2

Random Variables

2.1 Introduction

Statistics and data mining are concerned with data. How do we link sample spaces and events to data? The link is provided by the concept of a random variable.

2.1 Definition. *A* **random variable** *is a mapping*[1]

$$X : \Omega \to \mathbb{R}$$

that assigns a real number $X(\omega)$ *to each outcome* ω*.*

At a certain point in most probability courses, the sample space is rarely mentioned anymore and we work directly with random variables. But you should keep in mind that the sample space is really there, lurking in the background.

2.2 Example. Flip a coin ten times. Let $X(\omega)$ be the number of heads in the sequence ω. For example, if $\omega = HHTHHTHHTT$, then $X(\omega) = 6$. ■

[1] Technically, a random variable must be measurable. See the appendix for details.

2.3 Example. Let $\Omega = \Big\{(x,y);\ x^2+y^2 \le 1\Big\}$ be the unit disk. Consider drawing a point at random from Ω. (We will make this idea more precise later.) A typical outcome is of the form $\omega = (x,y)$. Some examples of random variables are $X(\omega) = x$, $Y(\omega) = y$, $Z(\omega) = x+y$, and $W(\omega) = \sqrt{x^2+y^2}$. ■

Given a random variable X and a subset A of the real line, define $X^{-1}(A) = \{\omega \in \Omega :\ X(\omega) \in A\}$ and let

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega;\ X(\omega) \in A\}) \\ \mathbb{P}(X = x) &= \mathbb{P}(X^{-1}(x)) = \mathbb{P}(\{\omega \in \Omega;\ X(\omega) = x\}). \end{aligned}$$

Notice that X denotes the random variable and x denotes a particular value of X.

2.4 Example. Flip a coin twice and let X be the number of heads. Then, $\mathbb{P}(X = 0) = \mathbb{P}(\{TT\}) = 1/4$, $\mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = 1/2$ and $\mathbb{P}(X = 2) = \mathbb{P}(\{HH\}) = 1/4$. The random variable and its distribution can be summarized as follows:

ω	$\mathbb{P}(\{\omega\})$	$X(\omega)$
TT	1/4	0
TH	1/4	1
HT	1/4	1
HH	1/4	2

x	$\mathbb{P}(X = x)$
0	1/4
1	1/2
2	1/4

Try generalizing this to n flips. ■

2.2 Distribution Functions and Probability Functions

Given a random variable X, we define the cumulative distribution function (or distribution function) as follows.

2.5 Definition. *The* **cumulative distribution function***, or* CDF*, is the function* $F_X : \mathbb{R} \to [0,1]$ *defined by*

$$F_X(x) = \mathbb{P}(X \le x). \tag{2.1}$$

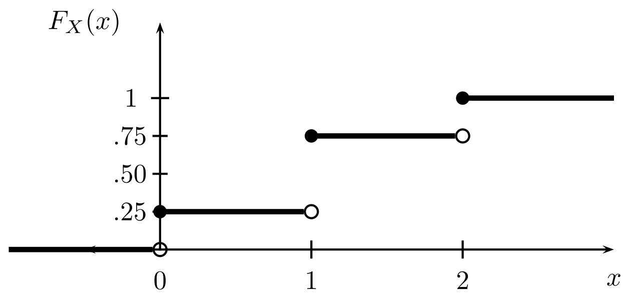


FIGURE 2.1. CDF for flipping a coin twice (Example 2.6.)

We will see later that the CDF effectively contains all the information about the random variable. Sometimes we write the CDF as F instead of F_X.

2.6 Example. Flip a fair coin twice and let X be the number of heads. Then $\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = 1/4$ and $\mathbb{P}(X = 1) = 1/2$. The distribution function is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

The CDF is shown in Figure 2.1. Although this example is simple, study it carefully. CDF's can be very confusing. Notice that the function is right continuous, non-decreasing, and that it is defined for all x, even though the random variable only takes values $0, 1$, and 2. Do you see why $F_X(1.4) = .75$? ∎

The following result shows that the CDF completely determines the distribution of a random variable.

2.7 Theorem. *Let X have CDF F and let Y have CDF G. If $F(x) = G(x)$ for all x, then $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ for all A.* [2]

2.8 Theorem. *A function F mapping the real line to $[0, 1]$ is a CDF for some probability $\mathbb{P}$ if and only if F satisfies the following three conditions:*

(i) F is non-decreasing: $x_1 < x_2$ implies that $F(x_1) \leq F(x_2)$.

(ii) F is normalized:

$$\lim_{x \to -\infty} F(x) = 0$$

[2] Technically, we only have that $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ for every measurable event A.

and

$$\lim_{x \to \infty} F(x) = 1.$$

(iii) F is right-continuous: $F(x) = F(x^+)$ for all x, where

$$F(x^+) = \lim_{\substack{y \to x \\ y > x}} F(y).$$

Proof. Suppose that F is a CDF. Let us show that (iii) holds. Let x be a real number and let $y_1, y_2, \ldots$ be a sequence of real numbers such that $y_1 > y_2 > \cdots$ and $\lim_i y_i = x$. Let $A_i = (-\infty, y_i]$ and let $A = (-\infty, x]$. Note that $A = \bigcap_{i=1}^{\infty} A_i$ and also note that $A_1 \supset A_2 \supset \cdots$. Because the events are monotone, $\lim_i \mathbb{P}(A_i) = \mathbb{P}(\bigcap_i A_i)$. Thus,

$$F(x) = \mathbb{P}(A) = \mathbb{P}\left(\bigcap_i A_i\right) = \lim_i \mathbb{P}(A_i) = \lim_i F(y_i) = F(x^+).$$

Showing (i) and (ii) is similar. Proving the other direction — namely, that if F satisfies (i), (ii), and (iii) then it is a CDF for some random variable — uses some deep tools in analysis. ■

2.9 Definition. *X is* **discrete** *if it takes countably[3] many values $\{x_1, x_2, \ldots\}$. We define the* **probability function** *or* **probability mass function** *for X by $f_X(x) = \mathbb{P}(X = x)$.*

Thus, $f_X(x) \geq 0$ for all $x \in \mathbb{R}$ and $\sum_i f_X(x_i) = 1$. Sometimes we write f instead of f_X. The CDF of X is related to f_X by

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i).$$

2.10 Example. The probability function for Example 2.6 is

$$f_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 2.2. ■

[3] A set is countable if it is finite or it can be put in a one-to-one correspondence with the integers. The even numbers, the odd numbers, and the rationals are countable; the set of real numbers between 0 and 1 is not countable.

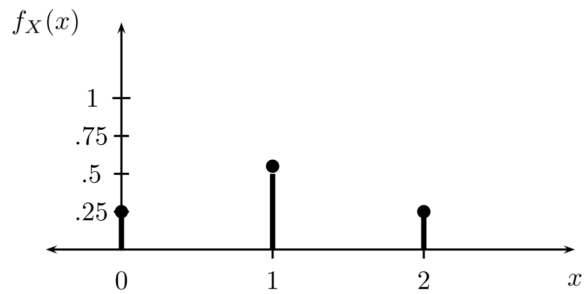


FIGURE 2.2. Probability function for flipping a coin twice (Example 2.6).

2.11 Definition. *A random variable X is* **continuous** *if there exists a function f_X such that $f_X(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f_X(x)dx = 1$ and for every $a \leq b$,*

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx. \tag{2.2}$$

The function f_X is called the **probability density function (PDF).** *We have that*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

and $f_X(x) = F'_X(x)$ at all points x at which F_X is differentiable.

Sometimes we write $\int f(x)dx$ or $\int f$ to mean $\int_{-\infty}^{\infty} f(x)dx$.

2.12 Example. Suppose that X has PDF

$$f_X(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $f_X(x) \geq 0$ and $\int f_X(x)dx = 1$. A random variable with this density is said to have a Uniform (0,1) distribution. This is meant to capture the idea of choosing a point at random between 0 and 1. The CDF is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

See Figure 2.3. ∎

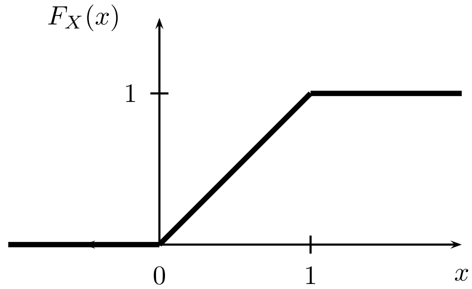


FIGURE 2.3. CDF for Uniform (0,1).

2.13 Example. Suppose that X has PDF

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{(1+x)^2} & \text{otherwise.} \end{cases}$$

Since $\int f(x)dx = 1$, this is a well-defined PDF. ■

Warning! Continuous random variables can lead to confusion. First, note that if X is continuous then $\mathbb{P}(X = x) = 0$ for every x. Don't try to think of $f(x)$ as $\mathbb{P}(X = x)$. This only holds for discrete random variables. We get probabilities from a PDF by integrating. A PDF can be bigger than 1 (unlike a mass function). For example, if $f(x) = 5$ for $x \in [0, 1/5]$ and 0 otherwise, then $f(x) \geq 0$ and $\int f(x)dx = 1$ so this is a well-defined PDF even though $f(x) = 5$ in some places. In fact, a PDF can be unbounded. For example, if $f(x) = (2/3)x^{-1/3}$ for $0 < x < 1$ and $f(x) = 0$ otherwise, then $\int f(x)dx = 1$ even though f is not bounded.

2.14 Example. Let

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{(1+x)} & \text{otherwise.} \end{cases}$$

This is not a PDF since $\int f(x)dx = \int_0^\infty dx/(1+x) = \int_1^\infty du/u = \log(\infty) = \infty$. ■

2.15 Lemma. *Let F be the CDF for a random variable X. Then:*

1. $\mathbb{P}(X = x) = F(x) - F(x^-)$ *where* $F(x^-) = \lim_{y \uparrow x} F(y)$*;*

2. $\mathbb{P}(x < X \leq y) = F(y) - F(x)$*;*

3. $\mathbb{P}(X > x) = 1 - F(x)$*;*

4. If X *is continuous then*

$$\begin{aligned} F(b) - F(a) &= \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) \\ &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b). \end{aligned}$$

It is also useful to define the inverse CDF (or quantile function).

2.16 Definition. *Let* X *be a random variable with* CDF F*. The* **inverse CDF** *or* **quantile function** *is defined by*[4]

$$F^{-1}(q) = \inf\left\{x : \; F(x) > q\right\}$$

for $q \in [0,1]$*. If* F *is strictly increasing and continuous then* $F^{-1}(q)$ *is the unique real number* x *such that* $F(x) = q$*.*

We call $F^{-1}(1/4)$ the **first quartile**, $F^{-1}(1/2)$ the **median** (or second quartile), and $F^{-1}(3/4)$ the **third quartile**.

Two random variables X and Y are **equal in distribution** — written $X \stackrel{d}{=} Y$ — if $F_X(x) = F_Y(x)$ for all x. This does not mean that X and Y are equal. Rather, it means that all probability statements about X and Y will be the same. For example, suppose that $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$. Let $Y = -X$. Then $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = 1/2$ and so $X \stackrel{d}{=} Y$. But X and Y are not equal. In fact, $\mathbb{P}(X = Y) = 0$.

2.3 Some Important Discrete Random Variables

Warning About Notation! It is traditional to write $X \sim F$ to indicate that X has distribution F. This is unfortunate notation since the symbol $\sim$ is also used to denote an approximation. The notation $X \sim F$ is so pervasive that we are stuck with it. Read $X \sim F$ as "X has distribution F" **not** as "X is approximately F".

[4]If you are unfamiliar with "inf", just think of it as the minimum.

The Point Mass Distribution. X has a point mass distribution at a, written $X \sim \delta_a$, if $\mathbb{P}(X = a) = 1$ in which case

$$F(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a. \end{cases}$$

The probability mass function is $f(x) = 1$ for $x = a$ and 0 otherwise.

The Discrete Uniform Distribution. Let $k > 1$ be a given integer. Suppose that X has probability mass function given by

$$f(x) = \begin{cases} 1/k & \text{for } x = 1, \ldots, k \\ 0 & \text{otherwise.} \end{cases}$$

We say that X has a uniform distribution on $\{1, \ldots, k\}$.

The Bernoulli Distribution. Let X represent a binary coin flip. Then $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ for some $p \in [0, 1]$. We say that X has a Bernoulli distribution written $X \sim \text{Bernoulli}(p)$. The probability function is $f(x) = p^x(1-p)^{1-x}$ for $x \in \{0, 1\}$.

The Binomial Distribution. Suppose we have a coin which falls heads up with probability p for some $0 \leq p \leq 1$. Flip the coin n times and let X be the number of heads. Assume that the tosses are independent. Let $f(x) = \mathbb{P}(X = x)$ be the mass function. It can be shown that

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{for } x = 0, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

A random variable with this mass function is called a Binomial random variable and we write $X \sim \text{Binomial}(n, p)$. If $X_1 \sim \text{Binomial}(n_1, p)$ and $X_2 \sim \text{Binomial}(n_2, p)$ then $X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$.

Warning! Let us take this opportunity to prevent some confusion. X is a random variable; x denotes a particular value of the random variable; n and p are **parameters**, that is, fixed real numbers. The parameter p is usually unknown and must be estimated from data; that's what statistical inference is all about. In most statistical models, there are random variables and parameters: don't confuse them.

The Geometric Distribution. X has a geometric distribution with parameter $p \in (0, 1)$, written $X \sim \text{Geom}(p)$, if

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

We have that

$$\sum_{k=1}^{\infty} \mathbb{P}(X=k) = p\sum_{k=1}^{\infty}(1-p)^k = \frac{p}{1-(1-p)} = 1.$$

Think of X as the number of flips needed until the first head when flipping a coin.

THE POISSON DISTRIBUTION. X has a Poisson distribution with parameter λ, written $X \sim \text{Poisson}(\lambda)$ if

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad x \geq 0.$$

Note that

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

The Poisson is often used as a model for counts of rare events like radioactive decay and traffic accidents. If $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Warning! We defined random variables to be mappings from a sample space Ω to $\mathbb{R}$ but we did not mention the sample space in any of the distributions above. As I mentioned earlier, the sample space often "disappears" but it is really there in the background. Let's construct a sample space explicitly for a Bernoulli random variable. Let $\Omega = [0,1]$ and define $\mathbb{P}$ to satisfy $\mathbb{P}([a,b]) = b - a$ for $0 \leq a \leq b \leq 1$. Fix $p \in [0,1]$ and define

$$X(\omega) = \begin{cases} 1 & \omega \leq p \\ 0 & \omega > p. \end{cases}$$

Then $\mathbb{P}(X=1) = \mathbb{P}(\omega \leq p) = \mathbb{P}([0,p]) = p$ and $\mathbb{P}(X=0) = 1-p$. Thus, $X \sim \text{Bernoulli}(p)$. We could do this for all the distributions defined above. In practice, we think of a random variable like a random number but formally it is a mapping defined on some sample space.

2.4 Some Important Continuous Random Variables

THE UNIFORM DISTRIBUTION. X has a Uniform(a,b) distribution, written $X \sim \text{Uniform}(a,b)$, if

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

where $a < b$. The distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b. \end{cases}$$

Normal (Gaussian). X has a Normal (or Gaussian) distribution with parameters μ and σ, denoted by $X \sim N(\mu, \sigma^2)$, if

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}, \quad x \in \mathbb{R} \tag{2.3}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$. The parameter μ is the "center" (or mean) of the distribution and σ is the "spread" (or standard deviation) of the distribution. (The mean and standard deviation will be formally defined in the next chapter.) The Normal plays an important role in probability and statistics. Many phenomena in nature have approximately Normal distributions. Later, we shall study the Central Limit Theorem which says that the distribution of a sum of random variables can be approximated by a Normal distribution.

We say that X has a **standard Normal distribution** if $\mu = 0$ and $\sigma = 1$. Tradition dictates that a standard Normal random variable is denoted by Z. The PDF and CDF of a standard Normal are denoted by $\phi(z)$ and $\Phi(z)$. The PDF is plotted in Figure 2.4. There is no closed-form expression for Φ. Here are some useful facts:

(i) If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

(ii) If $Z \sim N(0, 1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.

(iii) If $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$ are independent, then

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2 \right).$$

It follows from (i) that if $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma} \right) \\ &= \Phi\left(\frac{b-\mu}{\sigma} \right) - \Phi\left(\frac{a-\mu}{\sigma} \right). \end{aligned}$$

Thus we can compute any probabilities we want as long as we can compute the CDF $\Phi(z)$ of a standard Normal. All statistical computing packages will

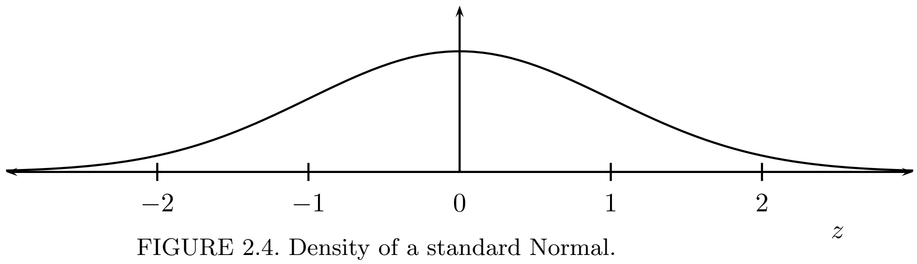


FIGURE 2.4. Density of a standard Normal.

compute $\Phi(z)$ and $\Phi^{-1}(q)$. Most statistics texts, including this one, have a table of values of $\Phi(z)$.

2.17 Example. Suppose that $X \sim N(3, 5)$. Find $\mathbb{P}(X > 1)$. The solution is

$$\mathbb{P}(X > 1) = 1 - \mathbb{P}(X < 1) = 1 - \mathbb{P}\left(Z < \frac{1-3}{\sqrt{5}}\right) = 1 - \Phi(-0.8944) = 0.81.$$

Now find $q = \Phi^{-1}(0.2)$. This means we have to find q such that $\mathbb{P}(X < q) = 0.2$. We solve this by writing

$$0.2 = \mathbb{P}(X < q) = \mathbb{P}\left(Z < \frac{q-\mu}{\sigma}\right) = \Phi\left(\frac{q-\mu}{\sigma}\right).$$

From the Normal table, $\Phi(-0.8416) = 0.2$. Therefore,

$$-0.8416 = \frac{q-\mu}{\sigma} = \frac{q-3}{\sqrt{5}}$$

and hence $q = 3 - 0.8416\sqrt{5} = 1.1181$. ■

Exponential Distribution. X has an Exponential distribution with parameter β, denoted by $X \sim \text{Exp}(\beta)$, if

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

where $\beta > 0$. The exponential distribution is used to model the lifetimes of electronic components and the waiting times between rare events.

Gamma Distribution. For $\alpha > 0$, the **Gamma function** is defined by $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$. X has a Gamma distribution with parameters α and

β, denoted by $X \sim \text{Gamma}(\alpha, \beta)$, if

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0$$

where $\alpha, \beta > 0$. The exponential distribution is just a Gamma$(1, \beta)$ distribution. If $X_i \sim \text{Gamma}(\alpha_i, \beta)$ are independent, then $\sum_{i=1}^n X_i \sim \text{Gamma}(\sum_{i=1}^n \alpha_i, \beta)$.

The Beta Distribution. X has a Beta distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted by $X \sim \text{Beta}(\alpha, \beta)$, if

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1.$$

t and Cauchy Distribution. X has a t distribution with ν degrees of freedom — written $X \sim t_\nu$ — if

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{(\nu+1)/2}}.$$

The t distribution is similar to a Normal but it has thicker tails. In fact, the Normal corresponds to a t with $\nu = \infty$. The Cauchy distribution is a special case of the t distribution corresponding to $\nu = 1$. The density is

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

To see that this is indeed a density:

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d\tan^{-1}(x)}{dx} \\ &= \frac{1}{\pi}\left[\tan^{-1}(\infty) - \tan^{-1}(-\infty)\right] = \frac{1}{\pi}\left[\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)\right] = 1. \end{aligned}$$

The χ^2 distribution. X has a χ^2 distribution with p degrees of freedom — written $X \sim \chi^2_p$ — if

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad x > 0.$$

If $Z_1, \ldots, Z_p$ are independent standard Normal random variables then $\sum_{i=1}^p Z_i^2 \sim \chi^2_p$.

2.5 Bivariate Distributions

Given a pair of discrete random variables X and Y, define the **joint mass function** by $f(x, y) = \mathbb{P}(X = x \text{ and } Y = y)$. From now on, we write $\mathbb{P}(X = x \text{ and } Y = y)$ as $\mathbb{P}(X = x, Y = y)$. We write f as $f_{X,Y}$ when we want to be more explicit.

2.18 Example. Here is a bivariate distribution for two random variables X and Y each taking values 0 or 1:

	$Y = 0$	$Y = 1$	
X=0	1/9	2/9	1/3
X=1	2/9	4/9	2/3
	1/3	2/3	1

Thus, $f(1, 1) = \mathbb{P}(X = 1, Y = 1) = 4/9$. ■

2.19 Definition. *In the continuous case, we call a function $f(x, y)$ a* PDF *for the random variables (X, Y) if*

(i) $f(x, y) \geq 0$ for all (x, y),

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$ and,

(iii) for any set $A \subset \mathbb{R} \times \mathbb{R}$, $\mathbb{P}((X, Y) \in A) = \int\int_A f(x, y)dxdy$.

In the discrete or continuous case we define the joint CDF as $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$.

2.20 Example. Let (X, Y) be uniform on the unit square. Then,

$$f(x, y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $\mathbb{P}(X < 1/2, Y < 1/2)$. The event $A = \{X < 1/2, Y < 1/2\}$ corresponds to a subset of the unit square. Integrating f over this subset corresponds, in this case, to computing the area of the set A which is $1/4$. So, $\mathbb{P}(X < 1/2, Y < 1/2) = 1/4$. ■

2.21 Example. Let (X, Y) have density

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \int_0^1 \int_0^1 (x+y) dx dy &= \int_0^1 \left[\int_0^1 x\, dx \right] dy + \int_0^1 \left[\int_0^1 y\, dx \right] dy \\ &= \int_0^1 \frac{1}{2} dy + \int_0^1 y\, dy = \frac{1}{2} + \frac{1}{2} = 1 \end{aligned}$$

which verifies that this is a PDF ■

2.22 Example. If the distribution is defined over a non-rectangular region, then the calculations are a bit more complicated. Here is an example which I borrowed from DeGroot and Schervish (2002). Let (X, Y) have density

$$f(x,y) = \begin{cases} c\, x^2 y & \text{if } x^2 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note first that $-1 \le x \le 1$. Now let us find the value of c. The trick here is to be careful about the range of integration. We pick one variable, x say, and let it range over its values. Then, for each fixed value of x, we let y vary over its range, which is $x^2 \le y \le 1$. It may help if you look at Figure 2.5. Thus,

$$\begin{aligned} 1 &= \int \int f(x,y) dy dx = c \int_{-1}^1 \int_{x^2}^1 x^2 y\, dy\, dx \\ &= c \int_{-1}^1 x^2 \left[\int_{x^2}^1 y\, dy \right] dx = c \int_{-1}^1 x^2 \frac{1-x^4}{2} dx = \frac{4c}{21}. \end{aligned}$$

Hence, $c = 21/4$. Now let us compute $\mathbb{P}(X \ge Y)$. This corresponds to the set $A = \{(x,y); 0 \le x \le 1, x^2 \le y \le x\}$. (You can see this by drawing a diagram.) So,

$$\begin{aligned} \mathbb{P}(X \ge Y) &= \frac{21}{4} \int_0^1 \int_{x^2}^x x^2\, y\, dy\, dx = \frac{21}{4} \int_0^1 x^2 \left[\int_{x^2}^x y\, dy \right] dx \\ &= \frac{21}{4} \int_0^1 x^2 \left(\frac{x^2 - x^4}{2} \right) dx = \frac{3}{20}. \ ■ \end{aligned}$$

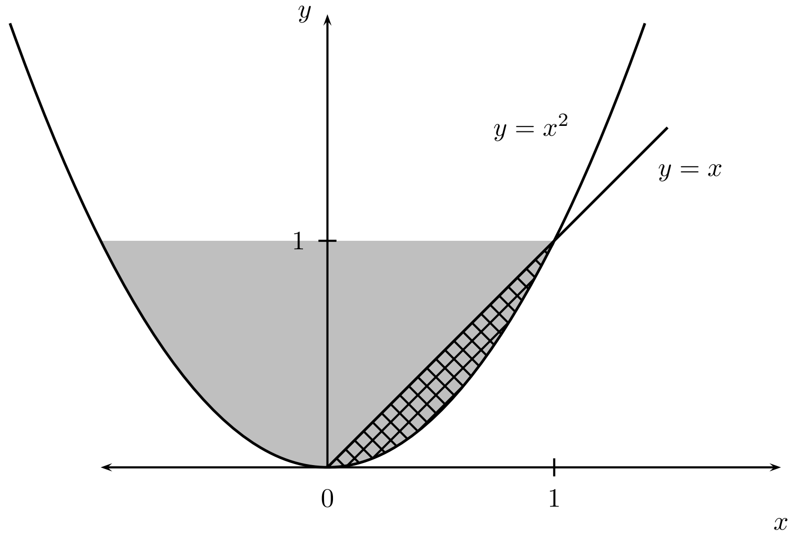


FIGURE 2.5. The light shaded region is $x^2 \le y \le 1$. The density is positive over this region. The hatched region is the event $X \ge Y$ intersected with $x^2 \le y \le 1$.

2.6 Marginal Distributions

2.23 Definition. *If* (X, Y) *have joint distribution with mass function* $f_{X,Y}$, *then the* **marginal mass function for** X *is defined by*

$$f_X(x) = \mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y f(x, y) \tag{2.4}$$

and the **marginal mass function for** Y *is defined by*

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x \mathbb{P}(X = x, Y = y) = \sum_x f(x, y). \tag{2.5}$$

2.24 Example. Suppose that $f_{X,Y}$ is given in the table that follows. The marginal distribution for X corresponds to the row totals and the marginal distribution for Y corresponds to the columns totals.

	$Y = 0$	$Y = 1$	
X=0	1/10	2/10	3/10
X=1	3/10	4/10	7/10
	4/10	6/10	1

For example, $f_X(0) = 3/10$ and $f_X(1) = 7/10$. ■

2.25 Definition. *For continuous random variables, the marginal densities are*

$$f_X(x) = \int f(x,y)dy, \quad \text{and} \quad f_Y(y) = \int f(x,y)dx. \tag{2.6}$$

The corresponding marginal distribution functions are denoted by F_X and F_Y.

2.26 Example. Suppose that

$$f_{X,Y}(x,y) = e^{-(x+y)}$$

for $x, y \geq 0$. Then $f_X(x) = e^{-x} \int_0^\infty e^{-y} dy = e^{-x}$. ■

2.27 Example. Suppose that

$$f(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$f_Y(y) = \int_0^1 (x+y)\, dx = \int_0^1 x\, dx + \int_0^1 y\, dx = \frac{1}{2} + y. \quad ■$$

2.28 Example. Let (X,Y) have density

$$f(x,y) = \begin{cases} \frac{21}{4} x^2 y & \text{if } x^2 \leq y \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$f_X(x) = \int f(x,y) dy = \frac{21}{4} x^2 \int_{x^2}^1 y\, dy = \frac{21}{8} x^2 (1 - x^4)$$

for $-1 \leq x \leq 1$ and $f_X(x) = 0$ otherwise. ■

2.7 Independent Random Variables

2.29 Definition. *Two random variables X and Y are* **independent** *if, for every A and B,*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \tag{2.7}$$

and we write $X \amalg Y$. Otherwise we say that X and Y are **dependent** *and we write $X \not\amalg Y$.*

In principle, to check whether X and Y are independent we need to check equation (2.7) for all subsets A and B. Fortunately, we have the following result which we state for continuous random variables though it is true for discrete random variables too.

2.30 Theorem. *Let X and Y have joint* PDF *$f_{X,Y}$. Then $X \amalg Y$ if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all values x and y.* [5]

2.31 Example. Let X and Y have the following distribution:

	$Y=0$	$Y=1$	
X=0	1/4	1/4	1/2
X=1	1/4	1/4	1/2
	1/2	1/2	1

Then, $f_X(0) = f_X(1) = 1/2$ and $f_Y(0) = f_Y(1) = 1/2$. X and Y are independent because $f_X(0)f_Y(0) = f(0,0)$, $f_X(0)f_Y(1) = f(0,1)$, $f_X(1)f_Y(0) = f(1,0)$, $f_X(1)f_Y(1) = f(1,1)$. Suppose instead that X and Y have the following distribution:

	$Y=0$	$Y=1$	
X=0	1/2	0	1/2
X=1	0	1/2	1/2
	1/2	1/2	1

These are not independent because $f_X(0)f_Y(1) = (1/2)(1/2) = 1/4$ yet $f(0,1) = 0$. ■

2.32 Example. Suppose that X and Y are independent and both have the same density

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let us find $\mathbb{P}(X + Y \le 1)$. Using independence, the joint density is

$$f(x,y) = f_X(x)f_Y(y) = \begin{cases} 4xy & \text{if } 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

[5]The statement is not rigorous because the density is defined only up to sets of measure 0.

Now,

$$\begin{aligned}\mathbb{P}(X+Y\leq 1) &= \int\int_{x+y\leq 1} f(x,y)dydx \\ &= 4\int_0^1 x\left[\int_0^{1-x} ydy\right]dx \\ &= 4\int_0^1 x\frac{(1-x)^2}{2}dx = \frac{1}{6}. \quad \blacksquare\end{aligned}$$

The following result is helpful for verifying independence.

2.33 Theorem. *Suppose that the range of X and Y is a (possibly infinite) rectangle. If $f(x,y) = g(x)h(y)$ for some functions g and h (not necessarily probability density functions) then X and Y are independent.*

2.34 Example. Let X and Y have density

$$f(x,y) = \begin{cases} 2e^{-(x+2y)} & \text{if } x>0 \text{ and } y>0 \\ 0 & \text{otherwise.}\end{cases}$$

The range of X and Y is the rectangle $(0,\infty)\times(0,\infty)$. We can write $f(x,y) = g(x)h(y)$ where $g(x) = 2e^{-x}$ and $h(y) = e^{-2y}$. Thus, $X \amalg Y$. ■

2.8 Conditional Distributions

If X and Y are discrete, then we can compute the conditional distribution of X given that we have observed $Y=y$. Specifically, $\mathbb{P}(X=x|Y=y) = \mathbb{P}(X=x, Y=y)/\mathbb{P}(Y=y)$. This leads us to define the conditional probability mass function as follows.

2.35 Definition. *The* **conditional probability mass function** *is*

$$f_{X|Y}(x|y) = \mathbb{P}(X=x|Y=y) = \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

if $f_Y(y) > 0$.

For continuous distributions we use the same definitions. [6] The interpretation differs: in the discrete case, $f_{X|Y}(x|y)$ is $\mathbb{P}(X=x|Y=y)$, but in the continuous case, we must integrate to get a probability.

[6]We are treading in deep water here. When we compute $\mathbb{P}(X\in A|Y=y)$ in the continuous case we are conditioning on the event $\{Y=y\}$ which has probability 0. We

2.36 Definition. *For continuous random variables, the* **conditional probability density function** *is*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

assuming that $f_Y(y) > 0$. Then,

$$\mathbb{P}(X \in A|Y = y) = \int_A f_{X|Y}(x|y)dx.$$

2.37 Example. Let X and Y have a joint uniform distribution on the unit square. Thus, $f_{X|Y}(x|y) = 1$ for $0 \le x \le 1$ and 0 otherwise. Given $Y = y$, X is Uniform$(0,1)$. We can write this as $X|Y = y \sim \text{Uniform}(0,1)$. ■

From the definition of the conditional density, we see that $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x)$. This can sometimes be useful as in example 2.39.

2.38 Example. Let

$$f(x,y) = \begin{cases} x + y & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let us find $\mathbb{P}(X < 1/4|Y = 1/3)$. In example 2.27 we saw that $f_Y(y) = y + (1/2)$. Hence,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{x+y}{y+\frac{1}{2}}.$$

So,

$$\begin{aligned} P\left(X < \frac{1}{4} \,\middle|\, Y = \frac{1}{3}\right) &= \int_0^{1/4} f_{X|Y}\left(x \,\middle|\, \frac{1}{3}\right) dx \\ &= \int_0^{1/4} \frac{x + \frac{1}{3}}{\frac{1}{3} + \frac{1}{2}} dx = \frac{\frac{1}{32} + \frac{1}{12}}{\frac{1}{3} + \frac{1}{2}} = \frac{11}{80}. \end{aligned}$$ ■

2.39 Example. Suppose that $X \sim \text{Uniform}(0,1)$. After obtaining a value of X we generate $Y|X = x \sim \text{Uniform}(x,1)$. What is the marginal distribution

avoid this problem by defining things in terms of the PDF. The fact that this leads to a well-defined theory is proved in more advanced courses. Here, we simply take it as a definition.

of Y? First note that,

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

So,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x) = \begin{cases} \frac{1}{1-x} & \text{if } 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The marginal for Y is

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)dx = \int_0^y \frac{dx}{1-x} = -\int_1^{1-y} \frac{du}{u} = -\log(1-y)$$

for $0 < y < 1$. ■

2.40 Example. Consider the density in Example 2.28. Let's find $f_{Y|X}(y|x)$. When $X = x$, y must satisfy $x^2 \le y \le 1$. Earlier, we saw that $f_X(x) = (21/8)x^2(1-x^4)$. Hence, for $x^2 \le y \le 1$,

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)} = \frac{2y}{1-x^4}.$$

Now let us compute $\mathbb{P}(Y \ge 3/4 | X = 1/2)$. This can be done by first noting that $f_{Y|X}(y|1/2) = 32y/15$. Thus,

$$\mathbb{P}(Y \ge 3/4 | X = 1/2) = \int_{3/4}^1 f(y|1/2)dy = \int_{3/4}^1 \frac{32y}{15} dy = \frac{7}{15}. \quad ■$$

2.9 Multivariate Distributions and IID Samples

Let $X = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are random variables. We call X a **random vector**. Let $f(x_1, \ldots, x_n)$ denote the PDF. It is possible to define their marginals, conditionals etc. much the same way as in the bivariate case. We say that $X_1, \ldots, X_n$ are independent if, for every $A_1, \ldots, A_n$,

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i). \tag{2.8}$$

It suffices to check that $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_{X_i}(x_i)$.

2.41 Definition. *If $X_1, \ldots, X_n$ are independent and each has the same marginal distribution with* CDF *F, we say that $X_1, \ldots, X_n$ are* IID *(independent and identically distributed) and we write*

$$X_1, \ldots X_n \sim F.$$

If F has density f we also write $X_1, \ldots X_n \sim f$. We also call $X_1, \ldots, X_n$ a **random sample of size n from F.**

Much of statistical theory and practice begins with IID observations and we shall study this case in detail when we discuss statistics.

2.10 Two Important Multivariate Distributions

MULTINOMIAL. The multivariate version of a Binomial is called a Multinomial. Consider drawing a ball from an urn which has balls with k different colors labeled "color 1, color 2, ..., color k." Let $p = (p_1, \ldots, p_k)$ where $p_j \geq 0$ and $\sum_{j=1}^k p_j = 1$ and suppose that p_j is the probability of drawing a ball of color j. Draw n times (independent draws with replacement) and let $X = (X_1, \ldots, X_k)$ where X_j is the number of times that color j appears. Hence, $n = \sum_{j=1}^k X_j$. We say that X has a Multinomial (n,p) distribution written $X \sim \text{Multinomial}(n, p)$. The probability function is

$$f(x) = \binom{n}{x_1 \ldots x_k} p_1^{x_1} \cdots p_k^{x_k} \tag{2.9}$$

where

$$\binom{n}{x_1 \ldots x_k} = \frac{n!}{x_1! \cdots x_k!}.$$

2.42 Lemma. *Suppose that $X \sim \text{Multinomial}(n, p)$ where $X = (X_1, \ldots, X_k)$ and $p = (p_1, \ldots, p_k)$. The marginal distribution of X_j is Binomial (n,p_j).*

MULTIVARIATE NORMAL. The univariate Normal has two parameters, μ and σ. In the multivariate version, μ is a vector and σ is replaced by a matrix Σ. To begin, let

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_k \end{pmatrix}$$

where $Z_1, \ldots, Z_k \sim N(0,1)$ are independent. The density of Z is [7]

$$\begin{aligned} f(z) &= \prod_{i=1}^{k} f(z_i) = \frac{1}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{k} z_j^2 \right\} \\ &= \frac{1}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2} z^T z \right\}. \end{aligned}$$

We say that Z has a standard multivariate Normal distribution written $Z \sim N(0, I)$ where it is understood that 0 represents a vector of k zeroes and I is the $k \times k$ identity matrix.

More generally, a vector X has a multivariate Normal distribution, denoted by $X \sim N(\mu, \Sigma)$, if it has density [8]

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2} |(\Sigma)|^{1/2}} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x - \mu) \right\} \tag{2.10}$$

where $|\Sigma|$ denotes the determinant of Σ, μ is a vector of length k and Σ is a $k \times k$ symmetric, positive definite matrix. [9] Setting $\mu = 0$ and $\Sigma = I$ gives back the standard Normal.

Since Σ is symmetric and positive definite, it can be shown that there exists a matrix $\Sigma^{1/2}$ — called the square root of Σ — with the following properties: (i) $\Sigma^{1/2}$ is symmetric, (ii) $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ and (iii) $\Sigma^{1/2}\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2} = I$ where $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$.

2.43 Theorem. *If $Z \sim N(0, I)$ and $X = \mu + \Sigma^{1/2} Z$ then $X \sim N(\mu, \Sigma)$. Conversely, if $X \sim N(\mu, \Sigma)$, then $\Sigma^{-1/2}(X - \mu) \sim N(0, I)$.*

Suppose we partition a random Normal vector X as $X = (X_a, X_b)$ We can similarly partition $\mu = (\mu_a, \mu_b)$ and

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}.$$

2.44 Theorem. *Let $X \sim N(\mu, \Sigma)$. Then:*

(1) The marginal distribution of X_a is $X_a \sim N(\mu_a, \Sigma_{aa})$.

(2) The conditional distribution of X_b given $X_a = x_a$ is

$$X_b | X_a = x_a \sim N\left(\mu_b + \Sigma_{ba}\Sigma_{aa}^{-1}(x_a - \mu_a),\ \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab} \right).$$

(3) If a is a vector then $a^T X \sim N(a^T \mu, a^T \Sigma a)$.

(4) $V = (X - \mu)^T \Sigma^{-1} (X - \mu) \sim \chi^2_k$.

---

[7] If a and b are vectors then $a^T b = \sum_{i=1}^{k} a_i b_i$.

[8] Σ^{-1} is the inverse of the matrix Σ.

[9] A matrix Σ is positive definite if, for all nonzero vectors x, $x^T \Sigma x > 0$.

2.11 Transformations of Random Variables

Suppose that X is a random variable with PDF f_X and CDF F_X. Let $Y = r(X)$ be a function of X, for example, $Y = X^2$ or $Y = e^X$. We call $Y = r(X)$ a transformation of X. How do we compute the PDF and CDF of Y? In the discrete case, the answer is easy. The mass function of Y is given by

$$\begin{aligned} f_Y(y) &= \mathbb{P}(Y = y) = \mathbb{P}(r(X) = y) \\ &= \mathbb{P}(\{x;\ r(x) = y\}) = \mathbb{P}(X \in r^{-1}(y)). \end{aligned}$$

2.45 Example. Suppose that $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = 1/4$ and $\mathbb{P}(X = 0) = 1/2$. Let $Y = X^2$. Then, $\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = 1/2$ and $\mathbb{P}(Y = 1) = \mathbb{P}(X = 1) + \mathbb{P}(X = -1) = 1/2$. Summarizing:

x	$f_X(x)$
-1	1/4
0	1/2
1	1/4

y	$f_Y(y)$
0	1/2
1	1/2

Y takes fewer values than X because the transformation is not one-to-one. ■

The continuous case is harder. There are three steps for finding f_Y:

Three Steps for Transformations

1. For each y, find the set $A_y = \{x :\ r(x) \le y\}$.
2. Find the CDF

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(r(X) \le y) \\ &= \mathbb{P}(\{x;\ r(x) \le y\}) \\ &= \int_{A_y} f_X(x)dx. \end{aligned} \tag{2.11}$$

3. The PDF is $f_Y(y) = F'_Y(y)$.

2.46 Example. Let $f_X(x) = e^{-x}$ for $x > 0$. Hence, $F_X(x) = \int_0^x f_X(s)ds = 1 - e^{-x}$. Let $Y = r(X) = \log X$. Then, $A_y = \{x :\ x \le e^y\}$ and

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(\log X \le y) \\ &= \mathbb{P}(X \le e^y) = F_X(e^y) = 1 - e^{-e^y}. \end{aligned}$$

Therefore, $f_Y(y) = e^y e^{-e^y}$ for $y \in \mathbb{R}$. ■

2.47 Example. Let $X \sim \text{Uniform}(-1,3)$. Find the PDF of $Y = X^2$. The density of X is

$$f_X(x) = \begin{cases} 1/4 & \text{if } -1 < x < 3 \\ 0 & \text{otherwise.} \end{cases}$$

Y can only take values in $(0,9)$. Consider two cases: (i) $0 < y < 1$ and (ii) $1 \leq y < 9$. For case (i), $A_y = [-\sqrt{y}, \sqrt{y}]$ and $F_Y(y) = \int_{A_y} f_X(x)dx = (1/2)\sqrt{y}$. For case (ii), $A_y = [-1, \sqrt{y}]$ and $F_Y(y) = \int_{A_y} f_X(x)dx = (1/4)(\sqrt{y}+1)$. Differentiating F we get

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & \text{if } 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & \text{if } 1 < y < 9 \\ 0 & \text{otherwise.} \end{cases} \blacksquare$$

When r is strictly monotone increasing or strictly monotone decreasing then r has an inverse $s = r^{-1}$ and in this case one can show that

$$f_Y(y) = f_X(s(y)) \left| \frac{ds(y)}{dy} \right|. \tag{2.12}$$

2.12 Transformations of Several Random Variables

In some cases we are interested in transformations of several random variables. For example, if X and Y are given random variables, we might want to know the distribution of X/Y, $X+Y$, $\max\{X,Y\}$ or $\min\{X,Y\}$. Let $Z = r(X,Y)$ be the function of interest. The steps for finding f_Z are the same as before:

Three Steps for Transformations

1. For each z, find the set $A_z = \{(x,y) : \ r(x,y) \leq z\}$.
2. Find the CDF

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}(r(X,Y) \leq z) \\ &= \mathbb{P}(\{(x,y);\ r(x,y) \leq z\}) = \int\int_{A_z} f_{X,Y}(x,y)\,dx\,dy. \end{aligned}$$

3. Then $f_Z(z) = F'_Z(z)$.

2.48 Example. Let $X_1, X_2 \sim \text{Uniform}(0,1)$ be independent. Find the density of $Y = X_1 + X_2$. The joint density of (X_1, X_2) is

$$f(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $r(x_1, x_2) = x_1 + x_2$. Now,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(r(X_1, X_2) \le y) \\ &= \mathbb{P}(\{(x_1, x_2) : \ r(x_1, x_2) \le y\}) = \int\int_{A_y} f(x_1, x_2) dx_1 dx_2. \end{aligned}$$

Now comes the hard part: finding A_y. First suppose that $0 < y \le 1$. Then A_y is the triangle with vertices $(0,0), (y,0)$ and $(0,y)$. See Figure 2.6. In this case, $\int\int_{A_y} f(x_1, x_2) dx_1 dx_2$ is the area of this triangle which is $y^2/2$. If $1 < y < 2$, then A_y is everything in the unit square except the triangle with vertices $(1, y-1), (1,1), (y-1, 1)$. This set has area $1 - (2-y)^2/2$. Therefore,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{y^2}{2} & 0 \le y < 1 \\ 1 - \frac{(2-y)^2}{2} & 1 \le y < 2 \\ 1 & y \ge 2. \end{cases}$$

By differentiation, the PDF is

$$f_Y(y) = \begin{cases} y & 0 \le y \le 1 \\ 2 - y & 1 \le y \le 2 \\ 0 & \text{otherwise.} \end{cases} \quad \blacksquare$$

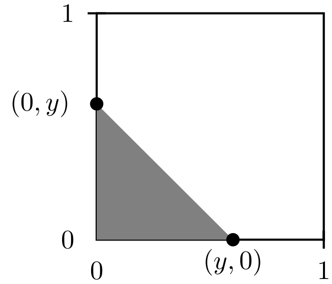
2.13 Appendix

Recall that a probability measure $\mathbb{P}$ is defined on a σ-field $\mathcal{A}$ of a sample space Ω. A random variable X is a **measurable** map $X : \Omega \to \mathbb{R}$. Measurable means that, for every x, $\{\omega : \ X(\omega) \le x\} \in \mathcal{A}$.

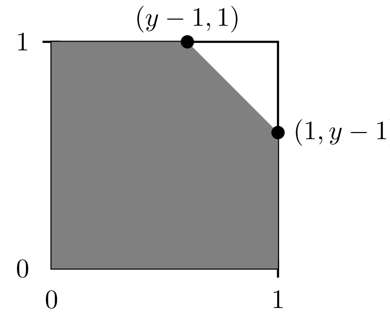
2.14 Exercises

1. Show that

$$\mathbb{P}(X = x) = F(x^+) - F(x^-).$$



This is the case $0 \le y < 1$.



This is the case $1 \le y \le 2$.

FIGURE 2.6. The set A_y for example 2.48. A_y consists of all points (x_1, x_2) in the square below the line $x_2 = y - x_1$.

2. Let X be such that $\mathbb{P}(X = 2) = \mathbb{P}(X = 3) = 1/10$ and $\mathbb{P}(X = 5) = 8/10$. Plot the CDF F. Use F to find $\mathbb{P}(2 < X \le 4.8)$ and $\mathbb{P}(2 \le X \le 4.8)$.

3. Prove Lemma 2.15.

4. Let X have probability density function

$$f_X(x) = \begin{cases} 1/4 & 0 < x < 1 \\ 3/8 & 3 < x < 5 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X.

(b) Let $Y = 1/X$. Find the probability density function $f_Y(y)$ for Y. Hint: Consider three cases: $\frac{1}{5} \le y \le \frac{1}{3}$, $\frac{1}{3} \le y \le 1$, and $y \ge 1$.

5. Let X and Y be discrete random variables. Show that X and Y are independent if and only if $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all x and y.

6. Let X have distribution F and density function f and let A be a subset of the real line. Let $I_A(x)$ be the indicator function for A:

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Let $Y = I_A(X)$. Find an expression for the cumulative distribution of Y. (Hint: first find the probability mass function for Y.)

7. Let X and Y be independent and suppose that each has a Uniform$(0,1)$ distribution. Let $Z = \min\{X, Y\}$. Find the density $f_Z(z)$ for Z. Hint: It might be easier to first find $\mathbb{P}(Z > z)$.

8. Let X have CDF F. Find the CDF of $X^+ = \max\{0, X\}$.

9. Let $X \sim \text{Exp}(\beta)$. Find $F(x)$ and $F^{-1}(q)$.

10. Let X and Y be independent. Show that $g(X)$ is independent of $h(Y)$ where g and h are functions.

11. Suppose we toss a coin once and let p be the probability of heads. Let X denote the number of heads and let Y denote the number of tails.

    (a) Prove that X and Y are dependent.

    (b) Let $N \sim \text{Poisson}(\lambda)$ and suppose we toss a coin N times. Let X and Y be the number of heads and tails. Show that X and Y are independent.

12. Prove Theorem 2.33.

13. Let $X \sim N(0,1)$ and let $Y = e^X$.

    (a) Find the PDF for Y. Plot it.

    (b) (Computer Experiment.) Generate a vector $x = (x_1, \ldots, x_{10{,}000})$ consisting of 10,000 random standard Normals. Let $y = (y_1, \ldots, y_{10{,}000})$ where $y_i = e^{x_i}$. Draw a histogram of y and compare it to the PDF you found in part (a).

14. Let (X, Y) be uniformly distributed on the unit disk $\{(x, y) : x^2 + y^2 \leq 1\}$. Let $R = \sqrt{X^2 + Y^2}$. Find the CDF and PDF of R.

15. (A universal random number generator.) Let X have a continuous, strictly increasing CDF F. Let $Y = F(X)$. Find the density of Y. This is called the probability integral transform. Now let $U \sim \text{Uniform}(0,1)$ and let $X = F^{-1}(U)$. Show that $X \sim F$. Now write a program that takes Uniform (0,1) random variables and generates random variables from an Exponential (β) distribution.

16. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ and assume that X and Y are independent. Show that the distribution of X given that $X + Y = n$ is Binomial(n, π) where $\pi = \lambda/(\lambda + \mu)$.

Hint 1: You may use the following fact: If $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, and X and Y are independent, then $X+Y \sim \text{Poisson}(\mu+\lambda)$.

Hint 2: Note that $\{X = x,\ X + Y = n\} = \{X = x,\ Y = n - x\}$.

17. Let

$$f_{X,Y}(x, y) = \begin{cases} c(x + y^2) & 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $P\left(X < \frac{1}{2} \mid Y = \frac{1}{2}\right)$.

18. Let $X \sim N(3, 16)$. Solve the following using the Normal table and using a computer package.

    (a) Find $\mathbb{P}(X < 7)$.

    (b) Find $\mathbb{P}(X > -2)$.

    (c) Find x such that $\mathbb{P}(X > x) = .05$.

    (d) Find $\mathbb{P}(0 \le X < 4)$.

    (e) Find x such that $\mathbb{P}(|X| > |x|) = .05$.

19. Prove formula (2.12).

20. Let $X, Y \sim \text{Uniform}(0, 1)$ be independent. Find the PDF for $X - Y$ and X/Y.

21. Let $X_1, \ldots, X_n \sim \text{Exp}(\beta)$ be IID. Let $Y = \max\{X_1, \ldots, X_n\}$. Find the PDF of Y. Hint: $Y \le y$ if and only if $X_i \le y$ for $i = 1, \ldots, n$.