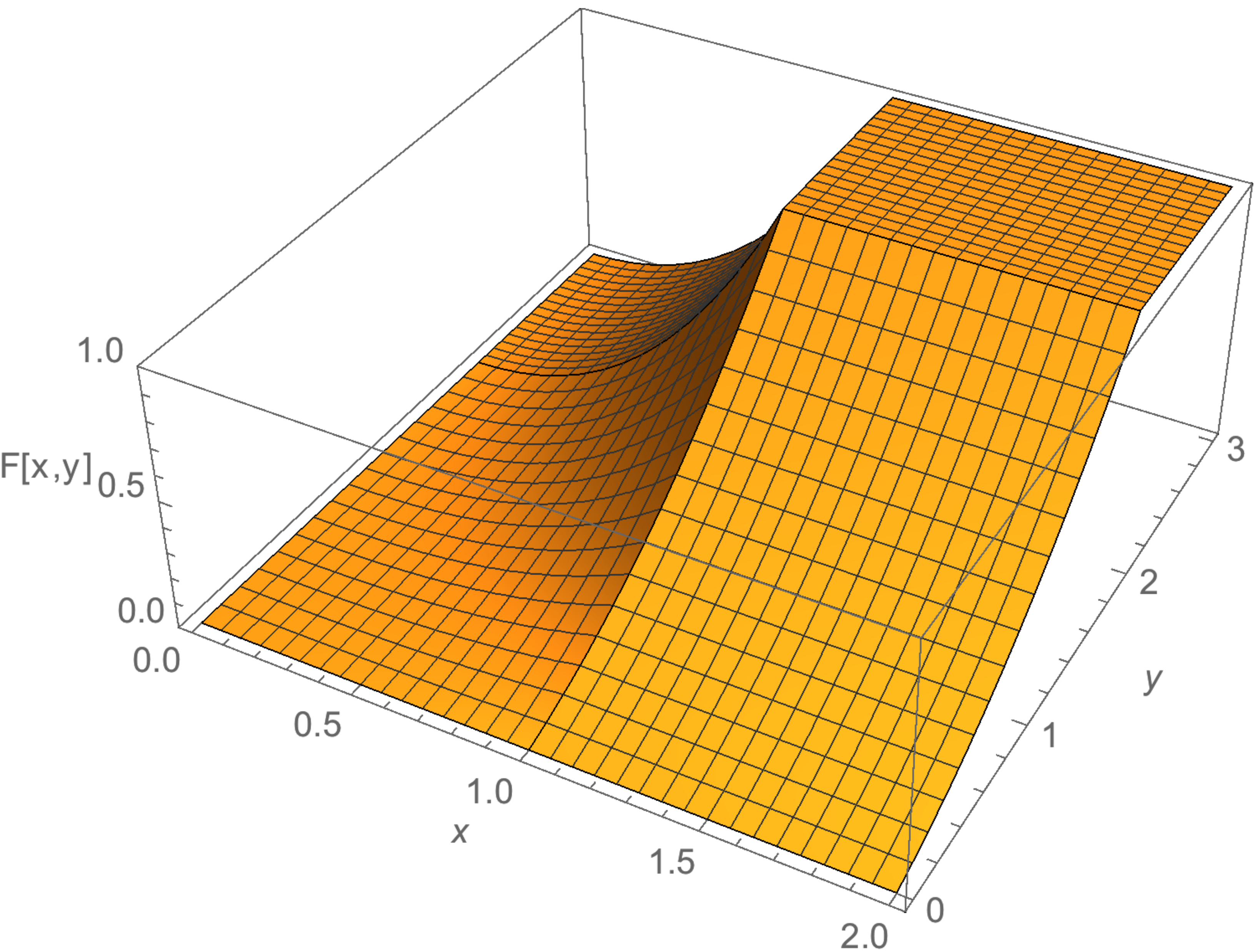


1.0
F[x,y]
0.5
0.0
0.0
0.5
1.0
x
1.5
2.0
0
1
y
2
3

Mathematical Statistics I

Based on course notes developed using Freund's
Mathematical Statistics

Mathematical Statistics I

Based on course notes developed using Freund's Mathematical Statistics

Sean M. Laverty
University of Central Oklahoma

Mathematical Statistics I

March 20, 2021

Sean M. Laverty is writing these notes to (hopefully) give us all a better chance at making it through the semester.

Cover Design: Sean M. Laverty

Edition: Emergency Edition 2021

Website: www.math.uco.edu

To students of mathematical and life sciences
everywhere but specifically at UCO.

Acknowledgements

PreTexT group.

Preface

See the Table of Contents for more.

Sean M. Laverty
Edmond, Oklahoma 2020

Contributors to the 0^{th} Edition

Many individuals have made this book possible. We will try to thank a few of them here, and hope we have not forgotten anybody really important.

Sean M. Laverty
Department of Mathematics and Statistics
University of Central Oklahoma
Edmond, Oklahoma, USA
slaverty@uco.edu

Contents

Chapter 1

Counting

We will learn how to count.

1.1 Counting

Some words.

Definition 1.1.1 $1, 2, 3, \ldots$ $\Diamond$

Chapter 2

Probability

We will learn about probability concepts.

2.1 Probability

Some words.

Definition 2.1.1 Sample text. $\Diamond$

Chapter 3

Discrete Random Variables

In this chapter we will step through using mathematical functions to describe probabilities associated with outcomes of experiments where the resulting measurements are discrete in nature. We will essentially revisit all of the ideas from Chapter **??**, but rely on the convenience of mathematical functions to replicate and advance our thinking and capabilities.

Given that, we will then start to make calculations about certain statistically and practically important properties of random variables using the concept of *mathematical expectation.* Following that, we will look at specific mathematical formulas for specific kinds of discrete random variables that describe common situations.

3.1 Probability distributions

Some words.

3.1.1 Random variables and probability distributions

See sections 3.1, 3.2. Recommended problems: (pg80)1,3,4ab,5,6,7a

Definition 3.1.1 probability distribution. If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the **probability distribution**. $\Diamond$

Theorem 3.1.2 conditions for probability distribution. *A function can serve as the probability distribution for a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions*

1. $f(x) \geq 0$ *for each value within its domain;*
2. $\sum_x f(x) = 1$, *where the summation extends over all the values within its domain*

Definition 3.1.3 distribution function. If X is a discrete random variable, the function given by where $f(t)$ is the value of the probability distribution of X at t, is called the **distribution function**, or the **cumulative distribution (function)**, of X. $\Diamond$

Theorem 3.1.4 properties of a distribution function. *The values $F(x)$ of the distribution function of a discrete random variable X satisfy the condi-*

tions

1. $F(-\infty) = 0$ *and* $F(\infty) = 1$ *(more carefully stated as* $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$*);*
2. *if* $a < b$*, then* $F(a) \leq F(b)$ *for any real numbers* a *and* b

Example 3.1.5 Verifying a simple probability mass function. Identify whether the function $f(x) = \frac{2x-1}{5}$ is a probability mass function for the discrete random variable X with values $x = 1, 2, 3, 4, 5$. If it is not a valid distribution "fix it".

Solution. Notice that values of $f(x) \geq 0$ for all possible x, but

$$\sum_{x=1}^{5} f(x) = 5 \neq 1.$$

Table 3.1.6

x	$f(x)$
1	$f(1) = \frac{1}{5}$
2	$f(2) = \frac{3}{5}$
3	$f(3) = \frac{5}{5}$
4	$f(4) = \frac{7}{5}$
5	$f(5) = \frac{9}{5}$

The function must be rescaled so that probabilities sum to one. The function $f(x) = \frac{2x-1}{25}$ is a valid probability mass function for the values of the random variable mentioned above.

Table 3.1.7

x	$f(x) = \frac{2x-1}{25}$
1	$f(1) = \frac{1}{25}$
2	$f(2) = \frac{3}{25}$
3	$f(3) = \frac{5}{25}$
4	$f(4) = \frac{7}{25}$
5	$f(5) = \frac{9}{25}$

□

Example 3.1.8 CDF of a simple probability mass function. The function $f(x) = \frac{2x-1}{25}$ is a probability mass function for the discrete random variable X with values $x = 1, 2, 3, 4, 5$. Find the cumulative distribution function.

Solution.

Table 3.1.9

x	$\Pr(X \leq x)$	$F(x)$
1	$f(1)$	$\frac{1}{25}$
2	$f(1) + f(2)$	$\frac{1}{25} + \frac{3}{25}$
3	$f(1) + f(2) + f(3)$	$\frac{4}{25} + \frac{5}{25}$
4	$f(1) + f(2) + f(3) + f(4)$	$\frac{9}{25} + \frac{7}{25}$
5	$f(1) + f(2) + f(3) + f(4) + f(5)$	$\frac{16}{25} + \frac{9}{25}$

We have

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{1}{25}, & 1 \leq x < 2 \\ \frac{4}{25}, & 2 \leq x < 3 \\ \frac{9}{25}, & 3 \leq x < 4 \\ \frac{16}{25}, & 4 \leq x < 5 \\ 1, & 5 \leq x \end{cases}$$

The function $F(x)$, is piecewise constant, accumulating new probability at the values of $x = 1, 2, 3, 4, 5$. □

3.1.2 Exercises

1. **Problem 3.1.** For each of the following, determine whether the given values can serve as the values of a probability distribution of a random variable with the range $x = 1, 2, 3,$ and 4:

   (a) $f(1) = 0.25, f(2) = 0.75, f(3) = 0.25,$ and $f(4) = 0.25;$

   (b) $f(1) = 0.15, f(2) = 0.27, f(3) = 0.29,$ and $f(4) = 0.29;$

   (c) $f(1) = \dfrac{1}{19}, f(2) = \dfrac{10}{19}, f(3) = \dfrac{2}{19},$ and $f(4) = \dfrac{5}{19}.$

2. **Problem 3.3.** Verify that $f(x) = \dfrac{2x}{k(k+1)}$ for $x = 1, 2, 3, \ldots, k$ can serve as the probability distribution of a random variable with the given range. *Bonus:* Find or describe $F(x)$.}

3. **Problem 3.4b.** Determine c so that the function $f(x) = c\dbinom{5}{x}$ for $x = 0, 1, 2, 3, 4, 5$ can serve as a probability distribution of a random variable with the given range.

4. **Problem 3.5.** For what values of k an

$$f(x) = (1-k)k^x$$

   serve as the values of the probability distribution of a random variable with countably infinite range $x = 0, 1, 2, \ldots$?

5. **Problem 3.7a.** Show that there are no values of c such that

$$f(x) = \frac{c}{x}$$

   can serve as the values of the probability distribution of a random variable with countably infinite range $x = 0, 1, 2, \ldots$

6. **Problem 3.6.** Construct a probability histogram for the probability distribution

$$f(x) = \frac{\binom{2}{x}\binom{4}{3-x}}{\binom{6}{3}} \text{ for } x = 0, 1, 2$$

7. **Problem 3.87.** The probability distribution of V, the number of weekly accidents at a certain intersection, is given by $g(0) = 0.4, g(1) = 0.3, g(2) = 0.2, g(3) = 0.1$. Construct the distribution function of V and draw its graph.

8. **Problem 3.88.** Using the statement and result of the previous problem, find the probability that there will be at least two accidents in any one

week using

(a) the original probabilities, and

(b) the values of the distribution function.

3.2 Mathematical expectation of discrete random variables

This is the start of Chapter 4 in Freund's Mathematical Statistics. In the first pass we will study the major topics of this chapter with a focus on those applying to discrete random variables.

3.2.1 Expected value

See section 4.1 - 4.2. Recommended problems: (pg 136) 7, 9, 10, 11, (pg 161) 60

Definition 3.2.1 expected value. If X is a discrete random variable and $f(x)$ is the value of its probability distribution at X, the **expected value** of X is given by

$$E[X] = \sum_x x \cdot f(x)$$

$\diamondsuit$

Theorem 3.2.2 expected value of a function of a random variable. *If X is a discrete random variable and $f(x)$ is the value of its probability distribution at X, the expected value of $g(X)$ is given by*

$$E[g(X)] = \sum_x g(x) \cdot f(x)$$

Theorem 3.2.3 expectation of a linear function. *If a and b are constants, then $E[aX + b] = aE[X] + b$.*

Corollary 3.2.4 *If a is a constant, then $E[aX] = aE[X]$.*

Corollary 3.2.5 *If b is a constant, then $E[b] = b$.*

3.2.2 Exercises

1. **Problem 4.9.** Suppose that X takes on values $0, 1, 2, 3$ with probabilities $\frac{1}{125}, \frac{12}{125}, \frac{48}{125}, \frac{64}{125}$

   (a) Find $E[X]$ and $E[X^2]$

   (b) Determine the value of $E[(3X + 2)^2]$

3.2.3 Expected value

See section 4.3. Recommended problems: 4.3 (pg 146) 20, 22, 23, 31, 33, 34, 40, (pg 162) 69, 73, 75

Definition 3.2.6 moments about the origin. The r^{th} **moment about the origin** of a random variable X, denoted by μ'_r, is the expected value of

$(X)^r$; symbolically

$$\mu'_r = E[(X)^r] = \sum_x x^r \cdot f(x)$$

for $r = 0, 1, 2, \ldots$ when X is discrete. $\diamondsuit$

Definition 3.2.7 mean of a discrete random variable. μ'_1 is called the **mean** of the distribution of X, or simply the **mean** of X; and it is denoted by μ. $\diamondsuit$

Definition 3.2.8 moments about the mean. The r^{th} **moment about the mean** of a random variable X, denoted by μ_r, is the expected value of $(X - \mu)^r$; symbolically

$$\mu_r = E[(X - \mu)^r] = \sum_x (x - \mu)^r \cdot f(x)$$

for $r = 0, 1, 2, \ldots$ when X is discrete. $\diamondsuit$

Now, you could imagine in some cases the moments being difficult to calculate as sums. We sometimes take the approach of building what are called moment-generating functions. These are mathematical functions whose purpose is to generate the moments of a distribution that we might need.

3.2.4 Exercises

1. **Problem 4.18.** Find μ, μ'_2, and σ^2 for the random variable X that has probability distribution $f(x) = 0.5$ for $x = \pm 2$.

3.2.5 Moment-generating functions

See section 4.5.

Definition 3.2.9 moment-generating function. The **moment-generating function** of a random variable X, where it exists, is given by

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} \cdot f(x)$$

when X is discrete. $\diamondsuit$

Notce that a moment-generating function $M_X(t)$ itself is a function of the variable t not X. As it turns out, we are most interested in values of the function at or near $t = 0$.

Example 3.2.10 moment-generating function via Taylor series. Recall that the Maclaurin series (Taylor series around zero) for e^{tx} is

$$e^{tx} = 1 + tx + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \cdots + \frac{1}{r!}(tx)^r + \cdots$$

This means (in the discrete case), that

$$\begin{aligned}
M_X(t) &= \sum_x e^{tx} f(x) \\
&= \sum_x \left(1 + tx + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \cdots + \frac{1}{r!}(tx)^r + \cdots\right) f(x) \\
&= \sum_x (f(x) + txf(x) + \frac{t^2}{2!}x^2 f(x) + \frac{t^3}{3!}x^3 f(x) + \cdots + \frac{t^r}{r!}x^r f(x) + \cdots) \\
&= \sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \cdots + \frac{t^r}{r!}\sum_x x^r f(x) + \cdots
\end{aligned}$$

Looking closely, at

$$M_X(t) = \sum_x f(x) + \left(\sum_x xf(x)\right)t + \left(\sum_x x^2 f(x)\right)\frac{t^2}{2!} + \left(\sum_x x^3 f(x)\right)\frac{t^3}{3!} + \cdots + \left(\sum_x x^r f(x)\right)\frac{t^r}{r!} + \cdots$$

coefficients of the terms $\frac{t^r}{r!}$ are the moments about the origin $\mu'_r = \sum_x x^r f(x)$

□

Example 3.2.11 moment-generating function for three cards. Recall the probability distribution $f(x) = P(X = x) = \frac{\binom{3}{x}}{8}$ for $x = 0, 1, 2, 3$ (this was used earlier to determine the probabilities of x heads on three flips of a coin).

$$\begin{aligned}
M_X(t) &= \sum_x e^{tx} f(x) \\
&= 1 \cdot \frac{1}{8} + e^t \cdot \frac{3}{8} + e^{2t} \cdot \frac{3}{8} + e^{3t} \cdot \frac{1}{8} \\
&= \frac{1}{8}\left(1 + 3e^t + 3e^{2t} + e^{3t}\right) \\
M_X(t) &= \frac{1}{8}(1 + e^t)^3
\end{aligned}$$

□

Theorem 3.2.12 moments via differentiation. *The r^{th} moment about the origin, μ'_r, can be written*

$$\mu'_r = \frac{d^r M_X(t)}{dt^r}\Big|_{t=0}$$

Example 3.2.13 moment-generating function for three cards, via differentiation. Referencing Example Example **??**, the mean of the random variable, given by μ'_1, whose MGF is $M_X(t) = \frac{1}{8}(1 + e^t)^3$ is found as follows

$$\begin{aligned}
\mu'_1 &= \left(\frac{d}{dt}\left(\frac{1}{8}(1 + e^t)^3\right)\right)\Big|_{t=0} \\
&= \left(\frac{3}{8}(1 + e^t)^2 e^t\right)\Big|_{t=0} \\
&= \left(\frac{3}{8}(2)^2\right) \\
\mu'_1 &= \frac{3}{2}
\end{aligned}$$

□

So, *given* a moment-generating function, a relatively simple application of calculus allows us to replace a more tedious calculation of the moment from its definition.

Theorem 3.2.14 moment-generating function of functions of a random variable. *If a and b are constants, then*

1. $M_{X+a}(t) = E[e^{(X+a)t}] = e^{at} \cdot M_X(t)$;
2. $M_{bX}(t) = E[e^{bXt}] = M_X(bt)$;
3. $M_{\frac{X+a}{b}}(t) = E\left[e^{\left(\frac{X+a}{b}\right)t}\right] = e^{(a/b)t} \cdot M_X\left(\frac{t}{b}\right)$;

The rules in Theorem **??** allow us to calculate moment-generating functions of simple functions of a random variable.

3.2.6 Exercises

1. **Problem 4.33.** Find the moment-generating function of the discrete random variable X that has the probability distribution

$$f(x) = 2\left(\frac{1}{3}\right)^x \quad \text{for } x = 1, 2, 3, \ldots$$

and use it to determine the values of μ_1' and μ_2'.

2. **Problem 4.40.** Given the moment-generating function $X_X(t) = e^{3t+8t^2}$, find the moment generating function of the random variable $Z = \frac{1}{4}(X-3)$ and use it to determine the mean and variance of Z.

3.2.7 Moments of linear combinations of random variables

See section 4.7. Recommended problems: 4.7-8 (pg 158) 48, 49, 57

Theorem 3.2.15 variance. *If $X_1, X_2, \ldots, X_n$ are random variables and $Y = \sum_{i=1}^{n} a_i X_i$ where $a_1, a_2, \ldots, a_n$ are constants, then*

$$E[Y] = \sum_{i=1}^{n} a_i E[X_i]$$

and

$$\mathrm{var}[Y] = \sum_{i=1}^{n} a_i^2 \,\mathrm{var}[X_i] + 2\sum_{i<j}\sum a_i a_j \,\mathrm{cov}[X_i, X_j].$$

Corollary 3.2.16 variance of independent random variables. *If $X_1, X_2, \ldots, X_n$ are independent random variables and $Y = \sum_{i=1}^{n} a_i X_i$ where $a_1, a_2, \ldots, a_n$ are constants, then*

$$\mathrm{var}[Y] = \sum_{i=1}^{n} a_i^2 \,\mathrm{var}[X_i]$$

Example 3.2.17 covariances of linear combinations. Consider three random variables X, Y, and Z with $\mu_X = 2$, with $\mu_Y = -3$, with $\mu_Z = 4$;

with $\sigma_X^2 = 1$, $\sigma_Y^2 = 5$, $\sigma_Z^2 = 2$; and $\text{cov}(X, Y) = -2$, $\text{cov}(X, Z) = -1$, and $\text{cov}(Y, Z) = 1$.

Find μ_W and $\text{var}(W) = \sigma_W^2$ for $W = 3X - Y + 2Z$.**Solution**. First, $\mu_W = (3)\mu_X + (-1)\mu_Y + (2)\mu_Z = 17$.

We could apply the theorem directly, but we can do this more directly with linear algebra. The idea is that we can picture the linear combination $W = 3X - Y + 2Z$ as

$$W = \begin{bmatrix} 3 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = (3)X + (-1)Y + (2)Z$$

Let a be the row vector $a = \begin{bmatrix} 3 & -1 & 2 \end{bmatrix}$, its transpose be the column vector a^T, and the matrix Σ be defined as follows,

$$\Sigma = \begin{bmatrix} 1 & -2 & -1 \\ -2 & 5 & 1 \\ -1 & 1 & 2 \end{bmatrix}$$

This approach can be justified by expanding the sums in Theorem **??** with a sum of 2 random variables.

We can calculate the variance of W by $\text{var}(W) = a\Sigma a^T$. Specifically,

$$a\Sigma a^T = \begin{bmatrix} 3 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & -2 & -1 \\ -2 & 5 & 1 \\ -1 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

Notice *Sigma* is symmetric and that the covariances lie in order along the main diagonal and the variances off-diagonal.

Multiplying the square matrix and column vector first, we have

$$a\Sigma a^T = \begin{bmatrix} 3 & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ -9 \\ 0 \end{bmatrix}$$

And finally, $a\Sigma a^T = 18$. □

Theorem 3.2.18 covariance of two linear combinations. *If $X_1, X_2, \ldots, X_n$ are random variables and $Y_1 = \sum_{i=1}^{n} a_i X_i$ and $Y_2 = \sum_{i=1}^{n} b_i X_i$ where $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ are constants, then*

$$\text{cov}[Y_1, Y_2] = \sum_{i=1}^{n} a_i b_i \,\text{var}[X_i] + \sum_{i<j}\sum (a_i b_j + a_j b_i) \,\text{cov}[X_i, X_j].$$

Corollary 3.2.19 *If $X_1, X_2, \ldots, X_n$ are independent random variables and $Y_1 = \sum_{i=1}^{n} a_i X_i$ and $Y_2 = \sum_{i=1}^{n} b_i X_i$, then*

$$\text{cov}[Y_1, Y_2] = \sum_{i=1}^{n} a_i b_i \,\text{var}[X_i]$$

The same logic used in Example **??** allows us to compute the covariance between two linear combinations of random variables directly also. Instead of calculating $a\Sigma a^T$ we will calculate $a\Sigma b^T$ where b is the vector of coefficients of the second linear combination.

3.2.8 Exercises

1. **Problem 4.48.** If X_1, X_2, X_3 are independent and have the means $4, 9, 3$ and the variances $3, 7, 5$, find the mean and variance of show that

(a) $Y = 2X_1 - 3X_2 + 4X_3$;

(b) $Z = X_1 + 2X_2 - X_3$.

2. **Problem 4.49.** Repeat both parts of the previous exercise after dropping the assumption of independence and using instead that $\text{cov}(X_1, X_2) = 1$, $\text{cov}(X_2, X_3) = -2$, $\text{cov}(X_1, X_3) = -3$.

3.3 Special probability distributions

Now we look at a collection of formulas for probability distributions. These describe how probabilities are assigned across a range of discrete values of a random variable. It turns out to be the case that values of random variables associated with certain kinds of experiments follow certain formulas. In this chapter, we will look at the formulas that describe certain discrete random variables.

For the purposes of generality we will emphasize that certain formulas include one or more parameters, these are symbols to which we will associate particular numerical values, much like we do m and b in the formula $y = f(x) = mx + b$. We should have a general understanding of something that follows this relationship and an understanding of the roles of both m and b in specifying that relationship.

3.3.1 Discrete uniform

Definition 3.3.1 discrete uniform distribution. A discrete random variable X has a **discrete uniform distribution** and is referred to as a **discrete uniform random variable** if and only if its probability distribution is given by

$$f(x) = \frac{1}{k} \text{ for } x = x_1, x_2, \ldots, x_k \text{ where } x_i \neq x_j \text{ for } i \neq j.$$

◊

This is sometimes written as DU(k), where k is the parameter of the distribution.

Example 3.3.2 mean of a discrete uniform distribution. Find the mean of a discrete uniform random variable (or of the discrete uniform distribution).**Solution**. Recall that

$$f(x) = \frac{1}{k} \text{ for } x = x_1, x_2, \ldots, x_k \text{ where } x_i \neq x_j \text{ for } i \neq j$$

and that

$$\mu = \sum_x x \cdot f(x).$$

This becomes

$$\mu = \sum_{i=1}^{k} x_i \left(\frac{1}{k}\right).$$

This is about as far as we can go in general without knowing more about k or the values of x_i. □

Similar to the previous "Example", the variance of a discrete uniform random variable (or of the discrete uniform distribution) is given by

$$\sigma^2 = \sum_x (x-\mu)^2 \cdot f(x) = \sum_x (x-\mu)^2 \left(\frac{1}{k}\right).$$

Once again, this is about as far as we can go in general.

Example 3.3.3 simple discrete uniform distribution. Find the mean of a discrete uniform random variable (or of the discrete uniform distribution) for which $x_i = i$.**Solution**. Recall that

$$f(x) = \frac{1}{k} \text{ for } x = 1, 2, \ldots, k$$

and that

$$\mu = \sum_x x \cdot f(x).$$

This becomes

$$\begin{aligned} \mu &= \sum_{i=1}^{k} i \left(\frac{1}{k}\right) \qquad . \\ &= \frac{1}{k}\left(\sum_{i=1}^{k} i\right) \\ &= \frac{1}{k}\left(\frac{k(k+1)}{2}\right) \\ \mu &= \frac{k+1}{2} \end{aligned}$$

Now, verify our earlier work on the mean number of dots to show on a balanced, 6-sided die. □

3.3.2 Exercises

1. **5.2.** If X has a discrete uniform distribution $f(x) = \frac{1}{k}$ for $x = 1, 2, \ldots, k$, show that its moment-generating function is given by

$$M_x(t) = \frac{e^t(1-e^{kt})}{k(1-e^t)}.$$

3.3.3 Bernoulli and Binomial distributions

Consider an experiment with two possible outcomes, flipping a coin for example. We consider one of those outcomes, say 'heads' as a success, and the other of those outcomes, in this case 'tails` as a failure. We allow these outcomes to each occur with a given probability. Here θ gives the probability of 'success' and $1-\theta$ is the corresponding probability of 'failure'.

Though the outcome of a single event may be useful, the Bernoulli distribution is perhaps most useful as a building block to describe the results of more complex experiments.

You will find tools for visualization at the following link.

Definition 3.3.4 Bernoulli distribution. A discrete random variable X has a **Bernoulli distribution** and is referred to as a **Bernoulli random variable** if and only if its probability distribution is given by

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} \text{ for } x = 0, 1.$$

◊

Example 3.3.5 mean and variance of a Bernoulli distribution. Find the mean and variance of a Bernoulli random variable (or of the Bernoulli distribution).**Solution**. Generally we calculate the mean as

$$\mu = \sum_x x \cdot f(x).$$

Specifically this becomes

$$\begin{aligned} \mu &= \sum_{i=0}^{1} x f(x;\theta) \\ &= 0\left(\theta^0(1-\theta)^1\right) + 1\left(\theta^1(1-\theta)^0\right) \\ \mu &= \theta \end{aligned}$$

We calculate the variance as $\sigma^2 = E[X^2] - (E[X])^2$. Keep in mind that we now know that $E[X] = \theta$, so we know that $(E[X])^2 = \theta^2$

Now

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{1} x^2 f(x;\theta) \\ &= 0^2\left(\theta^0(1-\theta)^1\right) + 1^2\left(\theta^1(1-\theta)^0\right) \\ E[X^2] &= \theta \end{aligned}$$

Given this,

$$\begin{aligned} \sigma^2 &= E[X^2] - (E[X])^2 \\ &= \theta - \theta^2 \\ \sigma^2 &= \theta(1-\theta) \end{aligned}$$

□

Example 3.3.6 moment-generating function of a Bernoulli distribution. Find the moment-generating function of a Bernoulli random variable (or of the Bernoulli distribution).**Solution**. Recall that $M_X(t) = E[e^{tx}] = \sum_x e^{tx} f(x;\theta)$. This becomes

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{1} e^{tx} f(x;\theta) \\ &= e^{0\cdot t}\left(\theta^0(1-\theta)^1\right) + e^{1\cdot t}\left(\theta^1(1-\theta)^0\right) \\ &= (1-\theta) + \theta e^t \\ M_X(t) &= 1 + \theta(e^t - 1) \end{aligned}$$

□

You could use the moment-generating function and the earlier theorem to calculate moments about the origin used to find the mean and variance.

Consider now a new random variable Y whose value is the sum of independent Bernoulli trials. This new variable has a 'binomial distribution' as defined below.

Definition 3.3.7 binomial distribution. A discrete random variable X has a **binomial distribution** and is referred to as a **binomial random variable** if and only if its probability distribution is given by

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \text{ for } x = 0, 1, \ldots, n.$$

$\Diamond$

In the definition, the term $\binom{n}{x}$ gives the number of ways to select the order of the x successes in n trials. The values of $b(x; n, \theta)$ give the coefficients of terms in the expansion of $((1-\theta)+\theta)^n$.

Example 3.3.8 Binomial coin flips. Find the probability of 5 heads in 9 coin flips under each of the following situations.

1. The coin is balanced and $P(\text{success}) = P(\text{failure}) = 0.5$
2. The coin is unbalanced and $P(\text{success})$ is 3 times larger than $P(\text{failure})$

Solution. This problem can be done by evaluating the binomial using 'Table 1' of the 'Statistical Tables Appendix' (see book), or by using the definition of probability distribution of the binomial random variable.

1. The coin is balanced and $P(\text{success}) = P(\text{failure}) = 0.5$. We are looking to evaluate the binomial probability, sometimes written $\text{Bin}(n = 9, \theta = 0.5)$, specifically $b(5; 9, 0.5)$. This is,
$$b(5; 9, 0.5) = \binom{9}{5} 0.5^5 (1-0.5)^{9-5} \approx 0.2461$$
2. The coin is unbalanced and $P(\text{success})$ is 3 times larger than $P(\text{failure})$. We first have to figure out the value for $P(\text{success}) = \theta$. We have said that $P(\text{success}) = 3 \cdot P(\text{failure})$ or $\theta = 3 \cdot (1-\theta)$. This gives $\theta = 0.75$ and, as above,
$$b(5; 9, 0.75) = \binom{9}{5} 0.75^5 (1-0.75)^{9-5} \approx 0.1168$$

$\square$

Example 3.3.9 Free throws. What is the probability of making 2 in 5 free throws if the probability of making one is 0.86?**Solution**. We need to calculate

$$b(2; 5, 0.86) = \binom{5}{2} 0.86^2 (1-0.86)^{5-2} = 10(0.86)^2(0.14)^3 \approx 0.0203.$$

This calculation corresponds to the ten possible orderings of 2 makes in 5 shot attempts, the probabilities associated with 2 'makes', and the probabilities associated with 3 'misses'. $\square$

Switching focus from x successes in n trials with a probability of success θ, we have $n - x$ failures in n trials with a probability of failure $1 - \theta$.

Theorem 3.3.10 reparameterizing a binomial.

$$b(x; n, \theta) = b(n - x; n, 1 - \theta)$$

Example 3.3.11 Free throws - revisited. Using theorem Theorem **??**, find the probability of making 2 in 5 free throws if the probability of making one is 0.86?**Solution**. We need to calculate

$$\begin{aligned} b(5-2;5,1-0.86) &= b(3;5,0.14) \\ &= \binom{5}{3} 0.14^3 (1-0.14)^{5-3} \\ &= 10(0.14)^3(0.86)^2 \\ &\approx 0.020 \end{aligned}$$

This calculation corresponds to the ten possible orderings of 2 makes in 5 shot attempts, the probabilities associated with 2 'makes', and the probabilities associated with 3 'misses'. □

For certain calculations, it is helpful to use the theorem to reduce the sizes of numbers used in the factorial or to find parameterizations for which the probabilities are known by a table.

Theorem 3.3.12 Mean and variance of binomial distribution. *The mean and variance of a binomial random variable (or of the binomial distribution) are*

$$\mu = n\theta \text{ and } \sigma^2 = n\theta(1-\theta).$$

Proof. These can be proved using the expectation and some clever re-indexing in evaluating the sum. ■

Theorem 3.3.13 Proportion of binomial successes. *If X is a binomially-distributed random variable with parameters n, θ, and $Y = \frac{X}{n}$ gives the proportion of successes,*

$$E[Y] = \theta \text{ and } \sigma_Y^2 = \frac{\theta(1-\theta)}{n}.$$

Checkpoint 3.3.14 probabilities of dice rolls. Find the expected value of the number of times a 2 or 3 shows in 15 rolls of a standard 6-sided die.**Hint**. What is the probability of 'success'?

Example 3.3.15 moment-generating function of a binomial distribution. Find the moment-generating function of a binomial random variable (or of the binomial distribution).**Solution**. Recall that $M_X(t) = E[e^{tx}] = \sum_x e^{tx} f(x;\theta)$. This becomes

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} e^{tx} b(x;n,\theta) \\ &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} \left(\theta e^t\right)^x (1-\theta)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (1-\theta)^{n-x} \left(\theta e^t\right)^x \\ M_X(t) &= \left((1-\theta) + \theta e^t\right)^n \end{aligned}$$

That last step, though a big leap in print, comes from applying the binomial theorem in reverse, that is

$$\sum_{k=0}^{n} \binom{n}{r} a^{n-r} b^r = (a+b)^n$$

where $a = 1-\theta$ and $b = \theta e^t$. □

You might recognize this as n factors of the moment-generating function of a Bernoulli random variable.

You could use the moment-generating function and Theorem **??** to calculate moments about the origin used to find the mean and variance as a mechanism to prove Theorem **??**.

3.3.4 Exercises

1. **Problem 5.7.** Verify Theorem **??**.
2. **Problem 5.10.** If X is a binomial random variable, for what value of θ is the probability $b(x; n, \theta)$ a maximum? In other words, maximize $b(x; n, \theta)$ with respect to θ.

3.3.5 Negative Binomial and Geometric distributions

The binomial distribution describes the probability of a certain number of successes in a certain amount of trials. Sometimes, instead, we are interested in the trial on which a particular success occurs. This is described by the negative binomial distribution.

This situation requires obtaining $k-1$ successes across the first $x-1$ trials, with the k^{th} and final success to occur on the x^{th} trial.

You will find tools for visualization at the following link.

Definition 3.3.16 negative binomial distribution. A discrete random variable X has a **negative binomial distribution** and is referred to as a **negative binomial random variable** if and only if its probability distribution is given by

$$b^*(x; k, \theta) = \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k} \text{ for } x = k, k+1, \ldots.$$

◊

Sometimes we refer to the random variable as being distributed according to $\text{NegBin}(k, \theta)$ or $\text{NB}(k, \theta)$. The values of the random variable describe *binomial waiting-times*, since the result is the number of trials until arriving at a particular outcome of interest.

Theorem 3.3.17 negative binomial probability as a binomial probability.

$$b^*(x; k, \theta) = \frac{k}{x} b(k; x, \theta).$$

Proof. This can be shown, relatively quickly by relatively simple manipulation of the definitions. Notice that we are equating on the left the negative binomial and on the right the binomial. ■

Checkpoint 3.3.18 Free throws with the negative binomial. A player makes a free throw with probability 0.86. What is the probability that the shooter makes her 3^{rd} shot on her 5^{th} attempt?**Solution**. We need to calculate

$$b^*(5; 3, 0.86) = \binom{5-1}{3-1} 0.86^3 (1-0.86)^{5-3} = 6(0.86)^3(0.14)^2.$$

This means 2 shots were made in the first 4 attempts, followed by the 3^rd make on the 5^th attempt.

Theorem 3.3.19 Mean and variance of the negative binomial distribution. *The mean and variance of the negative binomial distribution are*

$$\mu = \frac{k}{\theta}$$

and

$$\sigma^2 = \frac{k}{\theta}\left(\frac{1}{\theta} - 1\right).$$

Example 3.3.20 moment generating function of the negative binomial distribution. Find the moment-generating function of a negative binomial random variable (or of the negative binomial distribution).**Solution**. Recall that $M_X(t) = E[e^{tx}] = \sum_x e^{tx} f(x;\theta)$. This becomes

$$M_X(t) = \left(\frac{\theta e^t}{1-(1-\theta)e^t}\right)^k.$$

□

The question about when the first success occurs is a common one. So common, in fact, that it gets its own name, the *geometric distribution*. Now it should first be noted that we sometimes view 'success' strangely. Often we are instead thinking of a particular 'outcome of interest' rather than the traditional interpretation of 'success' as 'a good thing'.

Definition 3.3.21 geometric distribution. A discrete random variable X has a **geometric distribution** and is referred to as a **geometric random variable** if and only if its probability distribution is given by

$$g(x;\theta) = \theta(1-\theta)^{x-1} \text{ for } x = 1, 2, \ldots.$$

◊

The geometric distribution answers the age-old question "if at first you don't succeed, how many times did you fail?"

Theorem 3.3.22 Mean and variance of the geometric distribution. *The mean and variance of the geometric distribution are*

$$\mu = \frac{1}{\theta}$$

and

$$\sigma^2 = \frac{1}{\theta}\left(\frac{1}{\theta} - 1\right).$$

Sometimes we refer to the random variable as being distributed according to Geom(θ).

3.3.6 Exercises

1. **Problem 5.18.** Prove Theorem **??**.

2. **Problem 5.20.** Show that the moment-generating function of the geometric distribution is given by

$$M_X(t) = \frac{\theta e^t}{1 - e^t(1-\theta)}.$$

3. **Problem 5.21.** Use Theorem **??** and

$$M_X(t) = \frac{\theta e^t}{1 - e^t(1-\theta)}$$

to find μ and σ^2 by differentiation.

3.3.7 Hypergeometric distribution

The Hypergeometric distribution describes the probabilities of sampling from a finite population without replacement. Unlike the binomial where it is assumed that the probability of success is a constant, with the hypergeometric distribution, the probability of success changes with the selection process.

You will find tools for visualization at the following link.

Definition 3.3.23 hypergeometric distribution. A discrete random variable X has a **hypergeometric distribution** and is referred to as a **hypergeometric random variable** if and only if its probability distribution is given by

$$h(x; n, N, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \text{ for } x = 1, 2, \ldots, n; x \leq M \text{ and } n - x \leq N - M.$$

$\diamondsuit$

Theorem 3.3.24 Mean and variance of the hypergeometric distribution. *The mean and variance of the hypergeometric distribution are*

$$\mu = \frac{nM}{N}$$

and

$$\sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}.$$

Remark 3.3.25 There is a typo in the printed book for the class in the second numerator term.

Example 3.3.26 hypergeometric distribution for truck inspections. You randomly choose 6 out of 24 trucks for a new fleet of work trucks. It is known that 4 of the 24 trucks have failed a recent emissions inspection. What is the probability that none of the trucks in your fleet are "polluters"?**Solution**. Let $x = \#$ of polluters selected. Then, we are looking for

$$h(0; 6, 24, 4) = \frac{\binom{4}{0}\binom{24-4}{6-0}}{\binom{24}{6}}.$$

Notice the denominator is the total number of ways that we can choose 6 of 24 trucks. The first term in the numerator is the number of ways that we can choose 0 trucks from the collection of 4 defective trucks. The second term in the numerator is the number of ways that we can choose $6-0$ trucks from the $24-4$ non-defective trucks. $\square$

Remark 3.3.27 Below let $\theta = \frac{M}{N}$ (think through this). The mean and variance of the hypergeometric distribution can be written

$$\mu = \frac{nM}{N} = n\theta$$

and

$$\sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)} = n\theta(1-\theta)\left(\frac{N-n}{N-1}\right).$$

Above, the term $\left(\frac{N-n}{N-1}\right)$ is called the "finite population correction factor".

With this we might interpret the fraction $\theta = \frac{M}{N}$ as the "probability of success", given that we are choosing M successes from N objects.

As indicated at the beginning of this section, the binomial and hypergeometric are related. The binomial describes a situation where sampling is done with replacement, while the hypergeometric describes a situation where sampling is done without replacement.

You will find tools for visualization at the following link.

3.3.8 Exercises

1. 4.xx. xx

3.3.9 Poisson distribution

The Poisson distribution describes the occurrence of events taking place at a constant rate in time or over space. For example, if car accidents take place at a rate of 3 per 100 miles, we would use a Poisson distribution to describe the probabilities of certain numbers of accidents occurring along a known distance of highway.

You will find tools for visualization at the following link.

Definition 3.3.28 Poisson distribution. A discrete random variable X has a **Poisson distribution** and is referred to as a **Poisson random variable** if and only if its probability distribution is given by

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

◊

Theorem 3.3.29 Mean, variance, and MGF of the Poisson distribution. *The mean and variance of the Poisson distribution are*

$$\mu = \lambda$$

and

$$\sigma^2 = \lambda.$$

The moment-generating function of the Poisson distribution is

$$M_X(t) = e^{\lambda(t-1)}$$

The Poisson distribution is derived from the binomial distribution (see pgs. 81-82 in Hogg, Tanis, Zimmerman, 10th ed. for a very good derivation).

Though this is the case, and though the Poisson can be used to approximate the binomial under certain circumstances when the binomial probabilities would be numerically challenging to calculate, many applications of the Poisson distribute have absolutely nothing to do with an underlying binomial process.

You will find tools for visualization at the following link.

As described above, it is generally regarded as safe to use the Poisson as a means of approximating binomial probabilities when $n \geq 20$ and $\theta \leq 0.05$ or if $n > 100$ and $\theta < 0.10$. In some cases the approximation will work quite well even in violation of these bounds. Approximation may be slightly less important in modern times than it was in the past due to the ubiquity of computers and software, though the connection is still worth remembering.

Beyond its use in approximation, the Poisson distribution has numerous applications for calculation probabilities of events occurring over time or across space.

3.3.10 Exercises

1. 4.xx. xx

3.4 Multivariate discrete random variables

Some words.

3.4.1 Multivariate, marginal, and conditional distributions

See sections 3.5, 3.6, 3.7. Recommended problems: (pg 101) 42, 44, 45, 49, 53, 54

Definition 3.4.1 joint probability distribution. If X and Y are discrete random variables, the function given by $f(x,y) = P(X = x, Y = y)$ for each pair of values (x,y) within the range of X and Y is called the **joint probability distribution** of X and Y. ♢

Theorem 3.4.2 conditions for a joint probability distribution. *A bivariate function can serve as a joint probability distribution for a pair of discrete random variables X and Y if and only if its values, $f(x,y)$, satisfy the conditions*

1. $f(x,y) \geq 0$ *for each pair of values* (x,y) *within its domain;*
2. $\sum_x \sum_y f(x,y) = 1$ *where the double summation extends over all possible pairs* (x,y).

Definition 3.4.3 joint distribution function. If X and Y are discrete random variables, the function given by

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{s \leq x} \sum_{t \leq y} f(s,t) \text{ for } \infty < x, y < \infty$$

where $f(s,t)$ is the value of the joint probability distribution of X and Y at (s,t), is called the **joint distribution function** or **joint cumulative distribution** of X and Y. ♢

Definition 3.4.4 marginal distribution. If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y), the function given by

$$g(x) = \sum_y f(x, y)$$

for each x within the range of X is called the **marginal distribution** of X. Correspondingly, the function given by

$$h(y) = \sum_x f(x, y)$$

for each y within the range of Y is called the **marginal distribution** of Y. ◊

Definition 3.4.5 conditional distribution. conditional distribution

$$f(x|y) = \frac{f(x, y)}{h(y)}, h(y) \neq 0$$

$$w(y|x) = \frac{f(x, y)}{g(x)}, g(x) \neq 0$$

◊

3.4.2 Exercises

1. **Problem 3.42.** If the values of the joint probability distribution of \(X\) and \(Y\) are shown below,

$$P[x = 0, y = 0] = \frac{1}{12}$$
$$P[x = 1, y = 0] = \frac{1}{6}$$
$$P[x = 2, y = 0] = \frac{1}{24}$$
$$P[x = 0, y = 1] = \frac{1}{4}$$
$$P[x = 1, y = 1] = \frac{1}{4}$$
$$P[x = 2, y = 1] = \frac{1}{40}$$
$$P[x = 0, y = 2] = \frac{1}{8}$$
$$P[x = 1, y = 2] = \frac{1}{20}$$
$$P[x = 0, y = 3] = \frac{1}{120}$$

find

(a) (b) $P(X = 0, 1 \leq Y < 3)$

(b) (c) $P(X + Y \leq 1)$

(c) (d) $P(X > Y)$

2. **Problem 3.44.** If the joint probability distribution of X and Y is given by

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

find the value of c.

3. **Problem 3.45.** With references to the previous problem find

(a) $P(X \leq 1, Y > 2)$

(b) $P(X = 0, Y \leq 2)$

(c) $P(X + Y > 2)$

4. **Problem 3.70.** With reference to 3.42, find

(a) the marginal distribution of X

(b) the marginal distribution of X

(c) the conditional distribution of X given $Y = 1$

(d) the conditional distribution of Y given $X = 0$

3.4.3

Theorem 3.4.6 expected value of joint random variables. *If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y), the expected value of $g(X, Y)$ is given by*

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot f(x,y)$$

Theorem 3.4.7 expected value of a linear combination of random variables. *If $c_1, c_2, \ldots, c_n$ are constants, then*

$$E\left[\sum_{i=1}^{n} c_i g_i(X_1, X_2, \ldots, X_k)\right] = \sum_{i=1}^{n} c_i E\left[g_i(X_1, X_2, \ldots, X_k)\right]$$

3.4.4 Product moments

See section 4.6.

Definition 3.4.8 product moments about the origin. The r^{th} **and** s^{th} **product moment about the origin** of the random variables X and Y, denoted by $\mu_{r,s}$, is the expected value of $X^r Y^s$; symbolically

$$\mu'_{r,s} = E[X^r Y^s] = \sum_x \sum_y x^r y^s \cdot f(x,y)$$

$r = 0, 1, 2, \ldots$ and $s = 0, 1, 2, \ldots$ when X and Y are discrete. ◊

Special cases of product moments are $\mu'_{1,0} = E[X^1 Y^0] = E[X] = \mu_X$ and $\mu'_{0,1} = E[X^0 Y^1] = E[Y] = \mu_Y$.

As complicated as the definitions of the product moments may be, they lead to a way to define and calculate the very important concept of covariance.

- If we have a high probability of large X paired with large Y and small X paired with small Y, $\text{cov}(X, Y) > 0$

- If we have a high probability of large X paired with small Y and small X paired with large Y, $\text{cov}(X, Y) < 0$

Definition 3.4.9 product moments about the mean. The r^{th} **and** s^{th} **product moment about the mean** of the random variables X and Y, denoted by $\mu'_{r,s}$, is the expected value of $(X-\mu_X)^r(Y-\mu_Y)^s$; symbolically

$$\mu_{r,s} = E[(X-\mu_X)^r(Y-\mu_Y)^s] = \sum_x \sum_y (x-\mu_X)^r(y-\mu_Y)^s \cdot f(x,y)$$

$r = 0, 1, 2, \ldots$ and $s = 0, 1, 2, \ldots$ when X and Y are discrete. $\Diamond$

Definition 3.4.10 covariance. $\mu_{1,1}$ is called the **covariance** of X and Y, and it is denoted by σ_{XY} or $\text{cov}(X,Y)$, or $C(X,Y)$. $\Diamond$

Theorem 3.4.11 covariance from moments about the origin. $\sigma_{XY} = \mu'_{1,1} - \mu_X\mu_Y$

Theorem 3.4.12 independence and covariance. *If X and Y are independent, then*

$$E[XY] = E[X] \cdot E[Y]$$

and $\sigma_{XY} = 0$.

Remark 3.4.13 In terms of moments,

$$\mu'_{1,1} = \mu'_{1,0} \cdot \mu'_{0,1}$$

Theorem 3.4.14 product moments of independent random variables. *If $X_1, X_2, \ldots, X_n$ are independent, then $E[X_1X_2\cdots X_n] = E[X_1] \cdot E[X_2] \cdot \cdots \cdot E[X_n]$.*

Independence means covariance is zero, but covariances of zero does not mean independence.

3.4.5 Exercises

1. **Problem 4.41.** If X and Y have the joint probability distribution $f(x,y) = \frac{1}{4}$ for $(-3,-5)$, $(-1,-1)$, $(1,1)$, $(3,5)$, find $\text{cov}(X,Y)$.
2. **Problem 4.45.** If X and Y have the joint probability distribution $f(-1,0) = 0$, $f(-1,1) = \frac{1}{4}$, $f(0,0) = \frac{1}{6}$, $f(1,0) = \frac{1}{12}$, $f(1,1) = \frac{1}{2}$ show that
   (a) $\text{cov}(X,Y) = 0$;
   (b) the two random variables are not independent.

3.4.6 Conditional expectation

See section 4.8.

Definition 3.4.15 conditional expectation. If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at X, the **conditional expectation** of $u(X)$ given $Y = y$ is

$$E[u(X)|y] = \sum_x u(x) \cdot f(x|y)$$

and the **conditional expectation** of $v(Y)$ given $X = x$ is

$$E[v(Y)|x] = \sum_y v(y) \cdot w(y|x)$$

$\Diamond$

Definition 3.4.16 conditional mean. If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at X, the **conditional mean** of $u(X) = X$ given $Y = y$ is

$$\mu_{X|y} = E[X|y] = \sum_x x \cdot f(x|y)$$

and the **conditional mean** of $v(Y) = Y$ given $X = x$ is

$$\mu_{Y|x} = E[Y|x] = \sum_y y \cdot w(y|x)$$

◊

Definition 3.4.17 conditional variance. If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at X, the **conditional variance** of X given $Y = y$ is

$$\sigma^2_{X|y} = E[(X - \mu_{X|y})^2|y] = E[X^2] - \mu^2_{X|y}$$

and the **conditional expectation** of Y given $X = x$ is

$$\sigma^2_{Y|x} = E[(Y - \mu_{Y|x})^2|y] = E[Y^2] - \mu^2_{Y|x}$$

◊

3.4.7 Multivariate distributions

See Sec. 5.8, 5.9The multinomial distribution is an extension of the binomial distribution that tracks the occurrence in number of multiple types of outcomes.

The multivariate hypergeometric distribution is an extension of the hypergeometric distribution that tracks the occurrence in number of multiple types of outcomes.

3.4.8 Exercises

1. 4.xx. xx

Chapter 4

Continuous Random Variables

We will study the theory of single-variable continuous random variables, their mathematical expectations, and specific examples of densities from practice. Following that we will briefly look at the theory of multivariable continuous random variables. We will largely repeat the ideas of Chapter **??**, but this time for continuous random variables. Terminology will be introduced, but the ideas are the same, only now the process of summation is replaced by integration.

4.1 Probability densities

Some words.

4.1.1 Univariate distributions

See sections 3.4. Recommended problems: (pg89) 16, 17, 19, 20, 29$*$

Probability distributions for discrete random variables assign probability to discrete values, but for continuous random variables, probabilities are only assigned to intervals.

Definition 4.1.1 Probability density function. A function with values $f(x)$, defined over the set of all real numbers is called *a* **probability density function** of the continuous random variable X if and only if

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

for any real a and b such that $a \leq b$. $\Diamond$

Remark 4.1.2 A probability density function is often referred to as simply a density function or a PDF

Do not assume that $f(c)$ gives $P(X = c)$. A density is not a probability, it is a probability density. Since probabilities of continuous random variables are associated with intervals,

$$\int_c^c f(x)\,dx = 0.$$

Remark 4.1.3 The value of a PDF can be changed for some values of a random variable without changing the associated probabilities. Notice that Definition **??** specifies *a*, not *the* PDF

Theorem 4.1.4 probability density at a point. *If X is a continuous random variable and a and b are real numbers such that $a < b$ then,*

$$P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b)$$

Theorem 4.1.5 conditions for probability density. *A function can serve as the probability density for a continuous random variable X if and only if its values, $f(x)$, satisfy the conditions*

1. *$f(x) \ge 0$ for each value within its domain;*
2. *$\int_{-\infty}^{\infty} f(x)\,dx = 1$.*

Example 4.1.6 specifying an exponential probability density. Find k such that

$$f(x) = \begin{cases} ke^{-3x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

can serve as the PDF for X.

Note in the definition of $f(x)$ that $f(x) = 0$ for $x \le 0$.**Solution**. The strategy is to integrate to find the total area then choose k appropriately.

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\,dx &= \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx \\ &= \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} ke^{-3x}\,dx \\ &= 0 + \int_{0}^{\infty} ke^{-3x}\,dx \\ &= \lim_{t\to\infty} \left(\frac{k}{-3} e^{-3x} \Big|_0^t \right) \\ &= \lim_{t\to\infty} \left(\frac{k}{-3} e^{-3t} - \frac{k}{-3} e^{0} \right) \\ \int_{-\infty}^{\infty} f(x)\,dx &= \frac{k}{3} \end{aligned}$$

For $f(x)$ to satisfy Theorem **??**, we much choose $k = 3$ so that the value of the integral is 1. □

Example 4.1.7 using exponential probability density. Using the value of k found in Example **??** which ensures that

$$f(x) = \begin{cases} ke^{-3x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

can serve as the PDF for X, find $P(X < 1/2)$.

Recall we found that $k = 3$.**Solution**. The strategy is to integrate to find the corresponding definite integral according to Definition **??**.

$$\begin{aligned} \int_{-\infty}^{1/2} f(x)\,dx &= \int_{-\infty}^{0} 0\,dx + \int_{0}^{1/2} 3e^{-3x}\,dx \\ &= 0 + \left(-e^{-3x} \Big|_0^{1/2} \right) \\ &= \left(-e^{-3(1/2)} \right) - \left(-e^{-3(0)} \right) \\ &= 1 - e^{-1.5} \approx 0.7769 \end{aligned}$$

Recall, this is the area under the curve traced by the graph of $f(x)$ between

$x = 0$ and $x = 1/2$. □

Example 4.1.8 a hat-shaped probability density. Consider

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

as the PDF for X, and use it to find $P(0 < X < 0.5)$, $P(X = 1)$, and $P(0.5 < X < 1.5)$.**Hint**. The strategy is to integrate to find the corresponding definite integral according to Definition **??**.

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

□

Definition 4.1.9 If X is a continuous random variable, the function given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt \text{ for } -\infty < x < \infty$$

where $f(t)$ is the value of the probability density of X at t, is called the **distribution function**, or the **cumulative distribution (function)**, of X. ♢

Checkpoint 4.1.10 finding an exponential cumulative distribution function. Using the value of k found in Example **??** which ensures that

$$f(x) = \begin{cases} ke^{-3x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

can serve as the PDF for X, find $F(x) = P(X < x)$.

Recall we found that $k = 3$.**Hint**. Set up and find the integral in Definition **??**.

Example 4.1.11 cumulative distribution for a hat-shaped probability density. Find $F(x)$ for

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

Hint. Set up and find the integral in Definition **??**.

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

Solution. For $x < 0$, $F(x) = 0$ (no area has yet been accumulated). Now things get interesting. For $0 \le x < 1$, we integrate $f(t)$ to a variable limit of integration x.

$$\begin{aligned}\int_0^x t\,dt &= \left(\frac{t^2}{2}\right)\Big|_0^x \\ &= \left(\frac{x^2}{2} - \frac{0^2}{2}\right) \\ \int_0^x t\,dt &= \frac{x^2}{2}\end{aligned}$$

For $1 \leq x < 2$, we integrate $f(t)$ to a variable limit of integration x.

$$\begin{aligned}\int_1^x 2 - t\,dt &= \left(2t - \frac{t^2}{2}\right)\Big|_1^x \\ &= \left(\left(2x - \frac{x^2}{2}\right) - \left(2(1) - \frac{1^2}{2}\right)\right) \\ \int_1^x 2 - t\,dt &= 2x - \frac{x^2}{2} - \frac{3}{2}\end{aligned}$$

It is important here to remember that in writing $F(x)$ for points in $1 \leq x < 2$, the complete area from $0 \leq x < 1$ has already been accumulated in full. Beyond $x = 2$, all area under $f(x)$ has been accumulated. In other words for a point in $1 \leq x < 2$,

$$\begin{aligned}\int_0^x f(t)\,dt &= \int_0^1 t\,dt + \int_1^x 2 - t\,dt \\ &= \frac{1}{2} + 2x - \frac{x^2}{2} - \frac{3}{2} \\ &= 1 + 2x - \frac{x^2}{2}\end{aligned}$$

With this in mind we have

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{2}, & 0 \leq x < 1 \\ 1 + 2x - \frac{x^2}{2}, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$

$\square$

The considerations made in calculating the formula for $F(x)$ in the interval $1 \leq x$ have tremendous implications for how we approach jointly-distributed random variables in Section **??**, let this serve as a ***warning***.

You will find tools for visualization at the following link.

Theorem 4.1.12 distribution function (continuous). *The values $F(x)$ of the distribution function of a continuous random variable X satisfy the conditions*

1. *$F(-\infty) = 0$ and $F(\infty) = 1$ (more carefully stated as $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$);*

2. *if $a \leq b$, then $F(a) \leq F(b)$ for any real numbers a and b*

Given a cumulative distribution function of a continuous random variable, the construction of a probability density function, is much easier than the construction of a probability mass function, from its cumulative distribution function. For continuous random variables we will assume that the function

$f(x)$ is continuous and also differentiable every but perhaps at a finite set of values.

Theorem 4.1.13 density from distribution function. *If $f(x)$ and $F(x)$ are the values of the PDF and CDF of X at x, then*

$$P(a < X < b) = F(b) - F(a)$$

for any real constants $a < b$ and

$$f(x) = \frac{dF}{dx}$$

where the derivative exists.

Remark 4.1.14 density from distribution function. Remember that if $F(x)$ is a piecewise function, we would expect $f(x)$ to be piecewise as well. Additionally, we need to be careful at the locations where the definition of $F(x)$ changes, which is where $f(x)$ is likely to change definition as well.

Example 4.1.15 a ramp-shaped cumulative distribution. Consider

$$F(x) = \begin{cases} 0, & x < -1 \\ \dfrac{x+1}{2}, & -1 \leq x < 1 \\ 1, & 1 \leq x \end{cases}.$$

Find the probability density function $f(x)$. Notice that $F(x)$ is not differentiable at $x = \pm 1$ where the graph displays corners.**Solution**. For $x < -1$,

$$f(x) = \frac{d}{dx}\Big(0\Big) = 0.$$

For $-1 \leq x < 1$,

$$f(x) = \frac{d}{dx}\left(\frac{x+1}{2}\right) = 1/2.$$

For $1 \leq x$,

$$f(x) = \frac{d}{dx}\Big(1\Big) = 0.$$

Based on the work above,

$$f(x) = \begin{cases} 0, & x < -1 \\ 1/2, & -1 \leq x < 1 \\ 0, & 1 \leq x \end{cases}.$$

Since the value of the function at a single point, such as an endpoint,does not affect the value of probabilities, we will adopt the convention of specifying the non-zero part of the density on an open interval. This means we take $f(-1) = f(1) = 0$. Thus we have,

$$f(x) = \begin{cases} 1/2, & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}.$$

□

4.1.2 Exercises

1. **Problem 3.20.** The probability density of the random variable Y is given by

$$f(y) = \begin{cases} \frac{1}{8}(y+1), & 2 < y < 4 \\ 0, & \text{elsewhere} \end{cases}$$

Find $P(Y < 3.2)$ and $P(2.9 < Y < 3.2)$.

4.2 Expectation of continuous random variables

This is the start of Chapter 4 in Freund's Mathematical Statistics. Now we will study the major topics of this chapter with a focus on those applying to continuous random variables.

4.2.1 Expected value

See section 4.1 - 4.2. Recommended problems: (pg 136) 7, 9, 10, 11, (pg 161) 60

Definition 4.2.1 expected value (continuous). If X is a continuous random variable and $f(x)$ is the value of its probability density at x, the **expected value** of X is given by

$$E[X] = \int x \cdot f(x)\, dx$$

◊

Example 4.2.2 expected value (continuous). Consider the probability density given below

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

Find $E[X]$.**Solution**.

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f(x)\, dx \\ &= \int_{-\infty}^{0} 0\, dx + \int_0^1 \underbrace{x \cdot x}_{x^2}\, dx + \int_1^2 \underbrace{x \cdot (2-x)}_{2x - x^2}\, dx + \int_2^{\infty} 0\, dx \\ &= 0 + \int_0^1 x^2\, dx + \int_1^2 2x - x^2\, dx + 0 \\ &= \left(\frac{x^3}{3}\right)\Big|_0^1 + \left(x^2 - \frac{x^3}{3}\right)\Big|_1^2 \\ &= \left(\frac{1}{3} - 0\right) + \left(\left(4 - \frac{8}{3}\right) - \left(1 - \frac{1}{3}\right)\right) \\ E[X] &= 1 \end{aligned}$$

□

Theorem 4.2.3 expected value of a function of a random variable (continuous). *If X is a continuous random variable and $f(x)$ is the value of its probability density at x, the expected value of $g(X)$ is given by*

$$E[g(X)] = \int g(x) \cdot f(x)\, dx$$

Now would be a good time to review Theorem **??**, Corollary **??**, Corollary **??**, and Theorem **??**, all of which apply here as well. Recall, certain choices of $g(x)$ in Theorem **??** give us the mean, variance, and other special moments.

While we are at it, consider Theorem **??** which suggested that $E[aX+b] = aE[X]+b$. Defining expectation by integration gives a more convenient means of proving this statement.

Proof.

$$\begin{aligned}
E[aX+b] &= \int_{-\infty}^{\infty} (ax+b) \cdot f(x)\, dx \\
&= \int_{-\infty}^{\infty} ax \cdot f(x)\, dx + \int_{-\infty}^{\infty} b \cdot f(x)\, dx \\
&= a \int_{-\infty}^{\infty} x \cdot f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx \\
&= aE[x] + b
\end{aligned}$$

■

Notice that the second of these integrals is 1 since $f(x)$ is a probability density function.

4.2.2 Exercises

1. **Problem 4.7.** Find the expected value of the random variable Y whose probability density function is given by

$$f(y) = \begin{cases} \frac{1}{8}(y+1), & 2 < y < 4 \\ 0, & \text{elsewhere} \end{cases}$$

2. **Problem 4.10.** If the probability density of X is given by

$$f(x) = \begin{cases} \frac{1}{x\ln(3)}, & 1 < x < 3 \\ 0, & \text{elsewhere} \end{cases}$$

   (a) find $E[X]$, $E[X^2]$, $E[X^3]$.

   (b) determine the value of $E[X^3 + 2X^2 - 3X + 1]$.

4.2.3 Moments

See section 4.3. Recommended problems: 4.3 (pg 146) 20, 22, 23, 31, 33, 34, 40, (pg 162) 69, 73, 75

Definition 4.2.4 moments about the origin (continuous). The r^{th} **moment about the origin** of a random variable X, denoted by μ'_r, is the expected value of $(X)^r$; symbolically

$$\mu_r' = E[(X)^r] = \int_{-\infty}^{\infty} x^r \cdot f(x)\,dx$$

for $r = 0, 1, 2, \ldots$ when X is continuous. ♢

Definition 4.2.5 moments about the mean (continuous). The r^{th} **moment about the mean** of a random variable X, denoted by μ_r, is the expected value of $(X - \mu)^r$; symbolically

$$\mu_r = E[(X-\mu)^r] = \int_{-\infty}^{\infty} (x-\mu)^r \cdot f(x)\,dx$$

for $r = 0, 1, 2, \ldots$ when X is discrete. ♢

Example 4.2.6 8^{th}-order polynomial density. Consider

$$f(x) = \begin{cases} 630x^4(1-x)^4, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find μ and σ^2.**Hint**. It is recommended that you expand the term $(1-x)^4$. Doing that, and carefully integrating, you should find $\mu = 1/2$. Additionally, by finding $\mu_2' = E[X^2] = 3/11$, you will be able to use

$$\sigma^2 = \mu_2' - (\mu)^2$$

to find the variance is $\sigma^2 = 1/44$. □

4.2.4 Exercises

1. **Problem 4.20.** Find μ_r' and σ^2 for the random variable X with probability density given by

$$f(x) = \begin{cases} \dfrac{1}{x\ln(3)}, & 1 < x < 3 \\ 0, & \text{elsewhere} \end{cases}$$

Hint. Use the definition of $\mu_r' = E[X^r]$ to begin.

4.2.5 Chebyshev's Theorem

See section 4.4.

Theorem 4.2.7 Chebyshev's Theorem. *If μ and σ are the mean and the standard deviation of a random variable X, then for any positive constant k the probability is* at least $1 - \frac{1}{k^2}$ *that X will take on a value within k standard deviations of the mean; symbolically*

$$P(|x-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}, \sigma \neq 0$$

Theorem **??** gives a lower bound on the probability that the value of the random variable is within a certain distance (specifically $\pm k\sigma$) of the mean.

4.2.6 Exercises

1. **Problem 4.31.** What is the smallest value of k Chebyshev's theorem for which the probability that a random variable wil take on a value between $\mu - k\sigma$ and $\mu k + \sigma$ is

   (a) at least 0.95?

   (b) at least 0.99?

4.2.7 Moment-generating functions

See section 4.5.

Definition 4.2.8 moment-generating function (continuous). The **moment-generating function** of a random variable X, where it exists, is given by

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx$$

when X is discrete. ♢

Notce that a moment-generating function $M_X(t)$ itself is a function of the variable t not X. As it turns out, we are most interested in values of the function at or near $t = 0$.

Now would be a good time to review Theorem **??** and Theorem **??** which apply here in the continuous case as well.

Example 4.2.9 moment-generating function by definition. Find the moment-generating function of

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

Solution.

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \int_0^{\infty} e^{tx} \cdot e^{-x}\, dx \\ &= \int_0^{\infty} e^{-(1-t)x}\, dx \\ &= \lim_{A\to\infty} \left(\left(-\frac{1}{1-t} e^{-(1-t)x} \right) \Big|_0^A \right) \\ &= \lim_{A\to\infty} \left(\underbrace{\left(-\frac{1}{1-t} e^{-(1-t)A} \right)}_{\text{needs } t<1} - \left(-\frac{1}{1-t} e^{-(1-t)0} \right) \right) \\ M_X(t) &= 0 + \frac{1}{1-t} \end{aligned}$$

For the first integral to converge we require $t < 1$. □

With the result from Example **??**, assuming $|t| < 1$ we can expand around $t = 0$ to find

$$M_X(t) = \frac{1}{1-t} = 1 + t + t^2 + \cdots + t^r + \ldots$$

Remembering that μ'_r is the coefficient of $\dfrac{t^r}{r!}$, we can rewrite the previous expansion as

$$M_X(t) = \frac{1}{1-t} = 1 + 1\Big(\frac{t^1}{1!}\Big) + 2!\Big(\frac{t^2}{2!}\Big) + \cdots + r!\Big(\frac{t^r}{r!}\Big) + \ldots$$

which shows that $\mu'_r = r!$.

Above the primary difficulty once the moment-generating function has been calculated is probably in the expansion of the Maclaurin series. One alternative to this requires calculating the integrals that define the necessary expectations

$$\mu'_r = E[X^r] = \int_{-\infty}^{\infty} x^r e^{-x}\, dx$$

by integration by parts or tabular integration. Instead, given the moment-generating function, we could apply Theorem **??**.

Example 4.2.10 moments of an exponential by differentiation. Recall the moment-generating function from Example **??** was $M_X(t) = \dfrac{1}{1-t}$ for $t < 1$. Find μ'_1 and μ'_2.**Solution**. By theorem

$$\begin{aligned}
\mu'_1 &= \frac{d}{dt}\Big(M_X(t)\Big)\Big|_{t=0} \\
&= \frac{d}{dt}\Big(\frac{1}{1-t}\Big)\Big|_{t=0} \\
&= \Big(-(1-t)^{-2}(-1)\Big)\Big|_{t=0} \\
\mu'_1 &= \Big(-(1-0)^{-2}(-1)\Big) = 1
\end{aligned}$$

To calculate μ'_2 we take the second derivative of $M_X(t)$. Starting from $\dfrac{dM_X(t)}{dt} = (1-t)^{-2}$, we find

$$\begin{aligned}
\mu'_2 &= \frac{d}{dt}\Big(\frac{dM_X(t)}{dt}\Big)\Big|_{t=0} \\
&= \frac{d}{dt}\Big((1-t)^{-2}\Big)\Big|_{t=0} \\
&= \Big(-2(1-t)^{-3}(-1)\Big)\Big|_{t=0} \\
\mu'_2 &= \Big(-2(1-0)^{-3}(-1)\Big) = 2
\end{aligned}$$

□

4.2.8 Exercises

1. **Problem 4.34.** Find the moment-generating function of the continuous random variable X that has the probability density

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

and use it to determine the values of μ'_1, μ'_2, and σ^2.

4.3 Special probability densities

...

4.3.1 Continuous uniform

...

Definition 4.3.1 continuous uniform distribution. A continuous random variable X has a **continuous uniform distribution** and is referred to as a **continuous uniform random variable** if and only if its probability density is given by

$$u(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta-\alpha}, & \text{for } \alpha < x < \beta \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha, \beta > 0$. ◊

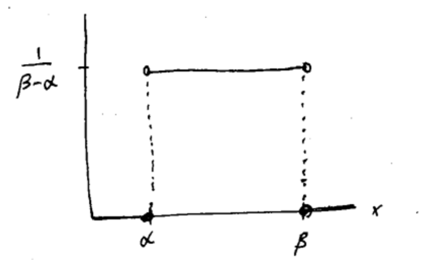


Figure 4.3.2 A folksy sketches of the continuous uniform distribution.

Theorem 4.3.3 mean and variance of continuous uniform. *The mean and variance of the uniform distribution are given by*

$$\mu = \frac{\alpha+\beta}{2} \text{ and } \sigma^2 = \frac{1}{12}(\beta-\alpha)^2$$

Checkpoint 4.3.4 continuous uniform moment-generating function. Find the moment generating function of the continuous uniform distribution.

4.3.1.1 Exercises

1. **Problem 6.1.** Show that if a random vriable has a uniform density with parameters α and β, the probability that it will take on a value less than $\alpha + p(\beta - \alpha)$ is p.
2. **Problem 6.3.** If a random vriable has a uniform density with parameters α and β, find its distribution function.

4.3.2 Exponential family

Two distributions, the exponential and chi-square, are special cases (i.e., parameterizations) of a very flexible distribution called the gamma.

We are at this point fairly familiar with the calculation of factorials. Factorials underlie our methods of counting, from permutations to combinations and partitions. For example, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. What would it mean to ask for 0.5!?

The gamma function allows us to generalize our concept of the factorial to noninteger, and even many nonpositive, values.

Definition 4.3.5 gamma function. The **gamma function** is given by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy$$

where $\alpha > 0$. $\Diamond$

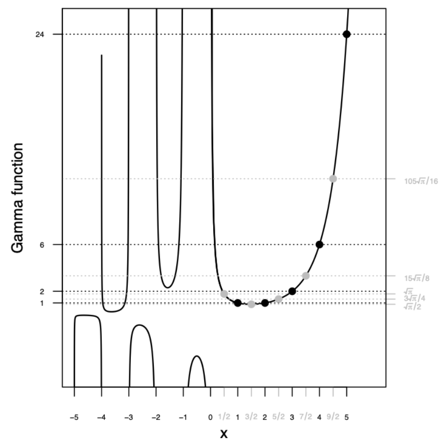


Figure 4.3.6 Graphs of the gamma function with a few key values.

Table 4.3.7 A table of common gamma function values.

α	$\Gamma(\alpha)$	α (cont.)	$\Gamma(\alpha)$ (cont.)
$1/2$	$\sqrt{\pi}$	$5/2$	$3\sqrt{\pi}/4$
1	$0! = 1$	3	$2! = 2$
$3/2$	$\sqrt{\pi}/2$	$7/2$	$15\sqrt{\pi}/8$
2	$1! = 1$	4	$3! = 6$

Definition 4.3.8 gamma distribution. A continuous random variable X has a **gamma distribution** and is referred to as a **gamma random variable** if and only if its probability density is given by

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & \text{for } 0 < x \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha, \beta > 0$. $\Diamond$

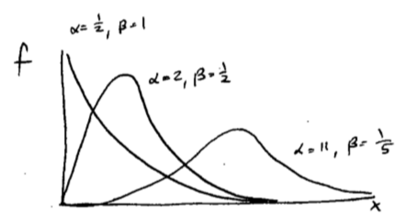


Figure 4.3.9 A few folksy sketches of the gamma distribution.

Theorem 4.3.10 moments about the origin of the gamma. *The r^{th} moment about the origin of the gamma distribution is given by*

$$\mu_r' = \frac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$$

Theorem 4.3.11 mean and variance of gamma. *The mean and variance of the gamma distribution are given by*

$$\mu = \alpha\beta \text{ and } \sigma^2 = \alpha\beta^2$$

Theorem 4.3.12 moment-generating function of the gamma. *The moment-generating function of the gamma distribution is given by*

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

If we are interested in events following the Poisson distribution (discretely-numbered events happening continuously over some time interval) we might be interested in the time to event (or between events)

Let Y describe the waiting time until first success with $\lambda = \alpha t$ being the Poisson rate.

$$\begin{aligned} F(Y) = P(Y \le y) &= 1 - P(Y > y) \\ &= 1 - P(\text{no successes/events in interval of length } y) \\ &= 1 - p(0; \alpha y) \\ &= 1 - \frac{(\alpha y)^0 e^{-\alpha y}}{0!} \\ F(Y) &= 1 - e^{-\alpha y} \qquad (F(Y) = 0,\ y \le 0) \end{aligned}$$

Figure 4.3.13 Derivation of the exponential distribution from a Poisson process (Part 1).

$$\begin{aligned} \text{Given } F(y),\quad f(y) &= \frac{dF}{dy} \\ &= \frac{d}{dy}\left(1 - e^{-\alpha y}\right) \\ f(y) &= \alpha e^{-\alpha y} \quad \text{for } y > 0. \end{aligned}$$

Letting $\alpha = \frac{1}{\theta}$ the event times are exponentially distributed.
(* $\alpha e^{-\alpha y}$ is another parametrization of $\frac{1}{\theta} e^{-x/\theta}$ in some sources).

Figure 4.3.14 Derivation of the exponential distribution from a Poisson process (Part 2).

Definition 4.3.15 exponential distribution. A continuous random variable X has an **exponential distribution** and is referred to as an **exponential random variable** if and only if its probability density is given by

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-x/\theta}, & \text{for } 0 < x \\ 0, & \text{elsewhere} \end{cases}$$

where $\theta > 0$. ♢

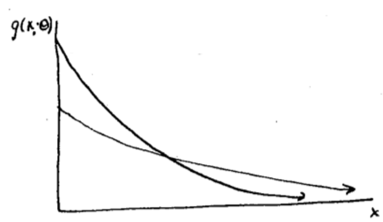


Figure 4.3.16 A few folksy sketches of the exponential distribution.

Corollary 4.3.17 mean and variance of exponential. *The mean and variance of the exponential distribution are given by*

$$\mu = \theta \text{ and } \sigma^2 = \theta^2$$

Checkpoint 4.3.18 By using the appropriate mathematical expectation, verify that the mean of the exponential distribution is as stated above.

Remark 4.3.19 The function used in Example **??** and Example **??** was an exponential distribution with $\lambda = {}^1\!/\!_\theta = 3$. Occasionally the exponential is parameterized by $\lambda > 0$ with

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } 0 < x \\ 0, & \text{elsewhere} \end{cases}$$

Definition 4.3.20 chi-suqare distribution. A continuous random variable X has a **chi-square distribution** and is referred to as a chi-square random variable if and only if its probability density is given by

$$f(x;\nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-x/2}, & \text{for } 0 < x \\ 0, & \text{elsewhere} \end{cases}$$

where ν is a parameter describing the *number of degrees of freedom.* ♢

4.3.2.1 Exercises

1. **Problem 6.10(a).** Find the probability that a random variable will exceed 1 if it has a gamma distribution with $\alpha = 2, \beta = 3$.
2. **Problem 6.23.** A random variable X has a *Weibull distribution* if and only if its probability density is given by

$$f(x;\alpha,\beta) = \begin{cases} kx^{\beta-1} e^{-\alpha x^\beta}, & \text{for } 0 < x \\ 0, & \text{elsewhere} \end{cases}$$

for $\alpha, \beta > 0$. Express k in terms of α, β. See also: WeBWorK.

4.3.3 Beta distribution

In some applications, the beta distribution (which is a probability density itself) is actually used to described the distributions of *other* probabilities.

Definition 4.3.21 Beta distribution. A continuous random variable X has a **Beta distribution** and is referred to as a Beta random variable if and only if its probability density is given by

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha, \beta > 0$. ♢

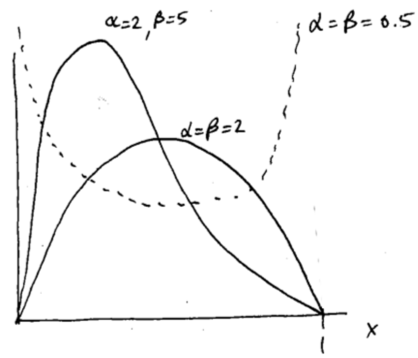


Figure 4.3.22 A few folksy sketches of the beta distribution. Notice that for $\alpha = \beta = 0.5$, the PDF has vertical asymptotes at $x = 0, 1$. For $\alpha = \beta = 2$, the PDF simplifies to $f(x; 2, 2) = 6x(1 - x)$.

Remark 4.3.23 The function used in Example **??** was a beta distribution with $\alpha = \beta = 5$.

4.3.3.1 Exercises

1. Problem 6.25(a). Verify that the integral of the beta density function from $-\infty$ to ∞ equals 1 if it has a gamma distribution with $\alpha = 2, \beta = 4$. See also: WeBWorK.

4.3.4 Normal distribution

...

Definition 4.3.24 Normal distribution. A continuous random variable X has a **normal distribution** and is referred to as a normal random variable if and only if its probability density is given by

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \text{for } -\infty < x < \infty$$

where $\sigma > 0$. ◊

Theorem 4.3.25 mean and variance of normal. *The mean and variance of the normal distribution are given by*

$$\mu = \mu \textit{ and } \sigma^2 = \sigma^2$$

Theorem 4.3.26 moment-generating function of the normal. *The moment-generating function of the normal distribution is given by*

$$M_X(t) = e^{\mu t + 1/2\sigma^2 t^2}$$

Remark 4.3.27 The PDF is parameterized by μ and $\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var}(X)}$.

Corollary 4.3.28 *The **standard normal** has $\mu = 0$ and $\sigma = 1$ and is given by*

$$f(x; \mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad \textit{for } -\infty < x < \infty$$

Remark 4.3.29 It is the standard normal distribution for which areas under the curve given by

$$\int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$$

are tabulated.

Table III: Standard Normal Distribution

hundredths

ones and tenths

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

$P(0 \le z \le 0.59)$

Also, for $z = 4.0$, 5.0, and 6.0, the probabilities are 0.49997, 0.4999997, and 0.499999999.

Figure 4.3.30 Table of standard normal probabilities. The highlighted entry "0.2224" corresponds to the probability $P(z < 0.59) = 0.2224$ and gives the area under the graph of the standard normal curve between $z = 0$ and $z = 0.59$.

Other areas can be calculated by symmetry.

Theorem 4.3.31 transforming to standard normal. *If X has a normal distribution with μ and σ, then $Z = \dfrac{X - \mu}{\sigma}$ has the standard normal distribution.*

A few special probabilities for the standard normal are given by

1. $P(-\infty < z < 0) = 0.5$
2. $P(0 < z < \infty) = 0.5$
3. $P(-\infty < z < \infty) = 1$
4. $P(-a < z < a) = 2 \cdot P(0 < z < a)$

Example 4.3.32 probabilities from the standard normal. Find the following probabilities from the standard normal distribution.

1. $P(|z| < 0.3)$
2. $P(0.3 < z < 0.6)$
3. $P(0.6 < z)$

Solution. To find $P(|z| < 0.3)$, we must first find $P(0 < z < 0.3)$. Once we have that, we can double this value by symmetry of the PDF.

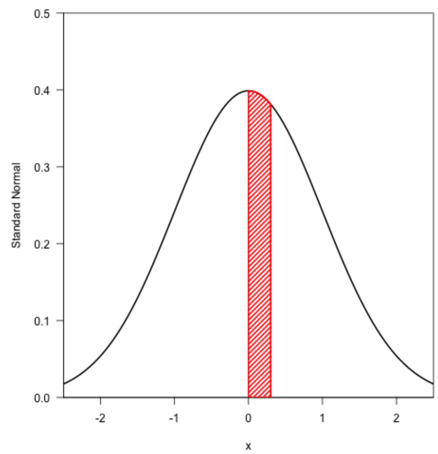


Figure 4.3.33 Area under the curve corresponding to desired probability.

Now to find $P(|z| < 0.3)$, we have the following image. By table we have that $P(0 < z < 0.3) = 0.1179$, so $P(|z| < 0.3) = 2(0.1179) = 0.2358$. For simplicity we will use an equality here, even though the tabulated areas are numerical approximations.

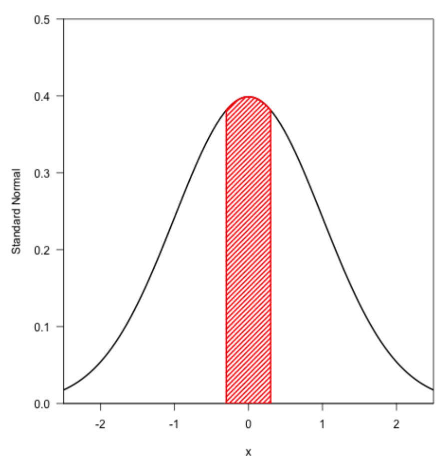


Figure 4.3.34 Area under the curve corresponding to desired probability.

Now to find $P(0.3 < z < 0.6)$, we have the following image. To solve this problem we find the area from $z = 0$ to $z = 0.6$, then subtract the area from $z = 0$ to $z = 0.3$.

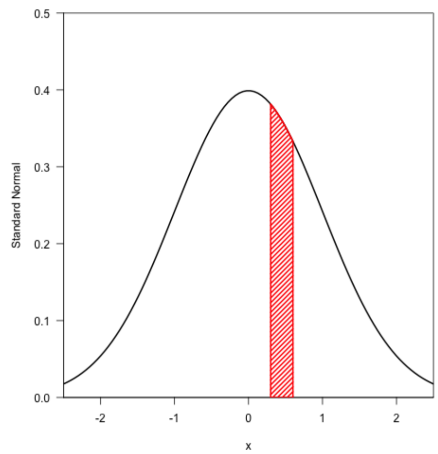


Figure 4.3.35 Area under the curve corresponding to desired probability.

Now to find $P(0.6 < z)$, we have the following image. To solve this problem we find the area from $z = 0$ to $z = 0.6$, then subtract this from the area under the entire right half of the curve. So $P(0.6 < z) = 0.5$ - $0.2257 = 0.2743$

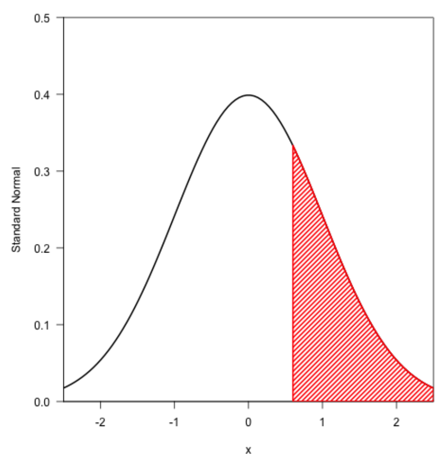


Figure 4.3.36 Area under the curve corresponding to desired probability.

□

Example 4.3.37 probabilities from the standard normal. Find the associated standard normal random variable Z for a normal random variable X with $\mu = 19.7$ and $\sigma = 9.1$ and use it to calculate the probability that $Z > 25$.**Hint**. Transform to Z then use the table. □

4.3.4.1 Exercises

1. **Problem 6.31.** Show that the normal distribution has a relative maximum at $x = \mu$ and inflection points at $x = \mu \pm \sigma$. See

also: WeBWorK.

4.3.5 Normal approximation to the binomial distribution

...

Theorem 4.3.38 normal approximation to the binomial. *If X has a binomial distribution with parameters n and θ, then the moment-generating* *function of* $\dfrac{X - n\theta}{\sqrt{n\theta(1-\theta)}}$ *approaches* $M_X(t) = \underbrace{e^{1/2t^2}}_{std.\ normal}$ *as* $n \to \infty$.

This is because of a 1 : 1 correspondence between moment-generating functions and PMFs/PDFs and the (unstated) fact that if $M_X(t) \to M_Y(t)$, then the distribution of $X \to$ the distribution of Y.

The approximation is true as $n \to \infty$ and can be used when both $n\theta, n(1-\theta) > 5$. This is theoretically important and was of great historical importance, but is less practically important

Since calculating densities reqires intervals, a *continuity correction* represents the binomial integer k by the *standard normal* interval $(k - 1/2, k + 1/2)$ for the purposes of integration.

4.3.5.1 Exercises

1. **Problem 6.75(a, b).** Make the normal approximation to $b(1; 150, 0.05)$. Is the approximation reasonable? How does it compare to a (rounded) true value of $b(1; 150, 0.05) = 0.0036$? See

also: WeBWorK.

4.4 Multivariate continuous random variables

Though we will not specifically look at special jointly-distributed continuous random variables, we certainly could. Instead here we will focus on the fundamental definitions and properties. This will be handy for example if we someday encountered a situation with a special multivariable probability density (e.g., the bivariate normal).

4.4.1 Multivariate distributions

See Sec. 3.5 through 3.7.

Definition 4.4.1 joint probability density. If X and Y are continuous random variables, the function given by $f(x, y) = P(X = x, Y = y)$ for each point in the xy-plane X and Y is called the **joint probability density** of X and Y. ♢

Example 4.4.2 jointly-distributed random variables. Consider

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

and use it to find $P(0 < x < 0.5, 1 < y < 1.5)$.

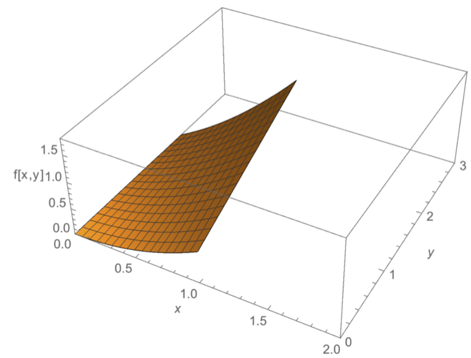


Figure 4.4.3 Graph of a probability density function.

Hint. Set up and evaluate a double integral. Here instead of area under a curve, we are looking for the volume under a surface that has been accumulated over the region of interest. □

Theorem 4.4.4 conditions for a joint probability density. *A bivariate function can serve as a joint probability density for a pair of continuous random variables X and Y if and only if its values, $f(x, y)$, satisfy the conditions*

1. $f(x, y) \geq 0 \; -\infty < x, y < \infty$;
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1$.

Definition 4.4.5 joint distribution function (continuous). If X and Y are continuous random variables, the function given by

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt \text{ for } -\infty < x, y < \infty$$

where $f(s, t)$ is the value of the joint probability density of X and Y at (s, t), is called the **joint density function** or **joint cumulative density** of X and Y. ◊

Example 4.4.6 joint CDF. Consider

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

and use it to find $F(x, y) = P(X < x, Y < y)$.**Hint**. Set up and evaluate a double integral to an arbitrary point in the plane.**Solution**. Here instead of area under a curve, we are looking for the volume under a surface that has been accumulated up to the arbitrary point (x, y).

$$F(x, y) = \begin{cases} 0 & x \leq 0 \text{ or } y \leq 0 \\ \frac{3x^2y^2}{20} + \frac{x^3y}{5} & 0 < x < 1, 0 < y < 2 \\ \frac{3y^2}{20} + \frac{y}{5} & 1 \leq x, 0 < y < 2 \\ \frac{12x^2}{20} + \frac{6x^3}{15} & 0 < x < 1, 2 \leq y \\ 1 & 1 \leq x, 2 \leq y \end{cases}$$

Notice that for $x \geq 1$ and $0 < y < 2$, the result depends only on the y-value, we are no longer accumulating probability in the x-direction (lower right region in

Figure **??**). Reversing the roles of the variables, the is true for the upper left region where $0 < x < 1$ and $y \geq 2$.

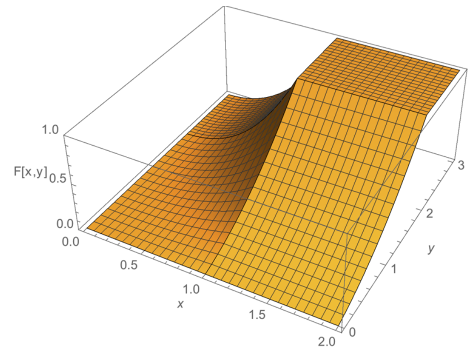


Figure 4.4.7 Graph of the joint cumulative distribution function.

□

Remark 4.4.8 Motivated by $f(x) = \dfrac{dF}{dx}$ we have

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

Definition 4.4.9 marginal density. If X and Y are continuous random variables and $f(x, y)$ is the value of their joint probability density at (x, y), the function given by

$$g(x) = \int f(x, y)\, dy$$

for each x within the range of X is called the **marginal density** of X. Correspondingly, the function given by

$$h(y) = \int f(x, y)\, dx$$

for each y within the range of Y is called the **marginal density** of Y. ♢

Checkpoint 4.4.10 marginal densities. Calculate the marginal densities of the joint probability distribution used in Example **??** and Example **??**

Definition 4.4.11 conditional density. conditional density

$$f(x|y) = \frac{f(x, y)}{h(y)}, h(y) \neq 0$$

$$w(y|x) = \frac{f(x, y)}{g(x)}, g(x) \neq 0$$

♢

4.4.2 Product moments

See section 4.6.

Definition 4.4.12 product moments about the origin. The $r^{\mathbf{th}}$ **and** $s^{\mathbf{th}}$ **product moment about the origin** of the random variables X and Y, denoted by $\mu_{r,s}$, is the expected value of $X^r Y^s$; symbolically

$$\mu'_{r,s} = E[X^r Y^s] = \int\int x^r y^s \cdot f(x,y)\,dy\,dx$$

$r = 0, 1, 2, \ldots$ and $s = 0, 1, 2, \ldots$ when X and Y are discrete. $\Diamond$

Special cases of product moments are $\mu'_{1,0} = E[X^1 Y^0] = E[X] = \mu_X$ and $\mu'_{0,1} = E[X^0 Y^1] = E[Y] = \mu_Y$.

Definition 4.4.13 product moments about the mean. The $r^{\mathbf{th}}$ **and** $s^{\mathbf{th}}$ **product moment about the mean** of the random variables X and Y, denoted by $\mu'_{r,s}$, is the expected value of $(X-\mu_X)^r(Y-\mu_Y)^s$; symbolically

$$\mu_{r,s} = E[(X-\mu_X)^r(Y-\mu_Y)^s] = \int\int (x-\mu_X)^r(y-\mu_Y)^s \cdot f(x,y)\,dy\,dx$$

$r = 0, 1, 2, \ldots$ and $s = 0, 1, 2, \ldots$ when X and Y are continuous. $\Diamond$

Now would be a good time to review Definition **??**, Theorem **??**, Theorem **??**, Remark **??**, and Theorem **??** all of which apply here as well.

Independence means covariance is zero, but covariances of zero does not mean independence.

4.4.3 Exercises

1. **Problem 4.41.** If X and Y have the joint probability distribution $f(x,y) = \frac{1}{4}$ for $(-3,-5)$, $(-1,-1)$, $(1,1)$, $(3,5)$, find $\operatorname{cov}(X,Y)$.
2. **Problem 4.45.** If X and Y have the joint probability distribution $f(-1,0) = 0$, $f(-1,1) = \frac{1}{4}$, $f(0,0) = \frac{1}{6}$, $f(1,0) = \frac{1}{12}$, $f(1,1) = \frac{1}{2}$ show that
   (a) $\operatorname{cov}(X,Y) = 0$;
   (b) the two random variables are not independent.

4.4.4 Conditional expectation

See section 4.8.

Definition 4.4.14 conditional expectation. If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at X, the **conditional expectation** of $u(X)$ given $Y = y$ is

$$E[u(X)|y] = \int u(x) \cdot f(x|y)\,dx$$

and the **conditional expectation** of $v(Y)$ given $X = x$ is

$$E[v(Y)|x] = \int v(y) \cdot w(y|x)\,dy$$

$\Diamond$

Definition 4.4.15 conditional mean. If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given

$Y = y$ at X, the **conditional mean** of $u(X) = X$ given $Y = y$ is

$$\mu_{X|y} = E[X|y] = \int x \cdot f(x|y)\, dx$$

and the **conditional mean** of $v(Y) = Y$ given $X = x$ is

$$\mu_{Y|x} = E[Y|x] = \int y \cdot w(y|x)\, dy$$

$\Diamond$

Definition 4.4.16 conditional variance. If X is a discrete random variable and $f(x|y)$ is the value of the conditional probability distribution of X given $Y = y$ at X, the **conditional variance** of X given $Y = y$ is

$$\sigma^2_{X|y} = E[(X - \mu_{X|y})^2|y] = E[X^2] - \mu^2_{X|y}$$

and the **conditional expectation** of Y given $X = x$ is

$$\sigma^2_{Y|x} = E[(Y - \mu_{Y|x})^2|y] = E[Y^2] - \mu^2_{Y|x}$$

$\Diamond$

Appendix A

Definitions

Section 1.1 Counting

Definition 1.1.1

Section 2.1 Probability

Definition 2.1.1

Section 3.1 Probability distributions

Definition 3.1.1 probability distribution
Definition 3.1.3 distribution function

Section 3.2 Mathematical expectation of discrete random variables

Definition 3.2.1 expected value
Definition 3.2.6 moments about the origin
Definition 3.2.7 mean of a discrete random variable
Definition 3.2.8 moments about the mean
Definition 3.2.9 moment-generating function

Section 3.3 Special probability distributions

Definition 3.3.1 discrete uniform distribution
Definition 3.3.4 Bernoulli distribution
Definition 3.3.7 binomial distribution
Definition 3.3.16 negative binomial distribution
Definition 3.3.21 geometric distribution
Definition 3.3.23 hypergeometric distribution
Definition 3.3.28 Poisson distribution

Section 3.4 Multivariate discrete random variables

Definition 3.4.1 joint probability distribution
Definition 3.4.3 joint distribution function

(Continued on next page)

Definition 3.4.4	marginal distribution
Definition 3.4.5	conditional distribution
Definition 3.4.8	product moments about the origin
Definition 3.4.9	product moments about the mean
Definition 3.4.10	covariance
Definition 3.4.15	conditional expectation
Definition 3.4.16	conditional mean
Definition 3.4.17	conditional variance

Section 4.1 Probability densities

Definition 4.1.1	Probability density function
Definition 4.1.9	

Section 4.2 Expectation of continuous random variables

Definition 4.2.1	expected value (continuous)
Definition 4.2.4	moments about the origin (continuous)
Definition 4.2.5	moments about the mean (continuous)
Definition 4.2.8	moment-generating function (continuous)

Section 4.3 Special probability densities

Definition 4.3.1	continuous uniform distribution
Definition 4.3.5	gamma function
Definition 4.3.8	gamma distribution
Definition 4.3.15	exponential distribution
Definition 4.3.20	chi-suqare distribution
Definition 4.3.21	Beta distribution
Definition 4.3.24	Normal distribution

Section 4.4 Multivariate continuous random variables

Definition 4.4.1	joint probability density
Definition 4.4.5	joint distribution function (continuous)
Definition 4.4.9	marginal density
Definition 4.4.11	conditional density
Definition 4.4.12	product moments about the origin
Definition 4.4.13	product moments about the mean
Definition 4.4.14	conditional expectation
Definition 4.4.15	conditional mean
Definition 4.4.16	conditional variance

Appendix B

Theorems

Section 3.1 Probability distributions

Theorem 3.1.2	conditions for probability distribution
Theorem 3.1.4	properties of a distribution function

Section 3.2 Mathematical expectation of discrete random variables

Theorem 3.2.2	expected value of a function of a random variable
Theorem 3.2.3	expectation of a linear function
Theorem 3.2.12	moments via differentiation
Theorem 3.2.14	moment-generating function of functions of a random variable
Theorem 3.2.15	variance
Theorem 3.2.18	covariance of two linear combinations

Section 3.3 Special probability distributions

Theorem 3.3.10	reparameterizing a binomial
Theorem 3.3.12	Mean and variance of binomial distribution
Theorem 3.3.13	Proportion of binomial successes
Theorem 3.3.17	negative binomial probability as a binomial probability
Theorem 3.3.19	Mean and variance of the negative binomial distribution
Theorem 3.3.22	Mean and variance of the geometric distribution
Theorem 3.3.24	Mean and variance of the hypergeometric distribution
Theorem 3.3.29	Mean, variance, and MGF of the Poisson distribution

Section 3.4 Multivariate discrete random variables

Theorem 3.4.2	conditions for a joint probability distribution
Theorem 3.4.6	expected value of joint random variables
Theorem 3.4.7	expected value of a linear combination of random variables
Theorem 3.4.11	covariance from moments about the origin
Theorem 3.4.12	independence and covariance
Theorem 3.4.14	product moments of independent random variables

(Continued on next page)

Section 4.1 Probability densities

Theorem 4.1.4	probability density at a point
Theorem 4.1.5	conditions for probability density
Theorem 4.1.12	distribution function (continuous)
Theorem 4.1.13	density from distribution function

Section 4.2 Expectation of continuous random variables

Theorem 4.2.3	expected value of a function of a random variable (continuous)
Theorem 4.2.7	Chebyshev's Theorem

Section 4.3 Special probability densities

Theorem 4.3.3	mean and variance of continuous uniform
Theorem 4.3.10	moments about the origin of the gamma
Theorem 4.3.11	mean and variance of gamma
Theorem 4.3.12	moment-generating function of the gamma
Theorem 4.3.25	mean and variance of normal
Theorem 4.3.26	moment-generating function of the normal
Theorem 4.3.31	transforming to standard normal
Theorem 4.3.38	normal approximation to the binomial

Section 4.4 Multivariate continuous random variables

Theorem 4.4.4	conditions for a joint probability density

Appendix C

Examples

Section 3.1 Probability distributions

Example 3.1.5 Verifying a simple probability mass function
Example 3.1.8 CDF of a simple probability mass function

Section 3.2 Mathematical expectation of discrete random variables

Example 3.2.10 moment-generating function via Taylor series
Example 3.2.11 moment-generating function for three cards
Example 3.2.13 moment-generating function for three cards, via differentiation
Example 3.2.17 covariances of linear combinations

Section 3.3 Special probability distributions

Example 3.3.2 mean of a discrete uniform distribution
Example 3.3.3 simple discrete uniform distribution
Example 3.3.5 mean and variance of a Bernoulli distribution
Example 3.3.6 moment-generating function of a Bernoulli distribution
Example 3.3.8 Binomial coin flips
Example 3.3.9 Free throws
Example 3.3.11 Free throws - revisited
Example 3.3.15 moment-generating function of a binomial distribution
Example 3.3.20 moment generating function of the negative binomial distribution
Example 3.3.26 hypergeometric distribution for truck inspections

Section 4.1 Probability densities

Example 4.1.6 specifying an exponential probability density
Example 4.1.7 using exponential probability density
Example 4.1.8 a hat-shaped probability density
Example 4.1.11 cumulative distribution for a hat-shaped probability density
Example 4.1.15 a ramp-shaped cumulative distribution

(Continued on next page)

Section 4.2 Expectation of continuous random variables

Example 4.2.2	expected value (continuous)
Example 4.2.6	8^{th}-order polynomial density
Example 4.2.9	moment-generating function by definition
Example 4.2.10	moments of an exponential by differentiation

Section 4.3 Special probability densities

Example 4.3.32	probabilities from the standard normal
Example 4.3.37	probabilities from the standard normal

Section 4.4 Multivariate continuous random variables

Example 4.4.2	jointly-distributed random variables
Example 4.4.6	joint CDF

Appendix D

GNU Free Documentation License

0. PREAMBLE. The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS. This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A “Secondary Section” is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document’s overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The “Invariant Sections” are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The “Cover Texts” are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A “Transparent” copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not “Transparent” is called “Opaque”.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The “Title Page” means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, “Title Page” means the text near the most prominent appearance of the work’s title, preceding the beginning of the body of the text.

The “publisher” means any person or entity that distributes copies of the Document to the public.

A section “Entitled XYZ” means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as “Acknowledgements”, “Dedications”, “Endorsements”, or “History”.) To “Preserve the Title” of such a section when you modify the Document means that it remains a section “Entitled XYZ” according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING. You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY. If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS. You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties — for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS. You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS. You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS. A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION. Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION. You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE. The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See `http://www.gnu.org/copyleft/`.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING. "Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents. To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

```
Copyright (C)  YEAR  YOUR NAME. Permission is granted to
       copy, distribute and/or modify this document under the terms of
       the GNU Free Documentation License, Version 1.3 or any later
       version published by the Free Software Foundation; with no
       Invariant Sections, no Front-Cover Texts, and no Back-Cover
       Texts. A copy of the license is included in the section entitled
       "GNU Free Documentation License".
```

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with… Texts." line with this:

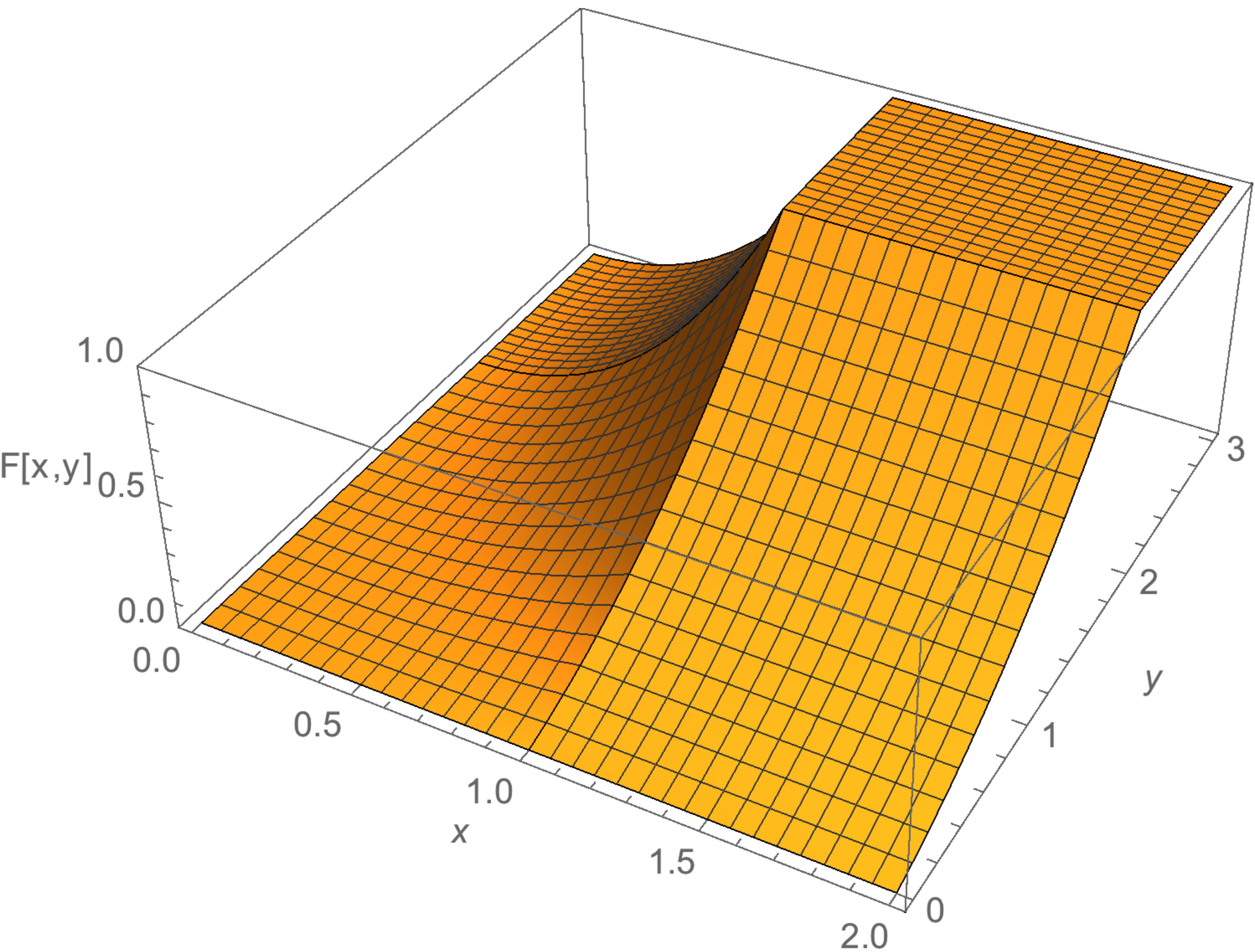
```
with the Invariant Sections being LIST THEIR TITLES, with
       the Front-Cover Texts being LIST, and with the Back-Cover Texts
       being LIST.
```

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Colophon

This book was authored in MathBook XML.



1.0
F[x,y]
0.5
0.0
0.0
0.5
1.0
x
1.5
2.0
0
1
y
2
3