6 Continuous Random Variables and Distributions

6.A Appendix: Continuous Distributions

We learned in Chapters 3 and 4 that random variables are functions defined on a probability space. But our subsequent analyses have shown that for many purposes, it is enough to focus on the random variables' distributions and joint distributions. In this appendix and the next, we provide a calculus-based treatment of the distributions of continuous random variables, explaining how the machinery from Chapters 3 and 4 is extended to random variables whose distributions are defined by density functions. As a rule, formulas involving expectations of continuous random variables are the same as ones for discrete random variables, but with integrals replacing sums. Our focus below is on properties for which the continuous analysis differs from the discrete one in a noteworthy way.

6.A.1 Cumulative distribution functions

As we explained in Section 6.2, the distribution of any random variable X can be completely described by specifying the cumulative probabilities $\mathrm{P}(X \le x)$ for all real numbers $x \in \mathbb{R}$. We therefore define the **(cumulative) distribution function** (or **cdf**) of a random variable X to be the function $F : (-\infty, \infty) \to [0, 1]$ given by

(6.A.1) $$F(x) = \mathrm{P}(X \le x) \text{ for all } x \in \mathbb{R}.$$

A distribution function completely describes a random variable's distribution, in that one can use it and the properties of probability measures to determine the ex ante probability that X lies in any naturally occurring set of real numbers.[1] For instance, if X has distribution function F, then the probability that X takes a value in the half-open interval $(a, b]$ is

(6.A.2) $$\mathrm{P}(X \in (a, b]) = \mathrm{P}(X \le b) - \mathrm{P}(X \le a) = F(b) - F(a),$$

where the first equality follows from the additivity axiom for disjoint events.

Cumulative distribution functions are characterized by three properties.

[1] By a "naturally occurring set of real numbers," we mean a *Borel set*. Borel sets include any set one would encounter in applications. For more details, see Patrick Billingsley, *Probability and Measure*, 3rd ed., Wiley, 1995, Sections 2 and 10.

Characterization of cumulative distribution functions.

A function F is the cumulative distribution function of some random variable if and only if it satisfies the following three properties:

(*i*) F *is* nondecreasing*: $x < y$ implies that $F(x) \leq F(y)$;*
(*ii*) $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$*;*
(*iii*) F *is* continuous from the right*: for all $x \in X$,* $\lim_{y\downarrow x} F(y) = F(x)$.

It is not too hard to understand why a cumulative distribution function must satisfy properties (i)–(iii). Property (i) says that if x is less than y, then $P(X \leq x)$ is at most $P(Y \leq y)$. Property (ii) amounts to the requirement that $P(X \in (-\infty, \infty))$ equal one. Property (iii) says if y approaches x *from above*, then $P(X \leq y)$ approaches $P(X \leq x)$. Notice that distribution functions need not be continuous from below: if z approaches x from below, we can only conclude that $\lim_{z\uparrow x} P(X \leq z) \leq P(X \leq x)$; the inequality will be strict if $P(X = x)$ is positive and will be an equality if $P(X = x)$ is zero.[2]

The italicized statement above also makes a converse claim: namely, that if a function F satisfies properties (i)–(iii), then there is a random variable whose distribution function is F. To prove this result, we must construct a suitable random variable as a function on a sample space—see Exercise 6.A.6.

6.A.2 Density functions

Continuous random variables are those whose distributions can be described by (**probability**) **density functions**, which are sometimes called **pdfs** for short. We now state the definition of density functions using calculus. For concision, we use the notations $\mathbb{R} = (-\infty, \infty)$ and $\mathbb{R}_+ = [0, \infty)$ for the sets of real numbers and nonnegative real numbers, respectively.

Definition.

The random variable X has **density function** $f : \mathbb{R} \to \mathbb{R}_+$ if

$$P(X \in (a, b]) = \int_a^b f(s)\, ds \quad \text{for all } (a, b] \subset \mathbb{R}. \tag{6.A.3}$$

For this definition to make sense, the density function must satisfy the total probability condition

$$\int_{-\infty}^{\infty} f(x)\, dx = 1. \tag{6.A.4}$$

Equations (6.A.2) and (6.A.3) imply that density and distribution functions are related by

$$F(b) - F(a) = \int_a^b f(s)\, ds.$$

[2] The fact that F is continuous from the right is a consequence of our having defined distribution functions (6.A.1) using a less-than-or-equal-to sign. Proving that distribution functions satisfy conditions (ii) and (iii) requires us to use the countable additivity axiom; see Exercises 6.A.4 and 6.A.5.

When the density function f is continuous, this equation and the fundamental theorem of calculus imply that $f(x) = \frac{d}{dx}F(x)$ for all $x \in \mathbb{R}$; if f is only piecewise continuous, as in the case of uniform density functions (see below), this relation holds at all points of continuity of f.

6.A.3 Expected values

The expected value of a discrete random variable is

$$E(X) = \sum_x x\,P(X = x).$$

In words, E(X) is the weighted average outcome of X.

The expected value of a continuous random variable is defined in an analogous way, but with the sum replaced by an integral, and with probability masses on discrete outcomes replaced by densities over continuous ones:

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx. \tag{6.A.5}$$

Suppose we define a new random variable as a function of X, say $Y = g(X)$. Then the expected value of Y is

$$E(Y) = \int_{-\infty}^{\infty} g(x) f(x)\,dx. \tag{6.A.6}$$

Since the variance of a random variable is the weighted average of the squared deviations of the outcomes from their mean, we can apply Q: which one formula to obtain a formula for the variance of a continuous random variable:

$$\mathrm{Var}(X) = E\big((X - E(X))^2\big) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\,dx. \tag{6.A.7}$$

Here are the basic examples of continuous distributions from Chapters 6 and 8.

Example *Uniform distributions.*

The random variable X has a *uniform distribution with parameters* l and h (denoted $X \sim uniform(l, h)$) if its density function is constant on the interval $[l, h]$ and is zero elsewhere. That is,

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

Evaluating formulas (6.A.5) and (6.A.7) shows that $E(X) = \frac{h+l}{2}$ and $\mathrm{Var}(X) = \frac{(h-l)^2}{12}$ (see Exercise 6.M.3).

Example *Exponential distributions.*

The random variable T has an *exponential distribution with rate* $\lambda > 0$ (denoted $T \sim$ *exponential*(λ)) if its density function is

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t \geq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{6.A.8}$$

Evaluating formulas (6.A.5) and (6.A.7) shows that $\mathrm{E}(T) = \frac{1}{\lambda}$ and $\mathrm{Var}(T) = \frac{1}{\lambda^2}$ (see Exercise 8.M.3).

Example *Normal distributions.*

The random variable X has a *normal distribution with mean* μ *and variance* σ^2 (denoted $X \sim N(\mu, \sigma^2)$) if its density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}.$$

The normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is the *standard normal distribution*. Standard normal random variables are typically denoted Z, and have density function

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}. \tag{6.A.9}$$

For the calculation verifying that a random variable with density function (6.A.9) has mean 0 and variance 1, see Exercise 6.M.8. We can proceed from this fact to the conclusion that a $N(\mu, \sigma^2)$ random variable has mean μ and variance σ^2 using formula (6.A.10) below; see Exercise 6.A.2.

6.A.4 Transformations of density functions

If X is a continuous random variable and ϕ is a smooth function from the real line to itself, then $Y = \phi(X)$ is also a continuous random variable. If X has density f, how do we determine the density of Y? We first consider the case in which ϕ is a monotone function (i.e., an increasing or decreasing function).

The density function of a transformed random variable.
Let the random variable X have density f, and let $\mathcal{X} = \{x : f(x) > 0\}$. Let $Y = \phi(X)$, where $\phi : \mathbb{R} \to \mathbb{R}$ is monotone and the derivative of its inverse is continuous. Then the density of Y is defined on $\phi(\mathcal{X})$ by

$$g(y) = \left|(\phi^{-1})'(y)\right| f(\phi^{-1}(y)) \tag{6.A.10}$$

and equals 0 elsewhere.

The second term in (6.A.10) evaluates the density function f at the point $\phi^{-1}(y)$, which is the x value that the function ϕ maps to y. Since ϕ is monotone, there is only one such x value.

To understand the initial scaling term, $|(\phi^{-1})'(y)|$, imagine that $\phi'(x_0) > 0$ is large for some $x_0 \in \mathcal{X}$. Then when x is close to x_0, $y = \phi(x)$ is about $\phi'(x_0)$ times as far away from y_0 as x is from x_0. Because of this, the transformation ϕ takes the mass that density function f places near x_0 and spreads it more thinly over a wider interval around y_0. This thinning out is captured by the factor $(\phi^{-1})'(y_0)$. By a standard calculus result this factor equals $1/\phi'(x_0)$; since we assumed at the start that $\phi'(x_0)$ is large, this factor is small, as we anticipated.

Example Let $T \sim \textit{exponential}(\lambda)$, so that the density of T on $[0, \infty)$ is $f(t) = \lambda e^{-\lambda t}$. Let $U = cT$ for some $c > 0$. Then $\phi(t) = ct$, so the density of U on $[0, \infty)$ is

$$g(u) = \frac{1}{c}\, f\left(\frac{u}{c}\right) = \frac{\lambda}{c}\, e^{-(\lambda/c)u}.$$

Thus U has an $\textit{exponential}(\frac{\lambda}{c})$ distribution.

If the transformation ϕ is not monotone, then there will be multiple x values that are mapped by ϕ to a single y value. Thus to determine the density of Y at y, we must add up terms like (6.A.10) corresponding to each x value that is mapped to y. Specifically, if $\mathcal{X}$ can be partitioned into intervals $\mathcal{X}_1, \dots, \mathcal{X}_k$ on which ϕ is monotone, then the density of Y at $y \in \phi(\mathcal{X})$ is

$$g(y) = \sum_{i:\, y \in \phi(\mathcal{X}_i)} |(\phi_i^{-1})'(y)| \cdot f(\phi_i^{-1}(y)). \tag{6.A.11}$$

Here ϕ_i denotes the restriction of ϕ to $\mathcal{X}_i$, a function that is monotone by construction.

6.B Appendix: Continuous Joint Distributions

Here we discuss joint distributions of multiple continuous random variables. To keep the notation simple, we focus on the bivariate case. Moving to the case of n random variables does not introduce too many novelties, but does complicate the notation.[3]

[3] For further discussion and examples, see George Casella and Roger L. Berger, *Statistical Inference*, 2nd edition, Duxbury/Thomson, 2002, Chapter 4.

6.B.1 Joint distribution functions

Let $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$ denote the set of points in the plane. The **joint distribution function** of random variables X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ defined by

$$F(x, y) = \mathrm{P}(X \le x, Y \le y). \tag{6.B.1}$$

A distribution function completely describes the joint distribution of a pair of random variables, in that one can use it and the properties of probability measures to determine the ex ante probability that the pair (X, Y) lies in any naturally occurring set in the plane.[4] For instance, one can use the additivity axiom for disjoint events to show that

(6.B.2)

$$\mathrm{P}(X \in (a_1, b_1], Y \in (a_2, b_2]) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

(See Exercise 6.B.1).

Joint distribution functions are characterized by analogues of the three properties that characterize distribution functions of a single variable, except that these properties now must hold for each component of F separately: specifically, for the functions $F(\cdot, y) : \mathbb{R} \to [0, 1]$ for each fixed $y \in \mathbb{R}$ and the functions $F(x, \cdot) : \mathbb{R} \to [0, 1]$ for each fixed $x \in \mathbb{R}$.

6.B.2 Joint density functions

The joint distribution of a pair of continuous random variables is described by its joint density function.

Definition.

The random variables X and Y have **joint density function** $f : \mathbb{R}^2 \to \mathbb{R}_+$ if

$$\mathrm{P}\big(X \in (a_1, b_1], Y \in (a_2, b_2]\big) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y)\,\mathrm{d}x\,\mathrm{d}y \tag{6.B.3}$$

for all $(a_1, b_1], (a_2, b_2] \subset \mathbb{R}$.

For this definition to make sense, the density function must satisfy the total probability condition

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1. \tag{6.B.4}$$

Equations (6.B.1) and (6.B.3) together imply that

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\,\mathrm{d}s\,\mathrm{d}t.$$

[4] "Naturally occurring set" again means Borel set; see footnote 1.

Thus if f is continuous, the fundamental theorem of calculus implies that it is equal to the cross partial derivative of F:

$$f(x,y) = \frac{\partial^2 F}{\partial x\,\partial y}(x,y).$$

6.B.3 Marginal density functions

Suppose we know the joint density function f of the pair (X, Y) but are only interested in the properties of X and Y in isolation. This information is contained in the marginal density functions of X and Y, which are defined as follows.

Definition.
The **marginal density functions** of X and Y are

(6.B.5) $$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x.$$

It follows easily from the total probability condition (6.B.4) that f_X and f_Y are legitimate density functions, in that they satisfy the total probability condition (6.A.4). Put differently, marginal density functions are instances of the univariate density functions from Section 6.A.

6.B.4 Expected values

If the pair of random variables (X, Y) has joint density function f, then the expected values of X and Y can be computed as

$$\mathrm{E}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,\mathrm{d}x\,\mathrm{d}y \quad \text{and} \quad \mathrm{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Substituting in definition (6.B.5) allows us to write these expected values in terms of the marginal densities of X and Y:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} xf_X(x)\,\mathrm{d}x \quad \text{and} \quad \mathrm{E}(Y) = \int_{-\infty}^{\infty} yf_Y(y)\,\mathrm{d}y.$$

Of course, these equations are just instances of equation (6.A.5) from the univariate setting.

Suppose we define a new random variable that is a function of both X and Y, say $Z = g(X, Y)$. Then the expected value of Z is

(6.B.6) $$\mathrm{E}(Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

Since the distribution of Z generally depends on the comovements of X and Y, one generally cannot replace formula (6.B.6) with one stated in terms of the marginal densities f_X and f_Y.

Formula (6.B.6) provides us with an expression for the covariance of X and Y in terms of their joint density function:

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{E}\big((X - \mathrm{E}(X))(Y - E(Y))\big)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mathrm{E}(X))(y - \mathrm{E}(Y))f(x, y)\,\mathrm{d}x\,\mathrm{d}y.\end{aligned}$$

6.B.5 Conditional density, mean, and variance functions

Sometimes we are interested in the behavior of one random variable in the pair (X, Y) after having learned the realization of the other. In the context of continuous random variables, such information is provided by conditional density functions. These in turn allow us to define conditional mean functions, which play a basic role in regression analysis and econometrics (see Chapters 19 and 20).

Definitions.

Let $\mathcal{X} = \{x : f_X(x) > 0\}$ and let $\mathcal{Y} = \{y : f_Y(y) > 0\}$. For $x \in \mathcal{X}$, the **conditional density function** of Y given $X = x$ is the function $f(\cdot\,|\,x) : \mathbb{R} \to \mathbb{R}_+$ defined by

$$f(y\,|\,x) = \frac{f(x, y)}{f_X(x)}. \tag{6.B.7}$$

The **conditional mean function** $\mathrm{E}(Y\,|\,X = \cdot\,) : \mathcal{X} \to \mathbb{R}$ and the **conditional variance function** $\mathrm{Var}(Y\,|\,X = \cdot\,) : \mathcal{X} \to \mathbb{R}$ are defined by

$$\begin{aligned}\mathrm{E}(Y\,|\,X = x) &= \int_{-\infty}^{\infty} y\,f(y\,|\,x)\,\mathrm{d}y \quad \text{and}\\ \mathrm{Var}(Y\,|\,X = x) &= \int_{-\infty}^{\infty} \big(y - \mathrm{E}(Y\,|\,X = x)\big)^2 f(y\,|\,x)\,\mathrm{d}y.\end{aligned}$$

Symmetrically, the conditional density of X given $Y = y$ is the function $f(\cdot\,|\,y) : \mathbb{R} \to \mathbb{R}_+$ defined by

$$f(x\,|\,y) = \frac{f(x, y)}{f_Y(y)}.$$

The conditional mean function $\mathrm{E}(X\,|\,Y = \cdot\,) : \mathcal{Y} \to \mathbb{R}$ and the conditional variance function $\mathrm{Var}(X\,|\,Y = \cdot\,) : \mathcal{Y} \to \mathbb{R}$ are defined by

$$\begin{aligned}\mathrm{E}(X\,|\,Y = y) &= \int_{-\infty}^{\infty} x\,f(x\,|\,y)\,\mathrm{d}x \quad \text{and}\\ \mathrm{Var}(X\,|\,Y = y) &= \int_{-\infty}^{\infty} \big(x - \mathrm{E}(X\,|\,Y = y)\big)^2 f(x\,|\,y)\,\mathrm{d}x.\end{aligned}$$

Integrating (6.B.7) with respect to x shows that conditional density functions are themselves univariate density functions in that they satisfy total probability condition (6.A.4).

Intuitively, the conditional density function $f(\cdot \mid x)$ describes the distribution of the random variable Y conditional on the event that $X = x$. Notice, however, that we are here conditioning on an event with probability zero, in violation of the definition of conditional probability from Chapter 2. Thus this intuitive interpretation of conditional density should be regarded with caution.[5]

As their names indicate, the conditional expectation function $\mathrm{E}(Y \mid X = \cdot)$ and the conditional variance function $\mathrm{Var}(Y \mid X = \cdot)$ describe how the conditional expectation and conditional variance of Y vary as the realization of X changes. These functions are of fundamental importance in econometrics.

Example Suppose that the random variable pair (X, Y) has joint density function

$$f(x, y) = \begin{cases} \mathrm{e}^{-y} & \text{if } y \geq x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

To verify that this is a legitimate density function, we observe that

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x\,\mathrm{d}y &= \int_{0}^{\infty} \mathrm{e}^{-y} \int_{0}^{y} \mathrm{d}x\,\mathrm{d}y = \int_{0}^{\infty} y\mathrm{e}^{-y}\,\mathrm{d}y \\ &= -y\mathrm{e}^{-y}\Big|_{0}^{\infty} - \int_{0}^{\infty} (-\mathrm{e}^{-y})\,\mathrm{d}y = -\mathrm{e}^{-y}\Big|_{0}^{\infty} = 1. \end{aligned}$$

The marginal density of X is 0 for $x < 0$ and is

$$\int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y = \int_{x}^{\infty} \mathrm{e}^{-y}\,\mathrm{d}y = -\mathrm{e}^{-y}\Big|_{x}^{\infty} = \mathrm{e}^{-x}$$

for $x \geq 0$. In the latter case, the conditional density of Y given $X = x$ is

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)} = \begin{cases} \mathrm{e}^{-(y-x)} & \text{if } y \geq x, \\ 0 & \text{otherwise.} \end{cases} \tag{6.B.8}$$

Comparing this to (6.A.8) (or applying transformation rule (6.A.10)) shows that (6.B.8) is the density of a random variable of the form $T + x$, where T has an *exponential*(1) distribution. It follows from this or from a direct calculation that

$$\mathrm{E}(Y \mid X = x) = 1 + x \quad \text{and} \quad \mathrm{Var}(Y \mid X = x) = 1.$$

[5]Being precise about these matters requires the measure-theoretic definition of conditional probability, a definition that is a cornerstone of modern probability theory. An excellent account can be found in Section 33 of Billingsley (see footnote 1); see especially Theorem 33.3 and Examples 33.5 and 33.12.

Example *Bivariate normal distributions.*

The random variable pair (X, Y) has a **bivariate normal distribution** with parameters μ_X, μ_Y, σ_X^2, σ_Y^2, and $\rho \in (-1, 1)$ if its joint density function is

(6.B.9)
$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right).$$

For some computations, it is more convenient to express this joint density in terms of the covariance $\sigma_{X,Y} = \rho\sigma_X\sigma_Y$ rather than in terms of ρ. To do so compactly, we write $|\Sigma| = \sigma_X^2\sigma_Y^2 - \sigma_{X,Y}^2$ and rearrange (6.B.9) to obtain

(6.B.10)
$$f(x, y) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2|\Sigma|}\left(\sigma_Y^2(x-\mu_X)^2 - 2\sigma_{X,Y}(x-\mu_X)(y-\mu_Y) + \sigma_X^2(y-\mu_Y)^2\right)\right).$$

Bivariate normal random variables enjoy the following properties:

(i) The marginal distributions of X and Y are $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$.
(ii) The correlation between X and Y is ρ.
(iii) Any linear combination $aX + bY$ of X and Y has a univariate normal distribution with the appropriate mean and variance:

$$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{X,Y}).$$

(iv) The conditional distribution of Y given $X = x$ is normal.
Specifically,

$$Y|_{X=x} \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2)\right).$$

Thus the conditional mean and conditional variance functions for Y are $\mathrm{E}(Y\,|\,X = x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$ and $\mathrm{Var}(Y\,|\,X = x) = \sigma_Y^2(1-\rho^2)$.
Symmetrically, the conditional distribution of X given $Y = y$ is

$$X|_{Y=y} \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X^2(1-\rho^2)\right).$$

For derivations of these properties, see Exercise 6.B.11.

To obtain an expression for the bivariate normal density that generalizes to settings with more than two random variables requires us to use matrices. Define the **covariance matrix** of the pair (X, Y) by

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{X,Y} \\ \sigma_{X,Y} & \sigma_Y^2 \end{pmatrix}.$$

Then the determinant of Σ is $|\Sigma| = \sigma_X^2\sigma_Y^2 - \sigma_{X,Y}^2$ (as above), and the inverse of Σ is

$$\Sigma^{-1} = \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_Y^2 & -\sigma_{X,Y} \\ -\sigma_{X,Y} & \sigma_X^2 \end{pmatrix}.$$

We can thus write (6.B.10) as

$$f(x,y) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}}\,\exp\left(-\tfrac{1}{2}\begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}'\Sigma^{-1}\begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}\right).$$ ■

6.B.6 Transformations of joint density functions

Let (X,Y) be a pair of random variables with joint density f. If ϕ is a function from $\mathbb{R}^2$ to itself, then we can define a new pair of random variables U and V by setting $(U,V) = \phi(X,Y)$ (i.e., by setting $U = \phi_1(X,Y)$ and $V = \phi_2(X,Y)$). If the function ϕ is smooth, then the random variables U and V will be continuous, and, in particular, will admit a joint density function.

As in the univariate case, there is a formula that expresses the joint density of (U,V) in terms of the joint density of (X,Y). Because we are now in a multivariate setting, this formula makes use of Jacobian determinants.[6] The result below focuses on the case in which the transformation ϕ is invertible.

The joint density function of a transformed random variable pair. *Let the random variable pair (X,Y) have joint density f, and let $\mathcal{A} = \{(x,y)\colon f(x,y) > 0\}$. Let $(U,V) = \phi(X,Y)$, where $\phi\colon \mathcal{A} \to \mathbb{R}^2$ is one-to-one with range $\mathcal{B} = \{\phi(x,y)\colon (x,y) \in \mathcal{A}\}$ and with differentiable inverse. Let*

$$D\phi^{-1}(u,v) = \begin{pmatrix} \dfrac{\partial\phi_1^{-1}}{\partial u}(u,v) & \dfrac{\partial\phi_1^{-1}}{\partial v}(u,v) \\ \dfrac{\partial\phi_2^{-1}}{\partial u}(u,v) & \dfrac{\partial\phi_2^{-1}}{\partial v}(u,v) \end{pmatrix}.$$

be the Jacobian of ϕ^{-1} at (u,v). Then the joint density function of (U,V) is defined on $\mathcal{B}$ by

$$\text{(6.B.11)} \qquad \begin{aligned} g(u,v) &= \left|\det D\phi^{-1}(u,v)\right| f(\phi^{-1}(u,v)) \\ &= \left|\frac{\partial\phi_1^{-1}}{\partial u}(u,v)\frac{\partial\phi_2^{-1}}{\partial v}(u,v) - \frac{\partial\phi_1^{-1}}{\partial v}(u,v)\frac{\partial\phi_2^{-1}}{\partial u}(u,v)\right| f(\phi^{-1}(u,v)) \end{aligned}$$

and equals 0 *elsewhere.*

As in the univariate case, the second term in (6.B.11) evaluates the original density f at the (x,y) value that is mapped to (u,v). The initial scaling term also follows the pattern from the univariate case; its role is to thin out or concentrate

[6]Since the formula includes the absolute value of a determinant, we will use the notation $\det A$ for the determinant of matrix A.

probability mass to account for the expansion or contraction of neighborhoods under transformation ϕ.[7]

If the transformation ϕ is not one-to-one, then to obtain the joint density of (U, V), we must partition the domain of ϕ into regions on which ϕ is one-to-one, as in (6.A.11) above.

We present an example applying transformation formula (6.B.11) in the next section.

6.B.7 Independent random variables

In Chapter 3, we defined the random variables X and Y to be independent if their joint distribution satisfies the product formula

$$P(X \in A, Y \in B) = P(X \in A)\,P(Y \in B).$$

When X and Y are continuous random variables, this product formula is required to hold for all naturally occurring sets A and B.[8] It turns out that this product rule is equivalent to a product rule for the random variables' joint and marginal density functions.

Independence of continuous random variables.

Suppose the random variable pair (X, Y) has joint density f and marginal densities f_X and f_Y. X and Y are independent if and only if

$$f(x, y) = f_X(x) f_Y(y) \quad \textit{for all } (x, y) \in \mathbb{R}^2.$$

The following fact, verified in Exercise 6.B.5, is useful for determining whether random variables with a given joint distribution are independent.

Checking independence of continuous random variables.

Suppose the random variable pair (X, Y) has joint density f. To show that X and Y are independent, it is enough to show that their joint density can be written as

$$f(x, y) = g(x)\,h(y) \quad \textit{for all } (x, y) \in \mathbb{R}^2$$

for some functions $g : \mathbb{R} \to \mathbb{R}_+$ and $h : \mathbb{R} \to \mathbb{R}_+$.

[7]In more detail: The absolute value of the determinant of a matrix $A \in \mathbb{R}^{2\times 2}$ describes how much area is expanded (for $|\det A| > 1$) or contracted (for $|\det A| < 1$) when A is used as a linear map from the plane $\mathbb{R}^2$ to itself. By Taylor's theorem, the Jacobian $D\phi(x, y)$ provides a linear approximation to the action of the function ϕ in a neighborhood of point (x, y). Thus $|\det D\phi(x, y)|$ describes the degree to which the transformation ϕ expands or contracts area when it maps a neighborhood of (x, y) to a larger or smaller neighborhood of $(u, v) = \phi(x, y)$. Since $D\phi^{-1}(u, v) = (D\phi(x, y))^{-1}$ (i.e., since the Jacobian matrix of the inverse function is the inverse of the Jacobian matrix) and since $\det A^{-1} = 1/(\det A)$, we have $|\det D\phi^{-1}(u, v)| = 1/|\det D\phi(x, y)|$. Thus the second term in (6.B.11) dilutes (if $|\det D\phi(x, y)| > 1$, so that $|\det D\phi^{-1}(u, v)| < 1$) or concentrates (if $|\det D\phi(x, y)| < 1$, so that $|\det D\phi^{-1}(u, v)| > 1$) probability mass in a way that counterbalances the expansion or contraction of area under ϕ in the vicinity of (x, y).

[8]This again means the Borel subsets; see footnote 1.

Example Continuing an example from the previous section, suppose that the random variable pair (X, Y) has joint density function

$$f(x, y) = \begin{cases} e^{-y} & \text{if } y \geq x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let $U = X$ and $V = Y - X$. What is the joint density function of (U, V)?

Here $\mathcal{A} = \{(x, y) \colon f(x, y) > 0\} = \{(x, y) \colon y \geq x \geq 0\}$ and

$$\phi(x, y) = \begin{pmatrix} x \\ y - x \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

and so $\mathcal{B} = \{\phi(x, y) \colon (x, y) \in \mathcal{A}\} = \{(u, v) \colon u, v \geq 0\} \equiv \mathbb{R}^2_+$. In addition, we have

$$\phi^{-1}(u, v) = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u \\ u + v \end{pmatrix}, \quad \text{so}$$

$$D\phi^{-1}(u, v) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and } \det D\phi^{-1}(u, v) = 1.$$

Thus the transformation formula (6.B.11) shows that the joint density of (U, V) is defined on $\mathbb{R}^2_+$ by

$$g(u, v) = 1 \times f(\phi^{-1}(u, v)) = e^{-(u+v)} = e^{-u}e^{-v}$$

and equals 0 elsewhere. We conclude that U and V are independent random variables, each with an *exponential*(1) distribution.

In addition, considering the inverse transformation from (U, V) to (X, Y) reveals that the latter pair is comprised of an *exponential*(1) random variable X and the sum of X and an independent *exponential*(1) random variable. ■

Example *Independence of bivariate normal random variables.*

We emphasized in Chapter 4 that zero correlation does not imply independence. But there is an important special case in which this implication holds.

Suppose that (X, Y) is a pair of random variables with a bivariate normal random distribution, and suppose that X and Y are uncorrelated. Then according to equation (6.B.9), the joint density function of X and Y is

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{1}{2}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right) \\ &= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \exp\left(-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right). \end{aligned}$$

Thus the previous result implies that X and Y are independent. We could also have derived this fact by examining the formulas for the conditional distributions of bivariate normal random variables: if $\rho = 0$, then $Y|_{X=x}$ has a $N(\mu_Y, \sigma_Y^2)$ distribution regardless of the value of x.

The hypothesis that X and Y follow a *bivariate* normal distribution is important here. If all we know is that X and Y are each normally distributed, then zero correlation would not imply independence. It is a good exercise to construct a counterexample that illustrates this point. ■

All of the properties of independent random variables described in Chapters 3 and 4 remain true in the continuous case. In particular, functions of independent random variables are independent random variables:

If X and Y are independent, then $f(X)$ and $g(Y)$ are independent.

Also, the expectation of the product of independent random variables equals the product of their expectations:

If X and Y are independent, then $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

These properties of pairs of independent random variables generalize directly to collections of many independent random variables.

6.B.8 Sums of independent random variables

As we have seen, sums of independent random variables are used in a wide range of applications of probability theory. If X and Y are independent discrete random variables, the distribution of their sum $S = X + Y$ is described by

$$\text{(6.B.12)} \qquad \mathrm{P}(S = s) = \sum_x \mathrm{P}(X = x, Y = s - x) = \sum_x \mathrm{P}(X = x)\mathrm{P}(Y = s - x).$$

For each outcome s, the first equality aggregates the probabilities of the outcome pairs $(x, y) = (x, s - x)$ whose components add up to s. The second equality follows from the product rule for independent random variables.

When the random variables being added have densities, there is an expression analogous to (6.B.12) for the density of the sum.

The density of the sum of two independent random variables.

Let X and Y be independent random variables with densities f_X and f_Y. Then the sum $S = X + Y$ has density

$$\text{(6.B.13)} \qquad f_S(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s - x)\, dx.$$

This formula is a consequence of the transformation formula (6.B.11): see Exercise 6.B.9. Away from probability theory, the function f_S is called the **convolution** of the functions f_X and f_Y.

Example Let X and Y be independent normal random variables with mean 0 and variances σ_X^2 and σ_Y^2. By the convolution formula, the density function of the sum $X + Y$ is

$$f_S(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_X} \exp\left(\frac{-x^2}{2\sigma_Y^2}\right) \cdot \frac{1}{\sqrt{2\pi}\,\sigma_Y} \exp\left(\frac{-(s-x)^2}{2\sigma_Y^2}\right) \mathrm{d}x.$$

Evaluating this integral yields

$$f_S(s) = \frac{1}{\sqrt{2\pi\,(\sigma_X^2+\sigma_Y^2)}} \exp\left(\frac{-s^2}{2(\sigma_X^2+\sigma_Y^2)}\right)$$

(see Exercise 6.B.10). Thus S has a normal distribution with mean 0 and variance $\sigma_X^2 + \sigma_Y^2$, as stated in Section 6.6.1. We will see that this fact can be derived more easily using tools introduced in Appendix 7.A. ■

The fact that the convolution formula (6.B.13) contains a possibly difficult integral can make it cumbersome to work with, and in fact the discrete formula (6.B.12) can be just as much trouble. In Appendix 7.A, we introduce machinery that can make working with the distributions of sums of independent random variables strikingly simple.

KEY TERMS AND CONCEPTS

(cumulative) distribution function (cdf) (p. 1)
(probability) density function (pdf) (p. 2)
transformations of density functions (p. 4)
joint distribution function (p. 6)
joint density function (p. 6)
marginal density functions (p. 7)
conditional density function (p. 8)
conditional mean function (p. 8)
conditional variance function (p. 8)
bivariate normal distribution (p. 10)
covariance matrix (p. 10)
transformations of joint density functions (p. 11)
independent random variables (p. 12)
convolution (p. 15)
Box-Muller transform (p. 18)

Exercises

Appendix 6.A exercises

Exercise 6.A.1. Let X be a continuous random variable. Use definitions (6.A.5) and (6.A.7) and basic properties of integrals to prove that $\mathrm{E}(aX+b) = a\mathrm{E}(X)+b$ and that $\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$.

Exercise 6.A.2. Let Z be a standard normal random variable, so that its density function is $f(z) = \mathrm{e}^{-z^2/2}/\sqrt{2\pi}$. Exercise 6.M.8 shows that $\mathrm{E}(Z) = 0$ and that $\mathrm{Var}(Z) = 1$.

Let $X = \sigma Z + \mu$ for some real number μ and some $\sigma > 0$.

a. Use Exercise 6.A.1 to show that $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$.
b. Use transformation formula (6.A.10) to show that X has a $N(\mu, \sigma^2)$ distribution.

Together, parts (a) and (b) establish that a $N(\mu, \sigma^2)$ random variable really does have mean μ and variance σ^2.

Exercise 6.A.3. Let $X \sim$ *uniform*$(0, 1)$, and let $Y = -\frac{1}{\lambda}\ln X$.

a. Use transformation formula (6.A.10) to show that $Y \sim$ *exponential*(λ).
b. Show that $Y \sim$ *exponential*(λ) directly by computing its cumulative distribution function.
c. Now suppose that $Y \sim$ *exponential*(λ). What transformation of Y has a *uniform*$(0, 1)$ distribution?

Exercise 6.A.4. Recall the countable additivity axiom:

(A3*)
$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i) \text{ for any sequence of disjoint events } \{A_1, A_2, A_3, \ldots\}.$$

a. Use axiom (A3*) to show that probability measures are *continuous from below*: If the events $\{A_i\}_{i=1}^{\infty}$ satisfy $A_1 \subseteq A_2 \subseteq \cdots$ and $\cup_{i=1}^{\infty} A_i = A$, then $\lim_{i\to\infty} \mathrm{P}(A_i) = \mathrm{P}(A)$.
b. Show that probability measures are *continuous from above*: If the events $\{B_i\}_{i=1}^{\infty}$ satisfy $B_1 \supseteq B_2 \supseteq \cdots$ and $\cap_{i=1}^{\infty} B_i = B$, then $\lim_{i\to\infty} \mathrm{P}(B_i) = \mathrm{P}(B)$. (Hint: Use De Morgan's law ($\bigcap_i A_i^C = \left(\bigcup_i A_i\right)^C$; compare Exercise 2.2.7) part (a), and the complement rule.)

Exercise 6.A.5. Let F be the distribution function of some random variable X. In this exercise, we use the consequences of countable additivity derived in Exercise 6.A.4 to show that F must satisfy the properties listed in Section 6.A.1.

a. Use both parts of Exercise 6.A.4 to show that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
b. Use Exercise 6.A.4(b) to show that F is *continuous from the right*: for all $x \in X$, $\lim_{y\downarrow x} F(y) = F(x)$.

Exercise 6.A.6. Let F be a function that satisfies properties (i)-(iii) from Section 6.A.1. Let $S = (0, 1]$ and let P be the probability measure on S that agrees with length, in that $\mathrm{P}((a, b]) = b - a$ for $(a, b] \subseteq (0, 1]$. Define the random variable X by $X(s) = \sup\{y : F(y) < s\}$. This random variable has distribution function F. This follows immediately from the fact that $\{s \in S : X(s) \leq x\} = \{s \in S : s \leq F(x)\}$. Prove this claim. (Hint: One direction of inclusion is straightforward; the other becomes straightforward after an appeal to the right continuity of F.)

Appendix 6.B exercises

Exercise 6.B.1. Use the definition of the joint distribution function, $F(x, y) = \mathrm{P}(X \leq x, Y \leq y)$, and the additivity axiom for disjoint events to show that

(6.B.2)

$$\mathrm{P}(a_1 < X \leq b_1, a_2 < Y \leq b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2).$$

Exercise 6.B.2. A student goes through the following two-step procedure. First, he spins a *uniform*$(0, 2)$ spinner. Second, if he obtains outcome x from this first spin, he then spins a *uniform*$(0, x^2)$ spinner. Let the random variables X and Y represent the outcomes of the first and second spins, respectively.

a. What is the conditional density function of X given $Y = y$?
b. What is the conditional mean function $\mathrm{E}(X \mid Y = \cdot)$?

Exercise 6.B.3. Let the pair of random variables (X, Y) have joint density function

$$f(x, y) = \begin{cases} 6(x - y)^2 & \text{if } x, y \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

a. Confirm that f is a joint density function by verifying that equation (6.B.4) holds, and use a computer or graphing calculator to sketch its graph.
b. Compute the marginal density function of Y.
c. For each $x \in [0, 1]$, compute the conditional density of Y given x.
d. Compute the conditional expectation function $\mathrm{E}(Y \mid X = x)$, and use a computer or graphing calculator to sketch its graph.
e. As x increases, the areas under the vertical cross sections $f(x, \cdot)$ go from being concentrated on high values of y to being concentrated on low values of y. Explain why this seems to contradict your answer to part (d). Then explain why there is actually no contradiction. (Hint: Sketch the graph of f_X.)

Exercise 6.B.4. Suppose that X and Y are independent standard normal random variables, and define the random variables U and V by $U = aX + bY$ and $V = cX + dY$.

a. Use the trait formulas from Chapters 3 and 4 to compute $E(U)$, $E(V)$, $\mathrm{Var}(U)$, $\mathrm{Var}(V)$, and $\mathrm{Cov}(U, V)$.
b. Show that $|\,\mathrm{Corr}(U, V)| = 1$ if and only if $ad = bc$. (Hint: Use the identity $(ac + bd)^2 + (ad - bc)^2 = (a^2 + b^2)(c^2 + d^2)$.)
c. Suppose that $ad \neq bc$. Using the transformation formula (6.B.11) and the identity from part (b) to compute the joint density of (U, V), show that this pair has a bivariate normal distribution with $E(U) = E(V) = 0$ and with the variances and covariance you found in part (a).

Exercise 6.B.5. Show that if the joint density of X and Y can be written as $f(x, y) = g(x)h(y)$, then X and Y are independent, meaning that $f(x, y) = f_X(x)f_Y(y)$. (Hint: Show that $\left(\int_{-\infty}^{\infty} g(x)\,dx\right)\left(\int_{-\infty}^{\infty} h(y)\,dy\right) = 1$.)

Exercise 6.B.6. Let X and Y be independent *uniform*$(0, 1)$ random variables. Use the convolution formula (6.B.13) to compute the density of $X + Y$.

Exercise 6.B.7. Let X and Y be independent *uniform*$(0, 1)$ random variables. Let $(U, V) = \phi(X, Y)$, where $\phi_1(x, y) = \sqrt{-2 \ln x}\, \cos(2\pi y)$ and $\phi_2(x, y) = \sqrt{-2 \ln x}\, \sin(2\pi y)$. Use the transformation formula (6.B.11) to show that U and V are independent standard normal random variables. The function ϕ is known as the **Box-Muller transform**. (Hints: To invert ϕ, use the identity $\cos^2\theta + \sin^2\theta = 1$. To compute the partial derivatives of ϕ^{-1}, recall that $\frac{d}{dz}\tan^{-1}(z) = 1/(1 + z^2)$.)

Exercise 6.B.8. Continuing the second example from Section 6.B.5, suppose that the random variable pair (X, Y) has joint density function

$$f(x, y) = \begin{cases} e^{-y} & \text{if } y \geq x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

a. Let $y > 0$. What is the conditional distribution of X given $Y = y$?
b. What are the conditional expectation function and the conditional variance function of X?

Exercise 6.B.9. Use the transformation formula (6.B.11) to derive the convolution formula (6.B.13). (Hint: Use the transformation $\phi(x, y) = \binom{x}{x+y}$.)

Exercise 6.B.10. Let $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ be independent random variables. Use the convolution formula (6.B.13) to show that $X + Y \sim N(0, \sigma_X^2 + \sigma_Y^2)$. (Hint: By completing the square, show that the function obtained from the convolution formula reduces to the product of the $N(0, \sigma_X^2 + \sigma_Y^2)$ density and an integral that evaluates to 1.)

Exercise 6.B.11. This exercise derives the properties of the bivariate normal distribution stated in Section 6.B.5. Let Z_1 and Z_2 be independent standard normal random variables, and define X_1 and X_2 by

$$X_1 = \sigma_1 Z_1 + \mu_1 \quad \text{and}$$
$$X_2 = \sigma_2\left(\rho Z_1 + \sqrt{1-\rho^2}\, Z_2\right) + \mu_2,$$

where $\sigma_1, \sigma_2 > 0$ and $\rho \in (-1, 1)$.

a. Use the transformation formula (6.B.11) to show that the joint density function of (X_1, X_2) is the bivariate normal density (6.B.9).
b. Show that $\text{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$, and hence that $\text{Corr}(X_1, X_2) = \rho$.
c. Use Exercise 6.B.10 to show that $aX_1 + bX_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_1\sigma_2\rho)$. It follows immediately that X_1 and X_2 have marginal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.
d. Show that the conditional distribution of X_2 given that $X_1 = x_1$ is the $N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$ distribution.

Exercise 6.B.12. Suppose that the random variable X has a $N(\mu, \sigma^2)$, and that conditional on $X = x$, the random variable Y has a normal distribution with mean ax and variance τ^2, where $a, \tau > 0$. Use the definition of conditional density to show that (X, Y) has a bivariate normal distribution, and to determine $\text{E}(Y)$, $\text{Var}(Y)$, and $\text{Cov}(X, Y)$.

Exercise 6.B.13. As in Exercise 6.B.12, suppose that the random variable X has a $N(\mu, \sigma^2)$, and that conditional on $X = x$, the random variable Y has a normal distribution with mean ax and variance τ^2, where $a, \tau > 0$. Compute $\text{E}(Y)$, $\text{Var}(Y)$, and $\text{Cov}(X, Y)$ using the law of iterated expectation (4.22), the decomposition of variance formula (4.24), and the decomposition of covariance formula (4.25).