribution of the Markov chain.

Solution:

a)

$$\alpha = 1 - \frac{1}{4} = \frac{3}{4}; \quad \beta = 1 - (\alpha - \frac{1}{4}) = 1 - \frac{1}{2} = \frac{1}{2}; \quad \gamma = 1 - \frac{1}{2} = \frac{1}{2}. \ \square$$

b)

$$P = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \ \square$$

c)

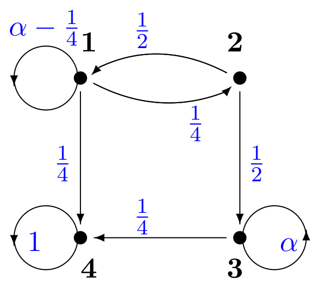
$$P(X_{2018}=3|X_{2015}=1)=\frac{1}{2}\cdot\alpha\cdot\frac{1}{4}=\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{4}=\frac{3}{32}.\ \square$$

d) Since for all n we have $p_{11}^{(n)}=p_{33}^{(n)}=0$, then Per(1) and Per(3) are undefined. Next, Per(2) = Per(4) = 1 and Per(5) = Per(6) = 2. $\square$

e) $C_1=\{5,6\}$ is an equivalence class with respect to $\longleftrightarrow$, so we may take a stationary state (p_5,p_6) of C_1, that is $(1/2,1/2)$, and complement it by $p_1=p_2=p_3=p_4=0$, obtaining

$$\vec{p}=\left(0,0,0,0,\frac{1}{2},\frac{1}{2}\right).\ \square$$

Exercise 7.5 *Given a directed graph:*



a) *Find the value of the parameter α for which the graph represents a Markov chain $(X_n)_{n=0}^{\infty}$ with the state space $E=\{1,2,3,4\}$.*

b Write the transition matrix P of the Markov chain $(X_n)_{n=0}^{\infty}$ from **a)**.

c Find the periods of states 1, 2, 3 and 4.

d Find the conditional probability of $X_{2017}=3$ given $X_{2015}=2$.

Solution:

a)

$$\begin{array}{ccccccccc} P(X_1=1|X_0=3) & + & P(X_1=2|X_0=3) & + & P(X_1=3|X_0=3) & + & P(X_1=4|X_0=3) & =1 \\ \| & & \| & & \| & & \| & \\ 0 & + & 0 & + & \alpha & + & 1/4 & =1 \end{array}$$

Hence $\alpha+1/4=1$ and then $\alpha=\frac{3}{4}$. $\square$

b)

$$P=\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}=\begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}.\ \square$$

c)

$$\text{Per}(1)=\text{Per}(2)=\text{Per}(3)=\text{Per}(4)=1.$$

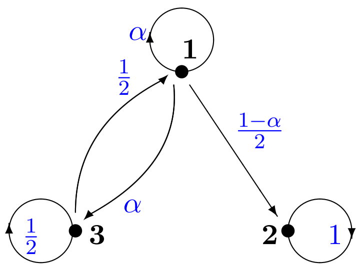
$$[\text{Per}(2)=\text{GCD}\{2,3,\dots\}=1].\ \square$$

d)

$$P(X_{2017}=3|X_{2015}=2)=P(X_2=3|X_0=2)=p_{23}^{(2)}=[p^2]_{13}=$$

$$\frac{1}{2}\cdot 0+0\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{3}{4}+0\cdot 0=\frac{1}{2}\cdot\frac{3}{4}=\frac{3}{8}.\ \square$$

Exercise 7.6 *A Markov chain $(X_n)_{n=0}^{\infty}$ with a state space $E = \{1, 2, 3\}$ is given by the following directed graph:*



a) Find the value of α.

b) Is the Markov chain ergodic? Explain

c) Calculate the 2-step transition matrix $P^{(2)}$ and find the conditional probabilities of $X_{401} = 1$ given $X_{399} = 2$ and of $X_{2017} = 3$ given $X_{2015} = 1$.

Solution:

a)

$$\alpha + \frac{1-\alpha}{2} + \alpha = 1 \;\Rightarrow\; 2\alpha + \frac{1}{2} - \frac{\alpha}{2} = 1 \;\Rightarrow\; \frac{3}{2}\alpha = \frac{1}{2} \;\Rightarrow\; \alpha = \frac{1}{3}. \;\square$$

b) There is no pass from 1 to 2. Hence $p_{21}^{(n)} = 0$ for all $n \geq 1$.

Therefore the Markov chain is not ergodic. $\square$

c)

$$P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix} \quad P^{(2)} = P = \begin{bmatrix} \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \\ 0 & 1 & 0 \\ \frac{5}{12} & \frac{1}{6} & \frac{5}{12} \end{bmatrix}.$$

$$P(X_{401} = 1 | X_{399} = 2) = P(X_2 = 1 | X_0 = 2) = p_{21}^{(2)} = 0.$$

$$P(X_{2017} = 3 | X_{2015} = 1) = P(X_2 = 3 | X_0 = 1) = p_{12}^{(2)} = \frac{5}{18}. \;\square$$

8 Appendix

8.1 $N_{0,1}$-distribution

Since $p_{N_{0,1}}(t)$ is even, then $F_{N_{0,1}}(-r) = 1 - F_{N_{0,1}}(r)$ *for every* $r \in \mathbb{R}$. So, we only include dates $F_{N_{0,1}}(r)$ for $r \geq 0$.

r \ $k \cdot \Delta r$	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.0**
0.0	**0.5000**	**0.5040**	**0.5080**	**0.5120**	**0.5160**	**0.5199**	**0.5239**	**0.5279**	**0.5319**	**0.53**
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.57
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.61
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.65
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.68
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.72
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.75
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.78
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.81
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.83
1.0	**0.8413**	**0.8438**	**0.8461**	**0.8485**	**0.8508**	**0.8531**	**0.8554**	**0.8577**	**0.8599**	**0.86**
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.88
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.90
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.91
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.93
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.94
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.95
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.96
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.97
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.97
2.0	**0.9772**	**0.9778**	**0.9783**	**0.9788**	**0.9793**	**0.9798**	**0.9803**	**0.9808**	**0.9812**	**0.98**
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.98
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.98
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.99
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.99
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.99
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.99
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.99
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.99
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.99
3.0	**0.9987**	**0.9987**	**0.9987**	**0.9988**	**0.9988**	**0.9989**	**0.9989**	**0.9989**	**0.9990**	**0.99**
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.99
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.99
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.99
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.99
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.99
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.99
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.99
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.99
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.00