Course: Theory of Probability I
Term: Fall 2013
Instructor: Gordan Zitkovic

Lecture 6
Basic Probability

Probability spaces

A mathematical setup behind a probabilistic model consists of a **sample space** Ω, a family of **events** and a **probability** $\mathbb{P}$. One thinks of Ω as being the set of all possible outcomes of a given random phenomenon, and the occurrence of a particular **elementary outcome** $\omega \in \Omega$ as depending on factors whose behavior is not fully known to the modeler. The family $\mathcal{F}$ is taken to be some collection of subsets of Ω, and for each $A \in \mathcal{F}$, the number $\mathbb{P}[A]$ is interpreted as the likelihood that some $\omega \in A$ occurs. Using the basic intuition that $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B]$, whenever A and B are disjoint (**mutually exclusive**) events, we conclude $\mathbb{P}$ should have all the properties of a finitely-additive measure. Moreover, a natural choice of normalization dictates that the likelihood of the **certain event** Ω be equal to 1. A regularity assumption[1] is often made and $\mathbb{P}$ is required to be σ-additive. All in all, we can single out probability spaces as a sub-class of measure spaces:

[1] Whether it is harmless or not leads to a very interesting philosophical discussion, but you will not get to read about it in these notes

Definition 6.1. A **probability space** is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is a non-empty set, $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

In many (but certainly not all) aspects, probability theory is a part of measure theory. For historical reasons and because of a different interpretation, some of the terminology/notation changes when one talks about measure-theoretic concepts in probability. Here is a list of what is different, and what stays the same:

1. We will always assume - often without explicit mention - that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given and fixed.
2. Continuity of measure is called **continuity of probability**, and, unlike the general case, does not require and additional assumptions in the case of a decreasing sequence (that is, of course, because $\mathbb{P}[\Omega] = 1 < \infty$.)

3. A measurable function from Ω to $\mathbb{R}$ is called a **random variable**[2]. Typically, the sample space Ω is too large for analysis, so we often focus our attention to families of real-valued functions[3] X on Ω. This way, $X^{-1}([a,b])$ is the set of all elementary outcomes $\omega \in \Omega$ with for which $X(\omega) \in [a,b]$. If we want to be able to compute the probability $\mathbb{P}[X^{-1}([a,b])]$, the set $X^{-1}([a,b])$ better be an event, i.e., $X^{-1}([a,b]) \in \mathcal{F}$. Hence the measurability requirement.

   Sometimes, it will be more convenient for random variables to take values in the extended set $\bar{\mathbb{R}}$ of real numbers. In that case we talk about **extended random variables** or **$\bar{\mathbb{R}}$-valued random variables**.

4. We use the measure-theoretic notation $\mathcal{L}^0$,$\mathcal{L}^0_+$, $\mathcal{L}^0(\bar{\mathbb{R}})$, etc. to denote the set of all random variables, non-negative random variables, extended random variables, etc.

5. Let $(S, \mathcal{S})$ be a measurable space. An $(\mathcal{F}, \mathcal{S})$-measurable map $X : \Omega \to S$ is called a **random element (of S)**.

   Random variables are random elements, but there are other important examples. If $(S, \mathcal{S}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, we talk about **random vectors**. More generally, if $S = \mathbb{R}^{\mathbb{N}}$ and $\mathcal{S} = \prod_n \mathcal{B}(\mathbb{R})$, the map $X : \Omega \to S$ is called a **discrete-time stochastic process**. Sometimes, the object of interest is a set (the area covered by a wildfire, e.g.) and then S is a collection of subsets of $\mathbb{R}^n$. There are many more examples.

6. The class of null-sets in $\mathcal{F}$ still plays the same role as it did in measure theory, but now we use the acronym **a.s.** (which stands for *almost surely*) instead of the measure-theoretic a.e.

7. The Lebesgue integral with respect to the probability $\mathbb{P}$ is now called **expectation** and is denoted by $\mathbb{E}$, so that we write

$$\mathbb{E}[X] \text{ instead of } \int X\, d\mathbb{P}, \text{ or } \int_\Omega X(\omega)\, \mathbb{P}[d\omega].$$

   For $p \in [1, \infty]$, the $\mathcal{L}^p$ spaces are defined just like before, and have the property that $\mathcal{L}^q \subseteq \mathcal{L}^p$, when $p \leq q$.

8. The notion of a.e.-convergence is now re-baptized as **a.s. convergence**, while convergence in measure is now called **convergence in probability**. We write $X_n \overset{a.s.}{\to} X$ if the sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges to a random variable X, a.s. Similarly, $X_n \overset{\mathbb{P}}{\to} X$ refers to convergence in probability. The notion of convergence in $\mathcal{L}^p$, for $p \in [1, \infty]$ is exactly the same as before. We write $X_n \overset{\mathcal{L}^p}{\to} X$ if $\{X_n\}_{n\in\mathbb{N}}$ converges to X in $\mathcal{L}^p$.

9. Since the constant random variable $X(\omega) = M$, for $\omega \in \Omega$ is integrable, a special case of the dominated convergence theorem,

[2] Random variables are usually denoted by capital letters such as X, Y Z, etc.

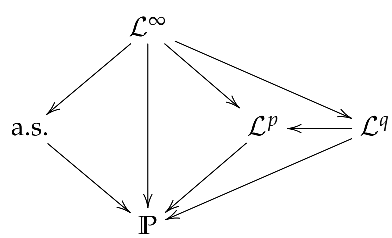
[3] If we interpret the knowledge of $\omega \in \Omega$ as the information about the true state of all parts of the model, $X(\omega)$ will typically correspond to a single numerical aspect ofi it.

known as the **bounded convergence theorem** holds in probability spaces:

Theorem 6.2 (Bounded convergence)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables such that there exists $M \geq 0$ such that $|X_n| \leq M$, a.s., and $X_n \to X$, a.s., then*

$$\mathbb{E}[X_n] \to \mathbb{E}[X].$$

10. The relationship between various forms of convergence can now be represented diagramatically as



where $1 \leq p \leq q < \infty$ and an arrow $A \to B$ means that A implies B, but that B *does not* imply A in general.

Distributions of random variables, vectors and elements

As we have already mentioned, Ω typically too big to be of direct use. Luckily, if we are only interested in a single random variable, all the useful probabilistic information about it is contained in the probabilities of the form[4] $\mathbb{P}[X \in B]$, for $B \in \mathcal{B}(\mathbb{R})$.

[4] It is standard to write $\mathbb{P}[X \in B]$ instead of the more precise $\mathbb{P}[\{X \in B\}]$ or $\mathbb{P}[\{\omega \in \Omega : X(\omega) \in B\}]$. Similarly, we will write $\mathbb{P}[X_n \in B_n$, i.o$]$ instead of $\mathbb{P}[\{X_n \in B_n\}$ i.o.$]$ and $\mathbb{P}[X_n \in B_n$, ev.$]$ instead of $\mathbb{P}[\{X_n \in B_n\}$ ev.$]$

The map $B \mapsto \mathbb{P}[X \in B]$ is, however, nothing but the push-forward of the measure $\mathbb{P}$ by the map X onto $\mathcal{B}(\mathbb{R})$:

Definition 6.3. The **distribution** of the random variable X is the probability measure μ_X on $\mathcal{B}(\mathbb{R})$, defined by

$$\mu_X(B) = \mathbb{P}[X^{-1}(B)],$$

that is the push-forward of the measure $\mathbb{P}$ by the map X.

In addition to be able to recover the information about various probabilities related to X from μ_X, one can evaluate any possible integral involving a function of X by integrating that function against μ_X (compare the statement to Problem 5.10):

Problem 6.1. Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel function. Then $g \circ X \in \mathcal{L}^{0-1}(\Omega, \mathcal{F}, \mathbb{P})$ if and only if $g \in \mathcal{L}^{0-1}(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_X)$ and, in that case,

$$\mathbb{E}[g(X)] = \int g \, d\mu_X.$$

In particular,

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \mu_X(dx).$$

Taken in isolation from everything else, two random variables X and Y for which $\mu_X = \mu_Y$ are the same from the probabilistic point of view. In that case we say that X and Y are **equally distributed** and write $X \stackrel{(d)}{=} Y$. On the other hand, if we are interested in their relationship with a third random variable Z, it can happen that X and Y have the same distribution, but that their relationship to Z is very different. It is the notion of **joint distribution** that sorts such things out. For a random vector $\boldsymbol{X} = (X_1, \dots, X_n)$, the measure $\mu_{\boldsymbol{X}}$ on $\mathcal{B}(\mathbb{R}^n)$ given by

$$\mu_{\boldsymbol{X}}(B) = \mathbb{P}[\boldsymbol{X} \in B],$$

is called the **distribution** of the random vector $\boldsymbol{X}$. Clearly, the measure $\mu_{\boldsymbol{X}}$ contains the information about the distributions of the individual components $X_1, \dots, X_n$, because

$$\begin{aligned} \mu_{X_1}(A) &= \mathbb{P}[X_1 \in A] = \mathbb{P}[X_1 \in A, X_2 \in \mathbb{R}, \dots, X_n \in \mathbb{R}] \\ &= \mu_{\boldsymbol{X}}(A \times \mathbb{R} \times \cdots \times \mathbb{R}). \end{aligned}$$

When $X_1, \dots, X_n$ are viewed as components in the random vector $\boldsymbol{X}$, their distributions are sometimes referred to as **marginal distributions**.

Example 6.4. Let $\Omega = \{1,2,3,4\}$, $\mathcal{F} = 2^\Omega$, with $\mathbb{P}$ characterized by $\mathbb{P}[\{\omega\}] = \frac{1}{4}$, for $\omega = 1, \dots, 4$. The map $X : \Omega \to \mathbb{R}$, given by $X(1) = X(3) = 0$, $X(2) = X(4) = 1$, is a random variable and its distribution is the measure $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ on $\mathcal{B}(\mathbb{R})$ (check that formally!), where δ_a denotes the Dirac measure on $\mathcal{B}(\mathbb{R})$, concentrated on $\{a\}$.

Similarly, the maps $Y : \Omega \to \mathbb{R}$ and $Z : \Omega \to \mathbb{R}$, given by $Y(1) = Y(2) = 0$, $Y(3) = Y(4) = 1$, and $Z(\omega) = 1 - X(\omega)$ are random variables with the same distribution as X. The joint distributions of the random vectors (X, Y) and (X, Z) are very different, though. The pair (X, Y) takes 4 different values $(0,0), (0,1), (1,0), (1,1)$, each with probability $\frac{1}{4}$, so that the distribution of (X, Y) is given by

$$\mu_{(X,Y)} = \tfrac{1}{4}\Big(\delta_{(0,0)} + \delta_{(0,1)} + \delta_{(1,0)} + \delta_{(1,1)}\Big).$$

On the other hand, it is impossible for X and Z to take the same value at the same time. In fact, there are only two values that the pair (X, Z) can take - $(0,1)$ and $(1,0)$. They happen with probability $\frac{1}{2}$ each, so

$$\mu_{(X,Z)} = \tfrac{1}{2}\Big(\delta_{(0,1)} + \delta_{(1,0)}\Big).$$

We will see later that the difference between (X, Y) and (X, Z) is best understood if we analyze the way the component random variables depend on each other. In the first case, even if the value of X is revealed, Y can still take the values 0 or 1 with equal probabilities. In the second case, as soon as we know X, we know Z.

More generally, if $X : \Omega \to S$, is a random element with values in the measurable space $(S, \mathcal{S})$, the **distribution of** X is the measure μ_X on $\mathcal{S}$, defined by $\mu_X(B) = \mathbb{P}[X \in B] = \mathbb{P}[X^{-1}(B)]$, for $B \in \mathcal{S}$.

Sometimes it is easier to work with a real-valued function F_X defined by

$$F_X(x) = \mathbb{P}[X \leq x],$$

which we call the **(cumulative) distribution function** (**cdf** for short), of the random variable[5] X. The following properties of F_X are easily derived by using continuity of probability from above and from below:

[5] A notion of a (cumulative) distribution function can be defined for random vectors, too, but it is not used as often as the single-component case, so we do not write about it here.

Proposition 6.5. *Let X be a random variable, and let F_X be its distribution function. Then,*

1. *F_X is non-decreasing and takes values in $[0,1]$,*
2. *F_X is right continuous,*
3. $\lim_{x\to\infty} F_X(x) = 1$ *and* $\lim_{x\to-\infty} F_X(x) = 0$.

The case when μ_X is absolutely continuous with respect to the Lebesgue measure is especially important:

Definition 6.6. A random variable X with the property that $\mu_X \ll \lambda$, where λ is the Lebesgue measure on $\mathcal{B}(\mathbb{R})$, is said to be **absolutely continuous**.

In that case, any Radon-Nikodym derivative $\frac{d\mu_X}{d\lambda}$ is called the **probability density function (pdf)** of X, and is denoted by f_X. Similarly, a random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is said to be **absolutely continuous** if $\mu_{\boldsymbol{X}} \ll \lambda$, where λ is the Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$, and the Radon-Nikodym derivative $\frac{d\mu_{\boldsymbol{X}}}{d\lambda}$, denoted by $f_{\boldsymbol{X}}$ is called the probability density function (pdf) of $\boldsymbol{X}$.

Problem 6.2.

1. Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be an absolutely-continuous random vector. Show that X_k is absolutely continuous, and that its pdf is given by

$$f_{X_k}(x) = \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{n-1 \text{ integrals}} f(\xi_1, \dots, \xi_{k-1}, x, \xi_{k+1}, \dots, \xi_n)\, d\xi_1 \dots d\xi_{k-1}\, d\xi_{k+1} \dots d\xi_n.$$

Note: As is should, $f_{X_k}(x)$ is defined only for almost all $x \in \mathbb{R}$; that is because $f_{\boldsymbol{X}}$ is defined only up to null sets in $\mathcal{B}(\mathbb{R}^n)$.

2. Let X be an absolutely-continuous random variable. Show that the random vector (X, X) is *not* absolutely continuous, even though both of its components are .

Problem 6.3. Let $\mathbf{X} = (X_1, \dots, X_n)$ be an absolutely-continuous random vector with density $f_{\mathbf{X}}$. For a Borel function $g : \mathbb{R}^n \to \mathbb{R}$ such that $g f_{\mathbf{X}} \in \mathcal{L}^{0-1}(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$, show that $g(\mathbf{X}) \in \mathcal{L}^{0-1}(\Omega, \mathcal{F}, \mathbb{P})$ and that

$$\mathbb{E}[g(\mathbf{X})] = \int g f_{\mathbf{X}} \, d\lambda = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} g(\xi_1, \dots, \xi_n) f_{\mathbf{X}}(\xi_1, \dots, \xi_n) \, d\xi_1 \dots d\xi_n.$$

Definition 6.7. A random variable X is said to be **discrete** if there exists a countable set $B \in \mathcal{B}(\mathbb{R})$ such that $\mu_X(B) = 1$.

Problem 6.4. Show that a sum of two discrete random variables is discrete, but that a sum of two absolutely-continuous random variables does not need to be absolutely continuous.

Definition 6.8. A distribution which has no atoms and is singular with respect to the Lebesgue measure is called **singular**.

Example 6.9. According to Problem 5.12, there exists a measure μ on $[0, 1]$, with the following properties

1. μ has no atoms, i.e., $\mu(\{x\}) = 0$, for all $x \in [0, 1]$,
2. μ and λ (the Lebesgue measure) are mutually singular
3. μ is supported by the Cantor set.

We set $(\Omega, \mathcal{F}, \mathbb{P}) = ([0, 1], \mathcal{B}([0, 1]), \mu)$, and define the random variable $X : \Omega \to \mathbb{R}$, by $X(\omega) = \omega$. It is clear that the distribution μ_X of X has the property that

$$\mu_X(B) = \mu(B \cap [0, 1]),$$

Thus, X is a random variable with a singular distribution.

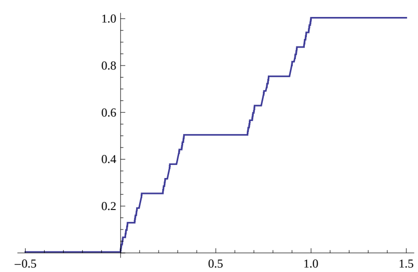


Figure 1: The CDF of the Cantor distribution.

Independence

The point at which probability departs from measure theory is when independence is introduced. As seen in Example 6.4, two random variables can "depend" on each other in different ways. One extreme (the case of X and Y) corresponds to the case when the dependence is very weak - the distribution of Y stays the same when the value of X is revealed:

Definition 6.10. Two random variables X and Y are said to be **independent** if

$$\mathbb{P}[\{X \in A\} \cap \{Y \in B\}] = \mathbb{P}[X \in A] \times \mathbb{P}[Y \in B] \text{ for all } A, B \in \mathcal{B}(\mathbb{R}).$$

It turns out that independence of random variables is a special case of the more-general notion of independence between families of sets.

Definition 6.11. Families $\mathcal{A}_1, \dots, \mathcal{A}_n$ of elements in $\mathcal{F}$ are said to be

1. **independent** if

$$\mathbb{P}[A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \times \mathbb{P}[A_{i_2}] \times \dots \times \mathbb{P}[A_{i_k}], \qquad (6.1)$$

   for all $k = 1, \dots, n$, $1 \leq i_1 < i_2 < \dots < i_k \leq n$, and all $A_{i_l} \in \mathcal{A}_{i_l}$, $l = 1, \dots, k$,

2. **pairwise independent** if

$$\mathbb{P}[A_{i_1} \cap A_{i_2}] = \mathbb{P}[A_{i_1}] \times \mathbb{P}[A_{i_2}],$$

   for all $1 \leq i_1 < i_2 \leq n$, and all $A_{i_1} \in \mathcal{A}_{i_1}$, $A_{i_2} \in \mathcal{A}_{i_2}$.

Problem 6.5.

1. Show, by means of an example, that the notion of independence would change if we asked for the product condition (6.1) to hold only for $k = n$ and $i_1 = 1, \dots, i_k = n$.
2. Show that, however, if $\Omega \in \mathcal{A}_i$, for all $i = 1, \dots, n$, it is enough to test (6.1) for $k = n$ and $i_1 = 1, \dots, i_k = n$ to conclude independence of $\mathcal{A}_i$, $i = 1, \dots, n$.

Problem 6.6. Show that random variables X and Y are independent if and only if the σ-algebras $\sigma(X)$ and $\sigma(Y)$ are independent.

Definition 6.12. Random variables $X_1, \dots, X_n$ are said to be **independent** if the σ-algebras $\sigma(X_1), \dots, \sigma(X_n)$ are independent. Events $A_1, \dots, A_n$ are called **independent** if the families $\mathcal{A}_i = \{A_i\}$, $i = 1, \dots, n$, are independent.

When only two families of sets are compared, there is no difference between pairwise independence and independence. For 3 or more, the difference is non-trivial:

Example 6.13. Let X_1, X_2, X_3 be independent random variables, each with the **coin-toss** distribution, i.e., $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$, for

$i = 1, 2, 3$. It is not hard to construct a probability space where such random variables may be defined explicitly: let $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $\mathcal{F} = 2^{\Omega}$, and let $\mathbb{P}$ be characterized by $\mathbb{P}[\{\omega\}] = \frac{1}{8}$, for all $\omega \in \Omega$. Define

$$X_i(\omega) = \begin{cases} 1, & \omega \in \Omega_i \\ -1, & \text{otherwise} \end{cases}$$

where $\Omega_1 = \{1, 3, 5, 7\}$, $\Omega_2 = \{2, 3, 6, 7\}$ and $\Omega_3 = \{5, 6, 7, 8\}$. It is easy to check that X_1, X_2 and X_3 are independent (X_i is the "i-th bit" in the binary representation of ω).

With X_1, X_2 and X_3 defined, we set

$$Y_1 = X_2 X_3,\ Y_2 = X_1 X_3 \text{ and } Y_3 = X_1 X_2,$$

so that Y_i has a coin-toss distribution, for each $i = 1, 2, 3$. Let us show that Y_1 and Y_2 (and then, by symmetry, Y_1 and Y_3, as well as Y_2 and Y_3) are independent:

$$\begin{aligned} \mathbb{P}[Y_1 = 1, Y_2 = 1] &= \mathbb{P}[X_2 = X_3, X_1 = X_3] = \mathbb{P}[X_1 = X_2 = X_3] \\ &= \mathbb{P}[X_1 = X_2 = X_3 = 1] + \mathbb{P}[X_1 = X_2 = X_3 = -1] \\ &= \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{4} = \mathbb{P}[Y_1 = 1] \times \mathbb{P}[Y_2 = 1]. \end{aligned}$$

We don't need to check the other possibilities, such as $Y_1 = 1, Y_2 = -1$, to conclude that Y_1 and Y_2 are independent (see Problem 6.7 below).

On the other hand, Y_1, Y_2 and Y_3 are not independent:

$$\begin{aligned} \mathbb{P}[Y_1 = 1, Y_2 = 1, Y_3 = 1] &= \mathbb{P}[X_2 = X_3, X_1 = X_3, X_1 = X_2] \\ &= \mathbb{P}[X_1 = X_2 = X_3] = \tfrac{1}{4} \\ &\neq \tfrac{1}{8} = \mathbb{P}[Y_1 = 1] \times \mathbb{P}[Y_2 = 1] \times \mathbb{P}[Y_3 = 1]. \end{aligned}$$

Problem 6.7. Show that if $A_1, \dots, A_n$ are independent, then so are the families $\mathcal{A}_i = \{A_i, A_i^c\}$, $i = 1, \dots, n$.

A more general statement is also true (and very useful):

Proposition 6.14. *Let $\mathcal{P}_i$, $i = 1, \dots, n$ be independent π-systems. Then, the σ-algebras $\sigma(\mathcal{P}_i)$, $i = 1, \dots, n$ are also independent.*

Proof. Let $\mathcal{F}_1$ denote the set of all $C \in \mathcal{F}$ such that

$$\mathbb{P}[C \cap A_{i_2} \cap \dots \cap A_{i_k}] = \mathbb{P}[C] \times \mathbb{P}[A_{i_2}] \times \dots \times \mathbb{P}[A_{i_k}],$$

for all $k = 2, \dots, n$, $1 < i_2 < \dots < i_k \leq n$, and all $A_{i_l} \in \mathcal{P}_{i_l}$, $l = 2, \dots, k$. It is easy to see that $\mathcal{F}_1$ is a λ-system which contains the π-system $\mathcal{P}_1$, and so, by the π-λ Theorem, it also contains $\sigma(\mathcal{P}_1)$. Consequently $\sigma(\mathcal{P}_1), \mathcal{P}_2, \dots \mathcal{P}_n$ are independent families.

A re-play of the whole procedure, but now with families $\mathcal{P}_2$, $\sigma(\mathcal{P}_1)$, $\mathcal{P}_3, \dots \mathcal{P}_n$, yields that the families $\sigma(\mathcal{P}_1), \sigma(\mathcal{P}_2), \mathcal{P}_3, \dots \mathcal{P}_n$ are also independent. Following the same pattern allows us to conclude after n steps that $\sigma(\mathcal{P}_1), \sigma(\mathcal{P}_2), \dots \sigma(\mathcal{P}_n)$ are independent. □

Remark 6.15. All notions of independence above extend to infinite families of objects (random variables, families of sets) by requiring that every finite sub-family be independent.

The result of Proposition 6.14 can be used to help us check independence of random variables:

Problem 6.8. Let $X_1, \dots, X_n$ be random variables.

1. Show that $X_1, \dots, X_n$ are independent if and only if

$$\mu_{\boldsymbol{X}} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n},$$

   where $\boldsymbol{X} = (X_1, \dots, X_n)$.

2. Show that $X_1, \dots, X_n$ are independent if and only if

$$\mathbb{P}[X_1 \le x_1, \dots, X_n \le x_n] = \mathbb{P}[X_1 \le x_1] \times \dots \times \mathbb{P}[X_n \le x_n],$$

   for all $x_1, \dots, x_n \in \mathbb{R}$.

Note: The family $\{\{X_i \le x\} : x \in \mathbb{R}\}$ does not include Ω, so that part (2) of Problem 6.5 cannot be applied directly.

3. Suppose that the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is absolutely continuous. Then $X_1, \dots, X_n$ are independent if and only if

$$f_{\boldsymbol{X}}(x_1, \dots, x_n) = f_{X_1}(x_1) \times \dots \times f_{X_n}(x_n), \ \lambda\text{-a.e.},$$

   where λ denotes the Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$.

4. Suppose that $X_1, \dots X_n$ are discrete with $\mathbb{P}[X_k \in C_k] = 1$, for countable subsets $C_1, \dots, C_n$ of $\mathbb{R}$. Show that $X_1, \dots, X_n$ are independent if and only if

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = \mathbb{P}[X_1 = x_1] \times \dots \times \mathbb{P}[X_n = x_n],$$

   for all $x_i \in C_i$, $i = 1, \dots, n$.

Problem 6.9. Let $X_1, \dots, X_n$ be independent random variables. Show that the random vector $\boldsymbol{X} = (X_1, \dots, X_n)$ is absolutely continuous if and only if each X_i, $i = 1, \dots, n$ is an absolutely-continuous random variable.

The usefulness of Proposition 6.14 is not exhausted, yet.

Problem 6.10.

1. Let $\mathcal{F}_{ij}$ $i = 1, \dots, n$, $j = 1, \dots, m_i$, be an independent collection of σ-algebras on Ω. Show that the σ-algebras $\mathcal{G}_1, \dots, \mathcal{G}_n$, where $\mathcal{G}_i = \sigma(\mathcal{F}_{i1}, \dots, \mathcal{F}_{im_i})$, are independent.
2. Let X_{ij} $i = 1, \dots, n$, $j = 1, \dots, m_i$, be an independent random variables, and let $f_i : \mathbb{R}^{m_i} \to \mathbb{R}$, $i = 1, \dots, n$, be Borel functions. Then the random variables $Y_i = f_i(X_{i\,1}, \dots, X_{i\,m_i})$, $i = 1, \dots, n$ are independent.

Hint: $\cup_{j=1,\dots,m_i} \mathcal{F}_{ij}$ generates $\mathcal{G}_i$, but is not quite a π-system.

Problem 6.11.

1. Let $X_1, \dots, X_n$ be random variables. Show that $X_1, \dots, X_n$ are independent if and only if
$$\prod_{i=1}^{n} \mathbb{E}[f_i(X_i)] = \mathbb{E}[\prod_{i=1}^{n} f_i(X_i)],$$
for all n-tuples $(f_1, \dots, f_n)$ of bounded continuous real functions.
2. Let $\{X_n^i\}_{n\in\mathbb{N}}$, $i = 1, \dots, m$ be sequences of random variables such that $X_n^1, \dots, X_n^m$ are independent for each $n \in \mathbb{N}$. If $X_n^i \overset{a.s.}{\to} X^i$, $i = 1, \dots, m$, for some $X^1, \dots, X^m \in \mathcal{L}^0$, show that $X^1, \dots, X^m$ are independent.

Hint: Approximate!

The idea "independent means multiply" applies not only to probabilities, but also to random variables:

Proposition 6.16. *Let X, Y be independent random variables, and let $h : \mathbb{R}^2 \to [0, \infty)$ be a measurable function. Then*
$$\mathbb{E}[h(X, Y)] = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x, y)\, \mu_X(dx) \right) \mu_Y(dy).$$

Proof. By independence and part 1. of Problem 6.8, the distribution of the random vector (X, Y) is given by $\mu_X \otimes \mu_Y$, where μ_X is the distribution of X and μ_Y is the distribution of μ_Y. Using Fubini's theorem, we get
$$\mathbb{E}[h(X, Y)] = \int h\, d\mu_{(X,Y)} = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x, y)\, \mu_X(dx) \right) \mu_Y(dy). \qquad \square$$

Proposition 6.17. *Let $X_1, X_2, \dots, X_n$ be independent random variables with $X_i \in \mathcal{L}^1$, for $i = 1, \dots, n$. Then*

1. $\prod_{i=1}^{n} X_i = X_1 \cdots X_n \in \mathcal{L}^1$, *and*

2. $\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \dots \mathbb{E}[X_n]$.

The product formula 2. remains true if we assume that $X_i \in \mathcal{L}^0_+$ (instead of $\mathcal{L}^1$), for $i = 1, \dots, n$.

Proof. Using the fact that X_1 and $X_2 \cdots X_n$ are independent random variables (use part 2. of Problem 6.10), we can assume without loss of generality that $n = 2$.

Focusing first on the case $X_1, X_2 \in \mathcal{L}^0_+$, we apply Proposition 6.16 with $h(x, y) = xy$ to conclude that

$$\begin{aligned}\mathbb{E}[X_1 X_2] &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x_1 x_2 \, \mu_{X_1}(dx_1) \right) \mu_{X_2}(dx_2) \\ &= \int_{\mathbb{R}} x_2 \mathbb{E}[X_1] \, \mu_{X_2}(dx_2) = \mathbb{E}[X_1]\mathbb{E}[X_2].\end{aligned}$$

For the case $X_1, X_2 \in \mathcal{L}^1$, we split $X_1 X_2 = X_1^+ X_2^+ - X_1^+ X_2^- - X_1^- X_2^+ + X_1^- X_2^-$ and apply the above conclusion to the 4 pairs $X_1^+ X_2^+$, $X_1^+ X_2^-$, $X_1^- X_2^+$ and $X_1^- X_2^-$. □

Problem 6.12 (Conditions for "independent-means-multiply")**.** Proposition 6.17 states that for independent X and Y, we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \tag{6.2}$$

whenever *both* $X, Y \in \mathcal{L}^1$ or *both* $X, Y \in \mathcal{L}^0_+$. Give an example which shows that (6.2) is no longer necessarily true in general if $X \in \mathcal{L}^0_+$ and $Y \in \mathcal{L}^1$.

Hint: Build your example so that $\mathbb{E}[(XY)^+] = \mathbb{E}[(XY)^-] = \infty$. Use $([0,1], \mathcal{B}([0,1]), \lambda)$ and take $Y(\omega) = \mathbf{1}_{[0,1/2]}(\omega) - \mathbf{1}_{(1/2,0]}(\omega)$. Then show that any random variable X with the property that $X(\omega) = X(1-\omega)$ is independent of Y.

Problem 6.13. Two random variables X, Y are said to be uncorrelated, if $X, Y \in \mathcal{L}^2$ and $\text{Cov}(X, Y) = 0$, where $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

1. Show that for $X, Y \in \mathcal{L}^2$, the expression for $\text{Cov}(X, Y)$ is well defined.
2. Show that independent random variables in $\mathcal{L}^2$ are uncorrelated.
3. Show that there exist uncorrelated random variables which are not independent.

Sums of independent random variables and convolution

Proposition 6.18. *Let X and Y be independent random variables, and let $Z = X + Y$ be their sum. Then the distribution μ_Z of Z has the following representation:*

$$\mu_Z(B) = \int_{\mathbb{R}} \mu_X(B - y) \, \mu_Y(dy), \text{ for } B \in \mathcal{B}(\mathbb{R}),$$

where $B - y = \{b - y : b \in B\}$.

Proof. We can view Z as a function $f(x,y) = x+y$ applied to the random vector (X,Y), and so, we have $\mathbb{E}[g(Z)] = \mathbb{E}[h(X,Y)]$, where $h(x,y) = g(x+y)$. In particular, for $g(z) = \mathbf{1}_B(z)$, Proposition 6.16 implies that

$$\mu_Z(B) = \mathbb{E}[g(Z)] = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\{x+y\in B\}}\, \mu_X(dx)\, \mu_Y(dy) =$$
$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{\{x\in B-y\}} \mu_X(dx) \right) \mu_Y(dy) = \int_{\mathbb{R}} \mu_X(B-y)\, \mu_Y(dy). \quad \square$$

One often sees the expression

$$\int_{\mathbb{R}} f(x)\, dF(x),$$

as notation for the integral $\int f\, d\mu$, where $F(x) = \mu((-\infty, x)]$. The reason for this is that such integrals - called the **Lebesgue-Stieltjes** integrals - have a theory parallel to that of the Riemann integral and the correspondence between $dF(x)$ and $d\mu$ is parallel to the correspondence between dx and $d\lambda$.

Corollary 6.19. *Let X, Y be independent random variables, and let Z be their sum. Then*

$$F_Z(z) = \int_{\mathbb{R}} F_X(z-y)\, dF_Y(y).$$

Definition 6.20. Let μ_1 and μ_2 be two probability measures on $\mathcal{B}(\mathbb{R})$. The **convolution** of μ_1 and μ_2 is the probability measure $\mu_1 * \mu_2$ on $\mathcal{B}(\mathbb{R})$, given by

$$(\mu_1 * \mu_2)(B) = \int_{\mathbb{R}} \mu_1(B-\xi)\, \mu_2(d\xi), \text{ for } B \in \mathcal{B}(\mathbb{R}),$$

where $B - \xi = \{x - \xi : x \in B\} \in \mathcal{B}(\mathbb{R})$.

Problem 6.14. Show that $*$ is a commutative and associative operation on the set of all probability measures on $\mathcal{B}(\mathbb{R})$.

Hint: Use Proposition 6.18

It is interesting to see how convolution relates to absolute continuity. To simplify the notation, we write $\int_A f(x)\, dx$ instead of (the more precise) $\int_A f(x)\, \lambda(dx)$ for the (Lebesgue) integral with respect to the Lebesgue measure on $\mathbb{R}$. When $A = [a,b] \in \bar{\mathbb{R}}$, we write $\int_a^b f(x)\, dx$.

Proposition 6.21. *Let X and Y be independent random variables, and suppose that X is absolutely continuous. Then their sum $Z = X+Y$ is also absolutely continuous and its density f_Z is given by*

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y)\, \mu_Y(dy).$$

Proof. Define $f(z) = \int_{\mathbb{R}} f_X(z-y)\,\mu_Y(dy)$, for some density f_X of X (remember, the density function is defined only λ-a.e.). The function f is measurable (why?) so it will be enough (why?) to show that

$$\mathbb{P}\Big[Z \in [a,b]\Big] = \int_{[a,b]} f(z)\,dz, \text{ for all } -\infty < a < b < \infty. \tag{6.2}$$

We start with the right-hand side of (6.2) and use Fubini's theorem to obtain

$$\begin{aligned}\int_{[a,b]} f(z)\,dz &= \int_{\mathbb{R}} \mathbf{1}_{[a,b]}(z) \left(\int_{\mathbb{R}} f_X(z-y)\,\mu_Y(dy) \right) dz \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{[a,b]}(z) f_X(z-y)\,dz \right) \mu_Y(dy)\end{aligned} \tag{6.3}$$

By the translation-invariance property of the Lebesgue measure, we have

$$\begin{aligned}\int_{\mathbb{R}} \mathbf{1}_{[a,b]}(z) f_X(z-y)\,dz &= \int_{\mathbb{R}} \mathbf{1}_{[a-y,b-y]}(z) f_X(z)\,dz \\ &= \mathbb{P}\Big[X \in [a-y,b-y]\Big] = \mu_X\Big([a,b]-y\Big).\end{aligned}$$

Therefore, by (6.3) and Proposition 6.18, we have

$$\begin{aligned}\int_{[a,b]} f(z)\,dz &= \int_{\mathbb{R}} \mu_X\Big([a,b]-y\Big)\,\mu_Y(dy) \\ &= \mu_Z\Big([a,b]\Big) = \mathbb{P}\Big[Z \in [a,b]\Big].\end{aligned}$$ □

Definition 6.22. The **convolution** of functions f and g in $\mathcal{L}^1(\mathbb{R})$ is the function $f * g \in \mathcal{L}^1(\mathbb{R})$ given by

$$(f * g)(z) = \int_{\mathbb{R}} f(z-x) g(x)\,dx.$$

Problem 6.15.

1. Use the reasoning from the proof of Proposition 6.21 to show that the convolution is well-defined operation on $\mathcal{L}^1(\mathbb{R})$.
2. Show that if X and Y are independent random variables and X is absolutely-continuous, then $X+Y$ is also absolutely continuous.

Do independent random variables exist?

We leave the most basic of the questions about independence for last: do independent random variable exist? We need a definition and two auxiliary results, first.

Definition 6.23. A random variable X is said to be **uniformly distributed on** (a,b), for $a < b \in \mathbb{R}$, if it is absolutely continuous with density

$$f_X(x) = \tfrac{1}{b-a} \mathbf{1}_{(a,b)}(x).$$

Our first result states a uniform random variable on $(0,1)$ can be transformed deterministically into any a random variable of prescribed distribution (cdf).

Proposition 6.24. *Let μ be a measure on $\mathcal{B}(\mathbb{R})$ with $\mu(\mathbb{R}) = 1$. Then, there exists a function $H_\mu : (0,1) \to \mathbb{R}$ such that the distribution of the random variable $X = H_\mu(U)$ is μ, whenever U is a uniform random variable on $(0,1)$.*

Note: this proposition is a basis for a technique used to simulate random variables. There are efficient algorithms for producing simulated values which resemble the uniform distribution in $(0,1)$ (so-called **pseudo-random numbers**). If a simulated value drawn with distribution μ is needed, one can simply apply the function H_μ to a pseudo-random number.

Proof. Let
F be the cdf corresponding to μ, i.e.,

$$F(x) = \mu((-\infty, x]).$$

The function F is non-decreasing, so it "almost" has an inverse: define

$$H_\mu(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

Since $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, $H_\mu(y)$ is well-defined and finite for all $y \in (0,1)$. Moreover, thanks to right-continuity and non-decrease of F, we have

$$H_\mu(y) \leq x \Leftrightarrow y \leq F(x), \text{ for all } x \in \mathbb{R}, y \in (0,1).$$

Therefore

$$\mathbb{P}[H_\mu(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x), \text{ for all } x \in \mathbb{R},$$

and the statement of the Proposition follows. □

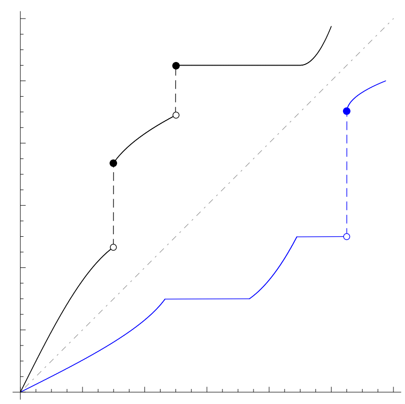


Figure 2: The right-continuous inverse H_μ (blue) of the CDF F (black)

Our next auxiliary result tells us how to construct a sequence of independent uniforms:

Proposition 6.25. *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and on it a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables such that*

1. *X_n has the uniform distribution on $(0,1)$, for each $n \in \mathbb{N}$, and*
2. *the sequence $\{X_n\}_{n\in\mathbb{N}}$ is independent.*

Proof. Set $(\Omega, \mathcal{F}, \mathbb{P}) = (\{-1,1\}^{\mathbb{N}}, \mathcal{S}, \mu_C)$ - the coin-toss space with the product σ-algebra and the coin-toss measure. Let $a : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be a bijection, i.e., $(a_{ij})_{i,j\in\mathbb{N}}$ is an arrangement of all natural numbers into a double array. For $i, j \in \mathbb{N}$, we define the map $\xi_{ij} : \Omega \to \{-1,1\}$, by

$$\xi_{ij}(s) = s_{a_{ij}},$$

i.e., ξ_{ij} is the natural projection onto the a_{ij}-th coordinate. It is straightforward to show that, under $\mathbb{P}$, the collection $(\xi_{ij})_{i,j\in\mathbb{N}}$ is independent; indeed, it is enough to check the equality

$$\mathbb{P}[\xi_{i_1 j_1} = 1, \ldots, \xi_{i_n j_n} = 1] = \mathbb{P}[\xi_{i_1 j_1} = 1] \times \cdots \times \mathbb{P}[\xi_{i_n j_n} = 1],$$

for all $n \in \mathbb{N}$ and all different $(i_1, j_1), \dots, (i_n, j_n) \in \mathbb{N} \times \mathbb{N}$.

At this point, we recycle the idea we used to construct the Lebesgue measure to construct an independent copy of a uniformly-distributed random variable from each row of $(\xi_{ij})_{i,j \in \mathbb{N}}$. We set

$$X_i = \sum_{j=1}^{\infty} \left(\tfrac{1+\xi_{ij}}{2} \right) 2^{-j}, \ i \in \mathbb{N}. \tag{6.4}$$

By second parts of Problems 6.10 and 6.11, we conclude that the sequence $\{X_i\}_{i \in \mathbb{N}}$ is independent. Moreover, thanks to (6.4), X_i is uniform on $(0, 1)$, for each $i \in \mathbb{N}$. □

Proposition 6.26. *Let $\{\mu_n\}_{n \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{B}(\mathbb{R})$. Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a sequence $\{X_n\}_{n \in \mathbb{N}}$ of random variables defined there such that*

1. *$\mu_{X_n} = \mu_n$, and*
2. *$\{X_n\}_{n \in \mathbb{N}}$ is independent.*

Proof. Start with the sequence of Proposition 6.25 and apply the function H_{μ_n} to X_n for each $n \in \mathbb{N}$, where H_{μ_n} is as in the proof of Proposition 6.24. □

An important special case covered by Proposition 6.26 is the following:

Definition 6.27. A sequence $\{X_n\}_{n \in \mathbb{N}}$ of random variables is said to be **independent and identically distributed (iid)** if $\{X_n\}_{n \in \mathbb{N}}$ is independent and all X_n have the same distribution.

Corollary 6.28. *Given a probability measure μ on $\mathbb{R}$, there exist a probability space supporting an iid sequence $\{X_n\}_{n \in \mathbb{N}}$ such that $\mu_{X_n} = \mu$.*

Additional Problems

Problem 6.16 (The standard normal distribution)**.** An absolutely continuous random variable X is said to have the **standard normal distribution** - denoted by $X \sim N(0, 1)$ - if it admits a density of the form

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

For a r.v. with such a distribution we write $X \sim N(0, 1)$.

1. Show that $\int_{\mathbb{R}} f(x)\, dx = 1$.
2. For $X \sim N(0, 1)$, show that $\mathbb{E}[|X|^n] < \infty$ for all $n \in \mathbb{N}$. Then compute the n^{th} moment $\mathbb{E}[X^n]$, for $n \in \mathbb{N}$.

Hint: Consider the double integral $\int_{\mathbb{R}^2} f(x) f(y)\, dx\, dy$ and pass to polar coordinates.

3. A random variable with the same distribution as X^2, where $X \sim N(0,1)$, is said to have the χ^2**-distribution**. Find an explicit expression for the density of the χ^2-distribution.

4. Let Y have the χ^2-distribution. Show that there exists a constant $\lambda_0 > 0$ such that $\mathbb{E}[\exp(\lambda Y)] < \infty$ for $\lambda < \lambda_0$ and $\mathbb{E}[\exp(\lambda Y)] = +\infty$ for $\lambda \geq \lambda_0$.

Note: For a random variable $Y \in \mathcal{L}^0_+$, the quantity $\mathbb{E}[\exp(\lambda Y)]$ is called the **exponential moment of order** λ.

5. Let $\alpha_0 > 0$ be a fixed, but arbitrary constant. Find an example of a random variable $X \geq 0$ with the property that $\mathbb{E}[X^\alpha] < \infty$ for $\alpha \leq \alpha_0$ and $\mathbb{E}[X^\alpha] = +\infty$ for $\alpha > \alpha_0$.

Hint: This is not the same situation as in 4. - this time the critical case α_0 is included in a different alternative. Try $X = \exp(Y)$, where $\mathbb{P}[Y \in \mathbb{N}] = 1$.

Problem 6.17 (The "memory-less" property of the exponential distribution)**.** A random variable is said to have **exponential distribution** with parameter $\lambda > 0$ - denoted by $X \sim \text{Exp}(\lambda)$ - if its distribution function F_X is given by

$$F_X(x) = 0 \text{ for } x < 0, \text{ and } F_X(x) = 1 - \exp(-\lambda x), \text{ for } x \geq 0.$$

1. Compute $\mathbb{E}[X^\alpha]$, for $\alpha \in (-1, \infty)$. Combine your result with the result of part 3. of Problem 6.16 to show that

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi},$$

where Γ is the Gamma function.

2. Remember that the conditional probability $\mathbb{P}[A|B]$ of A, given B, for $A, B \in \mathcal{F}$, $\mathbb{P}[B] > 0$ is given by

$$\mathbb{P}[A|B] = \mathbb{P}[A \cap B]/\mathbb{P}[B].$$

Compute $\mathbb{P}[X \geq x_2 | X \geq x_1]$, for $x_2 > x_1 > 0$ and compare it to $\mathbb{P}[X \geq (x_2 - x_1)]$.

conversely, suppose that y is a random variable with the property that $\mathbb{P}[Y > 0] = 1$ and $\mathbb{P}[Y > y] > 0$ for all $y > 0$. Assume further that

$$\mathbb{P}[Y \geq y_2 | Y \geq y_1] = \mathbb{P}[Y \geq y_2 - y_1], \text{ for all } y_2 > y_1 > 0. \quad (6.5)$$

Note: This can be interpreted as follows: the knowledge that the bulb stayed functional until x_1 does not change the probability that it will not explode in the next $x_2 - x_1$ units of time; bulbs have no memory.

Show that $Y \sim \text{Exp}(\lambda)$ for some $\lambda > 0$.

Hint: You can use the following fact: let $\phi : (0,\infty) \to \mathbb{R}$ be a Borel-measurable function such that $\phi(y) + \phi(z) = \phi(y+z)$ for all $y, z > 0$. Then there exists a constant $\mu \in \mathbb{R}$ such that $\phi(y) = \mu y$ for all $y > 0$.

Problem 6.18 (Some extensions of the Borel-Cantelli Lemma)**.**

1. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{L}^0_+$. Show that there exists a sequence of positive constants $\{c_n\}_{n\in\mathbb{N}}$ with the property that

$$\frac{X_n}{c_n} \to 0, \text{ a.s.}$$

Hint: Use the Borel-Cantelli lemma.

2. (The first) Borel-Cantelli lemma states that $\sum_{n\in\mathbb{N}} \mathbb{P}[A_n] < \infty$ implies $\mathbb{P}[A_n, \text{ i.o.}] = 0$. There are simple examples showing that the converse does not hold in general. Show that it *does* hold if the events $\{A_n\}_{n\in\mathbb{N}}$ are assumed to be independent. More precisely, show that, for an independent sequence $\{A_n\}_{n\in\mathbb{N}}$, we have

Hint: Use the inequality $(1-x) \leq e^{-x}$, $x \in \mathbb{R}$.

$$\sum_{n\in\mathbb{N}} \mathbb{P}[A_n] = \infty \text{ implies } \mathbb{P}[A_n, \text{ i.o.}] = 1.$$

This is often known as the **second Borel-Cantelli lemma**.

3. Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of events.
   (a) Show that $(\limsup A_n) \cap (\limsup A_n^c) \subseteq \limsup (A_n \cap A_{n+1}^c)$.
   (b) If $\liminf_{n\to\infty} \mathbb{P}[A_n] = 0$ and $\sum_n \mathbb{P}[A_n \cap A_{n+1}^c] < \infty$, show that

$$\mathbb{P}[\limsup_n A_n] = 0.$$

4. Let $\{X_n\}_{n\in\mathbb{N}}$ be an iid sequence in $\mathcal{L}^0$. Show that

$$\mathbb{E}[|X_1|] < \infty \text{ if and only if } \mathbb{P}[|X_n| \geq n, \text{ i.o.}] = 0.$$