Chapter 2. Random Variables and Probability Distributions

2.1. Introduction

In the previous chapter, we introduced common topics of probability. In this chapter, we translate those concepts into a mathematical framework. We invoke algebra for discrete variables and calculus for continuous variables. Every topic in this chapter is presented twice, once for discrete variables and again for continuous variables. The analogy in the two cases should be apparent and should reinforce the common underlying concepts. There is, in the second half of the chapter, another duplication of concepts in which we show that the same process of translation from the language of probability to that of mathematics can be performed not only when we have a single variable of interest, but also when we have two variables. Again, high-lighting this analogy between single and joint probability distributions explicitly reveals the common underlying concepts.

2.2. Random Variables & Sample Spaces

We begin with the introduction of a necessary vocabulary.

Random variable

A random variable is a function that associates a number, integer or real, with each element in a sample space.

Discrete Sample Space

If a sample space contains a finite number of possibilities or an unending sequences with as many elements as there are whole numbers, it is called a discrete sample space.

Example 2.1.: You flip two coins. Y is a random variable that counts the number of heads. The possible results and the value of the random variable associated with each result are given in the following table.

result	y
HH	2
HT	1
TH	1
TT	0

This sample space is discrete because there are a finite number of possible outcomes.

Example 2.2.: You subject a box containing *N* devices to a test. *Y* is a random variable that counts the number of defective devices. The value of the random variable ranges from 0 (no defects) to *N* (all defects). This sample space is discrete because there are a finite number of possible outcomes.

Continuous Sample Space

If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a continuous sample space.

Example 2.3.: You drive a car with five gallons of gas. *Y* is a random variable that represents the distance traveled. The possible results are infinite because even if the car averaged 20 miles per gallon, it could go 100.0 miles, 100.1, 100.01, 100.001, 100.0001 miles. The sample space is as infinite as real numbers.

2.3. Discrete Probability Distribution Functions (PDFs)

Probability distribution function (PDF)

The function, $f(x)$ is a probability distribution function of the discrete random variable x, if for each possible outcome a, the following three criteria are satisfied.

$$\begin{aligned} & f(x) \geq 0 \\ & \sum_x f(x) = 1 \\ & P(x = a) = f(a) \end{aligned} \tag{2.1}$$

The PDF is always non-negative. The PDF is normalized, meaning that the sum over all values of a discrete PDF is unity. The PDF evaluated at outcome a provides the probability of the occurrence of outcome a.

Example 2.4.: Eight devices are shipped to a retail outlet, 3 of which are defective. If a consumer purchases 2 computers, find the probability distribution for the number of defective devices bought by the consumer.

In order to solve this problem, first define the random variable and the range of the random variable. The random variable, x, is equal to the number of defective devices bought by the consumer. The random variable, x, can take on values of 0, 1, and 2. Those are the only number of defective devices the consumer can buy, given that they are only buying two devices.

The next step is to determine the size of the sample space. The number of ways that 2 can be taken from 8 without replacement is $\binom{8}{2} = 28$. We use the formula for combinations because the order of purchases does not matter. This is the total number of combinations of devices that the consumer can buy.

Third, the probability of a particular outcome is equal to the number of ways to get that outcome over the total number of ways:

$$f(x) = P(x = a) = \frac{\text{ways of getting a}}{\text{total ways}}$$

$$f(0) = P(x = 0) = \frac{\binom{3}{0}\binom{5}{2}}{28} = \frac{10}{28}$$

$$f(1) = P(x = 1) = \frac{\binom{3}{1}\binom{5}{1}}{28} = \frac{15}{28}$$

$$f(2) = P(x = 2) = \frac{\binom{3}{2}\binom{5}{0}}{28} = \frac{3}{28}$$

In each of these cases, we obtained the numerator, the number of ways of getting outcome a, by using the combination rule and the generalized multiplication rule. There are $\binom{3}{a}$ ways of choosing a defective devices from 3 defective devices. There are $\binom{5}{2\text{-}a}$ ways of choosing (2-a) good devices from 5 good devices. We use the generalized multiplication rule to get the number of ways of getting both of these outcomes in the numerator. As a preview, we will come to discover that this probability distribution is called the hypergeometric distribution in Chapter 4.

So we have the PDF, f(x), defined for all possible values of x. We have solved the problem.

Note: If someone asked for the probability for getting 3 (or any number other than 0, 1, or 2) defective devices, then the probability is zero and f(3)= 0.

Testing a discrete PDF for legitimacy

If you are asked to determine if a given PDF is legitimate, you are required to verify the three criteria in equation (2.1). Generally, the third criterion is given in the problem statement, so you only have to check the first two criteria.

The first criteria, $f(x) \geq 0$, can most easily be verified by plotting f(x) and showing that it is never negative. The second criteria, $\sum_x f(x) = 1$, can most easily be verified by direct summation of all *f*(*x*).

Normalizing a discrete PDF

Discrete PDF's must satisfy $\sum_x f(x) = 1$. Sometimes, you have the functional form of the PDF and you simply need to force it to satisfy this criterion. In that case you need to normalize the PDF so that it sums to unity.

If $f^*(x)$ is an unnormalized PDF, then it can be normalized by the multiplication of a constant,

$$f(x) = cf^*(x) = \frac{1}{\sum_x f^*(x)} f^*(x) \tag{2.2}$$

where that constant is the inverse of the sum of the unnormalized PDF.

Example 2.5.: Find the value of *c* that normalizes the following PDF.

$f(x) = c\left({}_4P_x\right)$ for *x* = 0, 1, 2, 3, & 4

To normalize, we sum the PDF over all values and set it to unity.

$$\sum_x f(x) = 1 = \sum_{x=0}^{4} f(x) = f(0) + f(1) + f(2) + f(3) + f(4)$$

$$\sum_x f(x) = 1 = \sum_{x=0}^{4} c\left({}_4P_x\right) = c\sum_{x=0}^{4}\left({}_4P_x\right) = c\left({}_4P_0 + {}_4P_1 + {}_4P_2 + {}_4P_3 + {}_4P_4\right) = c(1 + 4 + 12 + 24 + 24)$$

We then solve for simplify and solve for the normalization constant, *c*.

$$1 = c(65)$$

$$c = \frac{1}{65}$$

So the normalized PDF is

$$f(x) = \frac{1}{65}\left({}_4P_x\right)$$

Discrete Cumulative Distribution Function (CDF)

The discrete cumulative distribution function (CDF), $F(x)$ of a discrete random variable X with the probability distribution, $f(x)$, is given by

$$F(a) = P(x \le a) = \sum_{x \le a} f(x) \quad \text{for} \quad -\infty < x < \infty \tag{2.3}$$

The CDF is the probability that x is less than or equal to a.

Example 2.6.: In the above example, regarding the consumer purchasing devices, we can obtain the cumulative distribution directly:

$F(0) = f(0) = 10/28$,
$F(1) = f(0)+f(1)=25/28$,
$F(2)=f(0)+f(1)+f(2)=1$

Note: The cumulative distribution is always monotonically increasing, with x. The final value of the cumulative distribution is always unity, since the PDF is normalized.

Probability Histogram:

A probability histogram is a graphical representation of the distribution of a discrete random variable.

The histogram for the PDF and CDF for the Example 2.4. are given in Figure 2.1.

The histogram of the PDF provides a visual representation of the probability distribution, its most likely outcome and the shape of the distribution.

The histogram of the CDF provides a visual representation of the cumulative probability of an outcome. We observe that the CDF is monotonically increasing and ends at one, as it must since the PDF is normalized and sums to unity.

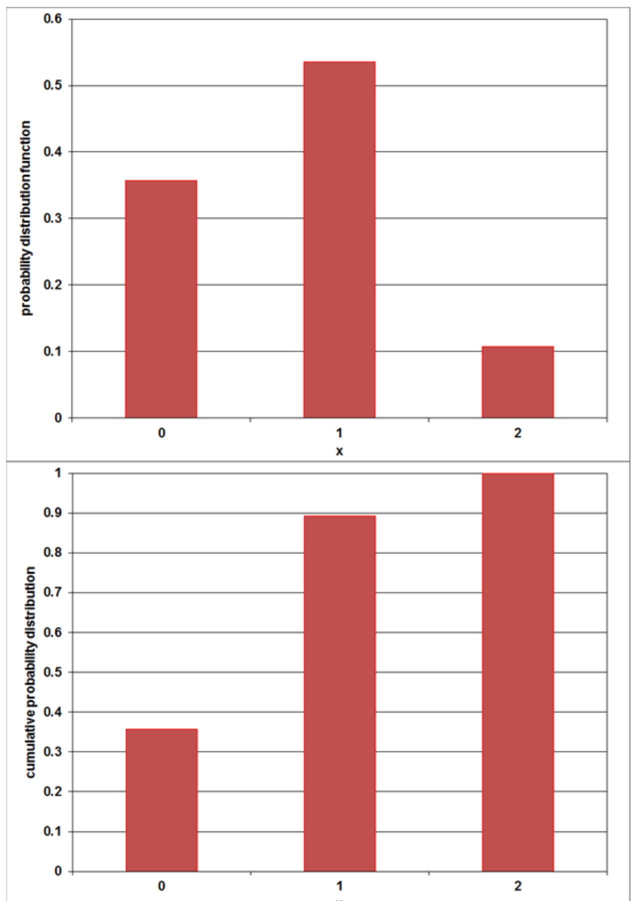

Figure 2.1. The histogram of the PDF (top) and CDF (bottom) for Example 2.4.

2.4. Continuous Probability Density Functions (PDFs)

Probability distribution functions of discrete random variables are called probability density functions when applied to continuous variables. Both have the same meaning and can be abbreviated commonly as PDF's. Probability density functions satisfy three criteria, which are analogous to those for discrete PDFs, namely

$$f(x) \geq 0 \quad \text{for all } x \in \mathrm{R}$$
$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{2.4}$$
$$P(a < x < b) = \int_a^b f(x)dx$$

The probability of finding an exact point on a continuous random variable is zero,

$$P(x = a) = P(a \leq x \leq a) = \int_a^a f(x)dx = 0$$

Consequently, the probability that a random variable is "greater than" or "greater than or equal to" a number is the same in for continuous random variables. The same is true of" less than" and "less than or equal to" signs for continuous random variables. This equivalence is absolutely not true for discrete random variables.

$$P(a < x < b) = P(a \leq x \leq b) \text{ and } P(a > x > b) = P(a \geq x \geq b)$$

Also it is important to note that substitution of a value into the PDF gives a probability only for a discrete random variable, in order words $P(x = a) = f(a)$ for discrete PDFs only. For a continuous random variable, $f(a)$ by itself doesn't provide a probability. Only the integral of $f(a)$ provides a probability from a continuous random variable.

Example 2.7.: A probability density function has the form

$$f(x) = \begin{cases} \dfrac{x^2}{3} & \text{for -1} < \text{x} < 2 \\ 0 & \text{otherwise} \end{cases}$$

A plot of the probability density distribution is shown in Figure 2.2. This plot is the continuous analog of the discrete histogram.

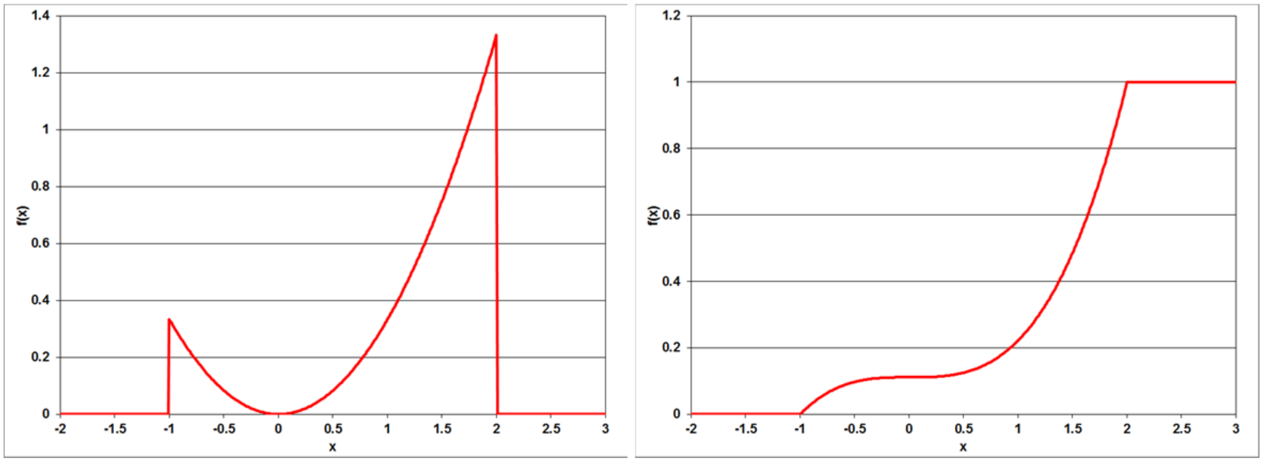

Figure 2.2. A plot of the PDF (left) and CDF (right) for Example 2.7.

The probability of finding an x between a and b is by (equation 2.4)

$$P(a < x < b) = \int_a^b f(x)dx = \begin{cases} \int_a^b \frac{x^2}{3} dx & \text{for -1} < \text{x} < 2 \\ \int_a^b 0 dx & \text{otherwise} \end{cases}$$

$$P(a < x < b) = \int_a^b f(x)dx = \begin{cases} \left(\frac{b^3}{9} - \frac{a^3}{9}\right) & \text{for -1} < \text{x} < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $P(-1 < x < 2)$

$$P(-1 < x < 2) = \int_{-1}^{2} f(x)dx = \left(\frac{2^3}{9} - \frac{(-1)^3}{9}\right) = 1$$

This result makes sense since the PDF is normalized and we have integrated over the entirety of the non-zero range of the random variable.

(b) Find $P(-\infty < x < \infty)$

We cannot integrate over discontinuities in a function. Therefore, we must break-up the integral over continuous parts.

$$P(-\infty < x < \infty) = \int_{-\infty}^{-1} f(x)dx + \int_{-1}^{2} f(x)dx + \int_{2}^{\infty} f(x)dx = 0 + \left(\frac{2^3}{9} - \frac{(-1)^3}{9}\right) + 0 = 1$$

Here we see that actually it is not practically necessary to integrate over the parts of the function where $f(x)$=0, because the integral over those ranges is also 0. In general practice, we just need to perform the integration over those ranges where the PDF, $f(x)$, is non-zero.

(c) Find $P(-\infty < x < 0)$

$$P(-\infty < x < 0) = \int_{-\infty}^{-1} f(x)dx + \int_{-1}^{0} f(x)dx = 0 + \left(\frac{0^3}{9} - \frac{(-1)^3}{9}\right) = \frac{1}{9}$$

Again, it is not necessary to explicitly integrate over anything but the non-zero portions of the PDF, as all other portions contribute nothing to the integral.

(d) Find $P(0 < x < 1)$

$$P(0 < x < 1) = \int_{0}^{1} f(x)dx = \left(\frac{1^3}{9} - \frac{0^3}{9}\right) = \frac{1}{9}$$

Testing a continuous PDF for legitimacy

If you are asked to determine if a given PDF is legitimate, you are required to verify the three criteria in equation (2.4). Generally, the third criterion is given in the problem statement, so you only have to check the first 2 criteria.

The first criteria, $f(x) \geq 0$, can most easily be verified by plotting $f(x)$ and showing that it is never negative. The second criteria, $\int_{-\infty}^{\infty} f(x)dx = 1$, can most easily be verified by direct integration of $f(x)$.

Normalizing a continuous PDF

Continuous PDF's must satisfy $\int_{-\infty}^{\infty} f(x)dx = 1$. Sometimes, you have the functional form of the PDF and you simply need to force it to satisfy this criterion. In that case you need to normalize the PDF so that it sums to unity.

If $f^*(x)$ is an unnormalized PDF, then it can be normalized by the multiplication of a constant,

$$f(x) = cf^*(x) = \frac{1}{\int_{-\infty}^{\infty} f^*(x)dx} f^*(x) \qquad (2.5)$$

where that constant is the inverse of the sum of the unnormalized PDF.

Example 2.8.: Find the value of c that normalizes the PDF.

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

To normalize:

$$\int_{-\infty}^{\infty} f(x)dx = 1 = \int_{-1}^{2} cx^2 dx = c\left.\frac{x^3}{3}\right|_{-1}^{2} = c\left(\frac{9}{3}\right) = 3c$$

$$c = \frac{1}{3}$$

So the normalized PDF is

$$f(x) = \begin{cases} \frac{x^2}{3} & \text{for -1} < \text{x} < 2 \\ 0 & \text{otherwise} \end{cases}$$

Continuous Cumulative distributions

The cumulative distribution *F*(*x*) of a continuous random variable *x* with density function *f*(*x*) is

$$F(a) = P(x \le a) = \int_{-\infty}^{a} f(x)d\text{x} \qquad \text{for} \quad -\infty < x < \infty \qquad (5.6)$$

This function gives the probability that a randomly selected value of the variable *x* is less than *a*. The implicit lower limit of cumulative distribution is negative infinity.

$$F(a) = P(-\infty \le x \le a) = P(x \le a)$$

Example 2.9.: Determine the cumulative distribution function for the PDF of Example 2.7.

$$F(a) = P(x \le a) = \int_{-\infty}^{a} f(t)dt = \begin{cases} 0 & \text{for } x < -1 \\ \left(\frac{a^3}{9} + \frac{1^3}{9}\right) & \text{for } -1 < x < 2 \\ 1 & \text{for } x > 2 \end{cases}$$

A plot of the cumulative distribution function for the PDF of Example 2.7. is shown in Figure 2.2. The CDF is again monotonically increasing. It begins at zero and ends at unity, since the PDF is normalized.

2.5. Relations between Inequalities

In the above section we have defined a specific function for the probability that *x* is less than or equal to *a*, namely the cumulative distribution. But what about when *x* is greater than *a*, or strictly less than *a*, etc.? Here, we discuss those possibilities.

Consider the fact that the probability of all outcomes must sum to one. Then we can write (regardless of whether the PDF is discrete or continuous)

$$P(\mathrm{x} < \mathrm{a}) + \mathrm{P}(\mathrm{x} = \mathrm{a}) + \mathrm{P}(\mathrm{x} > \mathrm{a}) = 1$$

Using the union rule we can write:

$$P(\mathrm{x} \le \mathrm{a}) = \mathrm{P}[(\mathrm{x} < \mathrm{a}) \cup (\mathrm{x} = \mathrm{a})] = P(\mathrm{x} < \mathrm{a}) + \mathrm{P}(\mathrm{x} = \mathrm{a}) + \mathrm{P}[(\mathrm{x} < \mathrm{a}) \cap (\mathrm{x} = \mathrm{a})]$$

The intersection is zero, because x cannot equal a and be less than a, so

$$P(\mathrm{x} \le \mathrm{a}) = \mathrm{P}[(\mathrm{x} < \mathrm{a}) \cup (\mathrm{x} = \mathrm{a})] = P(\mathrm{x} < \mathrm{a}) + \mathrm{P}(\mathrm{x} = \mathrm{a})$$

Similarly

$$P(\mathrm{x} \ge \mathrm{a}) = \mathrm{P}[(\mathrm{x} > \mathrm{a}) \cup (\mathrm{x} = \mathrm{a})] = P(\mathrm{x} > \mathrm{a}) + \mathrm{P}(\mathrm{x} = \mathrm{a})$$

Using these three rules, we can create a generalized method for obtaining any arbitrary probability. On the other hand, we can use the rules to create way to obtain any probability from just the cumulative distribution function. (This will be important later when we use PDF's for which only the cumulative distribution function is given.) Regardless of which method you use, you will obtain the same answer. In Table 2.1, we summarize the expression of each probability in terms of the cumulative PDF.

The continuous case has one important difference. In the continuous case, the probability of a random variable *x* equaling a single value *a* is zero. Why? Because the probability is a ratio of the number of ways of getting *a* over the total number of ways in the sample space. There is only one way to get a, namely x=a. But in the denominator, there is an infinite number of values of x, since x is continuous. Therefore, the P(x=a)=0. We can show this using the definition is we write,

$$P(x = a) = \int_a^a f(x)dx = 0 \quad \text{for continuous PDF's only.}$$

One consequence of this is that

$$P(\mathrm{x} \le \mathrm{a}) = P(\mathrm{x} < \mathrm{a}) + P(\mathrm{x} = \mathrm{a}) = P(\mathrm{x} < \mathrm{a})$$

$$P(\mathrm{x} \ge \mathrm{a}) = P(\mathrm{x} > \mathrm{a}) + P(\mathrm{x} = \mathrm{a}) = P(\mathrm{x} > \mathrm{a})$$

The probability of $\mathrm{x} \le \mathrm{a}$ is the same as $\mathrm{x} < \mathrm{a}$. Likewise, the probability of $\mathrm{x} \ge \mathrm{a}$ is the same as $\mathrm{x} > \mathrm{a}$. This fact makes the continuous case easy to generate. In Table 2.2, we summarize the expression of each probability in terms of the cumulative PDF.

Probability	Definition	from cumulative PDF
$P(\mathrm{x} = \mathrm{a})$	$f(\mathrm{a})$	$P(\mathrm{x} \le \mathrm{a}) - P(\mathrm{x} \le \mathrm{a}-1)$
$P(\mathrm{x} \le \mathrm{a})$	$\sum_{x \le a} f(x)$	$P(\mathrm{x} \le \mathrm{a})$
$P(\mathrm{x} < \mathrm{a})$	$\sum_{x < a} f(x)$	$P(\mathrm{x} \le \mathrm{a}-1)$
$P(\mathrm{x} \ge \mathrm{a})$	$\sum_{x \ge a} f(x)$	$1 - P(\mathrm{x} \le \mathrm{a}-1)$
$P(\mathrm{x} > \mathrm{a})$	$\sum_{x > a} f(x)$	$1 - P(\mathrm{x} \le \mathrm{a})$

Table 2.1. Relations between inequalities for discrete random variables.

Probability	Definition	from cumulative PDF
$P(\mathrm{x} = \mathrm{a})$	$\int_a^a f(x)dx = 0$	0
$P(\mathrm{x} \le \mathrm{a})$ or $P(\mathrm{x} < \mathrm{a})$	$\int_{-\infty}^a f(x)dx$	$P(\mathrm{x} \le \mathrm{a})$
$P(\mathrm{x} \ge \mathrm{a})$ or $P(\mathrm{x} > \mathrm{a})$	$\int_a^{\infty} f(x)dx$	1- $P(\mathrm{x} \le \mathrm{a})$

Table 2.2. Relations between inequalities for continuous random variables.

Let's close out this section with two more vocabulary words used to describe PDFs. The definitions of symmetric and skewed distributions are provided below. An example of each are plotted in Figure 2.3.

Symmetric

A probability density distribution is said to be symmetric if it can be folded along a vertical axis so that the two sides coincide.

Skew

A probability density distribution is said to be skewed if it is not symmetric.

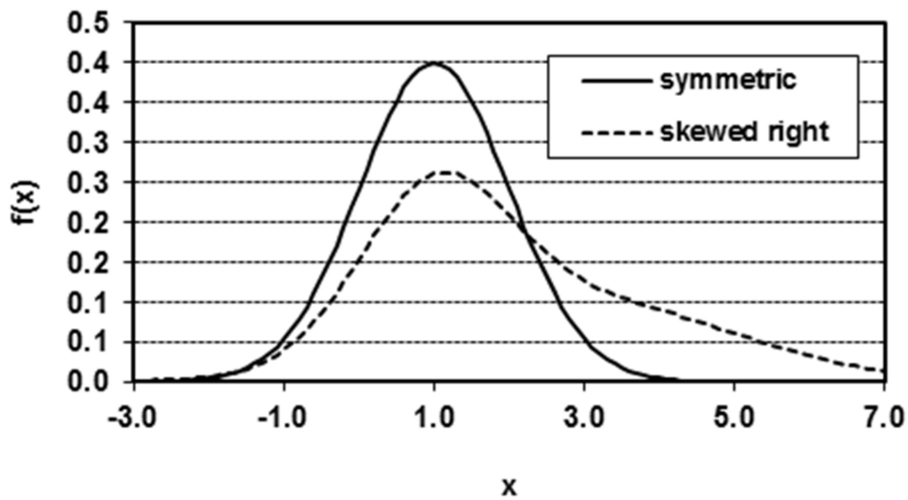

Figure 2.3. Examples of symmetric and skewed PDFs.

2.6. Discrete Joint Probability Distribution Functions

Thus far in this chapter, we have assumed that we have only one random variable. In many practical applications there are more than one random variable. The behavior of sets of random variables is described by Joint PDFs. In this book, we explicitly extend the formalism to two random variables. It can be extended to an arbitrary number of random variables. We will present this extension twice, once for discrete random variables and once for continuous random variables.

The function $f(x,y)$ is a joint probability distribution or probability mass function of the discrete random variable X and Y if

$$\begin{aligned} & f(x,y) \geq 0 \\ & \sum_x \sum_y f(x,y) = 1 \\ & P(x = a \cap y = b) = f(a,b) \end{aligned} \tag{2.7}$$

This is just the two variable extension of equation (2.1). The PDF is always positive. The PDF is normalized, summing to unity, over all combinations of *x* and *y*. The Joint PDF gives the intersection of the probability.

The extension of the cumulative discrete probability distribution, equation (2.4), is that for any region *A* in the *x*-*y* plane,

$$F(a,b) = P(x \le a \cap y \le b) = \sum_{x \le a} \sum_{y \le b} f(x,y) \tag{2.8}$$

That is to say, the probability that a result (*x*,*y*) is inside an arbitrary area, A, is equal to the sum of the probabilities for all of the discrete events inside *A*.

Example 2.10.: Consider the discrete Joint PDF, *f*(*x*,*y*), as given in the table below.

		x		
		1	2	3
y	0	1/20	2/20	3/20
	1	4/20	1/20	2/20
	2	2/20	2/20	3/20

Compute the probability that *x* is 1 and *y* is 2.

$$P(x = 1 \cap y = 2) = f(1,2) = \frac{2}{20}$$

Compute the probability that *x* is less than or equal to 1 and *y* is less than or equal to 2.

$$P(x = 1 \cap y \le 2) = \sum_{x=1}^{1} \sum_{y=0}^{2} f(x,y) = f(1,0) + f(1,1) + f(1,2) = \frac{1}{20} + \frac{4}{20} + \frac{2}{20} = \frac{7}{20}$$

2.7. Continuous Joint Probability Density Functions

The distribution of continuous variables can be extended in an exactly analogous manner as was done in the discrete case. The function *f*(*x*,*y*) is a Joint Density Function of the continuous random variables, *x* and *y*, if

$$f(x,y) \geq 0 \quad \text{for all x, y} \in \text{R}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \qquad (2.9)$$

$$P[(x,y) \in A] = \iint_A f(x,y)dxdy$$

This is just the two variable extension of equation (2.4). The PDF is always positive. The PDF is normalized, integrating to unity, over all combinations of *x* and *y*. The Joint PDF gives the intersection of the probability. That third equation takes a specific form, depending on the shape of the Area *A*. For a rectangle, it would look like:

$$P(a < x < b \cap c < y < d) = \int_c^d \int_a^b f(x,y)dxdy$$

Naturally, the cumulative distribution of the single variable case can also be extended to 2-variables.

$$F(x,y) = P(x \leq a \cap y \leq b) = \int_{-\infty}^{b}\int_{-\infty}^{a} f(x,y)dxdy \qquad (2.10)$$

Example 2.11.: Given the continuous Joint PDF, find $P(0 \leq x \leq 0.5 \cap 0.5 \leq y \leq 1)$

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & \text{for } 0 \leq \text{x} \leq 1, 0 \leq \text{y} \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$P(0 < x < \frac{1}{2} \cap \frac{1}{2} < y < 1) = \int_{\frac{1}{2}}^{1}\int_{0}^{\frac{1}{2}} \frac{2}{5}(2x+3y)dxdy$$

$$= \int_{0.5}^{1}\left[\frac{2x^2}{5} + \frac{6\text{xy}}{5}\right]_0^{0.5} dy = \int_{0.5}^{1}\left(\frac{1}{10} + \frac{3\text{y}}{5}\right)dy = \left[\frac{y}{10} + \frac{3\text{y}^2}{10}\right]_{0.5}^{1} = \frac{11}{40}$$

At this point, we should point out two things. First, we have presented four cases for discrete and continuous PDFs for one or two random variables. There are really only two core equations, the requirements for the probability distribution and the definition of the cumulative probability distribution. We have shown these 2 equations for 4 cases; (i) discrete, one variable, (ii) continuous one variable, (iii) discrete, two variable, and (iv) continuous 2 variable. Re-examine these eight equations to make sure that you see the similarities.

In this text, we are stopping at two variables. However, discrete and continuous probability distributions can be functions of an arbitrary number of variables.

2.8. Marginal Distributions and Conditional Probabilities

Marginal distributions give us the probability of obtaining one variable outcome regardless of the value of the other variable. Marginal distributions are needed to calculate conditional probabilities. The marginal distributions of *x* alone and of *y* alone are

$$g(x) = \sum_y f(x,y) \quad \text{and} \quad h(y) = \sum_x f(x,y) \tag{2.11}$$

for the discrete case and

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)dx \tag{2.12}$$

for the continuous case.

Example 2.12.: The discrete joint density function is given by the following table.

		x		
		1	2	3
y	0	1/20	2/20	3/20
	1	4/20	1/20	2/20
	2	2/20	2/20	3/20

Compute the marginal distribution of *x* at all possible values of *x*:

$g(x=1) = f(1,0) + f(1,1) + f(1,2) = 7/20$
$g(x=2) = f(2,0) + f(2,1) + f(2,2) = 5/20$
$g(x=3) = f(3,0) + f(3,1) + f(3,2) = 8/20$

Compute the marginal distribution of *y* at all possible values of *y*:

$h(y=0) = f(1,0) + f(2,0) + f(3,0) = 6/20$
$h(y=1) = f(1,1) + f(2,1) + f(3,1) = 7/20$
$h(y=2) = f(1,2) + f(2,2) + f(3,2) = 7/20$

We note that both marginal distributions are legitimate PDFs and satisfy the three requirements of equation (2.1), namely that they are non-negative, normalized and their evaluation yields probabilities.

Example 2.13.: The continuous joint density function is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & \text{for } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find g(x) and h(y) for this joint density function.

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{0} 0dy + \int_{0}^{1} \frac{2}{5}(2x+3y)dy + \int_{1}^{\infty} 0dy$$

$$= \frac{2}{5}\left(2xy + \frac{3}{2}y^2\right)\Bigg|_0^1 = \frac{4x}{5} + \frac{3}{5}$$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{1} \frac{2}{5}(2x+3y)dx + \int_{1}^{\infty} 0dx$$

$$= \frac{2}{5}\left(x^2 + 3yx\right)\Big|_0^1 = \frac{2}{5} + \frac{6\text{y}}{5}$$

These marginal distributions themselves satisfy all the properties of a probability density distribution, namely the requirements in equation (2.4). The physical meaning of the marginal distribution functions are that they give the individual effects of *x* and *y* separately.

Conditional Probability

We now relate the conditional probability to the marginal distributions defined above. We do this first for the discrete case and then for the continuous case.

Let *x* and *y* be two discrete random variables. The conditional distribution of the random variable *y*=*b*, given that *x*=*a*, is

$$f(y=b|x=a) = \frac{f(x=a, y=b)}{g(x=a)} \quad \text{where } g(a) > 0 \tag{2.13}$$

Similarly, the conditional distribution of the random variable *x*=*a*, given that *y*=*b*, is

$$f(x=a|y=b) = \frac{f(x=a, y=b)}{h(y=b)} \quad \text{where } h(b) > 0 \tag{2.14}$$

You should see that this conditional distribution is simply the application of the definition of the conditional probability, which we learned in Chapter 1,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad \text{for } P(A) > 0 \tag{1.38}$$

Example 2.14.: Given the discrete PDF in Example 2.12., calculate

(a) $f(y=2 | x=2)$

(b) $f(x=1 | y \le 2)$

(a) $f(y=2 | x=2)$. Using the conditional probability definition:

$$f(y=b | x=a) = \frac{f(x=a, y=b)}{g(x=a)}$$

We already have the denominator: $g(x{=}2) = 5/20$. The numerator is $f(x{=}2,y{=}2) = 2/20$. Therefore, the conditional probability is:

$$f(y=2 | x=2) = \frac{2/20}{5/20} = \frac{2}{5}$$

(b) $f(x=1 | y \le 2)$

$$f(x=1 | y \le 2) = \frac{f(x=1, y \le 2)}{h(y \le 2)}$$

The numerator is the sum over all values of $f(x,y)$ for which x=1, and $y \le 2$. So

$$f(x=1, y \le 2) = f(1,0) + f(1,1) + f(1,2) = \frac{1}{20} + \frac{4}{20} + \frac{2}{20} = \frac{7}{20}$$

The denominator is the sum over all $h(y)$ for $y \le 2$

$$h(y \le 2) = h(0) + h(1) + h(2) = \frac{6}{20} + \frac{7}{20} + \frac{7}{20} = \frac{20}{20} = 1$$

Therefore,

$$f(x=1|y\le 2)=\frac{7/20}{20/20}=\frac{7}{20}$$

A similar treatment can be done for the continuous case. Let *x* and *y* be two continuous variables. The conditional distribution of the random variable *c<y<d*, given that *a<x<b*, is

$$P(c<y<d|a<x<b)=\frac{\int_c^d\int_a^b f(x,y)dxdy}{\int_a^b g(x)dx}\quad where \int_a^b g(x)dx\ >0 \tag{2.15}$$

Similarly, the conditional distribution of the random variable *a<x<b*, given that *c<y<d*, is

$$P(a<x<b|c<y<d)=\frac{\int_c^d\int_a^b f(x,y)dxdy}{\int_c^d h(y)dy}\quad where \int_c^d h(y)dy\ >0 \tag{2.16}$$

Example 2.15.: Consider the continuous joint PDF in problem 2.13. Calculate $P(0<X<0.5|0.5<y<1)$.

$$P(0<X<0.5|0.5<y<1)=\frac{P(0<X<0.5\cap 0.5<y<1)}{P(0<X<0.5)}$$

$$P(0<X<0.5|0.5<y<1)=\frac{\int_{0.5}^{1}\int_{0}^{0.5} f(x,y)dxdy}{\int_{0.5}^{1} h(y)dy}$$

We calculated the numerator in Example 2.11. and it had a numerical value of 11/40.

The denominator is:

$$\int_{0.5}^{1} h(y)dy=\int_{0.5}^{1}\left(\frac{2}{5}+\frac{6y}{5}\right)dy=\left[\frac{2y}{5}+\frac{6y^2}{10}\right]_{0.5}^{1}=\frac{13}{20}$$

The conditional probability is then

$$P(0 < X < 0.5 | 0.5 < y < 1) == \frac{\frac{11}{40}}{\frac{13}{20}} = \frac{11}{26}$$

2.9. Statistical Independence

In Chapter 1, we used the conditional probability rule to as a check for independence of two outcomes. This same approach is repeated here for two random variables. Let *x* and *y* be two random variables, discrete or continuous, with joint probability distribution *f*(*x*,*y*) and marginal distributions *g*(*x*) and *h*(*y*). The random variables *x* and *y* are said to be statistically independent iff (if and only if)

$$f(x, y) = g(x)h(y) \text{ if and only if } x \text{ and } y \text{ are independent} \tag{2.17}$$

for all possible values of (*x*,*y*). This should be compared with the rule for independence of probabilities:

$$P(A \cap B) = P(A)P(B) \text{ iff and } A \text{ and } B \text{ are independent events} \tag{1.44}$$

Example 2.16.: In the continuous example given above, determine whether *x* and *y* are statistically independent random variables.

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y) & \text{for } 0 \le \text{x} \le 1, 0 \le \text{y} \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$g(x) = \frac{4x}{5} + \frac{3}{5} \text{ and } h(y) = \frac{2}{5} + \frac{6\text{y}}{5}$$

$$g(x)h(y) = \left(\frac{4x}{5} + \frac{3}{5}\right)\left(\frac{2}{5} + \frac{6\text{y}}{5}\right) = \frac{1}{25}(8x + 24xy + 6 + 18y)$$

The product of marginal distributions is not equal to the joint probability density distribution. Therefore, the variables are not statistically independent.

2.10. Problems

Problem 2.1.

Determine the value of c so that the following functions can serve as a PDF of the discrete random variable *X*.

$$f(x) = c(x^2 + 4) \text{ where } x = 0,1,2,3;$$

Problem 2.2.

A shipment of 7 computer monitors contains 2 defective monitors. A business makes a random purchase of 3 monitors. If x is the number of defective monitors purchased by the company, find the probability distribution of X. (This means you need three numbers, f(x=0), f(x=1), and f(x=2) because the random variable, X = number of defective monitors purchased, has a range from 0 to 2. Also, find the cumulative PDF, F(x). Plot the PDF and the cumulative PDF. These two plots must be turned into class on the day the homework is due.

Problem 2.3.

A continuous random variable, X, that can assume values between x=2 and x=5 has a PDF given by

$$f(x) = \frac{2}{27}(1 + x)$$

Find (a) P(X<4) and find (b) P(3<X<4). Plot the PDF and the cumulative PDF.

Problem 2.4.

Consider a system of particles that sit in an electric field where the energy of interaction with the electric field is given by E(x) = 2477.572 + 4955.144x, where x is spatial position of the particles. The probability distribution of the particles is given by statistical mechanics to be f(x) = c*exp(-E(x)/(R*T)) for 0<x<1 and 0 otherwise, where R = 8.314 J/mol/K and T = 270.0 Kelvin.

(a) Find the value of c that makes this a legitimate PDF.
(b) Find the probability that a particles sits at x<0.25
(c) Find the probability that a particles sits at x>0.75
(d) Find the probability that a particles sits at 0.25<x<0.75

Problem 2.5.

Let X denote the reaction time, in seconds, to a certain stimulant and Y denote the temperature (reduced units) at which a certain reaction starts to take place. Suppose that the random variables X and Y have the joint PDF,

$$f(x,y) = \begin{cases} cxy \text{ for } 0 < x < 1; 0 < y < 2.1 \\ 0 \qquad \text{elsewhere} \end{cases}$$

where c = 0.907029. Find (a) $P\left(0 \le X \le \frac{1}{2} \text{ and } \frac{1}{4} \le Y \le \frac{1}{2}\right)$ and (b) $P(X < Y)$.

Problem 2.6.

Let X denote the number of times that a control machine malfunctions per day (choices: 1, 2, 3) and Y denote the number of times a technician is called. f(x,y) is given in tabular form.

f(x,y)	**x**	**1**	**2**	**3**
y	1	0.05	0.05	0.1
	2	0.05	0.1	0.35
	3	0.0	0.2	0.1

(a) Evaluate the marginal distribution of X.
(b) Evaluate the marginal distribution of Y.
(c) Find P(Y = 3|X = 2).