Chapter 12
Probability Measures

Probability theory has become increasingly important in multiple parts of science. Getting deeply into probability theory requires a full book, not just a chapter. For readers who intend to pursue further studies in probability theory, this chapter gives you a good head start. For readers not intending to delve further into probability theory, this chapter gives you a taste of the subject.

Modern probability theory makes major use of measure theory. As we will see, a probability measure is simply a measure such that the measure of the whole space equals 1. Thus a thorough understanding of the chapters of this book dealing with measure theory and integration provides a solid foundation for probability theory.

However, probability theory is not simply the special case of measure theory where the whole space has measure 1. The questions that probability theory investigates differ from the questions natural to measure theory. For example, the probability notions of independent sets and independent random variables, which are introduced in this chapter, do not arise in measure theory.

Even when concepts in probability theory have the same meaning as well-known concepts in measure theory, the terminology and notation can be quite different. Thus one goal of this chapter is to introduce the vocabulary of probability theory. This difference in vocabulary between probability theory and measure theory occurred because the two subjects had different historical developments, only coming together in the first half of the twentieth century.

Dice used in games of chance. The beginning of probability theory can be traced to correspondence in 1654 between Pierre de Fermat (1601–1665) and Blaise Pascal (1623–1662) about how to distribute fairly money bet on an unfinished game of dice.

S. Axler, *Measure, Integration & Real Analysis*, Graduate Texts in Mathematics 282, https://doi.org/10.1007/978-3-030-33143-6_12

Probability Spaces

We begin with an intuitive and nonrigorous motivation. Suppose we pick a real number at random from the interval $(0,1)$, with each real number having an equal probability of being chosen (whatever that means). What is the probability that the chosen number is in the interval $(\frac{9}{10},1)$? The only reasonable answer to this question is $\frac{1}{10}$. More generally, if $I_1, I_2, \ldots$ is a disjoint sequence of open intervals contained in $(0,1)$, then the probability that our randomly chosen real number is in $\bigcup_{n=1}^{\infty} I_n$ should be $\sum_{n=1}^{\infty} \ell(I_n)$, where $\ell(I)$ denotes the length of an interval I. Still more generally, if A is a Borel subset of $(0,1)$, then the probability that our random number is in A should be the Lebesgue measure of A.

With the paragraph above as motivation, we are now ready to define a probability measure. We will use the notation and terminology common in probability theory instead of the conventions of measure theory.

In particular, the set in which everything takes place is now called Ω instead of the usual X in measure theory. The σ-algebra on Ω is called $\mathcal{F}$ instead of $\mathcal{S}$, which we have used in previous chapters. Our measure is now called P instead of μ. This new notation and terminology can be disorienting when first encountered. However, reading this chapter should help you become comfortable with this notation and terminology, which are standard in probability theory.

12.1 Definition ***probability measure***

Suppose $\mathcal{F}$ is a σ-algebra on a set Ω.

- A *probability measure* on $(\Omega, \mathcal{F})$ is a measure P on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$.
- Ω is called the *sample space*.
- An *event* is an element of $\mathcal{F}$ ($\mathcal{F}$ need not be mentioned if it is clear from the context).
- If A is an event, then $P(A)$ is called the *probability* of A.
- If P is a probability measure on $(\Omega, \mathcal{F})$, then the triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

12.2 Example *probability measures*

- Suppose $n \in \mathbf{Z}^+$ and Ω is a set containing exactly n elements. Let $\mathcal{F}$ denote the collection of all subsets of Ω. Then

$$\frac{\text{counting measure on } \Omega}{n}$$

is a probability measure on $(\Omega, \mathcal{F})$.

- As a more specific example of the previous item, suppose that $\Omega = \{40, 41, \ldots, 49\}$ and $P =$ (counting measure on Ω)$/10$. Let $A = \{\omega \in \Omega : \omega \text{ is even}\}$ and $B = \{\omega \in \Omega : \omega \text{ is prime}\}$. Then $P(A)$ [which is the probability that an element of this sample space Ω is even] is $\frac{1}{2}$ and $P(B)$ [which is the probability that an element of this sample space Ω is prime] is $\frac{3}{10}$.

This example illustrates the common practice in probability theory of using lower case ω to denote a typical element of upper case Ω.

- Let λ denote Lebesgue measure on the interval $[0, 1]$. Then λ is a probability measure on $([0, 1], \mathcal{B})$, where $\mathcal{B}$ denotes the σ-algebra of Borel subsets of $[0, 1]$.

- Let λ denote Lebesgue measure on $\mathbf{R}$, and let $\mathcal{B}$ denote the σ-algebra of Borel subsets of $\mathbf{R}$. Define $h \colon \mathbf{R} \to (0, \infty)$ by $h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Then $h\,d\lambda$ is a probability measure on $(\mathbf{R}, \mathcal{B})$ [see 9.6 for the definition of $h\,d\lambda$].

In measure theory, we used the notation χ_A to denote the characteristic function of a set A. In probability theory, this function has a different name and different notation, as we see in the next definition.

12.3 Definition *indicator function;* 1_A

If Ω is a set and $A \subset \Omega$, then the *indicator function* of A is the function $1_A \colon \Omega \to \mathbf{R}$ defined by

$$1_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The next definition gives the replacement in probability theory for measure theory's phrase *almost every*.

12.4 Definition *almost surely*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. An event A is said to happen *almost surely* if the probability of A is 1, or equivalently if $P(\Omega \setminus A) = 0$.

12.5 Example *almost surely*

Let P denote Lebesgue measure on the interval $[0, 1]$. If $\omega \in [0, 1]$, then ω is almost surely an irrational number (because the set of rational numbers has Lebesgue measure 0).

This example shows that an event having probability 1 (equivalent to happening almost surely) does not mean that the event definitely happens. Conversely, an event having probability 0 does not mean that the event is impossible. Specifically, if a real number is chosen at random from $[0, 1]$ using Lebesgue measure as the probability, then the probability that the number is rational is 0, but that event can still happen.

The following result is frequently useful in probability theory. A careful reading of the proof of this result, as our first proof in this chapter, should give you good practice using some of the notation and terminology commonly used in probability theory. This proof also illustrates the point that having a good understanding of measure theory and integration can often be extremely useful in probability theory—here we use the Monotone Convergence Theorem.

12.6 *Borel–Cantelli Lemma*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A_1, A_2, \dots$ is a sequence of events such that $\sum_{n=1}^{\infty} P(A_n) < \infty$. Then

$$P(\{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \in \mathbf{Z}^+\}) = 0.$$

Proof Let $A = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \in \mathbf{Z}^+\}$. Then

$$A = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n.$$

Thus $A \in \mathcal{F}$, and hence $P(A)$ makes sense.

The Monotone Convergence Theorem (3.11) implies that

$$\int_{\Omega} \Big(\sum_{n=1}^{\infty} 1_{A_n} \Big)\, dP = \sum_{n=1}^{\infty} \int_{\Omega} 1_{A_n}\, dP = \sum_{n=1}^{\infty} P(A_n) < \infty.$$

Thus $\sum_{n=1}^{\infty} 1_{A_n}$ is almost surely finite. Hence $P(A) = 0$. ■

Independent Events and Independent Random Variables

The notion of independent events, which we now define, is one of the key concepts that distinguishes probability theory from measure theory.

12.7 Definition *independent events*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space.

- Two events A and B are called *independent* if

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

- More generally, a family of events $\{A_k\}_{k \in \Gamma}$ is called *independent* if

$$P(A_{k_1} \cap \cdots \cap A_{k_n}) = P(A_{k_1}) \cdots P(A_{k_n})$$

whenever $k_1, \dots, k_n$ are distinct elements of Γ.

The next two examples should help develop your intuition about independent events.

12.8 Example *independent events: coin tossing*

Suppose $\Omega = \{H, T\}^4$, where H and T are symbols that you can think of as denoting "heads" and "tails". Thus elements of Ω are 4-tuples of the form

$$\omega = (\omega_1, \omega_2, \omega_3, \omega_4),$$

where each ω_j is H or T. Let $\mathcal{F}$ be the collection of all subsets of Ω, and let $P = (\text{counting measure on } \Omega)/16$, as we expect from a fair coin toss.

Let

$$A = \{\omega \in \Omega : \omega_1 = \omega_2 = \omega_3 = H\} \quad \text{and} \quad B = \{\omega \in \Omega : \omega_4 = H\}.$$

Then A contains two elements and thus $P(A) = \frac{1}{8}$, corresponding to probability $\frac{1}{8}$ that the first three coin tosses are all heads. Also, B contains eight elements and thus $P(B) = \frac{1}{2}$, corresponding to probability $\frac{1}{2}$ that the fourth coin toss is heads.

Now

$$P(A \cap B) = \tfrac{1}{16} = P(A) \cdot P(B),$$

where the first equality holds because $A \cap B$ consists of only the one element (H, H, H, H) and the second equality holds because $P(A) = \frac{1}{8}$ and $P(B) = \frac{1}{2}$. The equation above shows that A and B are independent events.

If we toss a fair coin many times, we expect that about half the time it will be heads. Thus some people mistakenly believe that if the first three tosses of a fair coin are heads, then the fourth toss should have a higher probability of being tails, to balance out the previous heads. However, the coin cannot remember that it had three heads in a row, and thus the fourth coin toss has probability $\frac{1}{2}$ of being heads regardless of the results of the three previous coin tosses. The independence of the events A and B above captures the notion that the results of a fair coin toss do not depend upon previous results.

12.9 Example *independent events: product probability space*

Suppose $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ are probability spaces. Then

$$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \times P_2),$$

as defined in Chapter 5, is also a probability space.

If $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$, then $(A \times \Omega_2) \cap (\Omega_1 \times B) = A \times B$. Thus

$$\begin{aligned}
(P_1 \times P_2)\big((A \times \Omega_2) \cap (\Omega_1 \times B)\big) &= (P_1 \times P_2)(A \times B) \\
&= P_1(A) \cdot P_2(B) \\
&= (P_1 \times P_2)(A \times \Omega_2) \cdot (P_1 \times P_2)(\Omega_1 \times B),
\end{aligned}$$

where the second equality follows from the definition of the product measure, and the third equality holds because of the definition of the product measure and because P_1 and P_2 are probability measures.

The equation above shows that the events $A \times \Omega_2$ and $\Omega_1 \times B$ are independent events in $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Compare the next result to the Borel–Cantelli Lemma (12.6).

12.10 *relative of Borel–Cantelli Lemma*

Suppose $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space and $\{A_n\}_{n \in \mathbf{Z}^+}$ is an independent family of events such that $\sum_{n=1}^{\infty} P(A_n) = \infty$. Then

$$P(\{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \in \mathbf{Z}^+\}) = 1.$$

Proof Let $A = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n \in \mathbf{Z}^+\}$. Then

12.11 $$\Omega \setminus A = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} (\Omega \setminus A_n).$$

If $m, M \in \mathbf{Z}^+$ are such that $m \leq M$, then

$$\begin{aligned} P\Big(\bigcap_{n=m}^{M} (\Omega \setminus A_n)\Big) &= \prod_{n=m}^{M} P(\Omega \setminus A_n) \\ &= \prod_{n=m}^{M} (1 - P(A_n)) \end{aligned}$$

12.12 $$\leq e^{-\sum_{n=m}^{M} P(A_n)},$$

where the first line holds because the family $\{\Omega \setminus A_n\}_{n \in \mathbf{Z}^+}$ is independent (see Exercise 4) and the third line holds because $1 - t \leq e^{-t}$ for all $t \geq 0$.

Because $\sum_{n=1}^{\infty} P(A_n) = \infty$, by choosing M large we can make the right side of 12.12 as close to 0 as we wish. Thus

$$P\Big(\bigcap_{n=m}^{\infty} (\Omega \setminus A_n)\Big) = 0$$

for all $m \in \mathbf{Z}^+$. Now 12.11 implies that $P(\Omega \setminus A) = 0$. Thus we conclude that $P(A) = 1$, as desired. ∎

For the rest of this chapter, assume that $\mathbf{F} = \mathbf{R}$. Thus, for example, if $(\Omega, \mathcal{F}, P)$ is a probability space, then $\mathcal{L}^1(P)$ will always refer to the vector space of ***real-valued*** $\mathcal{F}$-measurable functions on Ω such that $\int_{\Omega} |f| \, dP < \infty$.

12.13 Definition *random variable; expectation;* EX

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space.

- A *random variable* on $(\Omega, \mathcal{F})$ is a measurable function from Ω to $\mathbf{R}$.
- If $X \in \mathcal{L}^1(P)$, then the *expectation* (sometimes called the *expected value*) of the random variable X is denoted EX and is defined by

$$EX = \int_{\Omega} X \, dP.$$

If $\mathcal{F}$ is clear from the context, the phrase "random variable on Ω" can be used instead of the more precise phrase "random variable on $(\Omega, \mathcal{F})$". If both Ω and $\mathcal{F}$ are clear from the context, then the phrase "random variable" has no ambiguity and is often used.

Because $P(\Omega) = 1$, the expectation EX of a random variable $X \in \mathcal{L}^1(P)$ can be thought of as the *average* or *mean value* of X.

The next definition illustrates a convention often used in probability theory: the variable is often omitted when describing a set. Thus, for example, $\{X \in U\}$ means $\{\omega \in \Omega : X(\omega) \in U\}$, where U is a subset of $\mathbf{R}$. Also, probabilists often also omit the set brackets, as we do for the first time in the second bullet point below, when appropriate parentheses are nearby.

12.14 Definition ***independent random variables***

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space.

- Two random variables X and Y are called *independent* if $\{X \in U\}$ and $\{Y \in V\}$ are independent events for all Borel sets U, V in $\mathbf{R}$.
- More generally, a family of random variables $\{X_k\}_{k \in \Gamma}$ is called *independent* if $\{X_k \in U_k\}_{k \in \Gamma}$ is independent for all families of Borel sets $\{U_k\}_{k \in \Gamma}$ in $\mathbf{R}$.

12.15 Example *independent random variables*

- Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A, B \in \mathcal{F}$. Then 1_A and 1_B are independent random variables if and only if A and B are independent events, as you should verify.
- Suppose $\Omega = \{H, T\}^4$ is the sample space of four coin tosses, with Ω and P as in Example 12.8. Define random variables X and Y by

  $$X(\omega_1, \omega_2, \omega_3, \omega_4) = \text{number of } \omega_1, \omega_2, \omega_3 \text{ that equal } H$$

  and

  $$Y(\omega_1, \omega_2, \omega_3, \omega_4) = \text{number of } \omega_3, \omega_4 \text{ that equal } H.$$

  Then X and Y are not independent random variables because $P(X = 3) = \frac{1}{8}$ and $P(Y = 0) = \frac{1}{4}$ but $P(\{X = 3\} \cap \{Y = 0\}) = P(\varnothing) = 0 \neq \frac{1}{8} \cdot \frac{1}{4}$.
- Suppose $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ are probability spaces, Z_1 is a random variable on Ω_1, and Z_2 is a random variable on Ω_2. Define random variables X and Y on $\Omega_1 \times \Omega_2$ by

  $$X(\omega_1, \omega_2) = Z_1(\omega_1) \quad \text{and} \quad Y(\omega_1, \omega_2) = Z_2(\omega_2).$$

Then X and Y are independent random variables on $\Omega_1 \times \Omega_2$ (with respect to the probability measure $P_1 \times P_2$), as you should verify.

If X is a random variable and $f : \mathbf{R} \to \mathbf{R}$ is Borel measurable, then $f \circ X$ is a random variable (by 2.44). For example, if X is a random variable, then X^2 and e^X are random variables. The next result states that compositions preserve independence.

12.16 ***functions of independent random variables are independent***

Suppose $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space, X and Y are independent random variables, and $f, g \colon \mathbf{R} \to \mathbf{R}$ are Borel measurable. Then $f \circ X$ and $g \circ Y$ are independent random variables.

Proof Suppose U, V are Borel subsets of $\mathbf{R}$. Then

$$\begin{aligned} P\big(\{f \circ X \in U\} \cap \{g \circ Y \in V\}\big) &= P\big(\{X \in f^{-1}(U)\} \cap \{Y \in g^{-1}(V)\}\big) \\ &= P\big(X \in f^{-1}(U)\big) \cdot P\big(Y \in g^{-1}(V)\big) \\ &= P(f \circ X \in U) \cdot P(g \circ Y \in V), \end{aligned}$$

where the second equality holds because X and Y are independent random variables. The equation above shows that $f \circ X$ and $g \circ Y$ are independent random variables. ■

If $X, Y \in \mathcal{L}^1(P)$, then clearly $E(X + Y) = E(X) + E(Y)$. The next result gives a nice formula for the expectation of XY when X and Y are independent. This formula has sometimes been called the dream equation of calculus students.

12.17 ***expectation of product of independent random variables***

Suppose $(\Omega, \mathcal{F}, \mathcal{P})$ is a probability space and X and Y are independent random variables in $\mathcal{L}^2(P)$. Then

$$E(XY) = EX \cdot EY.$$

Proof First consider the case where X and Y are each simple functions, taking on only finitely many values. Thus there are distinct numbers $a_1, \ldots, a_M \in \mathbf{R}$ and distinct numbers $b_1, \ldots, b_N \in \mathbf{R}$ such that

$$X = a_1 1_{\{X=a_1\}} + \cdots + a_M 1_{\{X=a_M\}} \quad \text{and} \quad Y = b_1 1_{\{Y=b_1\}} + \cdots + b_N 1_{\{Y=b_N\}}.$$

Now

$$XY = \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k 1_{\{X=a_j\}} 1_{\{Y=b_k\}} = \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k 1_{\{X=a_j\} \cap \{Y=b_k\}}.$$

Thus

$$\begin{aligned} E(XY) &= \sum_{j=1}^{M} \sum_{k=1}^{N} a_j b_k P\big(\{X = a_j\} \cap \{Y = b_k\}\big) \\ &= \Big(\sum_{j=1}^{M} a_j P(X = a_j)\Big) \Big(\sum_{k=1}^{N} b_k P(Y = b_k)\Big) \\ &= EX \cdot EY, \end{aligned}$$

where the second equality above comes from the independence of X and Y. The last equation gives the desired conclusion in the case where X and Y are simple functions.

Now consider arbitrary independent random variables X and Y in $\mathcal{L}^2(P)$. Let $f_1, f_2, \ldots$ be a sequence of Borel measurable simple functions from $\mathbf{R}$ to $\mathbf{R}$ that approximate the identity function on $\mathbf{R}$ (the function $t \mapsto t$) in the sense that $\lim_{n\to\infty} f_n(t) = t$ for every $t \in \mathbf{R}$ and $|f_n(t)| \le t$ for all $t \in \mathbf{R}$ and all $n \in \mathbf{Z}^+$ (see 2.89, taking f to be the identity function, for construction of this sequence). The random variables $f_n \circ X$ and $f_n \circ Y$ are independent (by 12.17). Thus the result in the first paragraph of this proof shows that

$$E\big((f_n \circ X)(f_n \circ Y)\big) = E(f_n \circ X) \cdot E(f_n \circ Y)$$

for each $n \in \mathbf{Z}^+$. The limit as $n \to \infty$ of the right side of the equation above equals $EX \cdot EY$ [by the Dominated Convergence Theorem (3.31)]. The limit as $n \to \infty$ of the left side of the equation above equals $E(XY)$ [use Hölder's inequality (7.9)]. Thus the equation above implies that $E(XY) = EX \cdot EY$. ■

Variance and Standard Deviation

The variance and standard deviation of a random variable, defined below, measure how much a random variable differs from its expectation.

12.18 Definition *variance; standard deviation; $\sigma(X)$*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $X \in \mathcal{L}^2(P)$ is a random variable.

- The *variance* of X is defined to be $E\big((X - EX)^2\big)$.
- The *standard deviation* of X is denoted $\sigma(X)$ and is defined by

  $$\sigma(X) = \sqrt{E\big((X - EX)^2\big)}.$$

  In other words, the standard deviation of X is the square root of the variance of X.

The notation $\sigma^2(X)$ means $\big(\sigma(X)\big)^2$. Thus $\sigma^2(X)$ is the variance of X.

12.19 Example *variance and standard deviation of an indicator function*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A \in \mathcal{F}$ is an event. Then

$$\begin{aligned}
\sigma^2(1_A) &= E\big((1_A - E1_A)^2\big) \\
&= E\big((1_A - P(A))^2\big) \\
&= E\big(1_A - 2P(A) \cdot 1_A + P(A)^2\big) \\
&= P(A) - 2\big(P(A)\big)^2 + \big(P(A)\big)^2 \\
&= P(A) \cdot \big(1 - P(A)\big).
\end{aligned}$$

Thus $\sigma(1_A) = \sqrt{P(A) \cdot \big(1 - P(A)\big)}$.

The next result gives a formula for the variance of a random variable. This formula is often more convenient to use than the formula that defines the variance.

12.20 *variance formula*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $X \in \mathcal{L}^2(P)$ is a random variable. Then
$$\sigma^2(X) = E(X^2) - (EX)^2.$$

Proof We have

$$\begin{aligned}\sigma^2(X) &= E\big((X - EX)^2\big)\\ &= E\big(X^2 - 2(EX)X + (EX)^2\big)\\ &= E(X^2) - 2(EX)^2 + (EX)^2\\ &= E(X^2) - (EX)^2,\end{aligned}$$

as desired. ∎

Our next result is called Chebyshev's inequality. It states, for example (take $t = 2$ below) that the probability that a random variable X differs from its average by more than twice its standard deviation is at most $\frac{1}{4}$. Note that $P\big(|X - EX| \geq t\sigma(X)\big)$ is shorthand for $P\big(\{\omega \in \Omega : |X(\omega) - EX| \geq t\sigma(X)\}\big)$.

12.21 *Chebyshev's inequality*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $X \in \mathcal{L}^2(P)$ is a random variable. Then
$$P\big(|X - EX| \geq t\sigma(X)\big) \leq \frac{1}{t^2}$$
for all $t > 0$.

Proof Suppose $t > 0$. Then

$$\begin{aligned}P\big(|X - EX| \geq t\sigma(X)\big) &= P\big(|X - EX|^2 \geq t^2\sigma^2(X)\big)\\ &\leq \frac{1}{t^2\sigma^2(X)}E\big((X - EX)^2\big)\\ &= \frac{1}{t^2},\end{aligned}$$

where the second line above comes from applying Markov's inequality (4.1) with $h = |X - EX|^2$ and $c = t^2\sigma^2(X)$. ∎

The next result gives a beautiful formula for the variance of the sum of independent random variables.

12.22 ***variance of sum of independent random variables***

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $X_1, \dots, X_n \in \mathcal{L}^2(P)$ are independent random variables. Then

$$\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n).$$

Proof Using the variance formula given by 12.20, we have

$$\begin{aligned}
\sigma^2\Big(\sum_{k=1}^{n} X_k\Big) &= E\Big(\big(\sum_{k=1}^{n} X_k\big)^2\Big) - \Big(E\big(\sum_{k=1}^{n} X_k\big)\Big)^2 \\
&= E\Big(\sum_{k=1}^{n} X_k{}^2\Big) + 2E\Big(\sum_{1\le j<k\le n} X_j X_k\Big) - \Big(\sum_{k=1}^{n} EX_k\Big)^2 \\
&= \sum_{k=1}^{n} E(X_k^2) - \sum_{k=1}^{n} (EX_k)^2 + 2\Big(\sum_{1\le j<k\le n} E(X_j X_k)\Big) - 2\Big(\sum_{1\le j<k\le n} EX_j \cdot EX_k\Big) \\
&= \sum_{k=1}^{n} \sigma^2(X_k),
\end{aligned}$$

where the last equality uses 12.20, 12.17, and the hypothesis that $X_1, \dots, X_n$ are independent random variables. ∎

Conditional Probability and Bayes' Theorem

The conditional probability $P_B(A)$ that we are about to define should be interpreted to mean the probability that ω will be in A given that $\omega \in B$. Because ω is in $A \cap B$ if and only if $\omega \in B$ and $\omega \in A$, and because we expect probabilities to multiply, it is reasonable to expect that

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

Thus we are led to the following definition.

12.23 Definition ***conditional probability;*** P_B

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and B is an event with $P(B) > 0$. Define $P_B \colon \mathcal{F} \to [0,1]$ by

$$P_B(A) = \frac{P(A \cap B)}{P(B)}.$$

If $A \in \mathcal{F}$, then $P_B(A)$ is called the *conditional probability* of A given B.

You should verify that with B as above, P_B is a probability measure on $(\Omega, \mathcal{F})$. If $A \in \mathcal{F}$, then $P_B(A) = P(A)$ if and only if A and B are independent events.

We now present two versions of what is called Bayes' Theorem. You should do a web search and read about the many uses of these results, including some controversial applications.

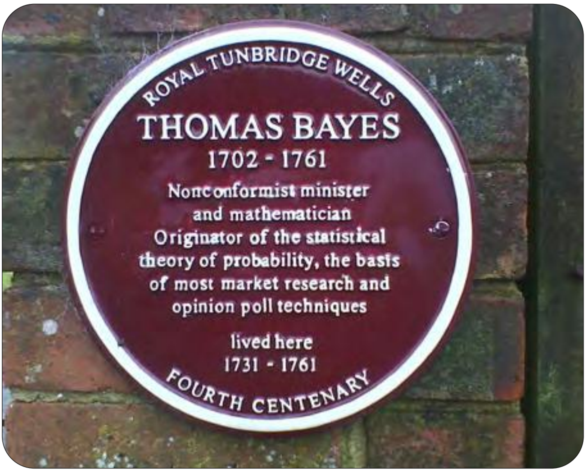
12.24 *Bayes' Theorem, first version*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and A, B are events with positive probability. Then

$$P_B(A) = \frac{P_A(B) \cdot P(A)}{P(B)}.$$

Proof We have

$$P_B(A) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B) \cdot P(A)}{P(A) \cdot P(B)} = \frac{P_A(B) \cdot P(A)}{P(B)}.$$



Plaque honoring Thomas Bayes in Tunbridge Wells, England.

12.25 *Bayes' Theorem, second version*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space, B is an event with positive probability, and $A_1, \ldots, A_n$ are pairwise disjoint events, each with positive probability, such that $A_1 \cup \cdots \cup A_n = \Omega$. Then

$$P_B(A_k) = \frac{P_{A_k}(B) \cdot P(A_k)}{\sum_{j=1}^{n} P_{A_j}(B) \cdot P(A_j)}$$

for each $k \in \{1, \ldots, n\}$.

Proof Consider the denominator of the expression above. We have

12.26
$$\sum_{j=1}^{n} P_{A_j}(B) \cdot P(A_j) = \sum_{j=1}^{n} P(A_j \cap B) = P(B).$$

Now suppose $k \in \{1, \ldots, n\}$. Then

$$P_B(A_k) = \frac{P_{A_k}(B) \cdot P(A_k)}{P(B)} = \frac{P_{A_k}(B) \cdot P(A_k)}{\sum_{j=1}^{n} P_{A_j}(B) \cdot P(A_j)},$$

where the first equality comes from the first version of Bayes's Theorem (12.24) and the second equality comes from 12.26.

Distribution and Density Functions of Random Variables

For the rest of this chapter, let $\mathcal{B}$ denote the σ-algebra of Borel subsets of $\mathbf{R}$.

Each random variable X determines a probability measure P_X on $(\mathbf{R}, \mathcal{B})$ and a function $\tilde{X}\colon \mathbf{R} \to [0,1]$ as in the next definition.

12.27 Definition ***probability distribution and distribution function; P_X; $\tilde{X}$***

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and X is a random variable.

- The *probability distribution* of X is the probability measure P_X defined on $(\mathbf{R}, \mathcal{B})$ by
$$P_X(B) = P(X \in B) = P\big(X^{-1}(B)\big).$$
- The *distribution function* of X is the function $\tilde{X}\colon \mathbf{R} \to [0,1]$ defined by
$$\tilde{X}(s) = P_X\big((-\infty, s]\big) = P(X \leq s).$$

You should verify that the probability distribution P_X as defined above is indeed a probability measure on $(\mathbf{R}, \mathcal{B})$. Note that the distribution function $\tilde{X}$ depends upon the probability measure P as well as the random variable X, even though P is not included in the notation $\tilde{X}$ (because P is usually clear from the context).

12.28 Example *probability distribution and distribution function of an indicator function*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A \in \mathcal{F}$ is an event. Then you should verify that
$$P_{1_A} = \big(1 - P(A)\big)\delta_0 + P(A)\delta_1,$$
where for $t \in \mathbf{R}$ the measure δ_t on $(\mathbf{R}, \mathcal{B})$ is defined by
$$\delta_t(B) = \begin{cases} 1 & \text{if } t \in B, \\ 0 & \text{if } t \notin B. \end{cases}$$

The distribution function of 1_A is the function $(1_A)\tilde{}\colon \mathbf{R} \to [0,1]$ given by
$$(1_A)\tilde{}(s) = \begin{cases} 0 & \text{if } s < 0, \\ 1 - P(A) & \text{if } 0 \leq s < 1, \\ 1 & \text{if } s \geq 1, \end{cases}$$
as you should verify.

One direction of the next result states that every probability distribution is a right-continuous increasing function, with limit 0 at $-\infty$ and limit 1 at ∞. The other direction of the next result states that every function with those properties is the distribution function of some random variable on some probability space. The proof shows that we can take the sample space to be $(0,1)$, the σ-algebra to be the Borel subsets of $(0,1)$, and the probability measure to be Lebesgue measure on $(0,1)$.

Your understanding of the proof of the next result should be enhanced by Exercise 13, which asserts that if the function $H\colon \mathbf{R} \to (0,1)$ appearing in the next result is continuous and injective, then the random variable $X\colon (0,1) \to \mathbf{R}$ in the proof is the inverse function of H.

12.29 *characterization of distribution functions*

Suppose $H\colon \mathbf{R} \to [0,1]$ is a function. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a random variable X on $(\Omega, \mathcal{F})$ such that $H = \tilde{X}$ if and only if the following conditions are all satisfied:

(a) $s < t \Rightarrow H(s) \leq H(t)$ (in other words, H is an increasing function);

(b) $\displaystyle\lim_{t \to -\infty} H(t) = 0$;

(c) $\displaystyle\lim_{t \to \infty} H(t) = 1$;

(d) $\displaystyle\lim_{t \downarrow s} H(t) = H(s)$ for every $s \in \mathbf{R}$ (in other words, H is right continuous).

Proof First suppose $H = \tilde{X}$ for some probability space $(\Omega, \mathcal{F}, P)$ and some random variable X on $(\Omega, \mathcal{F})$. Then (a) holds because $s < t$ implies $(-\infty, s] \subset (-\infty, t]$. Also, (b) and (d) follow from 2.60. Furthermore, (c) follows from 2.59, completing the proof in this direction.

To prove the other direction, now suppose that H satisfies (a) through (d). Let $\Omega = (0,1)$, let $\mathcal{F}$ be the collection of Borel subsets of the interval $(0,1)$, and let P be Lebesgue measure on $\mathcal{F}$. Define a random variable X by

12.30 $$X(\omega) = \sup\{t \in \mathbf{R} : H(t) < \omega\}$$

for $\omega \in (0,1)$. Clearly X is an increasing function and thus is measurable (in other words, X is indeed a random variable).

Suppose $s \in \mathbf{R}$. If $\omega \in \big(0, H(s)\big]$, then

$$X(\omega) \leq X\big(H(s)\big) = \sup\{t \in \mathbf{R} : H(t) < H(s)\} \leq s,$$

where the first inequality holds because X is an increasing function and the last inequality holds because H is an increasing function. Hence

12.31 $$\big(0, H(s)\big] \subset \{X \leq s\}.$$

If $\omega \in (0,1)$ and $X(\omega) \leq s$, then $H(t) \geq \omega$ for all $t > s$ (by 12.30). Thus

$$H(s) = \lim_{t \downarrow s} H(t) \geq \omega,$$

where the equality above comes from (d). Rewriting the inequality above, we have $\omega \in \big(0, H(s)\big]$. Thus we have shown that $\{X \leq s\} \subset \big(0, H(s)\big]$, which when combined with 12.31 shows that $\{X \leq s\} = \big(0, H(s)\big]$. Hence

$$\tilde{X}(s) = P(X \leq s) = P\Big(\big(0, H(s)\big]\Big) = H(s),$$

as desired. ∎

In the definition below and in the following discussion, λ denotes Lebesgue measure on $\mathbf{R}$, as usual.

12.32 Definition ***density function***

Suppose X is a random variable on some probability space. If there exists $h \in L^1(\mathbf{R})$ such that

$$\tilde{X}(s) = \int_{-\infty}^{s} h \, d\lambda$$

for all $s \in \mathbf{R}$, then h is called the *density function* of X.

If there is a density function of a random variable X, then it is unique [up to changes on sets of Lebesgue measure 0, which is already taken into account because we are thinking of density functions as elements of $L^1(\mathbf{R})$ instead of elements of $\mathcal{L}^1(\mathbf{R})$]; see Exercise 6 in Chapter 4.

If X is a random variable that has a density function h, then the distribution function $\tilde{X}$ is differentiable almost everywhere (with respect to Lebesgue measure) and $\tilde{X}'(s) = h(s)$ for almost every $s \in \mathbf{R}$ (by the second version of the Lebesgue Differentiation Theorem; see 4.19). Because $\tilde{X}$ is an increasing function, this implies that $h(s) \geq 0$ for almost every $s \in \mathbf{R}$. In other words, we can assume that a density function is nonnegative.

In the definition above of a density function, we started with a probability space and a random variable on it. Often in probability theory, the procedure goes in the other direction. Specifically, we can start with a nonnegative function $h \in L^1(\mathbf{R})$ such that $\int_{-\infty}^{\infty} h \, d\lambda = 1$. We use h to define a probability measure on $(\mathbf{R}, \mathcal{B})$ and then consider the identity random variable X on $\mathbf{R}$. The function h that we started with is then the density function of X. The following result formalizes this procedure and gives formulas for the mean and standard deviation in terms of the density function h.

12.33 ***mean and variance of random variable generated by density function***

Suppose $h \in L^1(\mathbf{R})$ is such that $\int_{-\infty}^{\infty} h \, d\lambda = 1$ and $h(x) \geq 0$ for almost every $x \in \mathbf{R}$. Let P be the probability measure on $(\mathbf{R}, \mathcal{B})$ defined by

$$P(B) = \int_B h \, d\lambda.$$

Let X be the random variable on $(\mathbf{R}, \mathcal{B})$ defined by $X(x) = x$ for each $x \in \mathbf{R}$. Then h is the density function of X. Furthermore, if $X \in \mathcal{L}^1(P)$ then

$$EX = \int_{-\infty}^{\infty} x h(x) \, d\lambda(x),$$

and if $X \in \mathcal{L}^2(P)$ then

$$\sigma^2(X) = \int_{-\infty}^{\infty} x^2 h(x) \, d\lambda(x) - \Big(\int_{-\infty}^{\infty} x h(x) \, d\lambda(x)\Big)^2.$$

Proof The equation $\tilde{X}(s) = \int_{-\infty}^{s} h\, d\lambda$ holds by the definitions of $\tilde{X}$ and P. Thus h is the density function of X.

Our definition of P to equal $h\, d\lambda$ implies that $\int_{-\infty}^{\infty} f\, dP = \int_{-\infty}^{\infty} fh\, d\lambda$ for all $f \in \mathcal{L}^1(P)$ [see Exercise 5 in Section 9A]. Thus the formula for the mean EX follows immediately from the definition of EX, and the formula for the variance $\sigma^2(X)$ follows from 12.20. ∎

The following example illustrates the result above with a few especially useful choices of the density function h.

12.34 Example *density functions*

- Suppose $h = 1_{[0,1]}$. This density function h is called the *uniform density* on $[0,1]$. In this case, $P(B) = \lambda(B \cap [0,1])$ for each Borel set $B \subset \mathbf{R}$. For the corresponding random variable $X(x) = x$ for $x \in \mathbf{R}$, the distribution function $\tilde{X}$ is given by the formula

$$\tilde{X}(s) = \begin{cases} 0 & \text{if } s \leq 0, \\ s & \text{if } 0 < s < 1, \\ 1 & \text{if } s \geq 1. \end{cases}$$

  The formulas in 12.33 show that $EX = \frac{1}{2}$ and $\sigma(X) = \frac{1}{2\sqrt{3}}$.

- Suppose $\alpha > 0$ and

$$h(x) = \begin{cases} 0 & \text{if } x < 0, \\ \alpha e^{-\alpha x} & \text{if } x \geq 0. \end{cases}$$

  This density function h is called the *exponential density* on $[0, \infty)$. For the corresponding random variable $X(x) = x$ for $x \in \mathbf{R}$, the distribution function $\tilde{X}$ is given by the formula

$$\tilde{X}(s) = \begin{cases} 0 & \text{if } s < 0, \\ 1 - e^{-\alpha s} & \text{if } s \geq 0. \end{cases}$$

  The formulas in 12.33 show that $EX = \frac{1}{\alpha}$ and $\sigma(X) = \frac{1}{\alpha}$.

- Suppose

$$h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

  for $x \in \mathbf{R}$. This density function is called the *standard normal density*. For the corresponding random variable $X(x) = x$ for $x \in \mathbf{R}$, we have $\tilde{X}(0) = \frac{1}{2}$. For general $s \in \mathbf{R}$, no formula exists for $\tilde{X}(s)$ in terms of elementary functions. However, the formulas in 12.33 show that $EX = 0$ and (with the help of some calculus) $\sigma(X) = 1$.

Weak Law of Large Numbers

Families of random variables all of which look the same in terms of their distribution functions get a special name, as we see in the next definition.

12.35 Definition ***identically distributed; i.i.d.***

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space.

- A family of random variables on $(\Omega, \mathcal{F})$ is called *identically distributed* if all the random variables in the family have the same distribution function.
- More specifically, a family $\{X_k\}_{k\in\Gamma}$ of random variables on $(\Omega, \mathcal{F})$ is called *identically distributed* if

$$P(X_j \leq s) = P(X_k \leq s)$$

  for all $j, k \in \Gamma$.
- A family of random variables that is independent and identically distributed is said to be *independent identically distributed*, often abbreviated as *i.i.d.*

12.36 Example *family of random variables for decimal digits is i.i.d.*

Consider the probability space $([0,1], \mathcal{B}, P)$, where $\mathcal{B}$ is the collection of Borel subsets of the interval $[0,1]$ and P is Lebesgue measure on $([0,1], \mathcal{B})$. For $k \in \mathbf{Z}^+$, define a random variable $X_k \colon [0,1] \to \mathbf{R}$ by

$$X_k(\omega) = k^{\text{th}}\text{-digit in decimal expansion of } \omega,$$

where for those numbers ω that have two different decimal expansions we use the one that does not end in an infinite string of 9s.

Notice that $P(X_k \leq \pi) = 0.4$ for every $k \in \mathbf{Z}^+$. More generally, the family $\{X_k\}_{k\in\mathbf{Z}^+}$ is identically distributed, as you should verify.

The family $\{X_k\}_{k\in\mathbf{Z}^+}$ is also independent, as you should verify. Thus $\{X_k\}_{k\in\mathbf{Z}^+}$ is an i.i.d. family of random variables.

Identically distributed random variables have the same expectation and the same standard deviation, as the next result shows.

12.37 ***identically distributed random variables have same mean and variance***

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\{X_k\}_{k\in\Gamma}$ is an identically distributed family of random variables in $\mathcal{L}^2(P)$. Then

$$EX_j = EX_k \quad \text{and} \quad \sigma(X_j) = \sigma(X_k)$$

for all $j, k \in \Gamma$.

Proof Suppose $j \in \mathbf{Z}^+$. Let $f_1, f_2, \dots$ be the sequence of simple functions converging pointwise to X_j as constructed in the proof of 2.89. The Dominated Convergence Theorem (3.31) implies that $EX_j = \lim_{n\to\infty} Ef_n$. Because of how each f_n is constructed, each Ef_n depends only on n and the numbers $P(c \le X_j < d)$ for $c < d$. However,

$$P(c \le X_j < d) = \lim_{m\to\infty} \Big(P\big(X_j \le d - \tfrac{1}{m}\big) - P\big(X_j \le c - \tfrac{1}{m}\big) \Big)$$

for $c < d$. Because $\{X_k\}_{k\in\Gamma}$ is an identically distributed family, the numbers above on the right are independent of j. Thus $EX_j = EX_k$ for all $j, k \in \mathbf{Z}^+$.

Apply the result from the paragraph above to the identically distributed family $\{{X_k}^2\}_{k\in\Gamma}$ and use 12.20 to conclude that $\sigma(X_j) = \sigma(X_k)$ for all $j, k \in \Gamma$. ∎

The next result has the nicely intuitive interpretation that if we repeat a random process many times, then the probability that the average of our results differs from our expected average by more than any fixed positive number ε has limit 0 as we increase the number of repetitions of the process.

12.38 *Weak Law of Large Numbers*

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\{X_k\}_{k\in\mathbf{Z}^+}$ is an i.i.d. family of random variables in $\mathcal{L}^2(P)$, each with expectation μ. Then

$$\lim_{n\to\infty} P\Big(\Big| \big(\tfrac{1}{n} \sum_{k=1}^{n} X_k\big) - \mu \Big| \ge \varepsilon \Big) = 0$$

for all $\varepsilon > 0$.

Proof Because the random variables $\{X_k\}_{k\in\mathbf{Z}^+}$ all have the same expectation and same standard deviation, by 12.37 there exist $\mu \in \mathbf{R}$ and $s \in [0, \infty)$ such that

$$EX_k = \mu \quad \text{and} \quad \sigma(X_k) = s$$

for all $k \in \mathbf{Z}^+$. Thus

$$E\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) = \mu \quad \text{and} \quad \sigma^2\Big(\frac{1}{n}\sum_{k=1}^{n} X_k\Big) = \frac{1}{n^2}\,\sigma^2\Big(\sum_{k=1}^{n} X_k\Big) = \frac{s^2}{n}, \tag{12.39}$$

where the last equality follows from 12.22 (this is where we use the independent part of the hypothesis).

Now suppose $\varepsilon > 0$. In the special case where $s = 0$, all the X_k are almost surely equal to the same constant function and the desired result clearly holds. Thus we assume $s > 0$. Let $t = \sqrt{n}\varepsilon/s$ and apply Chebyshev's inequality (12.21) with this value of t to the random variable $\frac{1}{n}\sum_{k=1}^{n} X_k$, using 12.39 to get

$$P\Big(\Big| \big(\frac{1}{n} \sum_{k=1}^{n} X_k\big) - \mu \Big| \ge \varepsilon \Big) \le \frac{s^2}{n\varepsilon^2}.$$

Taking the limit as $n \to \infty$ of both sides of the inequality above gives the desired result. ∎

EXERCISES 12

1 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A \in \mathcal{F}$. Prove that A and $\Omega \setminus A$ are independent if and only if $P(A) = 0$ or $P(A) = 1$.

2 Suppose P is Lebesgue measure on $[0,1]$. Give an example of two disjoint Borel subsets sets A and B of $[0,1]$ such that $P(A) = P(B) = \frac{1}{2}$, $[0, \frac{1}{2}]$ and A are independent, and $[0, \frac{1}{2}]$ and B are independent.

3 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A, B \in \mathcal{F}$. Prove that the following are equivalent:

- A and B are independent events.
- A and $\Omega \setminus B$ are independent events.
- $\Omega \setminus A$ and B are independent events.
- $\Omega \setminus A$ and $\Omega \setminus B$ are independent events.

4 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\{A_k\}_{k \in \Gamma}$ is a family of events. Prove the family $\{A_k\}_{k \in \Gamma}$ is independent if and only if the family $\{\Omega \setminus A_k\}_{k \in \Gamma}$ is independent.

5 Give an example of a probability space $(\Omega, \mathcal{F}, P)$ and events A, B_1, B_2 such that A and B_1 are independent, A and B_2 are independent, but A and $B_1 \cup B_2$ are not independent.

6 Give an example of a probability space $(\Omega, \mathcal{F}, P)$ and events A_1, A_2, A_3 such that A_1 and A_2 are independent, A_1 and A_3 are independent, and A_2 and A_3 are independent, but the family A_1, A_2, A_3 is not independent.

7 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space, $A \in \mathcal{F}$, and $B_1 \subset B_2 \subset \cdots$ is an increasing sequence of events such that A and B_n are independent events for each $n \in \mathbf{Z}^+$. Show that A and $\bigcup_{n=1}^{\infty} B_n$ are independent.

8 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\{A_t\}_{t \in \mathbf{R}}$ is an independent family of events such that $P(A_t) < 1$ for each $t \in \mathbf{R}$. Prove that there exists a sequence $t_1, t_2, \ldots$ in $\mathbf{R}$ such that $P\big(\bigcap_{n=1}^{\infty} A_{t_n}\big) = 0$.

9 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $B_1, \ldots, B_n \in \mathcal{F}$ are such that $P(B_1 \cap \cdots \cap B_n) > 0$. Prove that

$$P(A \cap B_1 \cap \cdots \cap B_n) = P(B_1) \cdot P_{B_1}(B_2) \cdots P_{B_1 \cap \cdots \cap B_{n-1}}(B_n) \cdot P_{B_1 \cap \cdots \cap B_n}(A)$$

for every event $A \in \mathcal{F}$.

10 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A \in \mathcal{F}$ is an event such that $0 < P(A) < 1$. Prove that

$$P(B) = P_A(B) \cdot P(A) + P_{\Omega \setminus A}(B) \cdot P(\Omega \setminus A)$$

for every event $B \in \mathcal{F}$.

11 Give an example of a probability space $(\Omega, \mathcal{F}, P)$ and $X, Y \in \mathcal{L}^2(P)$ such that $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$ but X and Y are not independent random variables.

12 Prove that if X and Y are random variables (possibly on two different probability spaces) and $\tilde{X} = \tilde{Y}$, then $P_X = P_Y$.

13 Suppose $H\colon \mathbf{R} \to (0,1)$ is a continuous one-to-one function satisfying conditions (a) through (d) of 12.29. Show that the function $X\colon (0,1) \to \mathbf{R}$ produced in the proof of 12.29 is the inverse function of H.

14 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and X is a random variable. Prove that the following are equivalent:

- $\tilde{X}$ is a continuous function on $\mathbf{R}$.
- $\tilde{X}$ is a uniformly continuous function on $\mathbf{R}$.
- $P(X = t) = 0$ for every $t \in \mathbf{R}$.
- $(\tilde{X} \circ X)\tilde{}(s) = s$ for all $s \in \mathbf{R}$.

15 Suppose $\alpha > 0$ and $h(x) = \begin{cases} 0 & \text{if } x < 0, \\ \alpha^2 x e^{-\alpha x} & \text{if } x \geq 0. \end{cases}$

Let $P = h\,d\lambda$ and let X be the random variable defined by $X(x) = x$ for $x \in \mathbf{R}$.

(a) Verify that $\int_{-\infty}^{\infty} h\,d\lambda = 1$.

(b) Find a formula for the distribution function $\tilde{X}$.

(c) Find a formula (in terms of α) for EX.

(d) Find a formula (in terms of α) for $\sigma(X)$.

16 Suppose $\mathcal{B}$ is the σ-algebra of Borel subsets of $[0,1)$ and P is Lebesgue measure on $([0,1], \mathcal{B})$. Let $\{e_k\}_{k \in \mathbf{Z}^+}$ be the family of functions defined by the fourth bullet point of Example 8.51 (notice that $k = 0$ is excluded). Show that the family $\{e_k\}_{k \in \mathbf{Z}^+}$ is an i.i.d.

17 Suppose $\mathcal{B}$ is the σ-algebra of Borel subsets of $(-\pi, \pi]$ and P is Lebesgue measure on $((-\pi, \pi], \mathcal{B})$ divided by 2π. Let $\{e_k\}_{k \in \mathbf{Z} \setminus \{0\}}$ be the family of trigonometric functions defined by the third bullet point of Example 8.51 (notice that $k = 0$ is excluded).

(a) Show that $\{e_k\}_{k \in \mathbf{Z} \setminus \{0\}}$ is not an independent family of random variables.

(b) Show that $\{e_k\}_{k \in \mathbf{Z} \setminus \{0\}}$ is an identically distributed family.