Chapter 2
Random Variables

All the definitions with sets presented in Chapter 1 are consistent, however if we wish to calculate and compute numerical values related to abstract spaces we need to standardize the spaces. The first step is to give the following definition.

Definition 2.1 (Measurable Function (m.f.)). Let $(\Omega_1, \mathscr{F}_1)$, $(\Omega_2, \mathscr{F}_2)$ be two measurable spaces. Let $f : \Omega_1 \longrightarrow \Omega_2$ be a function. f is called a measurable function if and only if for any set $B \in \mathscr{F}_2$ we have $f^{-1}(B) \in \mathscr{F}_1$. The inverse function is a set function defined in terms of the pre-image. Explicitly, for a given set $B \in \mathscr{F}_2$,

$$f^{-1}(B) = \{\omega_1 \in \Omega_1 : f(\omega_1) \in B\}$$

Note: This definition makes it possible to extend probability measures to other spaces. For instance, let f be a measurable function and assume that there exists a probability measure P_1 on the first space $(\Omega_1, \mathscr{F}_1)$. Then we can construct a probability measure on the second space $(\Omega_2, \mathscr{F}_2)$ by $(\Omega_2, \mathscr{F}_2, P_1 \circ f^{-1})$. Note that since f is measurable $f^{-1}(B)$ is in $\mathscr{F}_1$, thus $P_1 \circ f^{-1}(B) = P_1(f^{-1}(B))$ is well defined.

Reduction to $\mathbb{R}$. Random variables

Definition 2.2. Any measurable function with codomain $(\Omega_2, \mathscr{F}_2) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is called a random variable.

Consequence: Since the Borel sets in $\mathbb{R}$ are generated by $(-\infty, x]$ then we can have the definition of a random variable directly by:

$$f : \Omega_1 \longrightarrow \mathbb{R} \text{ such that } f^{-1}(-\infty, x] \in \mathscr{F} \text{ or } \{\omega : f(\omega) \leq x\} \in \mathscr{F}, \forall x \in \mathbb{R}.$$

We shall sometimes use $f(\omega) \leq x$ to denote $f^{-1}(-\infty, x)$. Traditionally, the random variables are denoted with capital letters from the end of the alphabet $X, Y, Z, \ldots$ and their values are denoted with corresponding small letters $x, y, z, \ldots$.

Definition 2.3 (Distribution of Random Variable). Assume that on the measurable space $(\Omega, \mathscr{F})$ we define a probability measure $\mathbf{P}$ so that it becomes a probability space $(\Omega, \mathscr{F}, \mathbf{P})$. If a random variable $X : \Omega \to \mathbb{R}$ is defined then we call its distribution, the set function μ defined on the Borel sets of $\mathbb{R}$: $\mathscr{B}(\mathbb{R})$, with values in $[0,1]$:

$$\mu(B) = \mathbf{P}(\{\omega : X(\omega) \in B\}) = \mathbf{P}\left(X^{-1}(B)\right) = \mathbf{P} \circ X^{-1}(B)$$

Remark 2.1. First note that the measure μ is defined on sets in $\mathbb{R}$ and takes values in the interval $[0,1]$. Therefore, the random variable X allows us to apparently eliminate the abstract space Ω. However, this is not the case since we still have to calculate probabilities using $\mathbf{P}$ in the definition of μ above.

However, there is one simplification we can make. If we recall the result of the exercises 1.6 and1.7, we know that all Borel sets are generated by the same type of sets. Using the same idea as before it is enough to describe how to calculate μ for the generators. We could of course specify any type of generating sets we wish (open sets, closed sets, etc) but it turns out the simplest way is to use sets of the form $(-\infty, x]$, since we only need to specify one end of the interval (the other is always $-\infty$). With this observation we only need to specify the measure $\mu = P \circ X^{-1}$ directly on the generators to completely characterize the probability measure.

Definition 2.4. [The distribution function of a random variable] The distribution function of a random variable X is $F : \mathbb{R} \to [0,1]$ with:

$$F(x) = \mu(-\infty, x] = \mathbf{P}(\{\omega : X(\omega) \in (-\infty, x]\}) = \mathbf{P}(\{\omega : X(\omega) \leq x\})$$

But wait a minute, this is exactly the definition of the cumulative distribution function (cdf) which you can find in any lower level probability classes. It is exactly the same thing except that in an effort to dumb down (in whomever opinion it was to teach the class that way) the meaning is lost and we cannot proceed with more complicated things. From the definition above we can deduce all the elementary properties of the cdf that you have learned (right-continuity, increasing, taking values between 0 and 1). In fact let me ask you to prove this in exercise .

Proposition 2.1. *The distribution function for any random variable X has the following properties:*

(i) F is increasing (i.e. if $x \leq y$ then $F(x) \leq F(y)$)[1]
(ii) F is right continuous (i.e. $\lim_{h \downarrow 0} F(x+h) = F(x)$*)*
(iii) $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$

Example 2.1 (Indicator random variable). Recall the indicator function from Definition 1.10. Let $\mathbf{1}_A$ be the indicator function of a set $A \subseteq \Omega$. This is a function

[1] In other math books a function with this property is called non-decreasing. I do not like the negation and I prefer to call a function like this increasing with the distinction that a function with the following property $x < y$ implies $F(x) < F(y)$ is going to be called a **strictly increasing** function

defined on Ω with values in $\mathbb{R}$. Therefore, it may be a random variable. According to the definition it is a random variable if the function is measurable. It is simple to show that this happens if and only if $A \in \mathscr{F}$ the σ-algebra associated with the probability space. Assuming that $A \in \mathscr{F}$, what is the distribution function of this random variable?

According to the definition we have to calculate $\mathbf{P} \circ \mathbf{1}_A^{-1}((-\infty, x])$ for any x. However, the function 1_A only takes two values 0 and 1. We can calculate immediately:

$$\mathbf{1}_A^{-1}((-\infty, x]) = \begin{cases} \emptyset & , \text{ if } x < 0 \\ A^c & , \text{ if } x \in [0,1) \\ \Omega & , \text{ if } x > 1 \end{cases} .$$

Therefore,

$$F(x) = \begin{cases} 0 & , \text{ if } x < 0 \\ \mathbf{P}(A^c) & , \text{ if } x \in [0,1) \\ 1 & , \text{ if } x \geq 1 \end{cases} .$$

Proving the following lemma is elementary using the properties of the probability measure (Proposition 1.3) and is left as an exercise.

Lemma 2.1. *Let F be the distribution function of X. Then:*

(i) $\mathbf{P}(X \geq x) = 1 - F(x)$
(ii) $\mathbf{P}(x < X \leq y) = F(x) - F(y)$
(iii) $\mathbf{P}(X = x) = F(x) - F(x-)$, *where* $F(x-) = \lim_{y \nearrow x} F(y)$ *the left limit of F at x.*

Above, we define a random variable as a measurable function with codomain $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. A more specific case is obtained when the random variable has the domain also equal to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. In this case the random variable is called a Borel function.

Definition 2.5 (Borel measurable function). A function $g : \mathbb{R} \to \mathbb{R}$ is called Borel (measurable) function if g is a measurable function from $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ into $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Example 2.2. Show that any continuous function $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable.

Solution 2.1. This is very simple. Recall that the Borel sets are generated by open sets. So it is enough to see what happens to the pre-image of a open set B. But g is a continuous function therefore $g^{-1}(B)$ is an open set and thus $g^{-1}(B) \in \mathscr{B}(\mathbb{R})$. Therefore by definition g is Borel measurable.

2.1 Discrete and Continuous Random Variables

Definition 2.6 (pdf pmf and all that). Note that the distribution function F always exists. In general the distribution function F is not necessarily derivable. However, if it is, we call its derivative $f(x)$ the *probability density function* (pdf):

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

Traditionally, a variable X with this property is called *a continuous random variable*.

Furthermore if F is piecewise constant (i.e., constant almost everywhere), or in other words there exist a countable sequence $\{a_1, a_2, \dots\}$ such that the function F is constant for every point except these a_i's and we denote $p_i = F(a_i) - F(a_i-)$, then the collection of p_i's is the traditional *probability mass function* (pmf) that characterizes a *discrete random variable*[2].

Remark 2.2. Traditional undergraduate textbooks segregate between discrete and continuous random variables. Because of this segregation they are the only variables presented and it appears that all the random variables are either discrete or continuous. In reality these are the only types that can be presented without following the general approach we take here. The definitions we presented here cover any random variable. Furthermore, the treatment of random variables is the same, no more segregation.

Important. So what is the point of all this? What did we just accomplish here?

The answer is: we successfully moved from the abstract space $(\Omega, \mathscr{F}, P)$ to something perfectly equivalent but defined on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Because of this we only need to define probability measures on $\mathbb{R}$ and show that anything coming from the original abstract space is equivalent with one of these distributions on $\mathbb{R}$. We have just constructed our first model.

Example 2.3 (Indicator r.v. (continued)). This indicator variable is also called the Bernoulli random variable. Notice that the variable only takes values 0 and 1 and the probability that the variable takes the value 1 may be easily calculated using the previous definitions:

$$\mathbf{P} \circ \mathbf{1}_A^{-1}(\{1\}) = \mathbf{P}\{\omega : \mathbf{1}_A(\omega) = 1\} = \mathbf{P}(A).$$

Therefore the variable is distributed as a Bernoulli random variable with parameter $p = \mathbf{P}(A)$. Alternately, we may obtain this probability using the previously computed distribution function:

$$\mathbf{P}\{\omega : \mathbf{1}_A(\omega) = 1\} = F(1) - F(1-) = 1 - \mathbf{P}(A^c) = \mathbf{P}(A)$$

Example 2.4. Roll a six sided fair die. Say $X(\omega) = 1$ if the die shows 1 ($\omega = 1$), $X = 2$ if the die shows 2, etc. Find $F(x) = \mathbf{P}(X \le x)$.

Solution 2.2 (Solution).

If $x < 1$ then $\mathbf{P}(X \le x) = 0$

[2] Again we used the notation $F(x-)$ for the left limit of function F at x or in a more traditional notation $\lim_{z \to x, z < x} F(z)$.

If $x \in [1,2)$ then $\mathbf{P}(X \leq x) = \mathbf{P}(X = 1) = 1/6$
If $x \in [2,3)$ then $\mathbf{P}(X \leq x) = \mathbf{P}(X(\omega) \in \{1,2\}) = 2/6$

We continue this way to get:

$$\mathbf{F}(x) = \begin{cases} 0 \text{ if } x < 1 \\ i/6 \text{ if } x \in [i, i+1) \text{ with } i = 1, \cdots, 5 \\ 1 \text{ if } x \geq 6 \end{cases}$$

Exercise 2.1 (Mixture of continuous and discrete random variable). Say a game asks you to toss a coin. If the coin lands Tail you lose 1$, if Head then you draw a number from $[1,2]$ at random and gain that number. Furthermore, suppose that the coins lands a Head with probability p. Let X be the amount of money won or lost after 1 game. Find the distribution of X.

Solution 2.3 (Solution). Let $\omega = (\omega_1, \omega_2)$ where $\omega_1 \in \{\text{Head}, \text{Tail}\}$ and ω_2 in the defining experiment space for the Uniform distribution. New define a random variable $Y(\omega_2)$ on the uniform $[1,2]$ space. Then the random variable X is defined as:

$$X(\omega) = \begin{cases} -1 & \text{, if } \omega_1 = \text{Tail} \\ Y(\omega_2) & \text{if } \omega_1 = \text{Head} \end{cases}$$

If $x \in [-1,1)$ we get :

$$\mathbf{P}(X \leq x) = \mathbf{P}(X = -1) = \mathbf{P}(\omega_1 = \text{Tail}) = 1 - p$$

If $x \in [1,2)$ we get:

$$\begin{aligned} \mathbf{P}(X \leq x) &= \underbrace{\mathbf{P}(X = -1 \text{ or } X \in [1,x))}_{\text{the two events are disjoint}} = 1 - p + \mathbf{P}(\omega_1 = \text{ heads }, Y \leq x) \\ &= 1 - p + p \underbrace{\mathbf{P}(Y \in [1,x))}_{\text{Uniform}[1,2]} \\ &= 1 - p + p \int_1^x 1 dy = 1 - p + p(x-1) \\ &= 1 - 2p + px. \end{aligned}$$

Note that if the two parts of the game are not independent of each-other we cannot calculate this distribution.

Finally, we obtain:

$$\mathbf{F}(x) = \begin{cases} 0 \text{ if } x < -1 \\ 1 - p \text{ if } x \in [-1,1) \\ 1 - 2p + px \text{ if } x \in [1,2) \\ 1 \text{ if } x \geq 2 \end{cases}$$

Checking that our calculation is correct It is always a good idea to check the result. We can verify the distribution function properties, and we can plot the function to confirm this.

Examples of commonly encountered Random Variables:

Discrete random variables

For discrete random variables we give the probability mass function and it will describe completely the distribution (recall that the distribution function is piecewise linear).

(i) *Bernoulli Distribution*, the random variable only takes two values:

$$\mathbf{X} = \begin{cases} 1 & \text{with } \mathbf{P}(X=1) = p \\ 0 & \text{with } \mathbf{P}(X=0) = 1-p \end{cases}$$

We denote a random variable X with this distribution with $X \sim Bernoulli(p)$.

(ii) *Binomial(n,p) distribution*, the random variable takes values in $\mathbb{N}$ with:

$$\mathbf{P}(X=k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} \text{ for any } k \in \{0,1,2,\dots,n\} \\ 0 \text{ otherwise} \end{cases}$$

Note: X has the same distribution as $Y_1 + \cdots Y_n$ where $Y_i \sim Bernoulli(p)$
We denote a random variable X with this distribution with $X \sim Binom(n,p)$.

(iii) *Geometric (p) distribution*:

$$\mathbf{P}(X=k) = \begin{cases} (1-p)^{k-1} p \text{ for any } k \in \{1,2\cdots\} \\ 0 \text{ otherwise} \end{cases}$$

This is sometimes called Geometric “number of trials” distribution. We can also talk about Geometric “number of failures distribution” distribution, defined:

$$\mathbf{P}(Y=k-1) = \begin{cases} (1-p)^{k-1} p \text{ for any } k \in \{1,2\cdots\} \\ 0 \text{ otherwise} \end{cases}$$

Most of the time when we write $X \sim Geometric(p)$ we mean that X has a Geometric number of trials distribution. In the rare cases when we use the other one we will specify very clearly.

(iv) *Negative Binomial (r,p) distribution*

$$\mathbf{P}(X=k) = \begin{cases} \binom{k-1}{r-1} (1-p)^{r-k} p^r \text{ for any } k \in \{r, r+1, \dots\} \\ 0 \text{ otherwise} \end{cases}$$

Similarly with the $Geometric(p)$ distribution we can talk about "number of failures" distribution, but I will not give that definition.

Let us stop for a moment and see where these distributions are coming from. Suppose we do a simple experiment, we repeat an experiment many times. This experiment only has two possible outcomes "success" with probability p and "failure" with probability $1-p$.

- The variable X that takes value 1 if the experiment is a success and 0 otherwise has a $Bernoulli(p)$ distribution.
- Repeat the experiment n times in such a way that no experiment influences the outcome of any other experiment[3] and we count how many of the n repetition actually resulted in success. Let Y be the variable denoting this number. Then $Y \sim Binom(n,p)$.
- If instead of repeating the experiment a fixed number of times we repeat the experiment as many times as are needed to see the first success, then the number of trials needed is going to be distributed as a $Geometric(p)$ random variable. If we count failures until the first success we obtain the $Geometric(p)$ "number of failures" distribution.
- If we repeat the experiment until we see r successes, the number of trials needed is a $NegativeBinomial(r,p)$

(v) *Hypergeometric distribution(N,m,n,p)*,

$$\mathbf{P}(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}} \; k \in \{0,1\cdots m\}$$

This may be thought of as drawing n balls from an urn containing m white balls and $N-m$ black balls, where X represents the number of white balls in the sample.

(vi) *Poisson Distribution*, the random variable takes values in $\mathbb{N}$,

$$\mathbf{P}(X=k) = \frac{\lambda^k}{k!}e^{-k}, \quad k = 0,1,2,\ldots$$

Continuous Random Variables.

In this case every random variable has a pdf and we will specify this function directly.

(i) *Uniform Distribution[a,b]*, the random variable represents the position of a point taken at random (without any preference) within the interval $[a,b]$.

[3] this is the idea of independence which we will discuss a bit later

$$f(x) = \begin{cases} \frac{1}{b-a} & , \text{ if } x \in [a,b] \\ 0 & , \text{otherwise} \end{cases}$$

(ii) *Exponential Distribution*(θ)

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0$$

(iii) *Normal Distribution*(μ, σ)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

There are many more distributions, for our purpose the few presented will suffice.

A special random variable: Diraç Delta distribution

For a fixed a real number, consider the following distribution function:

$$F_\delta(x) = \begin{cases} 0 & \text{if } x < a \\ 1 & \text{if } x \geq a \end{cases}$$

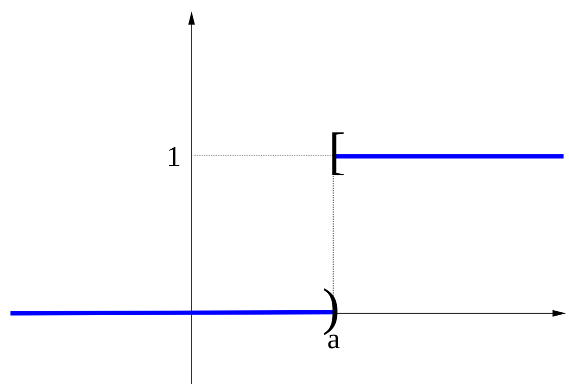


Fig. 2.1 A distribution function.

This function is plotted in Figure 2.1. Note that the function has all the properties of a distribution function (increasing, right continuous and limited by 0 and 1). However, the function is not derivable (the distribution does not have a pdf).

The random variable with this distribution is called a Diraç impulse function at a. It can only be described using measures. We will come back to this function when we develop the integration theory but for now let us say that if we define the associated set function:

$$\delta_{\{a\}}(A) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{otherwise} \end{cases}$$

this is in fact a probability measure with the property:

$$\int_{-\infty}^{\infty} f(x) d\delta_{\{a\}}(x) = f(a), \qquad \text{for all continuous functions } f$$

This will be written later as $\mathbf{E}^{\delta_{\{a\}}}[f] = f(a)$. (In other sciences: $\delta_{\{a\}}(f) = f(a)$).

Also note that $\delta_{\{a\}}(A)$ is a set function (a is fixed) and has the same value as the indicator $\mathbf{1}_A(a)$ which is a regular function (A is fixed).

2.2 Existence of random variables with prescribed distribution. Skorohod representation of a random variable

In the previous section we have seen that any random variable has a distribution function F, what is called in other classes the c.d.f. Recall the essential properties of this function from Proposition 2.1 on page 36: right-continuity, increasing, taking values between 0 and 1. An obvious question is given a function F with these properties can we construct a random variable with the desired distribution?

In fact yes we can and this is the first step in a very important theorem we shall see later in this course: the Skorohod representation theorem. However, recall that a random variable has to have as domain some probability space. It actually is true that we can construct random variables with the prescribed distribution on any space but recall that the purpose of creating random variables was to have a uniform way of treating probability. It is actually enough to give the Skorohod's construction on the probability space $([0,1], \mathscr{B}([0,1]), \lambda)$, where λ is the Lebesque measure.

On this space define the following random variables:

$$X^+(\omega) = \inf\{z \in \mathbb{R} : F(z) > \omega\}$$
$$X^-(\omega) = \inf\{z \in \mathbb{R} : F(z) \geq \omega\}$$

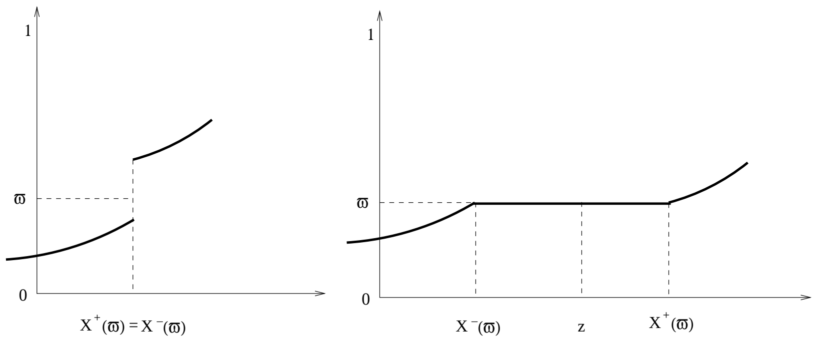
Note that in statistics X^- would be called the ω-quantile of the distribution F.

For most of the outcomes ω the two random variables are identical. Indeed, if at z with $\omega = F(z)$ the function F is non-constant then the two variables take the same values $X^+(\omega) = X^-(\omega) = z$. The two important cases when the variables take different values are depicted in Figure 2.2.

We need to show that the two variables have the desired distribution. To this end let $x \in \mathbb{R}$. Then we have:

$$\{\omega \in [0,1] : X^-(\omega) \leq x\} = [0, F(x)]$$

Indeed, if ω is in the left set then $X^-(\omega) \leq x$. By the definition of X^- then $\omega \leq F(x)$ and we have the inclusion $\subseteq$. If on the other hand $\omega \in [0, F(x)]$ then $\omega \leq F(x)$ and again by definition and right continuity of F, $X^-(\omega) \leq x$, thus we obtain $\supseteq$. Therefore, the distribution is:



(a) A point of discontinuity for F (b) An interval where the function F is constant

Fig. 2.2 Points where the two variables $X^{\pm}$ may have different outcomes

$$\lambda(\{\omega \in [0,1] : X^-(\omega) \leq x\}) = \lambda([0,F(x)]) = F(x) - 0 = F(x).$$

Finally, X^+ also has distribution function F and furthermore:

$$\lambda(X^+ \neq X^-) = 0.$$

By definition of X^+:

$$\{\omega \in [0,1] : X^-(\omega) \leq x\} \supseteq [0,F(x)),$$

and so $\lambda(X^+ \leq x) \geq F(x)$. Furthermore, since $X^- \leq X^+$ we have:

$$\{\omega \in \mathbb{R} : X^-(\omega) \neq X^+(\omega)\} = \bigcup_{x\in\mathbb{Q}} \{\omega \in \mathbb{R} : X^-(\omega) \leq x < X^+(\omega)\}$$

But for every such $x \in \mathbb{Q}$:

$$\lambda(\{\omega \in \mathbb{R} : X^-(\omega) \leq x < X^+(\omega)\}) = \lambda(\{X^- \leq x\} \setminus \{X^+ \leq x\}) \leq F(x) - F(x) = 0$$

Since $\mathbb{Q}$ is countable and any countable union of null sets is a null set the result follows.

2.3 Independence

In this section we extend the idea of independence originally defined for events to random variables. In order to do this we have to explain the joint distribution of several variables.

Example 2.5 (The idea of joint distribution). Suppose 2 points ξ_1, ξ_2 are tossed at random and independently onto a line segment of length L (ξ_1, ξ_2 are i.i.d.). What is the probability that the distance between the 2 points does not exceed 1?

Solution 2.4 (Solution). If $L \le 1$ then the probability is trivially equal to 1.

Assume that $L > 1$ (the following also works if 1 is substituted by a $l \le L$). What is the distribution of ξ_1 and ξ_2? They are both $Unif[0,L]$. We want to calculate $\mathbf{P}(|\xi_1 - \xi_2| \le 1)$.

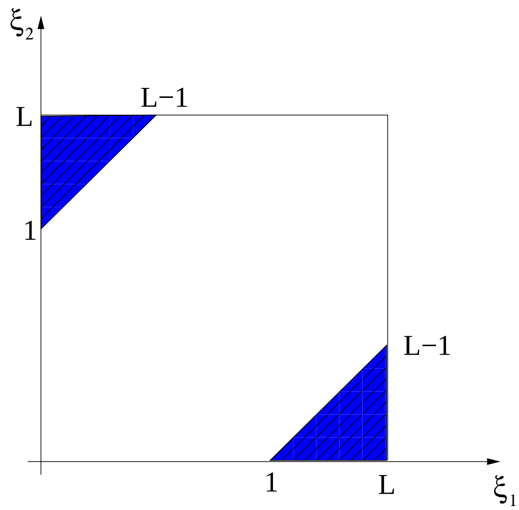


Fig. 2.3 The area we need to calculate. The blue parts need to be deleted.

We plot the surface we need to calculate in Figure 2.3. The area within the rectangle and not shaded is exactly the area we need. If we pick any point from within this area it will have the property that $|\xi_1 - \xi_2| \le 1$. Since the points are chosen uniformly from within the rectangle the chance of a point being chosen is the ratio between the “good” area and the total area.

The unshaded area from within the rectangle is: $L^2 - \frac{(L-1)^2}{2} - \frac{(L-1)^2}{2} = 2L - 1$. Therefore, the desired probability is:

$$\mathbf{P}(|\xi_1 - \xi_2| \le 1) = \frac{2L-1}{L^2}.$$

□

This geometrical proof works because the distribution is uniform and furthermore the points are chosen independently of each other. However if the distribution is anything else we need to go through the whole calculation. We shall see how to do this after we define joint probability. We need this to define the independence concept.

2.3.1 Joint distribution

We talked about σ-algebras in Chapter 1. Let us come back to them. If there is any hope of rigorous introduction into probability and stochastic processes, they are *unavoidable*. Later, when we will talk about stochastic processes we will find out the *crucial* role they play in quantifying the information available up to a certain time. For now let us play a bit with them.

Definition 2.7 (σ-algebra generated by a random variable). For a r.v. X we define the σ-algebra generated by X, denoted $\sigma(X)$ or sometime $\mathscr{F}_X$, the smallest σ-field $\mathscr{G}$ such that X is measurable on $(\Omega, \mathscr{G})$. It is the σ-algebra generated by the pre-images of Borel sets through X (recall that we have already presented this concept earlier in definition 1.3 on page 9). Because of this we can easily show[4]:

$$\sigma(X) = \sigma(\{\omega | X(\omega) \leq x\}, \text{ as } x \text{ varies in } \mathbb{R}).$$

Similarly, given $X_1, X_2, \dots, X_n$ random variables, we define the sigma algebra generated by them as the smallest sigma algebra such that all are measurable with respect to it. It turns out we can show easily that it is the sigma algebra generated by the union of the individual sigma algebras or put more specifically $\sigma(X_i, i \leq n)$ is the smallest sigma algebra containing all $\sigma(X_i)$, for $i = 1, 2, \dots, n$, or $\sigma(X_1) \vee \sigma(X_2) \vee \cdots \vee \sigma(X_n)$, again recall proposition 1.2 on page 10.

In Chapter 1 we defined Borel sigma algebras corresponding to any space Ω. We consider the special case when $\Omega = \mathbb{R}^n$. This allows us to define a random vector on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \mathbf{P})$ as $(X_1, X_2, \dots, X_n)$ where each X_i is a random variable. The probability $\mathbf{P}$ is defined on $\mathscr{B}(\mathbb{R}^n)$.

We can talk about its distribution (the *"joint distribution"* of the variables $(X_1, X_2, \dots, X_n)$) as the function:

$$\begin{aligned} F(x_1, x_2, \dots, x_n) &= \mathbf{P} \circ (X_1, X_2, \dots, X_n)^{-1} ((-\infty, x_1] \times \cdots \times (-\infty, x_n]) \\ &= \mathbf{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \end{aligned}$$

which is well defined for any $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$

In the special case when F can be written as:

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_X(t_1, \cdots, t_n) dt_1 \cdots dt_n,$$

we say that the vector X has a *joint density* and f_X is the joint probability density function of the random vector X.

[4] Remember that the Borel sets are generated by intervals of the type $(-\infty, x]$

Definition 2.8 (Marginal Distribution). Given the joint distribution of a random vector $X = (X_1, X_2, \dots, X_n)$ we define the marginal distribution of X_1:

$$F_{X_1}(x_1) = \lim_{\substack{x_2 \to \infty \\ \dots \\ x_n \to \infty}} F_X(x_1 \cdots x_n)$$

and similarly for all the other variables.[5]

2.3.2 Independence of random variables

We can now introduce the notions of independence and joint independence using the definition in Section 1.3, the probability measure $= \mathbf{P} \circ (X_1, X_2, \dots, X_n)^{-1}$ and any Borel sets. Writing more specifically that definition is transformed here:

Definition 2.9. The variables $(X_1, X_2, \dots, X_n, \dots)$ are independent if for every subset $J = \{j_1, j_2, \dots, j_k\}$ of $\{1, 2, 3, \dots\}$ we have:

$$\mathbf{P}\left(X_{j_1} \le x_{j_1}, X_{j_2} \le x_{j_2}, \dots, X_{j_k} \le x_{j_k}\right) = \prod_{j \in J} \mathbf{P}(X_j \le x_j)$$

Remark 2.3. The formula in the Definition 2.8 allows to obtain the marginal distributions from the joint distribution. The converse is generally false meaning that if we know the marginal distributions we cannot regain the joint.

However, there is one case when this is possible: when X_i are independent. In this case $F_X(x) = \prod_{i=1}^n F_{X_i}(x_i)$. That is why the i.i.d case is the most important in probability (we can regain the joint from the marginals without any other special knowledge).

Independence (specialized cases)

(i) If X and Y are discrete r.v.'s with joint probability mass function $p_{X,Y}(\cdot, \cdot)$ then they are independent if and only if

$$p_{X,Y}(x,y) = p_X(x) p_Y(y), \quad \forall x, y$$

(ii) If X and Y are continuous r.v.'s with joint probability density function f then they are independent if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \quad \forall x, y$$

where we used obvious notations for marginal distributions. The above definition can be extended to n dimensional vectors in an obvious way.

---

[5] We can also define it simpler as $\int_{-\infty}^{x_1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, \cdots, t_n) dt_1 \cdots dt_n$ if the joint pdf exists.

I.I.D. r.v.'s: (Independent Identically Distributed Random Variables). Many of the central ideas in probability involve sequences of random variables which are independent and identically distributed. That is a sequence of random variables $\{X_n\}$ such that X_n are independent and all have the same distribution function say $F(x)$.

Finally, we answer the question we asked in the earlier example: What to do if the variables ξ_1, ξ_2 are not uniformly distributed?

Suppose that ξ_1 had distribution F_{ξ_1} and ξ_2 had distribution F_{ξ_2}. Assuming that the two variables are independent we obtain the joint distribution:

$$F_{\xi_1\xi_2}(x_1,x_2) = F_{\xi_1}(x_1)F_{\xi_2}(x_2)$$

(If they are not independent we have to be given or infer the joint distribution).

The probability we are looking for is the area of the surface

$$\{(\xi_1,\xi_2)|\xi_1 \in [0,L], \xi_2 \in [0,L], \xi_1 - 1 \leq \xi_2 \leq \xi_1 + 1\}.$$

We shall find out how to calculate this probability using general distribution functions F_{ξ_1} and F_{ξ_2} in the next chapter. For now let us assume that the two variables have densities f_1 and f_2. Then, the desired probability is:

$$\int_0^L \int_0^L \mathbf{1}_{\{x_1-1\leq x_2\leq x_1+1\}}(x_1,x_2)\, f_{\xi_1}(x_1)\, f_{\xi_2}(x_2) dx_1 dx_2$$

which can be further calculated:

- When $L-1 < 1$ or $1 < L < 2$:

$$\int_0^{L-1}\int_0^{x_1+1} f_{\xi_1}(x_1)\, f_{\xi_2}(x_2) dx_2 dx_1 + (2-L)L + \int_1^L \int_{x_1-1}^L f_{\xi_1}(x_1)\, f_{\xi_2}(x_2) dx_2 dx_1$$

- When $L-1 > 1$ or $L > 2$:

$$\int_0^1 \int_0^{x_1+1} f_{\xi_1}(x_1)\, f_{\xi_2}(x_2) dx_2 dx_1 + \int_1^{L-1}\int_{x_1-1}^{x_1+1} f_{\xi_1}(x_1)\, f_{\xi_2}(x_2) dx_2 dx_1 + \int_{L-1}^L \int_{x_1-1}^L f_{\xi_1}(x_1)\, f_{\xi_2}(x_2) dx_2 dx_1$$

Above is given to remind about the calculation of a two dimensional integral.

2.4 Functions of random variables. Calculating distributions

Measurable functions allow us to construct new random variables. These new random variables possess their own distribution. This section is dedicated to calculating this new distribution. At this time it is not possible to work with abstract spaces (for that we will give a general theorem - the Transport formula in the next chapter) so all our calculations will be done in $\mathbb{R}^n$.

One dimensional functions

Let X be a random variable defined on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$. Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be a Borel measurable function. Let $Y = g(X)$ which is a new random variable. Its distribution is deduced as:

$$\begin{aligned}\mathbf{P}(Y \le y) &= \mathbf{P}(g(X) \le y) = \mathbf{P}(g(X) \in (-\infty, y]) = \mathbf{P}\left(X \in g^{-1}((-\infty, y])\right)\\ &= \mathbf{P}\left(\{\omega : X(\omega) \in g^{-1}((-\infty, y])\}\right)\end{aligned}$$

where $g^{-1}((-\infty, y])$ is the preimage of $(-\infty, y]$ through the function g, i.e.,:

$$\{x \in \mathbb{R} : g(x) \le y\}.$$

If the random variable X has p.d.f f then the probability has a simpler formula:

$$\mathbf{P}(Y \le y) = \int_{g^{-1}(-\infty, y]} f(x)dx$$

Example 2.6. Let X be a random variable distributed as a Normal (Gaussian) with mean zero and variance 1, $X \sim N(0,1)$. Let $g(x) = x^2$, and take $Y = g(X) = X^2$. Then:

$$\mathbf{P}(Y \le y) = \mathbf{P}(X^2 \le y) = \begin{cases} 0 & \text{if } y < 0 \\ \mathbf{P}(-\sqrt{y} \le X \le \sqrt{y}) & \text{if } y \ge 0 \end{cases}$$

Note that the preimage of $(-\infty, y]$ through the function $g(x) = x^2$ is either $\emptyset$ if $y < 0$ or $[-\sqrt{y}, \sqrt{y}]$ if $y \ge 0$. This is how we obtain above. In the nontrivial case $y \ge 0$ we get:

$$\mathbf{P}(Y \le y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \Phi(\sqrt{y}) - [1 - \Phi(\sqrt{y})] = 2\Phi(\sqrt{y}) - 1,$$

where Φ is the c.d.f of X, a $N(0,1)$ random variable. In this case $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$. Since the function Φ is derivable Y has a p.d.f. which can be obtained:

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy}[2\Phi(\sqrt{y})] = 2\Phi'(\sqrt{y})\frac{1}{2\sqrt{y}} \\
&= \frac{1}{\sqrt{y}}\Phi'(\sqrt{y}) = \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-y/2} \\
&= \frac{1}{\sqrt{2\pi y}}e^{-y/2}
\end{aligned}
$$

□

We note that a random variable Y with the p.d.f. described above is said to have a chi-squared distribution with one degree of freedom (the notation is χ_1^2).

Two and more dimensional functions

If the variable X does not have a p.m.f or a p.d.f there is not much we can do. The same relationship holds as in the 1 dimensional case. Specifically, if X is a n-dim random vector and $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a measurable function which defines a new random vector $Y = g(X)$ then its distribution is determined using:

$$\mathbf{P}(Y \leq y) = \mathbf{P}(g(X) \leq y) = \mathbf{P}\left(\{\omega : X(\omega) \in g^{-1}((-\infty, y])\}\right)$$

and this is the same relationship as before.

In the case when the vector X has a density then things become more specific. We will exemplify using R^2 but the same calculation works in n dimensions with no modification (other than the dimension of course). Suppose that a two dimensional random vector (X_1, X_2) has joint density f. Let $g : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be a measurable function:

$$g(x_1, x_2) = (g_1(x_1, x_2), g_2(x_1, x_2))$$

Suppose first that the function g is one-to-one[6]

Define a random vector $Y = (Y_1, Y_2) = g(X_1, X_2)$. First we find the support set of Y (i.e. the points where Y has nonzero probability). To this end let

$$
\begin{aligned}
&\mathcal{A} = \{(x_1, x_2) : f(x_1, x_2) > 0\} \\
&\mathcal{B} = \{(y_1, y_2) : y_1 = g_1(x_1, x_2) \text{ and } y_2 = g_2(x_1, x_2), \text{ for some } (x_1, x_2) \in \mathcal{A}\}
\end{aligned}
$$

This $\mathcal{B}$ is the image of $\mathcal{A}$ through g, it is also the support set of Y. Since g is one-to-one, when restricted to $g : \mathcal{A} \to \mathcal{B}$ it is also surjective, therefore forms a bijection between $\mathcal{A}$ and $\mathcal{B}$. Thus, the inverse function $g^{-1}(y_1, y_2) = (g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2))$ is a unique, well defined function.

---

[6] this is why we use the same dimension n for both X and Y vectors

To calculate the density of Y we need the derivative of this g^{-1} and that role is played by the Jacobian of the transformation (the determinant of the matrix of partial derivatives):

$$J = J_{g^{-1}}(y_1,y_2) = \begin{vmatrix} \frac{\partial g_1^{-1}}{\partial y_1}(y_1,y_2) & \frac{\partial g_2^{-1}}{\partial y_1}(y_1,y_2) \\ \frac{\partial g_1^{-1}}{\partial y_2}(y_1,y_2) & \frac{\partial g_2^{-1}}{\partial y_2}(y_1,y_2) \end{vmatrix}$$

Then, the joint p.d.f. of the vector Y is given by:

$$f_Y(y_1,y_2) = f\left(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2)\right)\ |J|\ \mathbf{1}_{\mathcal{B}}(y_1,y_2)$$

where we used the indicator notation and $|J|$ is the absolute value of the Jacobian.

Suppose that the function g is not one-to-one

In this case we recover the previous one-to-one case by restricting the function. Specifically, define the sets $\mathcal{A}$ and $\mathcal{B}$ as before. Now, the restricted function $g : \mathcal{A} \to \mathcal{B}$ is surjective. We partition $\mathcal{A}$ into $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_k$. The set $\mathcal{A}_0$ may contain several points which are difficult to deal with, the only condition is that $\mathbf{P}((X_1,X_2) \in \mathcal{A}_0) = 0$ (it is a null set). Furthermore, for all $i \neq 0$, each restriction $g : \mathcal{A}_i \to \mathcal{B}$ is one-to one. Thus, for each such $i \geq 1$, an inverse can be found $g_i^{-1}(y_1,y_2) = (g_{i1}^{-1}(y_1,y_2), g_{i2}^{-1}(y_1,y_2))$. This i-th inverse gives for any $(y_1,y_2) \in \mathcal{B}$ a unique $(x_1,x_2) \in \mathcal{A}_i$ such that $(y_1,y_2) = g(x_1,x_2)$. Let J_i be the Jacobian associated with the i-th inverse transformation. Then the joint p.d.f. of Y is:

$$f_Y(y_1,y_2) = \sum_{i=1}^{k} f\left(g_{i1}^{-1}(y_1,y_2), g_{i2}^{-1}(y_1,y_2)\right)\ |J_i|\ \mathbf{1}_{\mathcal{B}}(y_1,y_2)$$

Example 2.7. Let (X_1,X_2) have some joint p.d.f. $f(\cdot,\cdot)$. Calculate the density of X_1X_2.

Let us take $Y_1 = X_1X_2$ and $Y_2 = X_1$ i.e. $g(x_1,x_2) = (x_1x_2,\ x_1) = (y_1,y_2)$. The function thus constructed $g : \mathbb{R}^2 \to \mathbb{R}^2$ is bijective so $\mathcal{B} = \mathbb{R}^2$. To calculate its inverse:

$$\begin{aligned} x_1 &= y_2 \\ x_2 &= \frac{y_1}{x_1} = \frac{y_1}{y_2}, \end{aligned}$$

which gives:

$$g^{-1}(y_1,y_2) = \left(y_2, \frac{y_1}{y_2}\right)$$

We then get the Jacobian:

$$J_{g^{-1}}(y_1,y_2) = \begin{vmatrix} 0 & \frac{1}{y_2} \\ 1 & -\frac{y_1}{y_2^2} \end{vmatrix} = 0 - \frac{1}{y_2} = -\frac{1}{y_2}$$

Thus, the joint p.d.f of $Y = (Y_1, Y_2)$ is:

$$f_Y(y_1,y_2) = f\left(y_2, \frac{y_1}{y_2}\right)\left|\frac{1}{y_2}\right|,$$

where f is the given p.d.f. of X. To obtain the distribution of $X_1X_2 = Y_1$ we simply need the marginal p.d.f. obtained immediately by integrating out Y_2:

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f\left(y_2, \frac{1}{y_2}\right) \cdot \frac{1}{|y_2|} dy_2$$

□

Example 2.8 (A more specific example). Let X_1, X_2 be independent Exp(λ). Find the joint density of $Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{X_2}$. Also show that the variables Y_1 and Y_2 are independent.

Let $g(x_1,x_2) = \left(x_1 + x_2, \frac{x_1}{x_2}\right) = (y_1,y_2)$. Let us calculate the domain of the transformation.

Remember that the p.d.f of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x}\mathbf{1}_{(0,\infty)}(x),$$

thus $\mathcal{A} = (0,\infty) \times (0,\infty)$. Since $x_1, x_2 > 0$ we get that $x_1 + x_2 > 0$ and $\frac{x_1}{x_2} > 0$, and so $\mathcal{B} = (0,\infty)^2$ as well. The function g restricted to this sets is bijective as we can easily show by solving the equations: $y_1 = x_1 + x_2$ and $y_2 = \frac{x_1}{x_2}$. We obtain:

$$\begin{aligned} x_1 = x_2 y_2 &\Rightarrow y_1 = x_2 y_2 + x_2 \\ &\Rightarrow x_2 = \frac{y_1}{1+y_2} \\ &\Rightarrow x_1 = \frac{y_1 y_2}{1+y_2} \end{aligned}$$

Since the solution is unique the function g is one-to-one. Since the solution exists for all $(y_1,y_2) \in (0.\infty)^2$ the function is surjective. Its inverse is precisely:

$$g^{-1}(y_1,y_2) = \left(\frac{y_1 y_2}{1+y_2}, \frac{y_1}{1+y_2}\right)$$

Furthermore, the Jacobian is:

$$J_{g^{-1}}(y_1,y_2) = \begin{vmatrix} \frac{y_2}{1+y_2} & \frac{1}{1+y_2} \\ \frac{y_1}{(1+y_2)^2} & -\frac{y_1}{(1+y_2)^2} \end{vmatrix} = -\frac{y_1 y_2}{(1+y_2)^3} - \frac{y_1}{(1+y_2)^3} = -\frac{y_1}{(1+y_2)^2}$$

Thus the desired p.d.f is:

$$\begin{aligned}
f_Y(y_1,y_2) &= f\left(\frac{y_1y_2}{1+y_2},\frac{y_1}{1+y_2}\right)\left|-\frac{y_1}{(1+y_2)^2}\right|\mathbf{1}_{(y_1,y_2)\in(0,\infty)^2}\\
&= \lambda e^{-\lambda\frac{y_1y_2}{1+y_2}}\lambda e^{-\frac{y_1}{1+y_2}}\frac{y_1}{(1+y_2)^2}\mathbf{1}_{\{y_1,y_2>0\}}\\
&= \lambda^2 e^{-\lambda y_1}\frac{y_1}{(1+y_2)^2}\mathbf{1}_{\{y_1,y_2>0\}}
\end{aligned}$$

Finally, to end the example it is enough to recognize that the p.d.f. of Y can be decomposed into a product of two functions, one of them only in the variable y_1 and the other only a function of the variable y_2. Thus, if we apply the next lemma the example is solved. □

Lemma 2.2. *If the joint distribution f of a random vector (X,Y) factors as a product of functions of only x and y, i.e., there exist $g,h:\mathbb{R}\to\mathbb{R}$ such that $f(x,y)=g(x)h(y)$ then the variables X,Y are independent.*

Proof. Problem 2.12.

Example 2.9. Let X, Y be two random variables with joint p.d.f. $f(\cdot,\cdot)$. Calculate the density of $X+Y$.

Let $(U,V)=(X+Y,Y)$. We can easily calculate the domain and the inverse $g^{-1}(u,v)=(u-v,v)$. The Jacobian is:

$$J_{g^{-1}}(u,v)=\begin{vmatrix}1 & -1\\ 0 & 1\end{vmatrix}=1$$

As a result the desired p.d.f. is:

$$f_U(u)=\int_{-\infty}^{\infty}f(u-v,v)dv$$

We will observe this particular example later when we talk about convolutions.

Example 2.10. Let X_1 and X_2 be i.i.d. $N(0,1)$ random variables. Consider the function $g(x_1,x_2)=\left(\frac{x_1}{x_2},|x_2|\right)$. Calculate the joint distribution of $Y=g(X)$ and the distribution of the ratio of the two normals: X_1/X_2.

First, $\mathcal{A}=R^2$ and $\mathcal{B}=\mathbb{R}\times(0,\infty)$. Second, note that the transformation is not one-to-one. Also note that we have a problem when $x_2=0$[7]. Fortunately, we know

[7] 0 is in $\mathcal{A}$ since $f_{X_2}(0)>0$

how to deal with this situation. Take a partition of $\mathcal{A}$ as follows:

$$\mathcal{A}_0 = \{(x_1,0) : x_1 \in \mathbb{R}\},\ \mathcal{A}_1 = \{(x_1,x_2) : x_2 < 0\},\ \mathcal{A}_1 = \{(x_1,x_2) : x_2 > 0\}.$$

$\mathcal{A}_0$ has the desired property since $\mathbf{P}((X_1,X_2) \in \mathcal{A}_0) = \mathbf{P}(X_2 = 0) = 0$ (X_2 is a continuous random variable). Restricted to each A_i the function g is bijective and we can calculate its inverse in both cases:

$$g_1^{-1}(y_1,y_2) = (-y_1y_2,-y_2)$$
$$g_2^{-1}(y_1,y_2) = (y_1y_2,y_2)$$

In either case the Jacobian is identical $J_1 = J_2 = y_2$. Using the p.d.f. of a normal with mean zero and variance 1 ($f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$), and that X_1 and X_2 being independent the joint p.d.f. is the product of marginals we obtain:

$$\begin{aligned} f_Y(y_1,y_2) &= \left(\frac{1}{2\pi}e^{-(-y_1y_2)^2/2}e^{-(-y_2)^2/2}|y_2| + \frac{1}{2\pi}e^{-(y_1y_2)^2/2}e^{-(y_2)^2/2}|y_2|\right)\mathbf{1}_{\{y_2>0\}} \\ &= \frac{y_2}{\pi}e^{-\frac{(y_1^2+1)y_2^2}{2}}\mathbf{1}_{\{y_2>0\}}, \quad y_1 \in \mathbb{R}, \end{aligned}$$

and this is the desired joint distribution. To calculate the distribution of X_1/X_2 we calculate the marginal of Y_1 by integrating out y_2:

$$\begin{aligned} f_{Y_1}(y_1) &= \int_0^\infty \frac{y_2}{\pi}e^{-\frac{(y_1^2+1)y_2^2}{2}}dy_2 \quad (\text{ Change of variables } y_2^2 = t) \\ &= \int_0^\infty \frac{1}{2\pi}e^{-\frac{(y_1^2+1)}{2}t}dt = \frac{1}{2\pi}\frac{2}{y_1^2+1} \\ &= \frac{1}{\pi(y_1^2+1)}, \quad y_1 \in \mathbb{R} \end{aligned}$$

But this is the distribution of a Cauchy random variable. Thus we have just proven that the ratio of two independent $N(0,1)$ rv's has a Cauchy distribution. □

We conclude this chapter with a non-trivial application of the Borel-Cantelli lemmas. We have postponed this example until this point since we needed to learn about independent random variables first.

Example 2.11. Let $\{X_n\}$ a sequence of i.i.d. random variables, each exponentially distributed with rate 1, i.e.:

$$\mathbf{P}(X_n > x) = e^{-x}, \quad x > 0.$$

We wish to study how large are these variables when $n \to \infty$. To this end take $x = \alpha \log n$, for some $\alpha > 0$ and for any $n \geq 1$. Substitute into the probability above to obtain:

$$\mathbf{P}(X_n > \alpha \log n) = e^{-\alpha \log n} = n^{-\alpha} = \frac{1}{n^\alpha}.$$

But we know that the sum $\sum_n \frac{1}{n^\alpha}$ is divergent for the exponent $\alpha \leq 1$ and convergent for $\alpha > 1$. So we can apply the Borel-Cantelli lemmas since the events in question are independent. Thus,

If $\alpha \leq 1$ the sum is divergent and so $\sum_n \mathbf{P}(X_n > \alpha \log n) = \infty$, thus:

$$\mathbf{P}\left(\frac{X_n}{\log n} > \alpha \text{ i.o.}\right) = 1$$

If $\alpha > 1$ the sum is convergent, and $\sum_n \mathbf{P}(X_n > \alpha \log n) < \infty$, thus:

$$\mathbf{P}\left(\frac{X_n}{\log n} > \alpha \text{ i.o.}\right) = 0$$

We can express the same thing in terms of lim sup like so:

$$\mathbf{P}\left(\limsup_n \frac{X_n}{\log n} > \alpha\right) = \begin{cases} 0 & \text{, if } \alpha > 1 \\ 1 & \text{, if } \alpha \leq 1 \end{cases}$$

Since for all $\alpha \leq 1$ we have that $\mathbf{P}\left(\limsup_n \frac{X_n}{\log n} > \alpha\right) = 1$, then we necessarily have:

$$\mathbf{P}\left(\limsup_n \frac{X_n}{\log n} \geq 1\right) = 1$$

Take $\alpha = 1 + \frac{1}{k}$ and look at the other implication: $\mathbf{P}\left(\limsup_n \frac{X_n}{\log n} > 1 + \frac{1}{k}\right) = 0$, and this happens for all $k \in \mathbb{N}$, . But we can write:

$$\left\{\limsup_n \frac{X_n}{\log n} > 1\right\} = \bigcup_{k \in \mathbb{N}} \left\{\limsup_n \frac{X_n}{\log n} > 1 + \frac{1}{k}\right\},$$

and since any countable union of null sets is itself a null set, the probability of the event on the left must be zero. Therefore, $\limsup_n \frac{X_n}{\log n} \leq 1$ a.s. and combining with the finding above:

$$\limsup_n \frac{X_n}{\log n} = 1, \quad a.s.$$

This is very interesting since as we will see in the chapter dedicated to the Poisson process, these X_n are the inter-arrival times of this process. The example above tells us that if we look at the realizations of such a process then they form a sequence of numbers that has the upper limiting point equal to 1, or put differently there is no subsequence of inter-arrival times that in the limit is greater than the $\log n$.

Problems

2.1. Prove the Proposition 2.1. That is prove that the function F in Definition 2.4 is increasing, right continuous and taking values in the interval $[0,1]$, using only proposition 1.3 on page 13.

2.2. Show that any piecewise constant function is Borel measurable. (see description of piecewise constant functions in Definition 2.6

2.3. Give an example of two distinct random variables with the same distribution function.

2.4. Buffon's needle problem.
Suppose that a needle is tossed at random onto a plane ruled with parallel lines a distance L apart, where by a "needle" we mean a line segment of length $l \leq L$. What is the probability of the needle intersecting one of the parallel lines?

Hint: Consider the angle that is made by the needle with the parallel lines as a random variable α uniformly distributed in the interval $[0,2\pi]$ and the position of the midpoint of the needle as another random variable ξ also uniform on the interval $[0,L]$. Then express the condition "needle intersects the parallel lines" in terms of the position of the midpoint of the needle and the angle α. Do a calculation similar with example 2.5.

2.5. A random variable X has distribution function

$$F(x) = a + b \arctan \frac{x}{2} \quad , -\infty < x < \infty$$

Find:
a) The constants a and b
b) The probability density function of X

2.6. What is the probability that two randomly chosen numbers between 0 and 1 will have a sum no greater than 1 and a product no greater than $\frac{15}{64}$?

2.7. We know that the random variables X and Y have joint density $f(x,y)$. Assume that $\mathbf{P}(Y=0)=0$. Find the densities of the following variables:
a) $X+Y$
b) $X-Y$
c) XY
d) $\frac{X}{Y}$

2.8. Choose a point A at random in the interval $[0,1]$. Let L_1 (respectively L_2) be the length of the bigger (respectively smaller) segment determined by A on $[0,1]$. Calculate:
a) $\mathbf{P}(L_1 \leq x)$ for $x \in \mathbb{R}$.
b) $\mathbf{P}(L_2 \leq x)$ for $x \in \mathbb{R}$.

2.9. Two friends decide to meet at the Castle gate of Stevens Institute. They each arrive at that spot at some random time between a and $a+T$. They each wait for 15 minutes then leave if the other did not appear. What is the probability that they meet?

2.10. Let $X_1, X_2, \dots, X_n$ be independent $U(0,1)$ random variables. Let $M = \max_{1\le i\le n} X_i$. Calculate the distribution function of M.

2.11. The random variable whose probability density function is given by:

$$f(x) = \begin{cases} \frac{1}{2}\lambda e^{\lambda x} & , \quad \text{if } x \le 0 \\ \frac{1}{2}\lambda e^{-\lambda x} & , \quad \text{if } x > 0, \end{cases}$$

is said to have a Laplace, sometimes called a *double exponential*, distribution.
a) Verify that the density above defines a proper probability distribution.
b) Find the distribution function $F(x)$ for a Laplace random variable.

Now, let X and Y be independent exponential random variables with parameter λ. Let I be independent of X and Y and equally likely to be 1 or -1.

c) Show that $X - Y$ is a Laplace random variable.
d) Show that IX is a Laplace random variable.
e) Show that W is a Laplace random variable where:

$$W = \begin{cases} X & , \quad \text{if } I = 1 \\ -Y & , \quad \text{if } I = -1. \end{cases}$$

2.12. Give a proof of the lemma 2.2 on page 53.