MUST-HAVE MATH TOOLS FOR GRADUATE STUDY IN ECONOMICS

William Neilson

Department of Economics
University of Tennessee – Knoxville

September 2009

web.utk.edu/~wneilson/mathbook.pdf

Acknowledgments

Valentina Kozlova, Kelly Padden, and John Tilstra provided valuable proofreading assistance on the first version of this book, and I am grateful. Other mistakes were found by the students in my class. Of course, if they missed anything it is still my fault. Valentina and Bruno Wichmann have both suggested additions to the book, including the sections on stability of dynamic systems and order statistics.

The cover picture was provided by my son, Henry, who also proofread parts of the book. I have always liked this picture, and I thank him for letting me use it.

CONTENTS

CHAPTER

1

Econ and math

Every academic discipline has its own standards by which it judges the merits of what researchers claim to be true. In the physical sciences this typically requires experimental verification. In history it requires links to the original sources. In sociology one can often get by with anecdotal evidence, that is, with giving examples. In economics there are two primary ways one can justify an assertion, either using empirical evidence (econometrics or experimental work) or mathematical arguments.

Both of these techniques require some math, and one purpose of this course is to provide you with the mathematical tools needed to make and understand economic arguments. A second goal, though, is to teach you to speak mathematics as a second language, that is, to make you comfortable talking about economics using the shorthand of mathematics. In undergraduate courses economic arguments are often made using graphs. In graduate courses we tend to use equations. But equations often have graphical counterparts and vice versa. Part of getting comfortable about using math to do economics is knowing how to go from graphs to the underlying equations, and part is going from equations to the appropriate graphs.

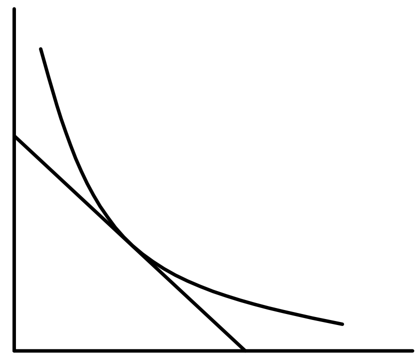


Figure 1.1: A constrained choice problem

1.1 Some important graphs

One of the fundamental graphs is shown in Figure 1.1. The axes and curves are not labeled, but that just amplifies its importance. If the axes are commodities, the line is a budget line, and the curve is an indifference curve, the graph depicts the fundamental consumer choice problem. If the axes are inputs, the curve is an isoquant, and the line is an iso-cost line, the graph illustrates the firm's cost-minimization problem.

Figure 1.1 raises several issues. How do we write the equations for the line and the curve? The line and curve seem to be tangent. How do we characterize tangency? At an even more basic level, how do we find slopes of curves? How do we write conditions for the curve to be curved the way it is? And how do we do all of this with equations instead of a picture?

Figure 1.2 depicts a different situation. If the upward-sloping line is a supply curve and the downward-sloping one is a demand curve, the graph shows how the market price is determined. If the upward-sloping line is marginal cost and the downward-sloping line is marginal benefit, the figure shows how an individual or firm chooses an amount of some activity. The questions for Figure 1.2 are: How do we find the point where the two lines intersect? How do we find the change from one intersection point to another? And how do we know that two curves will intersect in the first place?

Figure 1.3 is completely different. It shows a collection of points with a

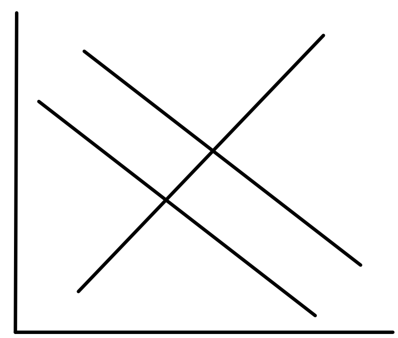


Figure 1.2: Solving simultaneous equations

line fitting through them. How do we fit the best line through these points? This is the key to doing empirical work. For example, if the horizontal axis measures the quantity of a good and the vertical axis measures its price, the points could be observations of a demand curve. How do we find the demand curve that best fits the data?

These three graphs are fundamental to economics. There are more as well. All of them, though, require that we restrict attention to two dimensions. For the first graph that means consumer choice with only two commodities, but we might want to talk about more. For the second graph it means supply and demand for one commodity, but we might want to consider several markets simultaneously. The third graph allows quantity demanded to depend on price, but not on income, prices of other goods, or any other factors. So, an important question, and a primary reason for using equations instead of graphs, is how do we handle more than two dimensions?

Math does more for us than just allow us to expand the number of dimensions. It provides *rigor*; that is, it allows us to make sure that our statements are true. All of our assertions will be logical conclusions from our initial assumptions, and so we know that our arguments are correct and we can then devote attention to the quality of the assumptions underlying them.

Figure 1.3: Fitting a line to data points

1.2 Math, micro, and metrics

The theory of microeconomics is based on two primary concepts: optimization and equilibrium. Finding how much a firm produces to maximize profit is an example of an optimization problem, as is finding what a consumer purchases to maximize utility. Optimization problems usually require finding maxima or minima, and calculus is the mathematical tool used to do this. The first section of the book is devoted to the theory of optimization, and it begins with basic calculus. It moves beyond basic calculus in two ways, though. First, economic problems often have agents simultaneously choosing the values of more than one variable. For example, consumers choose commodity bundles, not the amount of a single commodity. To analyze problems with several choice variables, we need multivariate calculus. Second, as illustrated in Figure 1.1, the problem is not just a simple maximization problem. Instead, consumers maximize utility subject to a budget constraint. We must figure out how to perform constrained optimization.

Finding the market-clearing price is an equilibrium problem. An equilibrium is simply a state in which there is no pressure for anything to change, and the market-clearing price is the one at which suppliers have no incentive to raise or lower their prices and consumers have no incentive to raise or lower their offers. Solutions to games are also based on the concept of equilibrium. Graphically, equilibrium analysis requires finding the intersection of two curves, as in Figure 1.2. Mathematically, it involves the solution of several equations in several unknowns. The branch of mathematics used for this is linear (or matrix) algebra, and so we must learn to manipulate matrices and use them to solve systems of equations.

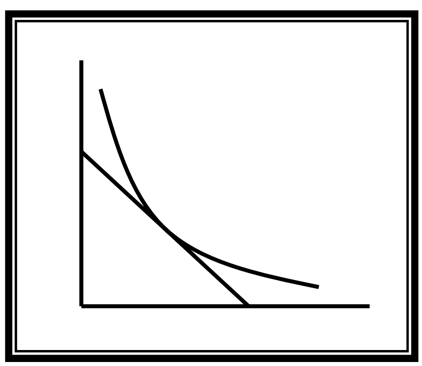
Economic exercises often involve *comparative statics analysis*, which involves finding how the optimum or equilibrium changes when one of the underlying parameters changes. For example, how does a consumer's optimal bundle change when the underlying commodity prices change? How does a firm's optimal output change when an input or an output price changes? How does the market-clearing price change when an input price changes? All of these questions are answered using comparative statics analysis. Mathematically, comparative statics analysis involves multivariable calculus, often in combination with matrix algebra. This makes it sound hard. It isn't really. But getting you to the point where you can perform comparative statics analysis means going through these two parts of mathematics.

Comparative statics analysis is also at the heart of empirical work, that is, econometrics. A typical empirical project involves estimating an equation that relates a dependent variable to a set of independent variables. The estimated equation then tells how the dependent variable changes, on average, when one of the independent variables changes. So, for example, if one estimates a demand equation in which quantity demanded is the dependent variable and the good's price, some substitute good prices, some complement good prices, and income are independent variables, the resulting equation tells how much quantity demanded changes when income rises, for example. But this is a comparative statics question. A good empirical project uses some math to derive the comparative statics results first, and then uses data to estimate the comparative statics results second. Consequently, econometrics and comparative statics analysis go hand-in-hand.

Econometrics itself is the task of fitting the best line to a set of data points, as in Figure 1.3. There is some math behind that task. Much of it is linear algebra, because matrices turn out to provide an easy way to present the relevant equations. A little bit of the math is calculus, because "best" implies "optimal," and we use calculus to find optima. Econometrics also requires a knowledge of probability and statistics, which is the third branch of mathematics we will study.

PART I

OPTIMIZATION



(multivariate calculus)

CHAPTER

2

Single variable optimization

One feature that separates economics from the other social sciences is the premise that individual actors, whether they are consumers, firms, workers, or government agencies, act rationally to make themselves as well off as possible. In other words, in economics everybody maximizes something. So, doing mathematical economics requires an ability to find maxima and minima of functions. This chapter takes a first step using the simplest possible case, the one in which the agent must choose the value of only a single variable. In later chapters we explore optimization problems in which the agent chooses the values of several variables simultaneously.

Remember that one purpose of this course is to introduce you to the mathematical tools and techniques needed to do economics at the graduate level, and that the other is to teach you to frame economic questions, and their answers, mathematically. In light of the second goal, we will begin with a graphical analysis of optimization and then find the math that underlies the graph.

Many of you have already taken calculus, and this chapter concerns single-variable, differential calculus. One difference between teaching calculus in

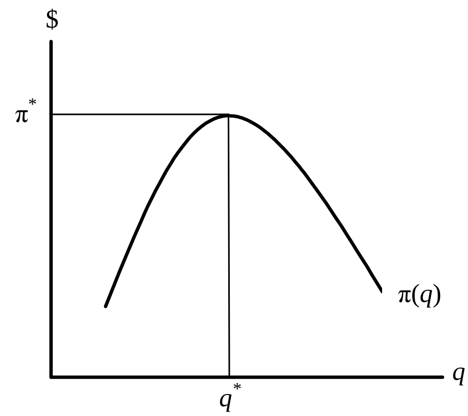


Figure 2.1: A profit function with a maximum

an economics course and teaching it in a math course is that economists almost never use trigonometric functions. The economy has cycles, but none regular enough to model using sines and cosines. So, we will skip trigonometric functions. We will, however, need logarithms and exponential functions, and they are introduced in this chapter.

2.1 A graphical approach

Consider the case of a competitive firm choosing how much output to produce. When the firm produces and sells q units it earns revenue $R(q)$ and incurs costs of $C(q)$. The profit function is

$$\pi(q) = R(q) - C(q).$$

The first term on the right-hand side is the firm's revenue, and the second term is its cost. Profit, as always, is revenue minus cost.

More importantly for this chapter, Figure 2.1 shows the firm's profit function. The maximum level of profit is π^*, which is achieved when output is q^*. Graphically this is very easy. The question is, how do we do it with equations instead?

Two features of Figure 2.1 stand out. First, at the maximum the slope of the profit function is zero. Increasing q beyond q^* reduces profit, and

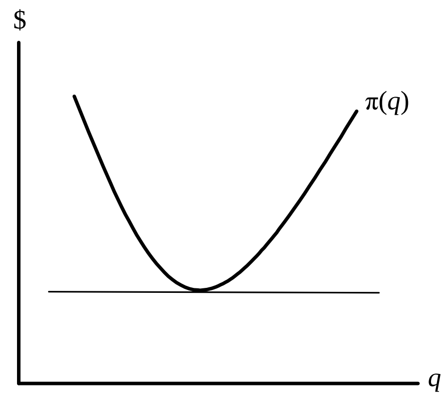


Figure 2.2: A profit function with a minimum

decreasing q below q^* also reduces profit. Second, the profit function rises up to q^* and then falls. To see why this is important, compare it to Figure 2.2, where the profit function has a minimum. In Figure 2.2 the profit function falls to the minimum then rises, while in Figure 2.1 it rises to the maximum then falls. To make sure we have a maximum, we have to make sure that the profit function is rising then falling.

This leaves us with several tasks. (1) We must find the slope of the profit function. (2) We must find q^* by finding where the slope is zero. (3) We must make sure that profit really is maximized at q^*, and not minimized. (4) We must relate our findings back to economics.

2.2 Derivatives

The derivative of a function provides its slope at a point. It can be denoted in two ways: $f'(x)$ or $df(x)/dx$. The **derivative** of the function f at x is defined as

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{2.1}$$

The idea is as follows, with the help of Figure 2.3. Suppose we start at x and consider a change to $x + h$. Then f changes from $f(x)$ to $f(x + h)$. The ratio of the change in f to the change in x is a measure of the slope:

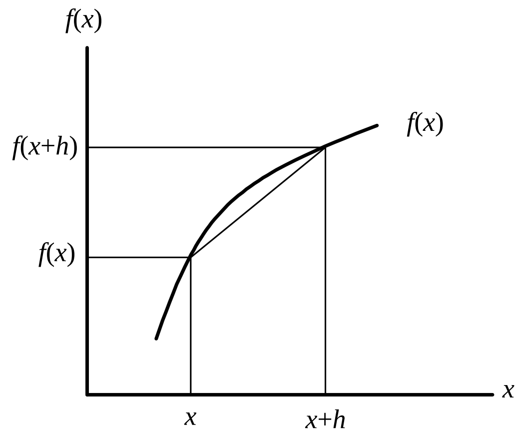


Figure 2.3: Approximating the slope of a function

$[f(x+h)-f(x)]/[(x+h)-x]$. Make the change in x smaller and smaller to get a more precise measure of the slope, and, in the limit, you end up with the derivative.

Finding the derivative comes from applying the formula in equation (2.1). And it helps to have a few simple rules in hand. We present these rules as a series of theorems.

Theorem 1 *Suppose* $f(x)=a$. *Then* $f'(x)=0$.

Proof.

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0}\frac{a-a}{h} \\ &= 0. \end{aligned}$$

■

Graphically, a *constant* function, that is, one that yields the same value for every possible x, is just a horizontal line, and horizontal lines have slopes of zero. The theorem says that the derivative of a constant function is zero.

Theorem 2 *Suppose* $f(x)=x$. *Then* $f'(x)=1$.

Proof.

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{(x+h) - x}{h} \\ &= \lim_{h \to 0} \frac{h}{h} \\ &= 1. \end{aligned}$$

■

Graphically, the function $f(x) = x$ is just a 45-degree line, and the slope of the 45-degree line is one. The theorem confirms that the derivative of this function is one.

Theorem 3 *Suppose* $f(x) = au(x)$. *Then* $f'(x) = au'(x)$.

Proof.

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{au(x+h) - au(x)}{h} \\ &= a \lim_{h \to 0} \frac{u(x+h) - u(x)}{h} \\ &= au'(x). \end{aligned}$$

■

This theorem provides a useful rule. When you multiply a function by a *scalar* (or constant), you also multiply the derivative by the same scalar. Graphically, multiplying by a scalar rotates the curve.

Theorem 4 *Suppose* $f(x) = u(x) + v(x)$. *Then* $f'(x) = u'(x) + v'(x)$.

Proof.

$$\begin{aligned} f'(x) &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \to 0} \frac{[u(x+h) + v(x+h)] - [u(x) + v(x)]}{h} \\ &= \lim_{h \to 0} \left[\frac{u(x+h) - u(x)}{h} + \frac{v(x+h) - u(x)}{h} \right] \\ &= u'(x) + v'(x). \end{aligned}$$

■

This rule says that the derivative of a sum is the sum of the derivatives.

The next theorem is the **product rule**, which tells how to take the derivative of the product of two functions.

Theorem 5 *Suppose* $f(x) = u(x)\cdot v(x)$. *Then* $f'(x) = u'(x)v(x)+u(x)v'(x)$.

Proof.

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0}\frac{[u(x+h)v(x+h)]-[u(x)v(x)]}{h} \\ &= \lim_{h\to 0}\left[\frac{[u(x+h)-u(x)]v(x)}{h}+\frac{u(x+h)[v(x+h)-v(x)]}{h}\right] \end{aligned}$$

where the move from line 2 to line 3 entails adding then subtracting $\lim_{h\to 0} u(x+h)v(x)/h$. Remembering that the limit of a product is the product of the limits, the above expression reduces to

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{[u(x+h)-u(x)]}{h}v(x)+\lim_{h\to 0}u(x+h)\frac{[v(x+h)-v(x)]}{h} \\ &= u'(x)v(x)+u(x)v'(x). \end{aligned}$$

■

We need a rule for functions of the form $f(x) = 1/u(x)$, and it is provided in the next theorem.

Theorem 6 *Suppose* $f(x) = 1/u(x)$. *Then* $f'(x) = -u'(x)/[u(x)]^2$.

Proof.

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0}\frac{\frac{1}{u(x+h)}-\frac{1}{u(x)}}{h} \\ &= \lim_{h\to 0}\frac{u(x)-u(x+h)}{h[u(x+h)u(x)]} \\ &= \lim_{h\to 0}\frac{-[u(x+h)-u(x)]}{h}\cdot\lim_{h\to 0}\frac{1}{u(x+h)u(x)} \\ &= -u'(x)\cdot\frac{1}{[u(x)]^2}. \end{aligned}$$

■

Our final rule concerns *composite functions*, that is, functions of functions. This rule is called the **chain rule**.

Theorem 7 *Suppose $f(x) = u(v(x))$. Then $f'(x) = u'(v(x)) \cdot v'(x)$.*

Proof. First suppose that there is some sequence $h_1, h_2, ...$ with $\lim_{i\to\infty} h_i = 0$ and $v(x + h_i) - v(x) \neq 0$ for all i. Then

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{u(v(x+h)) - u(v(x))}{h} \\ &= \lim_{h\to 0} \left[\frac{u(v(x+h)) - u(v(x))}{v(x+h) - v(x)} \cdot \frac{v(x+h) - v(x)}{h} \right] \\ &= \lim_{k\to 0} \frac{u(v(x)+k)) - u(v(x))}{k} \cdot \lim_{h\to 0} \frac{v(x+h) - v(x)}{h} \\ &= u'(v(x)) \cdot v'(x). \end{aligned}$$

Now suppose that there is no sequence as defined above. Then there exists a sequence $h_1, h_2, ...$ with $\lim_{i\to\infty} h_i = 0$ and $v(x + h_i) - v(x) = 0$ for all i. Let $b = v(x)$ for all x,and

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h\to 0} \frac{u(v(x+h)) - u(v(x))}{h} \\ &= \lim_{h\to 0} \frac{u(b) - u(b)}{h} \\ &= 0. \end{aligned}$$

But $u'(v(x)) \cdot v'(x) = 0$ since $v'(x) = 0$, and we are done. ■

Combining these rules leads to the following really helpful rule:

$$\frac{d}{dx} a[f(x)]^n = an[f(x)]^{n-1} f'(x). \tag{2.2}$$

This holds even if n is negative, and even if n is not an integer. So, for example, the derivative of x^n is nx^{n-1}, and the derivative of $(2x+1)^5$ is $10(2x+1)^4$. The derivative of $(4x^2-1)^{-.4}$ is $-.4(4x^2-1)^{-1.4}(8x)$.

Combining the rules also gives us the familiar division rule:

$$\frac{d}{dx}\left[\frac{u(x)}{v(x)}\right] = \frac{u'(x)v(x) - v'(x)u(x)}{[v(x)]^2}. \tag{2.3}$$

To get it, rewrite $u(x)/v(x)$ as $u(x)\cdot[v(x)]^{-1}$. We can then use the product rule and expression (2.2) to get

$$\begin{aligned}\frac{d}{dx}\left[u(x)v^{-1}(x)\right] &= u'(x)v^{-1}(x) + (-1)u(x)v^{-2}(x)v'(x) \\ &= \frac{u'(x)}{v(x)} - \frac{v'(x)u(x)}{v^2(x)}.\end{aligned}$$

Multiplying both the numerator and denominator of the first term by $v(x)$ yields (2.3).

Getting more ridiculously complicated, consider

$$f(x) = \frac{(x^3+2x)(4x-1)}{x^3}.$$

To differentiate this thing, split f into three component functions, $f_1(x) = x^3+2x$, $f_2(x) = 4x-1$, and $f_3(x) = x^3$. Then $f(x) = f_1(x)\cdot f_2(x)/f_3(x)$, and

$$f'(x) = \frac{f_1'(x)f_2(x)}{f_3(x)} + \frac{f_1(x)f_2'(x)}{f_3(x)} - \frac{f_1(x)f_2(x)f_3'(x)}{[f_3(x)]^2}.$$

We can differentiate the component functions to get $f_1(x) = 3x^2+2$, $f_2'(x) = 4$, and $f_3'(x) = 3x^2$. Plugging this all into the formula above gives us

$$f'(x) = \frac{(3x^2+2)(4x-1)}{x^3} + \frac{4(x^3+2x)}{x^3} - \frac{3(x^3+2x)(4x-1)x^2}{x^6}.$$

2.3 Uses of derivatives

In economics there are three major uses of derivatives.

The first use comes from the economics idea of "marginal this" and "marginal that." In principles of economics courses, for example, marginal cost is defined as the additional cost a firm incurs when it produces one more unit of output. If the cost function is $C(q)$, where q is quantity, marginal cost is $C(q+1) - C(q)$. We could divide output up into smaller units, though, by measuring in grams instead of kilograms, for example. Continually dividing output into smaller and smaller units of size h leads to the definition of marginal cost as

$$MC(q) = \lim_{h \to 0} \frac{c(q+h) - c(q)}{h}.$$

Marginal cost is simply the derivative of the cost function. Similarly, marginal revenue is the derivative of the revenue function, and so on.

The second use of derivatives comes from looking at their signs (the astrology of derivatives). Consider the function $y = f(x)$. We might ask whether an increase in x leads to an increase in y or a decrease in y. The derivative $f'(x)$ measures the change in y when x changes, and so if $f'(x) \geq 0$ we know that y increases when x increases, and if $f'(x) \leq 0$ we know that y decreases when x increases. So, for example, if the marginal cost function $MC(q)$ or, equivalently, $C'(q)$ is positive we know that an increase in output leads to an increase in cost.

The third use of derivatives is for finding maxima and minima of functions. This is where we started the chapter, with a competitive firm choosing output to maximize profit. The profit function is $\pi(q) = R(q) - C(q)$. As we saw in Figure 2.1, profit is maximized when the slope of the profit function is zero, or

$$\frac{d\pi}{dq} = 0.$$

This condition is called a **first-order condition**, often abbreviated as **FOC**. Using our rules for differentiation, we can rewrite the FOC as

$$\frac{d\pi}{dq} = R'(q^*) - C'(q^*) = 0, \tag{2.4}$$

which reduces to the familiar rule that a firm maximizes profit by producing where marginal revenue equals marginal cost.

Notice what we have done here. We have not used numbers or specific functions and, aside from homework exercises, we rarely will. Using general functions leads to expressions involving general functions, and we want to interpret these. We know that $R'(q)$ is marginal revenue and $C'(q)$ is marginal cost. We end up in the same place we do using graphs, which is a good thing. The power of the mathematical approach is that it allows us to apply the same techniques in situations where graphs will not work.

2.4 Maximum or minimum?

Figure 2.1 shows a profit function with a maximum, but Figure 2.2 shows one with a minimum. Both of them generate the same first-order condition: $d\pi/dq = 0$. So what property of the function tells us that we are getting a maximum and not a minimum?

In Figure 2.1 the slope of the curve decreases as q increases, while in Figure 2.2 the slope of the curve increases as q increases. Since slopes are just derivatives of the function, we can express these conditions mathematically by taking derivatives of derivatives, or *second derivatives*. The second derivative of the function $f(x)$ is denoted $f''(x)$ or d^2f/dx^2. For the function to have a maximum, like in Figure 2.1, the derivative should be decreasing, which means that the second derivative should be negative. For the function to have a minimum, like in Figure 2.2, the derivative should be increasing, which means that the second derivative should be positive. Each of these is called a **second-order condition** or **SOC**. The second-order condition for a maximum is $f''(x) \leq 0$, and the second-order condition for a minimum is $f''(x) \geq 0$.

We can guarantee that profit is maximized, at least locally, if $\pi''(q^*) \leq 0$. We can guarantee that profit is maximized *globally* if $\pi''(q) \leq 0$ for all possible values of q. Let's look at the condition a little more closely. The first derivative of the profit function is $\pi'(q) = R'(q) - C'(q)$ and the second derivative is $\pi''(q) = R''(q) - C''(q)$. The second-order condition for a maximum is $\pi''(q) \leq 0$, which holds if $R''(q) \leq 0$ and $C''(q) \geq 0$. So, we can guarantee that profit is maximized if the second derivative of the revenue function is nonpositive and the second derivative of the cost function is nonnegative. Remembering that $C'(q)$ is marginal cost, the condition $C''(q) \geq 0$ means that marginal cost is increasing, and this has an economic interpretation: each additional unit of output adds more to total cost than any unit preceding it. The condition $R''(q) \leq 0$ means that marginal revenue is decreasing, which means that the firm earns less from each additional unit it sells.

One special case that receives considerable attention in economics is the one in which $R(q) = pq$, where p is the price of the good. This is the revenue function for a price-taking firm in a perfectly competitive industry. Then $R'(q) = p$ and $R''(q) = 0$, and the first-order condition for profit maximization is $p - C'(q) = 0$, which is the familiar condition that price equals marginal cost. The second-order condition reduces to $-C''(q) \leq 0$, which says that marginal cost must be nondecreasing.

2.5 Logarithms and the exponential function

The functions $\ln x$ and e^x turn out to play an important role in economics. The first is the natural logarithm, and the second is the exponential function. They are related:

$$\ln e^x = e^{\ln x} = x.$$

The number $e \approx 2.718$. Without going into why these functions are special for economics, let me show you why they are special for math.

We know that

$$\frac{d}{dx}\left(\frac{x^n}{n}\right) = x^{n-1}.$$

We can get the function x^2 by differentiating $x^3/3$, the function x by differentiating $x^2/2$, the function x^{-2} by differentiating $-x^{-1}$, the function x^{-3} by differentiating $-x^{-2}/2$, and so on. But how can we get the function x^{-1}? We cannot get it by differentiating $x^0/0$, because that expression does not exist. We cannot get it by differentiating x^0, because $dx^0/dx = 0$. So how do we get x^{-1} as a derivative? The answer is the natural logarithm:

$$\frac{d}{dx}\ln x = \frac{1}{x}.$$

Logarithms have two additional useful properties:

$$\ln xy = \ln x + \ln y.$$

and

$$\ln(x^a) = a \ln x.$$

Combining these yields

$$\ln(x^a y^b) = a \ln x + b \ln y. \tag{2.5}$$

The left-hand side of this expression is non-linear, but the right-hand side is linear in the logarithms, which makes it easier to work with. Economists often use the form in (**??**) for utility functions and production functions.

The exponential function e^x also has an important differentiation property: it is its own derivative, that is,

$$\frac{d}{dx}e^x = e^x.$$

This implies that the derivative of $e^{u(x)} = u'(x)e^{u(x)}$.

2.6 Problems

1. Compute the derivatives of the following functions:

   (a) $f(x) = 12(x^3+1)^2 + 3\ln\frac{x}{2} - 5x^{-4}$

   (b) $f(x) = 1/(4x-2)^5$

   (c) $f(x) = e^{-14x^3+2x}$

   (d) $f(x) = (9\ln x)/x^{0.3}$

   (e) $f(x) = \frac{ax^2-b}{cx-d}$

2. Compute the derivative of the following functions:

   (a) $f(x) = 12(x-1)^2$

   (b) $g(x) = (\ln 3x)/(4x^2)$

   (c) $h(x) = 1/(3x^2-2x+1)^4$

   (d) $f(x) = xe^{-x}$

   (e)
   $$g(x) = \frac{(2x^2-3)\sqrt{5x^3+6}}{8-9x}$$

3. Use the definition of the derivative (expression 2.1) to show that the derivative of x^2 is $2x$.

4. Use the definition of a derivative to prove that the derivative of $1/x$ is $-1/x^2$.

5. Answer the following:

   (a) Is $f(x) = 2x^3 - 12x^2$ increasing or decreasing at $x = 3$?

   (b) Is $f(x) = \ln x$ increasing or decreasing at $x = 13$?

   (c) Is $f(x) = e^{-x}x^{1.5}$ increasing or decreasing at $x = 4$?

   (d) Is $f(x) = \frac{4x-1}{x+2}$ increasing or decreasing at $x = 2$?

6. Answer the following:

   (a) Is $f(x) = (3x - 2)/(4x + x^2)$ increasing or decreasing at $x = -1$?

   (b) Is $f(x) = 1/\ln x$ increasing or decreasing at $x = e$?

   (c) Is $f(x) = 5x^2 + 16x - 12$ increasing or decreasing at $x = -6$?

7. Optimize the following functions, and tell whether the optimum is a local maximum or a local minimum:

   (a) $f(x) = -4x^2 + 10x$

   (b) $f(x) = 120x^{0.7} - 6x$

   (c) $f(x) = 4x - 3\ln x$

8. Optimize the following functions, and tell whether the optimum is a local maximum or a local minimum:

   (a) $f(x) = 4x^2 - 24x + 132$

   (b) $f(x) = 20\ln x - 4x$

   (c) $f(x) = 36x - (x + 1)/(x + 2)$

9. Consider the function $f(x) = ax^2 + bx + c$.

   (a) Find conditions on a, b, and c that guarantee that $f(x)$ has a unique global maximum.

   (b) Find conditions on a, b, and c that guarantee that $f(x)$ has a unique global minimum.

10. Beth has a minion (named Henry) and benefits when the minion exerts effort, with minion effort denoted by m. Her benefit from m units of minion effort is given by the function $b(m)$. The minion does not like exerting effort, and his cost of effort is given by the function $c(m)$.

    (a) Suppose that Beth is her own minion and that her effort cost function is also $c(m)$. Find the equation determining how much effort she would exert, and interpret it.

    (b) What are the second-order condtions for the answer in (a) to be a maximum?

    (c) Suppose that Beth pays the minion w per unit of effort. Find the equation determining how much effort the minion will exert, and interpret it.

    (d) What are the second-order conditions for the answer in (c) to be a maximum?

11. A firm (Bilco) can use its manufacturing facility to make either widgets or gookeys. Both require labor only. The production function for widgets is

$$W = 20L^{1/2}$$

and the production function for gookeys is

$$G = 30L.$$

The wage rate is \$11 per unit of time, and the prices of widgets and gokeys are \$9 and \$3 per unit, repsectively. The manufacturing facility can accomodate 60 workers and no more. How much of each product should Bilco produce per unit of time? (Hint: If Bilco devotes L units of labor to widget production it has $60 - L$ units of labor to devote to gookey production, and its profit function is $\pi(L) = 9 \cdot 20L^{1/2} + 3 \cdot 30(60 - L) - 11 \cdot 60$.)

CHAPTER

3

Optimization with several variables

Almost all of the intuition behind optimization comes from looking at problems with a single choice variable. In economics, though, problems often involve more than one choice variable. For example, consumers choose bundles of commodities, so must choose amounts of several different goods simultaneously. Firms use many inputs and must choose their amounts simultaneously. This chapter addresses issues that arise when there are several variables.

The previous chapter used graphs to generate intuition. We cannot do that here because I am bad at drawing graphs with more than two dimensions. Instead, our intuition will come from what we learned in the last chapter.

3.1 A more complicated profit function

In the preceding chapter we looked at a profit function in which the firm chose how much output to produce. This time, instead of focusing on outputs, let's focus on inputs. Suppose that the firm can use n different inputs, and denote the amounts by $x_1, ..., x_n$. When the firm uses x_1 units of input 1, x_2 units of input 2, and so on, its output is given by the production function

$$Q = F(x_1, ..., x_n).$$

Inputs are costly, and we will assume that the firm can purchase as much of input i as it wants for price r_i, and it can sell as much of its output as it wants at the competitive price p. How much of each input should the firm use to maximize profit?

We know what to do when there is only one input ($n = 1$). Call the input labor (L) and its price the wage (w). The production function is then $Q = F(L)$. When the firm employs L units of labor it produces $F(L)$ units of output and sells them for p units each, for revenue of $pF(L)$. Its only cost is a labor cost equal to wL because it pays each unit of labor the wage w. Profit, then, is $\pi(L) = pF(L) - wL$. The first-order condition is

$$\pi'(L) = pF'(L) - w = 0,$$

which can be interpreted as the firm's profit maximizing labor demand equating the value marginal product of labor $pF'(L)$ to the wage rate. Using one additional unit of labor costs an additional w but increases output by $F'(L)$, which increases revenue by $pF'(L)$. The firm employs labor as long as each additional unit generates more revenue than it costs, and stops when the added revenue and the added cost exactly offset each other.

What happens if there are two inputs ($n = 2$), call them capital (K) and labor (L)? The production function is then $Q = F(K, L)$, and the corresponding profit function is

$$\pi(K, L) = pF(K, L) - rK - wL. \tag{3.1}$$

How do we find the first-order condition? That is the task for this chapter.

3.2 Vectors and Euclidean space

Before we can find a first-order condition for (3.1), we first need some terminology. A **vector** is an array of n numbers, written $(x_1, ..., x_n)$. In our example of the input-choosing profit-maximizing firm, the vector $(x_1, ..., x_n)$ is an *input vector*. For each i between 1 and n, the quantity x_i is the amount of the i-th input. More generally, we call x_i the i-th **component** of the vector $(x_1, ..., x_n)$. The number of components in a vector is called the **dimension** of the vector; the vector $(x_1, ..., x_n)$ is n-dimensional.

Vectors are collections of numbers. They are also numbers themselves, and it will help you if you begin to think of them this way. The set of real numbers is commonly denoted by $\mathbb{R}$, and we depict $\mathbb{R}$ using a number line. We can depict a 2-dimensional vector using a coordinate plane anchored by two real lines. So, the vector (x_1, x_2) is in $\mathbb{R}^2$, which can be thought of as $\mathbb{R} \times \mathbb{R}$. We call $\mathbb{R}^2$ the 2-dimensional **Euclidean space**. When you took plane geometry in high school, this was Euclidean geometry. When a vector has n components, it is in $\mathbb{R}^n$, or n-dimensional Euclidean space.

In this text vectors are sometimes written out as $(x_1, ..., x_n)$, but sometimes that is cumbersome. We use the symbol $\bar{x}$ to denote the vector whose components are $x_1, ..., x_n$. That way we can talk about operations involving two vectors, like $\bar{x}$ and $\bar{y}$.

Three common operations are used with vectors. We begin with addition:

$$\bar{x} + \bar{y} = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n).$$

Adding vectors is done component-by-component.

Multiplication is more complicated, and there are two notions. One is **scalar multiplication**. If $\bar{x}$ is a vector and a is a **scalar** (a real number), then

$$a\bar{x} = (ax_1, ax_2, ..., ax_n).$$

Scalar multiplication is achieved by multiplying each component of the vector by the same number, thereby either "scaling up" or "scaling down" the vector. Vector subtraction can be achieved through addition and using -1 as the scalar: $\bar{x} - \bar{y} = \bar{x} + (-1)\bar{y}$. The other form of multiplication is the **inner product**, sometimes called the **dot product**. It is done using the formula

$$\bar{x} \cdot \bar{y} = x_1y_1 + x_2y_2 + ... + x_ny_n.$$

Vector addition takes two vectors and yields another vector, and scalar multiplication takes a vector and a scalar and yields another vector. But the inner product takes two vectors and yields a scalar, or real number. You might wonder why we would ever want such a thing.

Here is an example. Suppose that a firm uses n inputs in amounts $x_1, ..., x_n$. It pays r_i per unit of input i. What is its total production cost? Obviously, it is $r_1x_1 + ... + r_nx_n$, which can be easily written as $\bar{r} \cdot \bar{x}$.

Similarly, if a consumer purchases a commodity bundle given by the vector $\bar{x} = (x_1, ..., x_n)$ and pays prices given by the vector $\bar{p} = (p_1, ..., p_n)$, her total expenditure is $\bar{p} \cdot \bar{x}$. Often it is more convenient to leave the "dot" out of the inner product, and just write $\bar{p}\bar{x}$. A second use comes from looking at $\bar{x} \cdot \bar{x} = x_1^2 + ... + x_n^2$. Then $\sqrt{\bar{x} \cdot \bar{x}}$ is the distance from the point $\bar{x}$ (remember, it's a number) to the origin. This is also called the **norm** of the vector $\bar{x}$, and it is written $\|\bar{x}\| = (\bar{x} \cdot \bar{x})^{\frac{1}{2}}$.

Both vector addition and the inner product are *commutative*, that is, they do not depend on the order in which the two vectors occur. This will contrast with matrices in a later chapter, where matrix multiplication is dependent on the order in which the matrices are written.

Vector analysis also requires some definitions for ordering vectors. For real numbers we have the familiar relations $>$, $\geq$, $=$, $\leq$, and $<$. For vectors,

$$\begin{aligned} \bar{x} &= \bar{y} \text{ if } x_i = y_i \text{ for all } i = 1, .., n; \\ \bar{x} &\geq \bar{y} \text{ if } x_i \geq y_i \text{ for all } i = 1, ..., n; \\ \bar{x} &> \bar{y} \text{ if } \bar{x} \geq \bar{y} \text{ but } \bar{x} \neq \bar{y}; \end{aligned}$$

and

$$\bar{x} \gg \bar{y} \text{ if } x_i > y_i \text{ for all } i = 1, ..., n.$$

From the third one it follows that $\bar{x} > \bar{y}$ if $x_i \geq y_i$ for all $i = 1, ..., n$ and $x_i > y_i$ for *some* i between 1 and n. The fourth condition can be read $\bar{x}$ is strictly greater than $\bar{y}$ *component-wise.*

3.3 Partial derivatives

The trick to maximizing a function of several variables, like (3.1), is to maximize it according to each variable separately, that is, by finding a first-order condition for the choice of K and another one for the choice of L. In general both of these conditions will depend on the values of both K and L, so we will have to solve some simultaneous equations. We will get to that later. The point is that we want to differentiate (3.1) once with respect to K and once with respect to L.

Differentiating a function of two or more variables with respect to only one of them is called *partial differentiation*. Let $f(x_1, ..., x_n)$ be a general function of n variables. The ***i*-th partial derivative** of f is

$$\frac{\partial f}{\partial x_i}(\bar{x}) = \lim_{h \to 0} \frac{f(x_1, x_2, ..., x_{i-1}, x_i + h, x_{i+1}, ..., x_n) - f(x_1, ..., x_n)}{h}. \quad (3.2)$$

This definition might be a little easier to see with one more piece of notation. The **coordinate vector** $\bar{e}^i$ is the vector with components given by $e^i_i = 1$ and $e^i_j = 0$ when $j \neq i$. The first coordinate vector is $\bar{e}^1 = (1, 0, ..., 0)$, the second coordinate vector is $\bar{e}^2 = (0, 1, 0, ..., 0)$, and so on through the n-th coordinate vector $\bar{e}^n = (0, ..., 0, 1)$. So, coordinate vector $\bar{e}^i$ has a one in the i-th place and zeros everywhere else. Using coordinate vectors, the definition of the i-th partial derivative in (3.2) can be rewritten

$$\frac{\partial f}{\partial x_i}(\bar{x}) = \lim_{h \to 0} \frac{f(\bar{x} + h\bar{e}^i) - f(\bar{x})}{h}.$$

The i-th partial derivative of the function f is simply the derivative one gets by holding all of the components fixed except for the i-th component. One takes the partial by pretending that all of the other variables are really just constants and differentiating as if it were a single-variable function. For example, consider the function $f(x_1, x_2) = (5x_1 - 2)(7x_2 - 3)^2$. The partial derivatives are $f_1(x_1, x_2) = 5(7x_2 - 3)^2$ and $f_2(x_1, x_2) = 14(5x_1 - 2)(7x_2 - 3)$.

We sometimes use the notation $f_i(\bar{x})$ to denote $\partial f(\bar{x})/\partial x_i$. When a function is defined over n-dimensional vectors it has n different partial derivatives.

It is also possible to take partial derivatives of partial derivatives, much like second derivatives in single-variable calculus. We use the notation

$$f_{ij}(\bar{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x}).$$

We call $f_{ii}(\bar{x})$ the *second partial* of f with respect to x_i, and we call $f_{ij}(\bar{x})$ the *cross partial* of $f(\bar{x})$ with respect to x_i and x_j. It is important to note that, in most cases,

$$f_{ij}(\bar{x}) = f_{ji}(\bar{x}),$$

that is, the order of differentiation does not matter. In fact, this result is important enough to have a name: **Young's Theorem**.

Partial differentiation requires a restatement of the chain rule:

Theorem 8 *Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by*

$$f(u_1(x), u_2(x), ..., u_n(x)),$$

where $u_1, ..., u_n$ are functions of the one-dimensional variable x. Then

$$\frac{df}{dx} = f_1 \cdot u_1'(x) + f_2 \cdot u_2'(x) + ... + f_n \cdot u_n'(x)$$

This rule is best explained using an example. Suppose that the function is $f(3x^2 - 2, 5\ln x)$. It's derivative with respect to x is

$$\frac{d}{dx}f(3x^2 - 2, 5\ln x) = f_1(3x^2 - 2, 5\ln x) \cdot (6x) + f_2(3x^2 - 2, 5\ln x) \cdot \left(\frac{5}{x}\right).$$

The basic rule to remember is that when variable we are differentiating with respect to appears in several places in the function, we differentiate with respect to each argument separately and then add them together. The following lame example shows that this works. Let $f(y_1, y_2) = y_1 y_2$, but the values y_1 and y_2 are both determined by the value of x, with $y_1 = 2x$ and $y_2 = 3x^2$. Substituting we have

$$\begin{aligned} f(x) &= (2x)(3x^2) = 6x^3 \\ \frac{df}{dx} &= 18x^2. \end{aligned}$$

But, if we use the chain rule, we get

$$\begin{aligned} \frac{df}{dx} &= \frac{dy_1}{dx} \cdot y_2 + \frac{dy_2}{dx} \cdot y_1 \\ &= (2)(3x^2) + (6x)(2x) = 18x^2. \end{aligned}$$

It works.

3.4 Multidimensional optimization

Let's return to our original problem, maximizing the profit function given in expression (3.1). The firm chooses both capital K and labor L to maximize

$$\pi(K, L) = pF(K, L) - rK - wL.$$

Think about the firm as solving two problems simultaneously: (i) given the optimal amount of labor, L^*, the firm wants to use the amount of capital that maximizes $\pi(K, L^*)$; and (ii) given the optimal amount of capital, K^*, the firm wants to employ the amount of labor that maximizes $\pi(K^*, L)$. Problem (i) translates into

$$\frac{\partial}{\partial K}\pi(K^*, L^*) = 0$$

and problem (ii) translates into

$$\frac{\partial}{\partial L}\pi(K^*, L^*) = 0.$$

Thus, optimization in several dimensions is just like optimization in each single dimension separately, with the provision that all of the optimization problems must be solved together. The two equations above are the first-order conditions for the profit-maximization problem.

To see how this works, suppose that the production function is $F(K, L) = K^{1/2} + L^{1/2}$, that the price of the good is $p = 10$, the price of capital is $r = 5$, and the wage rate is $w = 4$. Then the profit function is $\pi(K, L) = 10(K^{1/2} + L^{1/2}) - 5K - 4L$. The first-order conditions are

$$\frac{\partial}{\partial K}\pi(K, L) = 5K^{-1/2} - 5 = 0$$

and

$$\frac{\partial}{\partial L}\pi(K, L) = 5L^{-1/2} - 4 = 0.$$

The first equation gives us $K^* = 1$ and the second gives us $L^* = 25/16$. In this example the two first-order conditions were independent, that is, the FOC for K did not depend on L and the FOC for L did not depend on K. This is not always the case, as shown by the next example.

Example 1 *The production function is $F(K, L) = K^{1/4}L^{1/2}$, the price is 12, the price of capital is $r = 6$, and the wage rate is $w = 6$. Find the optimal values of K and L.*

Solution. The profit function is

$$\pi(K, L) = 12K^{1/4}L^{1/2} - 6K - 6L.$$

The first-order conditions are

$$\frac{\partial}{\partial K}\pi(K, L) = 3K^{-3/4}L^{1/2} - 6 = 0 \tag{3.3}$$

and

$$\frac{\partial}{\partial L}\pi(K, L) = 6K^{1/4}L^{-1/2} - 6 = 0. \tag{3.4}$$

To solve these, note that (3.4) can be rearranged to get

$$\begin{aligned} \frac{K^{1/4}}{L^{1/2}} &= 1 \\ K^{1/4} &= L^{1/2} \\ K &= L^2 \end{aligned}$$

where the last line comes from raising both sides to the fourth power. Plugging this into (3.3) yields

$$\begin{aligned} \frac{L^{1/2}}{K^{3/4}} &= 2 \\ \frac{L^{1/2}}{(L^2)^{3/4}} &= 2 \\ \frac{L^{1/2}}{L^{3/2}} &= 2 \\ \frac{1}{L} &= 2 \\ L &= \frac{1}{2}. \end{aligned}$$

Plugging this back into $K = L^2$ yields $K = 1/4$. ■

This example shows the steps for solving a multi-dimensional optimization problem.

Now let's return to the general problem to see what the first-order conditions tell us. The general profit-maximization problem is

$$\max_{x_1,\ldots,x_n} pF(x_1, ..., x_n) - r_1 x_1 - ... - r_n x_n$$

or, in vector notation,

$$\max_{\bar{x}} pF(\bar{x}) - r \cdot \bar{x}.$$

The first-order conditions are:

$$\begin{gathered} pF_1(x_1, ..., x_n) - r_1 = 0 \\ \vdots \\ pF_n(x_1, ..., x_n) - r_n = 0. \end{gathered}$$

The i-th FOC is $pF_i(\bar{x}) = r_i$, which is the condition that the value marginal product of input i equals its price. This is the same as the condition for a single variable, and it holds for every input.

3.5 Comparative statics analysis

Being able to do multivariate calculus allows us to do one of the most important tasks in microeconomics: *comparative statics analysis*. The standard comparative statics questions is, "How does the optimum change when one of the underlying variables changes?" For example, how does the firm's demand for labor change when the output price changes, or when the wage rate changes?

This is an important problem, and many papers (and dissertations) have relied on not much more than comparative statics analysis. If there is one tool you have in your kit at the end of the course, it should be comparative statics analysis.

To see how it works, let's return to the profit maximization problem with a single choice variable:

$$\max_L pF(L) - wL.$$

The FOC is

$$pF'(L) - w = 0. \tag{3.5}$$

The comparative statics question is, how does the optimal value of L change when p changes?

To answer this, let's first assume that the marginal product of labor is strictly decreasing, so that

$$F''(L) < 0.$$

Note that this guarantees that the second-order condition for a maximization is satisfied. The trick we now take is to **implicitly differentiate** equation (3.5) with respect to p, treating L as a function of p. In other words, rewrite the FOC so that L is replaced by the function $L^*(p)$:

$$pF'(L^*(p)) - w = 0$$

and differentiate both sides of the expression with respect to p. We get

$$F'(L^*(p)) + pF''(L^*(p))\frac{dL^*}{dp} = 0.$$

The comparative statics question is now simply the astrology question, "What is the sign of dL^*/dp?" Rearranging the above equation to isolate dL^*/dp on the left-hand side gives us

$$\frac{dL^*}{dp} = -\frac{F'(L^*)}{pF''(L^*)}.$$

We know that $F'(L) > 0$ because production functions are increasing, and we know that $pF''(L^*) < 0$ because we assumed strictly diminishing marginal product of labor, i.e. $F''(L) < 0$. So, dL^*/dp has the form of the negative of a ratio of a positive number to a negative number, which is positive. This tells us that the firm demands more labor when the output price rises, which makes sense: when the output price rises producing output becomes more profitable, and so the firm wants to expand its operation to generate more profit.

We can write the comparative statics problem generally. Suppose that the **objective function**, that is, the function the agent wants to optimize, is $f(x, s)$, where x is the choice variable and s is a shift parameter. Assume that the second-order condition holds strictly, so that $f_{xx}(x, s) < 0$ for a maximization problem and $f_{xx}(x, s) > 0$ for a minimization problem. These conditions guarantee that there is no "flat spot" in the objective function, so that there is a unique solution to the first-order condition. Let x^* denote the optimal value of x. The comparative statics question is, "What is the sign of dx^*/ds?" To get the answer, first derive the first-order condition:

$$f_x(x, s) = 0.$$

Next implicitly differentiate with respect to s to get

$$f_{xx}(x, s)\frac{dx^*}{ds} + f_{xs}(x, s) = 0.$$

Rearranging yields the comparative statics derivative

$$\frac{dx^*}{ds} = -\frac{f_{xs}(x, s)}{f_{xx}(x, s)}. \tag{3.6}$$

We want to know the sign of this derivative.

For a maximization problem we have $f_{xx} < 0$ by assumption and so the negative sign cancels out the sign of the denominator. Consequently, the sign of the comparative statics derivative dx^*/ds is the same as the sign of the numerator, $f_{xs}(x, s)$. For a minimization problem we have $f_{xx} > 0$, and so the comparative statics derivative dx^*/ds has the opposite sign from the partial derivative $f_{xs}(x, s)$.

Example 2 *A person must decide how much to work in a single 24-hour day. She gets paid w per hour of labor, but gets utility from each hour she* *does not spend at work. This utility is given by the function $u(t)$, where t is the amount of time spent away from work, and u has the properties $u'(t) > 0$ and $u''(t) < 0$. Does the person work more or less when her wage increases?*

Solution. Let L denote the amount she works, so that $24 - L$ is is the amount of time she does not spend at work. Her utility from this leisure time is therefore $u(24 - L)$, and her objective is

$$\max_L wL + u(24 - L).$$

The FOC is

$$w - u'(24 - L^*) = 0.$$

Write L^* as a function of w to get $L^*(w)$ and rewrite the FOC as:

$$w - u'(24 - L^*(w)) = 0.$$

Differentiate both sides with respect to w (this is implicit differentiation) to get

$$1 + u''(24 - L^*)\frac{dL^*}{dw} = 0.$$

Solving for the comparative statics derivative dL^*/dw yields

$$\frac{dL^*}{dw} = -\frac{1}{u''(24 - L^*)} > 0.$$

She works more when the wage rate w increases. ■

3.5.1 An alternative approach (that I don't like)

Many people use an alternative approach to comparative statics analysis. It gets to the same answers, but I do not like this approach as much. We will get to why later.

The approach begins with total differential, and the **total differential** of the function $g(x_1, ..., x_n)$ is

$$dg = g_1 dx_1 + ... + g_n dx_n.$$

We want to use total differentials to get comparative statics derivatives.

Remember our comparative statics problem: we choose x to optimize $f(x, s)$. The FOC is

$$f_x(x, s) = 0.$$

Let's take the total differential of both sides. The total differential of the right-hand side is zero, and the total differential of the left-hand side is

$$d[f_x(x, s)] = f_{xx}dx + f_{xs}ds.$$

Setting the above expression equal to zero yields

$$f_{xx}dx + f_{xs}ds = 0.$$

The derivative we want is the comparative statics derivative dx/ds. We can solve for this expression in the above equation:

$$\begin{aligned} f_{xx}dx + f_{xs}ds &= 0 \\ f_{xx}dx &= -f_{xs}ds \\ \frac{dx}{ds} &= -\frac{f_{xs}}{f_{xx}}. \end{aligned} \tag{3.7}$$

This is exactly the comparative statics derivative we found above in equation (3.6). So the method works, and many students find it straightforward and easier to use than implicit differentiation.

Let's stretch our techniques a little and have a problem with two shift parameters, s and r, instead of just one. The problem is to optimize $f(x, r, s)$, and the FOC is

$$f_x(x, r, s) = 0.$$

If we want to do comparative statics analysis using our (preferred) implicit differentiation approach, we would first write x as a function of the two shift parameters, so that

$$f_x(x(r, s), r, s) = 0.$$

To find the comparative statics derivative dx/dr, we implicitly differentiate with respect to r to get

$$\begin{aligned} f_{xx}\frac{dx}{dr} + f_{xr} &= 0 \\ \frac{dx}{dr} &= -\frac{f_{xr}}{f_{xx}}. \end{aligned}$$

This should not be a surprise, since it is just like expression (3.6) except it replaces s with r. Using total differentials, we would first take the total differential of $f_x(x, r, s)$ to get

$$f_{xx}dx + f_{xr}dr + f_{xs}ds = 0.$$

We want to solve for dx/dr, and doing so yields

$$\begin{aligned} f_{xx}dx + f_{xr}dr + f_{xs}ds &= 0 \\ f_{xx}dx &= -f_{xr}dr - f_{xs}ds \\ \frac{dx}{dr} &= -\frac{f_{xr}}{f_{xx}} - \frac{f_{xs}}{f_{xx}}\frac{ds}{dr}. \end{aligned}$$

On the face of it, this does not look like the same answer. But, both s and r are shift parameters, so s is not a function of r. That means that $ds/dr = 0$. Substituting this in yields

$$\frac{dx}{dr} = -\frac{f_{xr}}{f_{xx}}$$

as expected.

So what is the difference between the two approaches? In the implicit differentiation approach we recognized that s does not depend on r at the *beginning* of the process, and in the total differential approach we recognized it at the end. So both work, it's just a matter of when you want to do your remembering.

All of that said, I still like the implicit differentiation approach better. To see why, think about what the derivative dx/ds means. As we constructed it back in equation (2.1), dx/ds is the limit of $\Delta x/\Delta s$ as $\Delta s \to 0$. According to this intuition, ds is the limit of Δs as it goes to zero, so ds is zero. But we divided by it in equation (3.7), and you were taught very young that you cannot divide by zero. So, on a purely mathematical basis, I object to the total differential approach because it entails dividing by zero, and I prefer to think of dx/ds as a single entity with a long name, and not a ratio of dx and ds. On a practical level, though, the total differential approach works just fine. It's just not going to show up anywhere else in this book.

3.6 Problems

1. Consider the vectors $\bar{x} = (4, -3, 6, 2)$ and $\bar{y} = (6, 1, 7, 7)$.

(a) Write down the vector $2\bar{y} + 3\bar{x}$.

(b) Which of the following, if any, are true: $\bar{x} = \bar{y}$, $\bar{x} \leq \bar{y}$, $\bar{x} < \bar{y}$, or $\bar{x} \ll \bar{y}$?

(c) Find the inner product $\bar{x} \cdot \bar{y}$.

(d) Is $\sqrt{\bar{x} \cdot \bar{x}} + \sqrt{\bar{y} \cdot \bar{y}} \geq \sqrt{(\bar{x} + \bar{y}) \cdot (\bar{x} + \bar{y})}$?

2. Consider the vectors $\bar{x} = (5, 0, -6, -2)$ and $\bar{y} = (3, 2, 3, 2)$.

(a) Write down the vector $6\bar{x} - 4\bar{y}$.

(b) Find the inner product $\bar{x} \cdot \bar{y}$.

(c) Verify that $\sqrt{\bar{x} \cdot \bar{x}} + \sqrt{\bar{y} \cdot \bar{y}} > \sqrt{(\bar{x} + \bar{y}) \cdot (\bar{x} + \bar{y})}$.

3. Consider the function $f(x, y) = 4x^2 + 3y^2 - 12xy + 18x$.

(a) Find the partial derivative $f_x(x, y)$.

(b) Find the partial derivative $f_y(x, y)$.

(c) Find the critical point of f.

4. Consider the function $f(x, y) = 16xy - 4x + 2/y$.

(a) Find the partial derivative $f_x(x, y)$.

(b) Find the partial derivative $f_y(x, y)$.

(c) Find the critical point of f.

5. Consider the function $u(x, y) = 3 \ln x + 2 \ln y$.

(a) Write the equation for the indifference curve corresponding to the utility level k.

(b) Find the slope of the indifference curve at point (x, y).

6. A firm faces inverse demand function $p(q) = 120 - 4q$, where q is the firm's output. Its cost function is cq.

(a) Write down the firm's profit function.

(b) Find the profit-maximizing level of profit as a function of the unit cost c.

(c) Find the comparative statics derivative dq/dc. Is it positive or negative?

(d) Write the maximized profit function as a function of c.

(e) Find the derivative showing how profit changes when c changes.

(f) Show that $d\pi/dc = -q$.

7. Find dx/da from each of the following expressions.

(a)

$$15x^2 + 3xa - 5\frac{x}{a} = 20.$$

(b)

$$6x^2a = 5a - 5xa^2$$

8. Each worker at a firm can produce 4 units per hour, each worker must be paid \$$w$ per hour, and the firm's revenue function is $R(L) = 30\sqrt{L}$, where L is the number of workers employed (fractional workers are okay). The firm's profit function is $\pi(L) = 30\sqrt{4L} - wL$.

(a) Show that $L^* = 900/w^2$.

(b) Find dL^*/dw. What's its sign?

(c) Find $d\pi^*/dw$. What's its sign?

9. An isoquant is a curve showing the combinations of inputs that all lead to the same level of output. When the production function over capital K and labor L is $F(K, L)$, an isoquant corresponding to 100 units of output is given by the equation $F(K, L) = 100$.

(a) If capital is on the vertical axis and labor on the horizontal, find the slope of the isoquant.

(b) Suppose that production is increasing in both capital and labor. Does the isoquant slope upward or downward?

CHAPTER

4

Constrained optimization

Microeconomics courses typically begin with either consumer theory or producer theory. If they begin with consumer theory the first problem they face is the consumer's constrained optimization problem: the consumer chooses a commodity bundle, or vector of goods, to maximize utility without spending more than her budget. If the course begins with producer theory, the first problem it poses is the firm's cost minimization problem: the firm chooses a vector of inputs to minimize the cost of producing a predetermined level of output. Both of these problems lead to a graph of the form in Figure 1.1.

So far we have only looked at unconstrained optimization problems. But many problems in economics have constraints. The consumer's budget constraint is a classic example. Without a budget constraint, and under the assumption that more is better, a consumer would choose an infinite amount of every good. This obviously does not help us describe the real world, because consumers cannot purchase unlimited quantities of every good. The budget constraint is an extra condition that the optimum must satisfy. How do we make sure that we get the solution that maximizes utility while still letting the budget constraint hold?

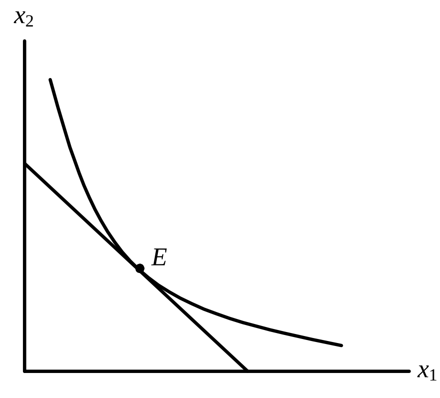


Figure 4.1: Consumer's problem

4.1 A graphical approach

Let's look more carefully at the consumer's two-dimensional maximization problem. The problem can be written as follows:

$$\max_{x_1,x_2} u(x_1, x_2)$$
$$\text{s.t. } p_1x_1 + p_2x_2 = M$$

where p_i is the price of good i and M is the total amount the consumer has to spend on consumption. The second line is the budget constraint, it says that total expenditure on the two goods is equal to M. The abbreviation "s.t." stands for "subject to," so that the problem for the consumer is to choose a commodity bundle to maximize utility subject to a budget constraint.

Figure 4.1 shows the problem graphically, and it should be familiar. What we want to do in this section is figure out what equations we need to characterize the solution. The optimal consumption point, E, is a tangency point, and it has two features: (i) it is where the indifference curve is tangent to the budget line, and (ii) it is on the budget line. Let's translate these into math.

To find the tangency condition we must figure out how to find the slopes of the two curves. We can do this easily using implicit differentiation. Begin with the budget line, because it's easier. Since x_2 is on the vertical axis, we want to find a slope of the form dx_2/dx_1. Treating x_2 as a function of x_1 and rewriting the budget constraint yields

$$p_1x_1 + p_2x_2(x_1) = M.$$

Implicit differentiation gives us

$$p_1 + p_2\frac{dx_2}{dx_1} = 0$$

because the derivative of M with respect to x_1 is zero. Rearranging yields

$$\frac{dx_2}{dx_1} = -\frac{p_1}{p_2}. \tag{4.1}$$

Of course, we could have gotten to the same place by rewriting the equation for the budget line in slope-intercept form, $x_2 = M/p_2 - (p_1/p_2)x_1$, but we have to use implicit differentiation anyway to find the slope of the indifference curve, and it is better to apply it first to the easier case.

Now let's find the slope of the indifference curve. The equation for an indifference curve is

$$u(x_1, x_2) = k$$

for some scalar k. Treat x_2 as a function of x_1 and rewrite to get

$$u(x_1, x_2(x_1)) = k.$$

Now implicitly differentiate with respect to x_1 to get

$$\frac{\partial u(x_1, x_2)}{\partial x_1} + \frac{\partial u(x_1, x_2)}{\partial x_2}\frac{dx_2}{dx_1} = 0.$$

Rearranging yields

$$\frac{dx_2}{dx_1} = -\frac{\frac{\partial u(x_1,x_2)}{\partial x_1}}{\frac{\partial u(x_1,x_2)}{\partial x_2}}. \tag{4.2}$$

The numerator is the marginal utility of good 1, and the denominator is the marginal utility of good 2, so the slope of the indifference curve is the negative of the ratio of marginal utilities, which is also known as the marginal rate of substitution.

Condition (i), that the indifference curve and budget line are tangent, requires that the slope of the budget line in (4.1) is the same as the slope of the indifference curve in (4.2), or

$$-\frac{u_1(x_1, x_2)}{u_2(x_1, x_2)} = -\frac{p_1}{p_2}. \tag{4.3}$$

The other condition, condition (ii), says that the bundle (x_1, x_2) must lie on the budget line, which is simply

$$p_1x_1 + p_2x_2 = M. \tag{4.4}$$

Equations (4.3) and (4.4) constitute two equations in two unknowns (x_1 and x_2), and so they completely characterize the solution to the consumer's optimization problem. The task now is to characterize the solution in a more general setting with more dimensions.

4.2 Lagrangians

The way that we solve constrained optimization problems is by using a trick developed by the 18-th century Italian-French mathematician Joseph-Louis Lagrange. (There is also a 1972 ZZ Top song called La Grange, so don't get confused.) Suppose that our objective is to solve an n-dimensional constrained utility maximization problem:

$$\max_{x_1,...,x_n} u(x_1, ..., x_n)$$
$$\text{s.t. } p_1x_1 + ... + p_nx_n = M.$$

Our first step is to set up the **Lagrangian**

$$\mathcal{L}(x_1, ..., x_n, \lambda) = u(x_1, ..., x_n) + \lambda(M - p_1x_1 - ... - p_nx_n).$$

This requires some interpretation. First of all, the variable λ is called the **Lagrange multiplier** (and the Greek letter is lambda). Second, let's think about the quantity $M - p_1x_1 - ... - p_nx_n$. It has to be zero according to the budget constraint, but suppose it was positive. What would it mean? M is income, and $p_1x_1 + ... + p_nx_n$ is expenditure on consumption. Income minus expenditure is simply *unspent income*. But unspent income is measured in dollars, and utility is measured in utility units (or *utils*), so we cannot simply add these together. The Lagrange multiplier converts the dollars into utils, and is therefore measured in utils/dollar. The expression $\lambda(M - p_1x_1 - ... - p_nx_n)$ can be interpreted as the *utility of unspent income.*

The Lagrangian, then, is the utility of consumption plus the utility of unspent income. The budget constraint, though, guarantees that there is no unspent income, and so the second term in the Lagrangian is necessarily zero. We still want it there, though, because it is important for finding the right set of first-order conditions.

Note that the Lagrangian has not only the x_i's as arguments, but also the Lagrange multiplier λ. The first-order conditions arise from taking $n+1$ partial derivatives of $\mathcal{L}$, one for each of the x_i's and one for λ:

$$\frac{\partial \mathcal{L}}{\partial x_1} = \frac{\partial u}{\partial x_1} - \lambda p_1 = 0 \tag{4.5a}$$

$$\vdots$$

$$\frac{\partial \mathcal{L}}{\partial x_n} = \frac{\partial u}{\partial x_n} - \lambda p_n = 0 \tag{4.5b}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = M - p_1x_1 - ... - p_nx_n = 0 \tag{4.5c}$$

Notice that the last FOC is simply the budget constraint. So, optimization using the Lagrangian guarantees that the budget constraint is satisfied. Also, optimization using the Lagrangian turns the n-dimensional constrained optimization problem into an $(n+1)$-dimensional *unconstrained* optimization problem. These two features give the Lagrangian approach its appeal.

4.3 A 2-dimensional example

The utility function is $u(x_1, x_2) = x_1^{0.5}x_2^{0.5}$, the prices are $p_1 = 10$ and $p_2 = 20$, and the budget is $M = 120$. The consumer's problem is then

$$\max_{x_1,x_2} x_1^{1/2}x_2^{1/2}$$
$$\text{s.t. } 10x_1 + 20x_2 = 120.$$

What are the utility maximizing levels of x_1 and x_2?

To answer this, we begin by setting up the Lagrangian

$$\mathcal{L}(x_1, x_2, \lambda) = x_1^{1/2}x_2^{1/2} + \lambda(120 - 10x_1 - 20x_2).$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0.5x_1^{-1/2}x_2^{1/2} - 10\lambda = 0 \tag{4.6a}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = 0.5x_1^{1/2}x_2^{-1/2} - 20\lambda = 0 \tag{4.6b}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 120 - 10x_1 - 20x_2 = 0 \tag{4.6c}$$

Of course, equation (4.6c) is simply the budget constraint.

We have three equations in three unknowns (x_1, x_2, and λ). To solve them, first rearrange (4.6a) and (4.6b) to get

$$\lambda = \frac{1}{20}\left(\frac{x_2}{x_1}\right)^{1/2}$$

and

$$\lambda = \frac{1}{40}\left(\frac{x_1}{x_2}\right)^{1/2}.$$

Set these equal to each other to get

$$\begin{aligned} \frac{1}{20}\left(\frac{x_2}{x_1}\right)^{1/2} &= \frac{1}{40}\left(\frac{x_1}{x_2}\right)^{1/2} \\ 2\left(\frac{x_2}{x_1}\right)^{1/2} &= \left(\frac{x_1}{x_2}\right)^{1/2} \\ 2x_2 &= x_1 \end{aligned}$$

where the last line comes from cross-multiplying. Substitute $x_1 = 2x_2$ into (4.6c) to get

$$\begin{aligned} 120 - 10(2x_2) - 20x_2 &= 0 \\ 40x_2 &= 120 \\ x_2 &= 3. \end{aligned}$$

Because $x_1 = 2x_2$, we have

$$x_1 = 6.$$

Finally, we know from the rearrangement of (4.6a) that

$$\begin{aligned}\lambda &= \frac{1}{20}\left(\frac{x_2}{x_1}\right)^{1/2} \\ &= \frac{1}{20}\left(\frac{3}{6}\right)^{1/2} \\ &= \frac{1}{20\sqrt{2}}.\end{aligned}$$

4.4 Interpreting the Lagrange multiplier

Remember that we said that the second term in the Lagrangian is the utility value of unspent income, which, of course, is zero because there is no unspent income. This term is $\lambda(M - p_1x_1 - p_2x_2)$. So, the Lagrange multiplier λ should be the *marginal utility of (unspent) income*, because it is the slope of the utility-of-unspent-income function. Let's see if this is true.

To do so, let's generalize the problem so that income is M instead of 120. All of the steps are the same as above, so we still have $x_1 = 2x_2$. Substituting into the budget constraint gives us

$$\begin{aligned}M - 10(2x_2) - 20x_2 &= 0 \\ x_2 &= \frac{M}{40} \\ x_1 &= \frac{M}{20} \\ \lambda &= \frac{1}{20\sqrt{2}}.\end{aligned}$$

Plugging these numbers back into the utility function gives us

$$u(x_1, x_2) = \left(\frac{M}{20}\right)^{0.5}\left(\frac{M}{40}\right)^{0.5} = \frac{M}{20\sqrt{2}}.$$

Differentiating this expression with respect to income M yields

$$\frac{du}{dM} = \frac{1}{20\sqrt{2}} = \lambda,$$

and the Lagrange multiplier really does measure the marginal utility of income.

In general, *the Lagrange multiplier measures the marginal value of relaxing the constraint*, where the units used to measure value are determined by the objective function. In our case the objective function is a utility function, so the marginal value is marginal utility. The constraint is relaxed by allowing income to be higher, so the Lagrange multiplier measures the marginal utility of income.

Now think instead about a firm's cost-minimization problem. Let x_i be the amount of input i employed by the firm, let r_i be its price, let $F(x_1, ..., x_n)$ be the production function, and let q be the desired level of output. The firm's problem would be

$$\min_{x_1,\ldots,x_n} r_1x_1 + ... + r_nx_n$$

$$\text{s.t. } F(x_1, ..., x_n) = q$$

The Lagrangian is then

$$\mathcal{L}(x_1, ..., x_n, \lambda) = r_1x_1 + ... + r_nx_n + \lambda(q - F(x_1, ..., x_n)).$$

Since the firm is minimizing cost, reducing cost from the optimum would require reducing the output requirement q. So, relaxing the constraint is lowering q. The interpretation of λ is the marginal cost of output, which was referred to simply as marginal cost way back in Chapter 2. So, using the Lagrangian to solve the firm's cost minimization problem gives you the firm's marginal output cost function for free.

4.5 A useful example - Cobb-Douglas

Economists often rely on the *Cobb-Douglas* class of functions which take the form

$$f(x_1, ..., x_n) = x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$$

where all of the a_i's are positive. The functional form arose out of a 1928 collaboration between economist Paul Douglas and mathematician Charles Cobb, and was designed to fit Douglas's production data.

To see its usefulness, consider a consumer choice problem with a Cobb-Douglas utility function $u(\bar{x}) = x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$:

$$\max_{x_1,\ldots,x_n} u(\bar{x})$$

$$\text{s.t. } p_1x_1 + ... + p_nx_n = M.$$

Form the Lagrangian

$$\mathcal{L}(x_1, ..., x_n, \lambda) = x_1^{a_1} \cdots x_n^{a_n} + \lambda(M - p_1x_1 - ... - p_nx_n).$$

The first-order conditions take the form

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{a_i}{x_i} x_1^{a_1} \cdots x_n^{a_n} - \lambda p_i = 0$$

for $i = 1, ..., n$ and

$$\frac{\partial \mathcal{L}}{\partial \lambda} = M - p_1x_1 - ... - p_nx_n = 0,$$

which is just the budget constraint. The expression for $\partial\mathcal{L}/\partial x_i$ can be rearranged to become

$$\frac{a_i}{x_i} x_1^{a_1} \cdots x_n^{a_n} - \lambda p_i = \frac{a_i}{x_i} u(\bar{x}) - \lambda p_i = 0.$$

This yields that

$$p_i = \frac{a_i u(\bar{x})}{\lambda x_i} \tag{4.7}$$

for $i = 1, ..., n$. Substitute these into the budget constraint:

$$\begin{aligned}
M - p_1x_1 - ... - p_nx_n &= 0 \\
M - \frac{a_1 u(\bar{x})}{\lambda x_1} x_1 - ... - \frac{a_n u(\bar{x})}{\lambda x_n} x_n &= 0 \\
M - \frac{u(\bar{x})}{\lambda}(a_1 + ... + a_n) &= 0.
\end{aligned}$$

Now solve this for the Lagrange multiplier λ:

$$\lambda = u(\bar{x}) \frac{a_1 + ... + a_n}{M}.$$

Finally, plug this back into (4.7) to get

$$\begin{aligned}
p_i &= \frac{a_i u(\bar{x})}{x_i} \cdot \frac{1}{\lambda} \\
&= \frac{a_i u(\bar{x})}{x_i} \cdot \frac{M}{u(\bar{x})(a_1 + ... + a_n)} \\
&= \frac{a_i}{a_1 + ... + a_n} \cdot \frac{M}{x_i}.
\end{aligned}$$

Finally, solve this for x_i to get the demand function for good i:

$$x_i = \frac{a_i}{a_1 + ... + a_n} \cdot \frac{M}{p_i}. \tag{4.8}$$

That was a lot of steps, but rearranging (4.8) yields an intuitive and easily memorizable expression. In fact, most graduate students in economics have memorized it by the end of their first semester because it turns out to be so handy. Rearrange (4.8) to

$$\frac{p_i x_i}{M} = \frac{a_i}{a_1 + ... + a_n}.$$

The numerator of the left-hand side is the amount spent on good i. The denominator of the left-hand side is the total amount spent. The left-hand side, then, is the *share* of income spent on good i. The equation says that the share of spending is determined entirely by the exponents of the Cobb-Douglas utility function. In especially convenient cases the exponents sum to one, in which case the spending share for good i is just equal to the exponent on good i.

The demand function in (4.8) lends itself to some particularly easy comparative statics analysis. The obvious comparative statics derivative for a demand function is with respect to its own price:

$$\frac{dx_i}{dp_i} = -\frac{a_i}{a_1 + ... + a_n} \cdot \frac{M}{p_i^2} \leq 0$$

and so demand is downward-sloping, as it should be. Another comparative statics derivative is with respect to income:

$$\frac{dx_i}{dM} = \frac{a_i}{a_1 + ... + a_n} \cdot \frac{1}{p_i} \geq 0.$$

All goods are normal goods when the utility function takes the Cobb-Douglas form. Finally, one often looks for the effects of changes in the prices of other goods. We can do this by taking the comparative statics derivative of x_i with respect to price p_j, where $j \neq i$.

$$\frac{dx_i}{dp_j} = 0.$$

This result holds because the other prices appear nowhere in the demand function (4.8), which is another feature that makes Cobb-Douglas special.

We can also use Cobb-Douglas functions in a production setting. Consider the firm's cost-minimization problem when the production function is Cobb-Douglas, so that $F(\bar{x}) = x_1^{a_1} \cdots x_n^{a_n}$. This time, though, we are going to assume that $a_1 + ... + a_n = 1$. The problem is

$$\min_{x_1,...,x_n} p_1x_1 + ... + p_nx_n$$

$$\text{s.t. } x_1^{a_1} \cdots x_n^{a_n} = q.$$

Set up the Lagrangian

$$\mathcal{L}(x_1, ..., x_n, \lambda) = p_1x_1 + ... + p_nx_n + \lambda(q - x_1^{a_1} \cdots x_n^{a_n}).$$

The first-order conditions take the form

$$\frac{\partial \mathcal{L}}{\partial x_i} = p_i - a_i\lambda\frac{F(\bar{x})}{x_i} = 0$$

for $i = 1, ..., n$ and

$$\frac{\partial \mathcal{L}}{\partial \lambda} = q - x_1^{a_1} \cdots x_n^{a_n} = 0,$$

which is just the production constraint. Rearranging the expression for $\partial\mathcal{L}/\partial x_i$ yields

$$x_i = a_i\lambda\frac{q}{p_i}, \tag{4.9}$$

because the production constraint tells us that $F(\bar{x}) = q$. Plugging this into the production constraint give us

$$\begin{aligned}\left(a_1\lambda\frac{q}{p_1}\right)^{a_1} \cdots \left(a_n\lambda\frac{q}{p_n}\right)^{a_n} &= q \\ \left(\frac{a_1}{p_1}\right)^{a_1} \cdots \left(\frac{a_n}{p_n}\right)^{a_n} \lambda^{a_1+...+a_n} q^{a_1+...+a_n} &= q.\end{aligned}$$

But $a_1 + ... + a_n = 1$, so the above expression reduces further to

$$\begin{aligned}\left(\frac{a_1}{p_1}\right)^{a_1} \cdots \left(\frac{a_n}{p_n}\right)^{a_n} \lambda q &= q \\ \left(\frac{a_1}{p_1}\right)^{a_1} \cdots \left(\frac{a_n}{p_n}\right)^{a_n} \lambda &= 1\end{aligned}$$

$$\lambda = \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n}. \tag{4.10}$$

We can substitute this back into (4.9) to get

$$\begin{aligned} x_i &= a_i \lambda \frac{q}{p_i} \\ &= \frac{a_i}{p_i} \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q. \end{aligned}$$

This is the input demand function, and it depends on the amount of output being produced (q), the input prices ($p_1, ..., p_n$), and the exponents of the Cobb-Douglas production function.

This doesn't look particularly useful or intuitive. It can be, though. Plug it back into the original objective function $p_1 x_1 + ... + p_n x_n$ to get the cost function

$$\begin{aligned} C(q) &= p_1 x_1 + ... + p_n x_n \\ &= p_1 \frac{a_1}{p_1} \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q + ... + p_n \frac{a_n}{p_n} \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q \\ &= a_1 \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q + ... + a_n \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q \\ &= (a_1 + ... + a_n) \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q \\ &= \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n} q, \end{aligned}$$

where the last equality holds because $a_1 + ... + a_n = 1$.

This one is pretty easy to remember. And it has a cool comparative statics result:

$$\frac{dC(q)}{dq} = \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n}. \tag{4.11}$$

Why is this cool? There are three reasons. First, $dC(q)/dq$ is marginal cost, and q appears nowhere on the right-hand side. This means that the Cobb-Douglas production function gives us constant marginal cost. Second, compare the marginal cost function to the original production function:

$$\begin{aligned} F(\bar{x}) &= x_1^{a_1} \cdots x_n^{a_n} \\ MC(q) &= \left(\frac{p_1}{a_1}\right)^{a_1} \cdots \left(\frac{p_n}{a_n}\right)^{a_n}. \end{aligned}$$

You can get the marginal cost function by replacing the x_i's in the production function with the corresponding p_i/a_i. And third, remember how, at the end of the the last section on interpreting the Lagrange multiplier, we said that in a cost-minimization problem the Lagrange multiplier is just marginal cost? Compare equations (4.10) and (4.11). They are the same. I told you so.

4.6 Problems

1. Use the Lagrange multiplier method to solve the following problem:

$$\max_{x,y} 12x^2y^4$$

$$\text{s.t. } 2x + 4y = 120$$

   [Hint: You should be able to check your answer against the general version of the problem in Section 4.5.]

2. Solve the following problem:

$$\max_{a,b} 3\ln a + 2\ln b$$
$$\text{s.t. } 12a + 14b = 400$$

3. Solve the following problem:

$$\min_{x,y} 16x + y$$
$$x^{1/4}y^{3/4} = 1$$

4. Solve the following problem:

$$\max_{x,y} 3xy + 4x$$

$$\text{s.t. } 4x + 12y = 80$$

5. Solve the following problem:

$$\min_{x,y} 5x + 2y$$

$$\text{s.t. } 3x + 2xy = 80$$

6. This is a lame but instructive problem. A farmer has 10 acres of land and uses it to grow corn. Profit from growing an acre of corn is given by $\pi(x) = 400x + 2x^2$, where x is the number of acres of corn planted. So, the farmer's problem is

$$\max_x 400x + 2x^2$$
$$\text{s.t. } x = 10$$

   (a) Find the first derivative of the profit function. Does its sign make sense?

   (b) Find the second derivative of the profit function. Does its sign make sense?

   (c) Set up the Lagrangian and use it to find the optimal value of x. (Hint: It had better be 10.)

   (d) Interpret the Lagrange multiplier.

   (e) Find the marginal value of an acre of land without using the Lagrange multiplier.

   (f) The second derivative of the profit function is positive. Does that mean that profit is minimized when $x = 10$?

7. Another lame but instructive problem: A firm has the capacity to use 4 workers at a time. Each worker can produce 4 units per hour, each worker must be paid $10 per hour, and the firm's revenue function is $R(L) = 30\sqrt{L}$, where L is the number of workers employed (fractional workers are okay). The firm's profit function is $\pi(L) = 30\sqrt{4L} - 10L$. It must solve the problem

$$\max_L 30\sqrt{4L} - 10L$$
$$\text{s.t. } L = 4$$

   (a) Find the first derivative of the profit function. Does its sign make sense?

   (b) Find the second derivative of the profit function. Does its sign make sense?

   (c) Set up the Lagrangian and use it to find the optimal value of L. [Hint: It had better be 4.]

(d) Interpret the Lagrange multiplier.

(e) Find the marginal profit from a worker without using the Lagrange multiplier.

(f) The second derivative of the profit function is negative. Does that mean profit is maximized when $L = 4$?

8. Here is the obligatory comparative statics problem. A consumer chooses x and y to

$$\max_{x,y} x^{\alpha}y^{1-\alpha}$$
$$\text{s.t. } p_x x + p_y y = M$$

where $p_x > 0$ is the price of good x, $p_y > 0$ is the price of good y, $M > 0$ is income, and $0 < \alpha < 1$.

(a) Show that $x^* = \alpha M/p_x$ and $y^* = (1-\alpha)M/p_y$.

(b) Find $\partial x^*/\partial M$ and $\partial y^*/\partial M$. Can you sign them?

(c) Find $\partial x^*/\partial p_x$ and $\partial y^*/\partial p_x$. Can you sign them?

9. This is the same as problem 2.11 but done using Lagrange multipliers. A firm (Bilco) can use its manufacturing facility to make either widgets or gookeys. Both require labor only. The production function for widgets is

$$W = 20w^{1/2},$$

where w denotes labor devoted to widget production, and the production function for gookeys is

$$G = 30g,$$

where g denotes labor devoted to gookey production. The wage rate is \$11 per unit of time, and the prices of widgets and gokeys are \$9 and \$3 per unit, repsectively. The manufacturing facility can accomodate 60 workers and no more.

(a) Use a Lagrangian to determine how much of each product Bilco should produce per unit of time.

(b) Interpret the Lagrange multiplier.

10. A farmer has a fixed amount F of fencing material that she can use to enclose a property. She does not yet have the property, but it will be a rectangle with length L and width W. Furthermore, state law dictates that every property must have a side smaller than S in length, and in this case $S < F/4$. [This last condition makes the constraint binding, and other than that you need not worry about it.] By convention, W is always the short side, so the state law dictates that $W \leq S$. The farmer wants to use the fencing to enclose the largest possible area, and she also wants to obey the law.

    (a) Write down the farmer's constrained maximization problem. [Hint: There should be two constraints.]

    (b) Write down the Lagrangian with two mulitpliers, one for each constraint, and solve the farmer's problem. [Hint: The solution will be a function of F and S.] Please use μ as the second multiplier.

    (c) Which has a greater impact on the area the farmer can enclose, a marginal increase in S or a marginal increase in F? Justify your answer.

CHAPTER

5

Inequality constraints

The previous chapter treated all constraints as equality constraints. Sometimes this is the right thing to do. For example, the firm's cost-minimization problem is to find the least-cost combination of inputs to produce a fixed amount of output, q. The constraint, then, is that output must be q, or, letting $F(x_1, ..., x_n)$ be the production function when the inputs are $x_1, ..., x_n$, the constraint is

$$F(x_1, ..., x_n) = q.$$

Other times equality constraints are not the right thing to do. The consumer choice problem, for example, has the consumer choosing a commodity bundle to maximize utility, subject to the constraint that she does not spend more than her income. If the prices are $p_1, ..., p_n$, the goods are $x_1, ..., x_n$, and income is M, then the budget constraint is

$$p_1 x_1 + ... + p_n x_n \leq M.$$

It may be the case that the consumer spends her entire income, in which case the constraint would hold with equality. If she gets utility from saving, though, she may not want to spend her entire income, in which case the budget constraint would not hold with equality.

Firms have capacity constraints. When they build manufacturing facilities, the size of the facility places a constraint on the maximum output the firm can produce. A capacity-constrained firm's problem, then, is to maximize profit subject to the constraint that output not exceed capacity, or $q \leq \bar{q}$. In the real world firms often have excess capacity, which means that the capacity constraint does not hold with equality.

Finally, economics often has implicit nonnegativity constraints. Firms cannot produce negative amounts by transforming outputs back into inputs. After all, it is difficult to turn a cake back into flour, baking powder, butter, salt, sugar, and unbroken eggs. Often we want to assume that we cannot consume negative amounts. As economists we must deal with these nonnegativity constraints.

The goal for this chapter is to figure out how to deal with inequality constraints. The best way to do this is through a series of exceptionally lame examples. What makes the examples lame is that the solutions are so transparent that it is hardly worth going through the math. The beauty of lame examples, though, is that this transparency allows you to see exactly what is going on.

5.1 Lame example - capacity constraints

Let's begin with a simple unconstrained profit maximization problem. The firm chooses an amount to produce x, the market price is fixed at 80, and the cost function is $4x^2$. The problem is

$$\max_x 80x - 4x^2.$$

The first-order condition is

$$80 - 8x = 0,$$

so the optimum is

$$x = 10.$$

The second-order condition is

$$-8 < 0$$

which obviously holds, so the optimum is actually a maximum. The problem is illustrated in Figure 5.1.

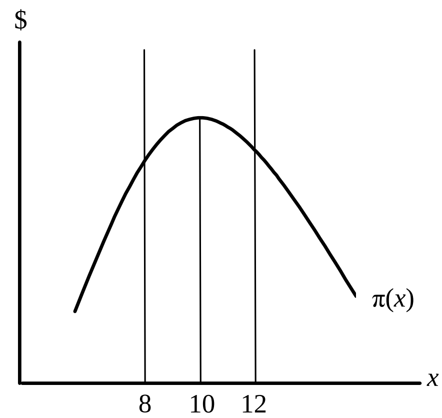


Figure 5.1: A lame example using capacity constraints

5.1.1 A binding constraint

Now let's put in a capacity constraint: $x \leq 8$. This will obviously restrict the firm's output because it would like to produce 10 units but can only produce 8. (See why it's a lame example?) The constraint will hold with equality, in which case we say that the constraint is **binding**. Let's look at the math.

$$\max_x 80x - 4x^2$$
$$\text{s.t. } x \leq 8$$

Form the Lagrangian

$$\mathcal{L}(x, \lambda) = 80x - 4x^2 + \lambda(8 - x).$$

The first-order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 80 - 8x - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 8 - x = 0. \end{aligned}$$

The second equation tells us that $x = 8$, and the first equation tells us that $\lambda = 80 - 64 = 16$.

So far we've done nothing new. The important step here is to think about λ. Remember that the interpretation of the Lagrange multiplier is that it is the marginal value of relaxing the constraint. In this case value is profit, so it is the marginal profit from relaxing the constraint. We can compute this directly:

$$\begin{aligned} \pi(x) &= 80x - 4x^2 \\ \pi'(x) &= 80 - 8x \\ \pi'(8) &= 80 - 64 = 16 \end{aligned}$$

Plugging the constrained value ($x = 8$) into the marginal profit function $\pi'(x)$ tells us that when output is 8, an increase in output leads to an increase in profit by 16. And this is exactly the Lagrange multiplier.

5.1.2 A nonbinding constraint

Now let's change the capacity constraint to $x \leq 12$ and solve the problem intuitively. First, we know that profit reaches its unconstrained maximum when $x = 10$. The constraint does not rule this level of production out, so the constrained optimum is also $x = 10$. Because of this the capacity constraint is **nonbinding**, that is, it does not hold with equality. Nonbinding constraints are sometimes called **slack**.

Let's think about the Lagrange multiplier. We know that it is the marginal value of relaxing the constraint. How would profit change if we relaxed the constraint from $x \leq 12$ to, say, $x \leq 13$? The unconstrained maximum is still feasible, so the firm would still produce 10 units and still generate exactly the same amount of profit. So, the marginal value of relaxing the constraint must be zero, and we have $\lambda = 0$.

Now that we know the answers, let's go back and look at the problem.

$$\max_x 80x - 4x^2$$
$$\text{s.t. } x \leq 12$$

This problem generates the Lagrangian

$$\mathcal{L}(x, \lambda) = 80x - 4x^2 + \lambda(12 - x).$$

Since we already know the answers, let's plug them in. In particular, we know that $\lambda = 0$, so the Lagrangian becomes

$$\mathcal{L}(x, \lambda) = 80x - 4x^2$$

which is just the original unconstrained profit function.

We arrived at our answers ($x = 10$, $\lambda = 0$) intuitively. How can we get them mechanically? After all, the purpose of the math is to make sure we get the answers right without relying solely on our intuition.

One thing for sure is that we will need a new approach. To see why, suppose we analyze our 12-unit constraint problem in the usual way. Differentiating the Lagrangian yields

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x} &= 80 - 8x - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 12 - x = 0\end{aligned}$$

The second equation obviously implies that $x = 12$, in which case the first equation tells us that $\lambda = 80 - 8x = 80 - 96 = -16$. If we solve the problem using our old approach we find that (1) the constraint is binding, which is wrong, and (2) the Lagrange multiplier is negative, which means that relaxing the constraint makes profit even lower. You can see this in Figure 5.1. When output is 12 the profit function is downward-sloping. Since the Lagrange multiplier is marginal profit, we get a negative Lagrange multiplier when we are past the profit-maximizing level of output.

5.2 A new approach

The key to the new approach is thinking about how the Lagrangian works. Suppose that the problem is

$$\max_{x_1,\ldots,x_n} f(x_1, ..., x_n)$$

$$\text{s.t. } g(x_1, ..., x_n) \le M$$

The Lagrangian is

$$\mathcal{L}(x_1, ..., x_n, \lambda) = f(x_1, ..., x_n) + \lambda[M - g(x_1, ..., x_n)]. \tag{5.1}$$

When the constraint is binding, the term $M - g(x_1, ..., x_n)$ is zero, in which case $\mathcal{L}(x_1, ..., x_n, \lambda) = f(x_1, ..., x_n)$. When the constraint is nonbinding the Lagrange multiplier is zero, in which case $\mathcal{L}(x_1, ..., x_n, \lambda) = f(x_1, ..., x_n)$ once again. So we need a condition that says

$$\lambda[M - g(x_1, ..., x_n)] = 0.$$

Note that
$$\frac{\partial[\lambda[M-g(x_1,...,x_n)]]}{\partial\lambda}=M-g(x_1,...,x_n)$$
and in the old approach we set this equal to zero. We can no longer do this when the constraint is nonbinding, but notice that
$$\lambda\frac{\partial[\lambda[M-g(x_1,...,x_n)]]}{\partial\lambda}=\lambda(M-g(x_1,...,x_n)).$$
This is exactly what we need to be equal to zero.

We also need to restrict the Lagrange multiplier to be nonnegative. Remember from the lame example when the capacity constraint was binding at $x=12$ we got a negative Lagrange multiplier, and that was the wrong answer. In fact, looking at expression (5.1) we can make $\mathcal{L}$ really large by making both λ and $(M-g(x_1,...,x_n))$ really negative. But when $(M-g(x_1,...,x_n))<0$ we have violated the constraint, so that is not allowed.

The condition that
$$\lambda\frac{\partial\mathcal{L}}{\partial\lambda}=\lambda[(M-g(x_1,...,x_n)]=0$$
is known as a **complementary slackness** condition. It says that one of two constraints must bind. One constraint is $\lambda\geq 0$, and it binds if $\lambda=0$, in which case the complementary slackness condition holds. The other constraint is $g(x_1,...,x_n)\leq M$, and it binds if $g(x_1,...,x_n)=M$, in which case the complementary slackness condition also holds. If one of the constraints is slack, the other one has to bind. The beauty of the complementary slackness condition is that it forces one of two constraints to bind using a single equation.

The first-order conditions for the inequality-constrained problem are
$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial x_1}&=0\\&\vdots\\\frac{\partial\mathcal{L}}{\partial x_n}&=0\\\lambda\frac{\partial\mathcal{L}}{\partial\lambda}&=0\\\lambda&\geq 0\end{aligned}$$

The first set of conditions ($\partial\mathcal{L}/\partial x_i = 0$) are the same as in the standard case. The last two conditions are the ones that are different. The second-last condition ($\lambda\partial\mathcal{L}/\partial\lambda = 0$) guarantees that either the constraint binds, in which case $\partial\mathcal{L}/\partial\lambda = 0$, or the constraint does not bind, in which case $\lambda = 0$. The last condition says that the Lagrange multiplier cannot be negative, which means that relaxing the constraint cannot reduce the value of the objective function.

We have a set of first-order conditions, but this does not tell us how to solve them. To do this, let's go back to our lamest example:

$$\max_x 80x - 4x^2$$
$$\text{s.t. } x \leq 12$$

which generates the Lagrangian

$$\mathcal{L}(x, \lambda) = 80x - 4x^2 + \lambda(12 - x).$$

The first-order conditions are

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial x} &= 80 - 8x - \lambda = 0 \\ \lambda\frac{\partial\mathcal{L}}{\partial\lambda} &= \lambda(12 - x) = 0 \\ \lambda &\geq 0 \end{aligned}$$

Now what?

The methodology for solving this system is tedious, but it works. The second equation ($\lambda(12 - x) = 0$) is true if either (1) $\lambda = 0$, (2) $12 - x = 0$, or (3) both. So what we have to do is find the solution when $\lambda = 0$ and find the solution when $12 - x = 0$. Let's see what happens.

Case 1: $\lambda = 0$. If $\lambda = 0$ then the second and third conditions obviously hold. Plugging $\lambda = 0$ into the first equation yields $80 - 8x - 0 = 0$, or $x = 10$.

Case 2: $12 - x = 0$. Then $x = 12$, and plugging this into the first equation yields $\lambda = 80 - 96 = -16$, which violates the last condition. So case 2 cannot be the answer.

We are left with only one solution, and it is the correct one: $x = 10$, $\lambda = 0$.

The general methodology for multiple constraints is as follows: Construct a Lagrange multiplier for each constraint. Each Lagrange multiplier can be either zero, in which case that constraint is nonbinding, or it can be positive, in which case its constraint is binding. Then try all possible combinations of zero/positive multipliers. Most of them will lead to violations. If only one does not lead to a violation, that is the answer. If several combinations do not lead to violations, then you must choose which one is the best. You can do this by plugging the values you find into the objective function. If you want to maximize the objective function, you choose the case that generates the highest value of the objective function.

5.3 Multiple inequality constraints

Let's look at another lame example:

$$\max_{x,y} x^{\frac{1}{3}} y^{\frac{2}{3}}$$

$$\begin{aligned} \text{s.t. } x + y &\leq 60 \\ x + y &\leq 120 \end{aligned}$$

Why is this a lame example? Because we know that the second constraint must be nonbinding. After all if a number is smaller than 60, it must also be strictly smaller than 120. The solution to this problem will be the same as the solution to the problem

$$\max_{x,y} x^{\frac{1}{3}} y^{\frac{2}{3}}$$

$$\text{s.t. } x + y \leq 60$$

This looks like a utility maximization problem, as can be seen in Figure 5.2. The consumer's utility function is $u(x, y) = x^{1/3}y^{2/3}$, the prices of the two goods are $p_x = p_y = 1$, and the consumer has 60 to spend. The utility function is Cobb-Douglas, and from what we learned in Section 4.5 we know that $x = 20$ and $y = 40$.

We want to solve it mechanically, though, to learn the steps. Assign a separate Lagrange multiplier to each constraint to get the Lagrangian

$$\mathcal{L}(x, y, \lambda_1, \lambda_2) = x^{\frac{1}{3}} y^{\frac{2}{3}} + \lambda_1(60 - x - y) + \lambda_2(120 - x - y).$$

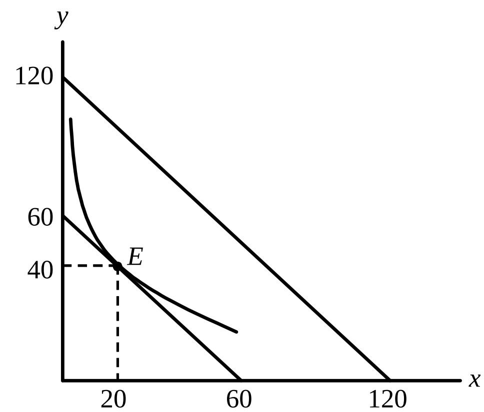


Figure 5.2: A lame example with two budget constraints

The first-order conditions are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= \frac{1}{3}\frac{y^{2/3}}{x^{2/3}} - \lambda_1 - \lambda_2 = 0 \\
\frac{\partial \mathcal{L}}{\partial y} &= \frac{2}{3}\frac{x^{1/3}}{y^{1/3}} - \lambda_1 - \lambda_2 = 0 \\
\lambda_1 \frac{\partial \mathcal{L}}{\partial \lambda_1} &= \lambda_1(60 - x - y) = 0 \\
\lambda_2 \frac{\partial \mathcal{L}}{\partial \lambda_2} &= \lambda_2(120 - x - y) = 0 \\
\lambda_1 &\geq 0 \\
\lambda_2 &\geq 0
\end{aligned}$$

This time we have four possible cases: (1) $\lambda_1 = \lambda_2 = 0$, so that neither constraint binds. (2) $\lambda_1 > 0$, $\lambda_2 = 0$, so that only the first constraint binds. (3) $\lambda_1 = 0$, $\lambda_2 > 0$, so that only the second constraint binds. (4) $\lambda_1 > 0$, $\lambda_2 > 0$, so that both constraints bind. As before, we need to look at all four cases.

Case 1: $\lambda_1 = \lambda_2 = 0$. In this case the first equation in the first-order conditions reduces to $\frac{1}{3}(y/x)^{2/3} = 0$, which implies that $y = 0$. The second equation reduces to $\frac{2}{3}(x/y)^{1/3} = 0$, but this cannot be true if $y = 0$ because we are not allowed to divide by zero. So Case 1 cannot be the answer. There is an easier way to see this, though. If neither constraint binds, the problem becomes

$$\max_{x,y} x^{\frac{1}{3}} y^{\frac{2}{3}}$$

The objective function is increasing in both arguments, and since there are no constraints we want both x and y to be as large as possible. So $x \to \infty$ and $y \to \infty$. But this obviously violates the constraints.

Case 2: $\lambda_1 > 0$, $\lambda_2 = 0$. The first-order conditions reduce to

$$\begin{aligned} \frac{1}{3}\frac{y^{2/3}}{x^{2/3}} - \lambda_1 &= 0 \\ \frac{2}{3}\frac{x^{1/3}}{y^{1/3}} - \lambda_1 &= 0 \\ 60 - x - y &= 0 \end{aligned}$$

and the solution to this system is $x = 20$, $y = 40$, and $\lambda_1 = \frac{1}{3}2^{2/3} > 0$. The remaining constraint, $x + y \leq 120$, is satisfied because $x + y = 60 < 120$. Case 2 works, and it corresponds to the case shown in Figure 5.2.

Case 3: $\lambda_1 = 0$, $\lambda_2 > 0$. Now the first-order conditions reduce to

$$\begin{aligned} \frac{1}{3}\frac{y^{2/3}}{x^{2/3}} - \lambda_2 &= 0 \\ \frac{2}{3}\frac{x^{1/3}}{y^{1/3}} - \lambda_2 &= 0 \\ 120 - x - y &= 0 \end{aligned}$$

The solution to this system is $x = 40$, $y = 80$, and $\lambda_2 = \frac{1}{3}2^{2/3} > 0$. The remaining constraint, $x + y \leq 60$, is violated because $x + y = 120$, so Case 3 does not work. There is an easier way to see this, though, just by looking at the constraints and exploiting the lameness of the example. If the second constraint binds, $x + y = 120$. But then the first constraint, $x + y \leq 60$ cannot possibly hold.

Case 4: $\lambda_1 > 0$, $\lambda_2 > 0$. The first of these conditions implies that the first constraint binds, so that $x + y = 60$. The second condition implies that $x + y = 120$, so that we are on the outer budget constraint in Figure 5.2. But then the inner budget constraint is violated, so Case 4 does not work.

Only Case 2 works, so we know the solution: $x^* = 20$, $y^* = 40$, $\lambda_1^* = \frac{1}{3}2^{2/3}$, and $\lambda_2^* = 0$.

5.4 A linear programming example

A linear programming problem is one in which the objective function and all of the constraints are linear, such as in the following example:

$$\max_{x,y} 2x + y$$

$$\begin{aligned} \text{s.t. } 3x + 4y &\leq 60 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

This problem has three constraints, so we must use the multiple constraint methodology from the preceding section. It is useful in what follows to refer to the first constraint, $3x + 4y \leq 60$, as a budget constraint. The other two constraints are nonnegativity constraints.

The Lagrangian is

$$\mathcal{L}(x, y, \lambda_1, \lambda_2, \lambda_3) = 2x + y + \lambda_1(60 - 3x - 4y) + \lambda_2(x - 0) + \lambda_3(y - 0).$$

Notice that we wrote all of the constraint terms, that is $(60 - 3x - 4y)$ and $(x - 0)$ and $(y - 0)$ so that they are nonnegative. We have been doing this throughout this book.

The first-order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 2 - 3\lambda_1 + \lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 1 - 4\lambda_1 + \lambda_3 = 0 \\ \lambda_1 \frac{\partial \mathcal{L}}{\partial \lambda_1} &= \lambda_1(60 - 3x - 4y) = 0 \\ \lambda_2 \frac{\partial \mathcal{L}}{\partial \lambda_2} &= \lambda_2 x = 0 \\ \lambda_3 \frac{\partial \mathcal{L}}{\partial \lambda_3} &= \lambda_3 y = 0 \\ \lambda_1 &\geq 0,\ \lambda_2 \geq 0,\ \lambda_3 \geq 0 \end{aligned}$$

Since there are three constraints, there are $2^3 = 8$ possible cases. We are going to narrow some of them down intuitively before going on. The first constraint is like a budget constraint, and the objective function is increasing in both of its arguments. The other two constraints are nonnegativity constraints, saying that the consumer cannot consume negative amounts of the goods. Since there is only one budget-type constraint, it has to bind, which means that $\lambda_1 > 0$. The only question is whether one of the other two constraints binds.

A binding budget constraint means that we cannot have both $x = 0$ and $y = 0$, because if we did then we would have $3x + 4y = 0 < 60$, and the budget constraint would not bind. We are now left with three possibilities: (1) $\lambda_1 > 0$ so the budget constraint binds, $\lambda_2 > 0$, and $\lambda_3 = 0$, so that $x = 0$ but $y > 0$. (2) $\lambda_1 > 0$, $\lambda_2 = 0$, and $\lambda_3 > 0$, so that $x > 0$ and $y = 0$. (3) $\lambda_1 > 0$, $\lambda_2 = 0$, and $\lambda_3 = 0$, so that both x and y are positive. We will consider these one at a time.

Case 1: $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 = 0$. Since $\lambda_2 > 0$ the constraint $x \geq 0$ must bind, so $x = 0$. For the budget constraint to hold we must have $y = 15$. This yields a value for the objective function of

$$2x + y = 2 \cdot 0 + 15 = 15.$$

Case 2: $\lambda_1 > 0$, $\lambda_2 = 0$, $\lambda_3 > 0$. This time we have $y = 0$, and the budget constraint implies that $x = 20$. The objective function then takes the value

$$2x + y = 2 \cdot 20 + 0 = 40.$$

Case 3: $\lambda_1 > 0$, $\lambda_2 = 0$, $\lambda_3 = 0$. In this case both x and y are positive. The first two equations in the first-order conditions become

$$\begin{aligned} 2 - 3\lambda_1 &= 0 \\ 1 - 4\lambda_1 &= 0 \end{aligned}$$

The first of these reduces to $\lambda_1 = 2/3$, and the second reduces to $\lambda_1 = 1/4$. These cannot both hold, so Case 3 does not work.

We got solutions in Case 1 and Case 2, but not in Case 3. So which is the answer? The one that generates a larger value for the objective function. In Case 1 the maximized objective function took a value of 15 and in Case 2 it took a value of 40, which is obviously higher. So Case 2 is the solution.

To see what we just did, look at Figure 5.3. The three constraints identify a triangular feasible set. Case 1 is the corner solution where the budget line meets the y-axis, and Case 2 is the corner solution where the budget line

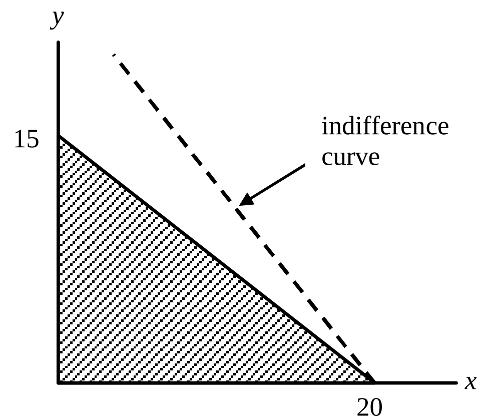


Figure 5.3: Linear programming problem

meets the x-axis. Case 3 is an "interior" solution that is on the budget line but not on either axis. The objective was to find the point in the feasible set that maximized the function $f(x, y) = 2x+y$. We did this by comparing the values we got at the two corner solutions, and we chose the corner solution that gave us the larger value.

The methodology for solving linear programming problems involves finding all of the corners and choosing the corner that yields the largest value of the objective function. A typical problem has more than two dimensions, so it involves finding $x_1, ..., x_n$, and it has more than one budget constraint. This generates lots and lots of corners, and the real issue in the study of linear programming is finding an algorithm that efficiently checks the corners.

5.5 Kuhn-Tucker conditions

Economists sometimes structure the first-order conditions for inequality-constrained optimization problems differently than the way we have done it so far. The alternative formulation was developed by Harold Kuhn and A.W. Tucker, and it is known as the Kuhn-Tucker formulation. The first-order conditions we will derive are known as the Kuhn-Tucker conditions.

Begin with the very general maximization problem, letting $\mathbf{x}$ be the vector

$\mathbf{x} = (x_1, ..., x_n)$:

$$\max_{\mathbf{x}} f(\mathbf{x})$$

$$\begin{aligned} \text{s.t. } g^1(\mathbf{x}) &\leq b_1, ..., g^k(\mathbf{x}) \leq b_k \\ x_1 &\geq 0, ..., x_n \geq 0. \end{aligned}$$

There are k "budget-type" constraints and n non-negativity constraints.

To solve this, form the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda_1, ..., \lambda_k, v_1, ..., v_k) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i [b_i - g^i(\mathbf{x})] + \sum_{j=1}^{n} v_j x_j.$$

We get the following first-order conditions:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= \frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g^1}{\partial x_1} - ... - \lambda_k \frac{\partial g^k}{\partial x_1} + v_1 = 0 \\ &\vdots \\ \frac{\partial \mathcal{L}}{\partial x_n} &= \frac{\partial f}{\partial x_n} - \lambda_1 \frac{\partial g^1}{\partial x_n} - ... - \lambda_k \frac{\partial g^k}{\partial x_n} + v_n = 0 \\ \lambda_1 \frac{\partial \mathcal{L}}{\partial \lambda_1} &= \lambda_1 [b_1 - g^1(\mathbf{x})] = 0 \\ &\vdots \\ \lambda_k \frac{\partial \mathcal{L}}{\partial \lambda_k} &= \lambda_k [b_k - g^k(\mathbf{x})] = 0 \\ v_1 \frac{\partial \mathcal{L}}{\partial v_1} &= v_1 x_1 = 0 \\ &\vdots \\ v_n \frac{\partial \mathcal{L}}{\partial v_n} &= v_n x_n = 0 \end{aligned}$$

$$\lambda_1, ..., \lambda_k, v_1, ..., v_n \geq 0$$

There are $2n + k$ conditions plus the $n + k$ nonnegativity constraints for the multipliers. It is useful to have some shorthand to shrink this system down:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= 0 \text{ for } i = 1, ..., n \\ \lambda_j \frac{\partial \mathcal{L}}{\partial \lambda_j} &= 0 \text{ for } j = 1, ..., k \\ v_i x_i &= 0 \text{ for } i = 1, ..., n \\ \lambda_j &\geq 0 \text{ for } j = 1, ..., k \\ v_i &\geq 0 \text{ for } i = 1, ..., n \end{aligned} \tag{5.2}$$

Suppose instead we had constructed a different Lagrangian:

$$K(\mathbf{x}, \lambda_1, ..., \lambda_k) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i [b_i - g^i(\mathbf{x})].$$

This Lagrangian, known as the Kuhn-Tucker Lagrangian, only has k multipliers for the k budget-type constraints, and no multipliers for the nonnegativity constraints. The two Lagrangians are related, with

$$\mathcal{L}(\mathbf{x}, \lambda_1, ..., \lambda_k, v_1, ..., v_k) = K(\mathbf{x}, \lambda_1, ..., \lambda_k) + \sum_{j=1}^{n} v_j x_j.$$

We can rewrite the system of first-order conditions (5.2) as

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_i} &= \frac{\partial K}{\partial x_i} + v_i = 0 \text{ for } i = 1, ..., n \\ \lambda_j \frac{\partial \mathcal{L}}{\partial \lambda_j} &= \lambda_j \frac{\partial K}{\partial \lambda_j} = 0 \text{ for } j = 1, ..., k \\ v_i x_i &= 0 \text{ for } i = 1, ..., n \\ \lambda_j &\geq 0 \text{ for } j = 1, ..., k \\ v_i &\geq 0 \text{ for } i = 1, ..., n \end{aligned}$$

Pay close attention to the first and third equations. If $v_i = 0$ then the first equation yields

$$v_i = 0 \Longrightarrow \frac{\partial K}{\partial x_i} = 0 \Longrightarrow x_i \frac{\partial K}{\partial x_i} = 0.$$

On the other hand, if $v_i > 0$ then the i-th inequality constraint, $x_i \geq 0$, is binding which means that

$$v_i > 0 \Longrightarrow x_i = 0 \Longrightarrow x_i \frac{\partial K}{\partial x_i} = 0.$$

Either way we have

$$x_i \frac{\partial K}{\partial x_i} = 0 \text{ for } i = 1, ..., n.$$

The **Kuhn-Tucker conditions** use this information. The new set of first-order conditions is

$$\begin{aligned} x_i \frac{\partial K}{\partial x_i} &= 0 \text{ for } i = 1, ..., n \\ \lambda_j \frac{\partial K}{\partial \lambda_j} &= 0 \text{ for } j = 1, ..., k \\ \lambda_j &\geq 0 \text{ for } j = 1, ..., k \\ x_i &\geq 0 \text{ for } i = 1, ..., n \end{aligned} \tag{5.3}$$

This is a system of $n + k$ equations in $n + k$ unknowns plus $n + k$ non-negativity constraints. Thus, it simplifies the original set of conditions by removing n equations and n unknowns. It is also a very symmetric-looking set of conditions. Remember that the Kuhn-Tucker Lagrangian is $K(x_1, ..., x_n, \lambda_1, ..., \lambda_k)$. Instead of distinguishing between x's and λ's, let them all be z's, in which case the Kuhn-Tucker Lagrangian is $K(z_1, ..., z_{n+k})$. Then the Kuhn-Tucker conditions reduce to $z_j \partial K / \partial z_j = 0$ and $z_j \geq 0$ for $j = 1, ..., n+k$. This is fairly easy to remember, which is an advantage. The key to Kuhn-Tucker conditions, though, is remembering that they are just a particular reformulation of the standard inequality-constrained optimization problem with multiple constraints.

5.6 Problems

1. Consider the following problem:

$$\begin{gathered} \max_{x,y} x^2 y \\ \text{s.t. } 2x + 3y \leq 24 \\ 4x + y \leq 20 \end{gathered}$$

The Lagrangian can be written

$$\mathcal{L}(x, y, \lambda_1, \lambda_2) = x^2 y + \lambda_1 (24 - 2x - 3y) + \lambda_2 (20 - 4x - y)$$

(a) Solve the alternative problem

$$\max_{x,y} x^2 y$$
$$\text{s.t. } 2x + 3y = 24$$

Do the resulting values of x and y satisfy $4x + y \leq 20$?

(b) Solve the alternative problem

$$\max_{x,y} x^2 y$$
$$\text{s.t. } 4x + y = 20$$

Do the resulting values of x and y satisfy $2x + 3y \leq 24$?

(c) Based on your answers to (a) and (b), which of the two constraints bind? What do these imply about the values of λ_1 and λ_2?

(d) Solve the original problem.

(e) Draw a graph showing what is going on in this problem.

2. Consider the following problem:

$$\max_{x,y} x^2 y$$
$$\text{s.t. } 2x + 3y \leq 24$$
$$4x + y \leq 36$$

The Lagrangian can be written

$$\mathcal{L}(x, y, \lambda_1, \lambda_2) = x^2 y + \lambda_1(24 - 2x - 3y) + \lambda_2(36 - 4x - y)$$

(a) Solve the alternative problem

$$\max_{x,y} x^2 y$$
$$\text{s.t. } 2x + 3y = 24$$

Do the resulting values of x and y satisfy $4x + y \leq 36$?

(b) Solve the alternative problem

$$\max_{x,y} x^2 y$$
$$\text{s.t. } 4x + y = 36$$

Do the resulting values of x and y satisfy $2x + 3y \leq 24$?

(c) Based on your answers to (a) and (b), which of the two constraints bind? What do these imply about the values of λ_1 and λ_2?

(d) Solve the original problem.

(e) Draw a graph showing what is going on in this problem.

3. Consider the following problem:

$$\max_{x,y} 4xy - 3x^2$$

$$\begin{aligned} \text{s.t. } x + 4y &\leq 36 \\ 5x + 2y &\leq 45 \end{aligned}$$

The Lagrangian can be written

$$\mathcal{L}(x, y, \lambda_1, \lambda_2) = 4xy - 3x^2 + \lambda_1(36 - x - 4y) + \lambda_2(45 - 5x - 2y)$$

(a) Solve the alternative problem

$$\max_{x,y} 4xy - 3x^2$$

$$\text{s.t. } x + 4y = 36$$

Do the resulting values of x and y satisfy $5x + 2y \leq 45$?

(b) Solve the alternative problem

$$\max_{x,y} 4xy - 3x^2$$

$$\text{s.t. } 5x + 2y = 45$$

Do the resulting values of x and y satisfy $x + 4y \leq 36$?

(c) Find the solution to the original problem, including the values of λ_1 and λ_2.

4. Consider the following problem:

$$\max_{x,y} 3xy - 8x$$

$$\begin{aligned} \text{s.t. } x + 4y &\leq 24 \\ 5x + 2y &\leq 30 \end{aligned}$$

The Lagrangian can be written

$$\mathcal{L}(x, y, \lambda_1, \lambda_2) = 3xy - 8x + \lambda_1(24 - x - 4y) + \lambda_2(30 - 5x - 2y)$$

(a) Solve the alternative problem

$$\max_{x,y} 3xy - 8x$$

$$\text{s.t. } x + 4y = 24$$

Do the resulting values of x and y satisfy $5x + 2y \leq 30$?

(b) Solve the alternative problem

$$\max_{x,y} 3xy - 8x$$

$$\text{s.t. } 5x + 2y = 30$$

Do the resulting values of x and y satisfy $x + 4y \leq 24$?

(c) Find the solution to the original problem, including the values of λ_1 and λ_2.

5. Consider the following problem:

$$\begin{array}{c} \max_{x,y} x^2 y \\ \text{s.t. } 4x + 2y \leq 42 \\ x \geq 0 \\ y \geq 0 \end{array}$$

(a) Write down the Kuhn-Tucker Lagrangian for this problem.

(b) Write down the Kuhn-Tucker conditions.

(c) Solve the problem.

6. Consider the following problem:

$$\max_{x,y} xy + 40x + 60y$$

$$\begin{aligned} \text{s.t. } x + y &\leq 12 \\ x, y &\geq 0 \end{aligned}$$

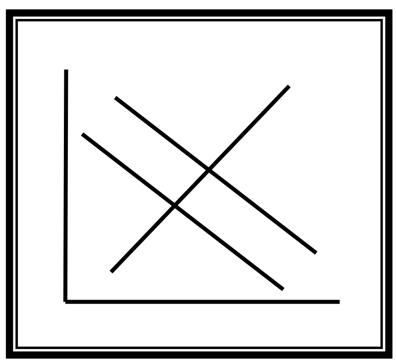
(a) Write down the Kuhn-Tucker Lagrangian for this problem.

(b) Write down the Kuhn-Tucker conditions.

(c) Solve the problem.

PART II

SOLVING SYSTEMS OF EQUATIONS



(linear algebra)

CHAPTER

6

Matrices

Matrices are 2-dimensional arrays of numbers, and they are useful for many things. They also behave differently that ordinary real numbers. This chapter tells how to work with matrices and what they are for.

6.1 Matrix algebra

A **matrix** is a rectangular array of numbers, such as the one below:

$$A = \begin{pmatrix} 6 & -5 \\ 2 & 3 \\ -1 & 4 \end{pmatrix}.$$

Matrices are typically denoted by capital letters. They have dimensions corresponding to the number of rows and number of columns. The matrix A above has 3 rows and 2 columns, so it is a 3×2 matrix. Matrix dimensions are always written as (# rows) $\times$ (# columns).

An **element** of a matrix is one of the entries. The element in row i and column j is denoted a_{ij}, and so in general a matrix looks like

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix}.$$

The matrix A above is an $n \times k$ matrix. A matrix in which the number of rows equals the number of columns is called a *square* matrix. In such a matrix, elements of the form a_{ii} are called *diagonal* elements because they land on the diagonal of the square matrix.

An n-dimensional vector can be thought of as an $n \times 1$ matrix. Therefore, in matrix notation vectors are written vertically:

$$\bar{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

When we write a vector as a column matrix we typically leave off the accent and write it simply as x.

Matrix addition is done element by element:

$$\begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1k}+b_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & \cdots & a_{nk}+b_{nk} \end{pmatrix}.$$

Before one can add matrices, though, it is important to make sure that the dimensions of the two matrices are identical. In the above example, both matrices are $n \times k$.

Just like with vectors, it is possible to multiply a matrix by a scalar. This is done element by element:

$$tA = t \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{pmatrix} = \begin{pmatrix} ta_{11} & \cdots & ta_{1k} \\ \vdots & \ddots & \vdots \\ ta_{n1} & \cdots & ta_{nk} \end{pmatrix}.$$

The big deal in matrix algebra is matrix multiplication. To multiply matrices A and B, several things are important. First, the order matters, as you will see. Second, the number of columns in the first matrix must equal the number of rows in the second. So, one must multiply an $n \times k$ matrix on the left by a $k \times m$ matrix on the right. The result will be an $n \times m$ matrix, with the k's canceling out. The formula for multiplying matrices is as follows. Let $C = AB$, with A an $n \times k$ matrix and B a $k \times m$ matrix. Then

$$c_{ij} = \sum_{s=1}^{k} a_{is} b_{sj}.$$

This is easier to see when we write the matrices A and B side-by-side:

$$C = AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k1} & b_{k2} & \cdots & b_{km} \end{pmatrix}.$$

Element c_{11} is

$$c_{11} = a_{11}b_{11} + a_{12}b_{21} + ... + a_{1s}b_{s1} + ... + a_{1k}b_{k1}.$$

So, element c_{11} is found by multiplying each member of row 1 in matrix A by the corresponding member of column 1 in matrix B and then summing. Element c_{ij} is found by multiplying each member of row i in matrix A by the corresponding member of column j in matrix B and then summing. For there to be the right number of elements for this to work, the number of columns in A must equal the number of rows in B.

As an example, multiply the two matrices below:

$$A = \begin{pmatrix} 6 & -1 \\ 4 & 3 \end{pmatrix}, B = \begin{pmatrix} 3 & 4 \\ -2 & 4 \end{pmatrix}.$$

Then

$$AB = \begin{pmatrix} 6 \cdot 3 + (-1)(-2) & 6 \cdot 4 + (-1)4 \\ 4 \cdot 3 + 3(-2) & 4 \cdot 4 + 3 \cdot 4 \end{pmatrix} = \begin{pmatrix} 20 & 20 \\ 6 & 28 \end{pmatrix}.$$

However,

$$BA = \begin{pmatrix} 3 \cdot 6 + 4 \cdot 4 & 3(-1) + 4 \cdot 3 \\ (-2)6 + 4 \cdot 4 & (-2)(-1) + 4 \cdot 3 \end{pmatrix} = \begin{pmatrix} 34 & 9 \\ 4 & 14 \end{pmatrix}.$$

Obviously, $AB \neq BA$, and matrix multiplication is *not* commutative. Because of this, we use the terminology that we **left-multiply** by B when we want BA and **right-multiply** by B when we want AB.

The square matrix

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

is special, and is called the **identity matrix**. To see why it is special, consider any $n \times k$ matrix A, and let I be the n-dimensional identity matrix. Letting $B = IA$, we get $b_{ij} = 0 \cdot a_{1j} + 0 \cdot a_{2j} + ... + 1 \cdot a_{ij} + ... + 0 \cdot a_{nj} = a_{ij}$. So, $IA = A$. The same thing happens when we right-multiply A by a k-dimensional identity matrix. Then $AI = A$. So, multiplying a matrix by the identity matrix is the same as multiplying an ordinary number by 1.

The **transpose** of the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} \end{pmatrix}.$$

is the matrix A^T given by

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{nk} \end{pmatrix}.$$

The transpose is generated by switching the rows and columns of the original matrix. Because of this, the transpose of an $n \times k$ matrix is a $k \times n$ matrix. Note that

$$(AB)^T = B^T A^T$$

so that the transpose of the product of two matrices is the product of the transposes of the two matrices, *but you have to switch the order of the matrices.* To check this, consider the following example employing the same two matrices we used above.

$$A = \begin{pmatrix} 6 & -1 \\ 4 & 3 \end{pmatrix}, B = \begin{pmatrix} 3 & 4 \\ -2 & 4 \end{pmatrix}.$$

$$AB = \begin{pmatrix} 20 & 20 \\ 6 & 28 \end{pmatrix}, \ (AB)^T = \begin{pmatrix} 20 & 6 \\ 20 & 28 \end{pmatrix}.$$

$$A^T = \begin{pmatrix} 6 & 4 \\ -1 & 3 \end{pmatrix}, \ B^T = \begin{pmatrix} 3 & -2 \\ 4 & 4 \end{pmatrix}.$$

$$B^T A^T = \begin{pmatrix} 3 & -2 \\ 4 & 4 \end{pmatrix} \begin{pmatrix} 6 & 4 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 20 & 6 \\ 20 & 28 \end{pmatrix} = (AB)^T,$$

as desired, but

$$A^T B^T = \begin{pmatrix} 6 & 4 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 4 & 4 \end{pmatrix} = \begin{pmatrix} 34 & 4 \\ 9 & 14 \end{pmatrix} = (BA)^T.$$

6.2 Uses of matrices

Suppose that you have a system of n equations in n unknowns, such as this one:

$$\begin{array}{c} a_{11}x_1 + ... + a_{1n}x_n = b_1 \\ a_{21}x_1 + ... + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + ... + a_{nn}x_n = b_n \end{array} \tag{6.1}$$

We can write this system easily using matrix notation. Letting

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \text{ and } b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

we can rewrite the system of equations (6.1) as

$$Ax = b. \tag{6.2}$$

The primary use of matrices is to solve systems of equations. As you have seen in the optimization examples, systems of equations regularly arise in economics.

Equation (6.2) raises two questions. First, when does a solution exist, that is, when can we find a vector x such that $Ax = b$? Second, how do we find the solution when it exists? The answers to both questions depend on the **inverse matrix** A^{-1}, which is a matrix having the property that

$$A^{-1}A = AA^{-1} = I,$$

that is, the matrix that you multiply A by to get the identity matrix. Remembering that the identity matrix plays the role of the number 1 in matrix multiplication, and that for ordinary numbers the inverse of y is the number y^{-1} such that $y^{-1} \cdot y = 1$, this formula is exactly what we are looking for. If A has an inverse (and that is a big if), then (6.2) can be solved by left-multiplying both sides of the equation by the inverse matrix A^{-1}:

$$\begin{aligned} A^{-1}Ax &= A^{-1}b \\ x &= A^{-1}b \end{aligned}$$

For ordinary real numbers, every number except 0 has a multiplicative inverse. Many more matrices than just one fail to have an inverse, though, so we must devote considerable attention to whether or not a matrix has an inverse.

A second use of matrices is for deriving second-order conditions for multivariable optimization problems. Recall that in a single-variable optimization problem with objective function f, the second-order condition is determined by the sign of the second derivative f''. When f is a function of several variables, however, we can write the vector of first partial derivatives

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

and the *matrix* of second partials

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

The relevant second-order conditions will come from conditions on the matrix of second partials, and we will do this in Chapter 9.

6.3 Determinants

Determinants of matrices are useful for determining (hence the name) whether a matrix has an inverse and also for solving equations such as (6.2). The **de-**

terminant of a square matrix A (and the matrix *must* be square) is denoted $|A|$. Defining it depends on the size of the matrix.

Start with a 1×1 matrix $A = (a_{11})$. The determinant $|A|$ is simply a_{11}. It can be either positive or negative, so don't confuse the determinant with the absolute value, even though they both use the same symbol.

Now look at a 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Here the determinant is defined as

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}.$$

For a 3×3 matrix we go through some more steps. Begin with the 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

We can get a **submatrix** of A by deleting a row and column. For example, if we delete the second row and the first column we are left with the submatrix

$$A_{21} = \begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix}.$$

In general, submatrix A_{ij} is obtained from A by deleting row i and column j. Note that there is one submatrix for each element, and you can get that submatrix by eliminating the element's row and column from the original matrix. Every element also has something called a *cofactor* which is based on the element's submatrix. Specifically, the **cofactor** of a_{ij} is the number c_{ij} given by

$$c_{ij} = (-1)^{i+j}|A_{ij}|,$$

that is, it is the determinant of the submatrix A_{ij} multiplied by -1 if $i + j$ is odd and multiplied by 1 if $i + j$ is even.

Using these definitions we can finally get the determinant of a 3×3 matrix, or any other square matrix for that matter. There are two ways to do it. The most common is to choose a column j. Then

$$|A| = a_{1j}c_{1j} + a_{2j}c_{2j} + ... + a_{nj}c_{nj}.$$

Before we see what this means for a 3×3 matrix let's check that it works for a 2×2 matrix. Choosing column $j = 1$ gives us

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}c_{11} + a_{12}c_{12} = a_{11}a_{22} + a_{12}(-a_{21}),$$

where $c_{11} = a_{22}$ because $1+1$ is even and $c_{21} = -a_{12}$ because $1+2$ is odd. We get exactly the same thing if we choose the second column, $j = 2$:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{12}c_{12} + a_{22}c_{22} = a_{12}(-a_{21}) + a_{22}a_{11}.$$

Finally, let's look at a 3×3 matrix, choosing $j = 1$.

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11}c_{11} + a_{21}c_{21} + a_{31}c_{31} \\ &= a_{11}(a_{22}a_{33} - a_{32}a_{23}) \\ &\quad -a_{21}(a_{12}a_{33} - a_{32}a_{13}) \\ &\quad +a_{31}(a_{12}a_{23} - a_{22}a_{13}). \end{aligned}$$

We can also find determinants by choosing a row. If we choose row i, then the determinant is given by

$$|A| = a_{i1}c_{i1} + a_{i2}c_{i2} + ... + a_{in}c_{in}.$$

The freedom to choose any row or column allows one to use zeros strategically. For example, when evaluating the determinant

$$\begin{vmatrix} 6 & 8 & -1 \\ 2 & 0 & 0 \\ -9 & 4 & 7 \end{vmatrix}$$

it would be best to choose the second row because it has two zeros, and the determinant is simply $a_{21}c_{21} = 2(-60) = -120$.

6.4 Cramer's rule

The process of using determinants to solve the system of equations given by $Ax = b$ is known as **Cramer's rule**. Begin with the matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Construct the matrix B_i from A by replacing the i-th column of A with the column vector b, so that

$$B_1 = \begin{pmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

and

$$B_i = \begin{pmatrix} a_{11} & \cdots & a_{1(i-1)} & b_1 & a_{1(i+1)} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2(i-1)} & b_2 & a_{2(i+1)} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n(i-1)} & b_n & a_{n(i+1)} & \cdots & a_{nn} \end{pmatrix}.$$

According to Cramer's rule, the solution to $Ax = b$ is the column vector x where

$$x_i = \frac{|B_i|}{|A|}.$$

Let's make sure this works using a simple example. The system of equations is

$$\begin{aligned} 4x_1 + 3x_2 &= 18 \\ 5x_1 - 3x_2 &= 9 \end{aligned}$$

Adding the two equations together yields

$$\begin{aligned} 9x_1 &= 27 \\ x_1 &= 3 \\ x_2 &= 2 \end{aligned}$$

Now let's do it using Cramer's rule. We have

$$A = \begin{pmatrix} 4 & 3 \\ 5 & -3 \end{pmatrix} \text{ and } b = \begin{pmatrix} 18 \\ 9 \end{pmatrix}.$$

Generate the matrices

$$B_1 = \begin{pmatrix} 18 & 3 \\ 9 & -3 \end{pmatrix} \text{ and } B_2 = \begin{pmatrix} 4 & 18 \\ 5 & 9 \end{pmatrix}.$$

Now compute determinants to get

$$|A| = -27, \ |B_1| = -81, \text{ and } |B_2| = -54.$$

Applying Cramer's rule we get

$$x = \begin{pmatrix} \frac{|B_1|}{|A|} \\ \frac{|B_2|}{|A|} \end{pmatrix} = \begin{pmatrix} \frac{-81}{-27} \\ \frac{-54}{-27} \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

6.5 Inverses of matrices

One important implication of Cramer's rule links the determinant of A to the existence of an inverse. To see why, recall that the solution, if it exists, to the system $Ax = b$ is $x = A^{-1}b$. Also, we know from Cramer's rule that $x_i = |B_i| / |A|$. For this number to exist, it must be the case that $|A| \neq 0$. This is sufficiently important that it has a name: the matrix A is **singular** if $|A| = 0$ and it is **nonsingular** if $|A| \neq 0$. Singular matrices do not have inverses, but nonsingular matrices do.

With some clever manipulation we can use Cramer's rule to invert the matrix A. To see how, recall that the inverse is defined so that

$$AA^{-1} = I.$$

We want to find A^{-1}, and it helps to define

$$A^{-1} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix}$$

and

$$x^i = \begin{pmatrix} x_{1i} \\ \vdots \\ x_{ni} \end{pmatrix}.$$

Remember that the coordinate vector $\bar{e}^i$ has zeros everywhere except for the i-th element, which is 1, and so

$$e^1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and e^i is the column vector with $e^i_i = 1$ and $e^i_j = 0$ when $j \neq i$. Note that e^i is the i-th column of the identity matrix. Then i-th column of the inverse matrix A^{-1} can be found by applying Cramer's rule to the system

$$Ax^i = e^i. \tag{6.3}$$

Construct the matrix B^i_j from A by replacing the j-th column of A with the column vector e^i. Then

$$B^i_j = \begin{pmatrix} a_{11} & \cdots & a_{1(j-1)} & 0 & a_{1(j+1)} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{(i-1)1} & \cdots & a_{(i-1)(j-1)} & 0 & a_{(i-1)(j+1)} & \cdots & a_{(i-1)n} \\ a_{i1} & \cdots & a_{i(j-1)} & 1 & a_{i(j+1)} & \cdots & a_{in} \\ a_{(i+1)1} & \cdots & a_{(i+1)(j-1)} & 0 & a_{(i+1)(j+1)} & \cdots & a_{(i+1)n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n(j-1)} & 0 & a_{n(j+1)} & \cdots & a_{nn} \end{pmatrix}$$

The solution to (6.3) is

$$x_{ji} = \frac{\left|B^i_j\right|}{|A|}.$$

Once again we can only get an inverse if $|A| \neq 0$.

We can simplify this further. Note that only one element of the j-th column of B^i_j is non-zero. So, we can take the determinant of B^i_j by using the j-th column and getting

$$\left|B^i_j\right| = (-1)^{i+j} |A_{ij}| = c_{ij},$$

which is a cofactor of the matrix A. So, we get the formula for the inverse:

$$x_{ji} = \frac{(-1)^{i+j} |A_{ij}|}{|A|}.$$

Note that the subscript on x is ji but the subscript on A is ij.

Let's check to see if this works for a 2×2 matrix. Use the matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We get

$$B^1_1 = \begin{pmatrix} 1 & b \\ 0 & d \end{pmatrix},\ B^2_1 = \begin{pmatrix} 0 & b \\ 1 & d \end{pmatrix},\ B^1_2 = \begin{pmatrix} a & 1 \\ c & 0 \end{pmatrix},\ \text{and}\ B^2_2 = \begin{pmatrix} a & 0 \\ c & 1 \end{pmatrix}.$$

The determinants are

$$\left|B_1^1\right| = d,\ \left|B_1^2\right| = -b,\ \left|B_2^1\right| = -c, \text{ and } \left|B_2^2\right| = a.$$

We also know that $|A| = ad - bc$. Thus,

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We can check this easily:

$$\begin{aligned} A^{-1}A &= \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} \\ &= \frac{1}{ad-bc}\begin{pmatrix} ad-bc & -ba+ba \\ -ca+ac & -cb+ad \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

6.6 Problems

1. Perform the following computations:

   (a) $4\begin{pmatrix} 6 & -4 & 2 \\ 3 & 3 & 9 \end{pmatrix} - 5\begin{pmatrix} 1 & 0 & 6 \\ 2 & 3 & -5 \end{pmatrix}$

   (b) $\begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}\begin{pmatrix} 3 & -2 & -1 \\ 4 & 4 & 1 \end{pmatrix}$

   (c) $\begin{pmatrix} 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 1 & 1 & 0 & 7 \end{pmatrix}\begin{pmatrix} 2 & 2 & 1 \\ 3 & -3 & 0 \\ 7 & 2 & -1 \\ -4 & -5 & 0 \end{pmatrix}$

   (d) $\begin{pmatrix} 2 & 1 & 1 \end{pmatrix}\begin{pmatrix} 3 & 1 & -1 \\ 2 & 1 & 1 \\ 1 & 2 & 2 \end{pmatrix}\begin{pmatrix} 5 \\ -1 \\ -1 \end{pmatrix}$

2. Perform the following computations:

   (a) $6\begin{pmatrix} 3 & -3 \\ -4 & 5 \\ 1 & 2 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 8 & 10 \\ -6 & -4 \\ -6 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 5 & 1 & 3 \\ 4 & 0 & 6 \\ -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} 0 & -4 \\ 1 & 5 \\ 2 & 5 \end{pmatrix}$

(c) $\begin{pmatrix} 6 & -1 & 3 & 0 \\ 1 & 5 & 2 & 1 \end{pmatrix} \begin{pmatrix} 5 & -1 & 4 \\ 0 & -2 & 4 \\ 1 & -3 & 4 \\ 0 & 6 & 4 \end{pmatrix}$

(d) $\begin{pmatrix} 10 & 5 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 1 \\ 1 & 3 & 1 \\ 5 & 0 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -1 \\ -3 \end{pmatrix}$

3. Find the determinants of the following matrices:

(a) $\begin{pmatrix} 2 & 1 \\ -4 & 5 \end{pmatrix}$

(b) $\begin{pmatrix} 3 & 1 & 0 \\ -2 & 7 & -2 \\ 2 & 0 & 6 \end{pmatrix}$

4. Find the determinants of the following matrices:

(a) $\begin{pmatrix} 3 & 6 \\ -4 & -1 \end{pmatrix}$

(b) $\begin{pmatrix} 2 & 0 & -1 \\ 1 & 3 & 0 \\ 0 & 6 & -1 \end{pmatrix}$

5. Use Cramer's rule to solve the following system of equations:

$$\begin{aligned} 6x - 2y - 3z &= 1 \\ 2x + 4y + z &= -2 \\ 3x - z &= 8 \end{aligned}$$

6. Use Cramer's rule to solve the following system of equations:

$$\begin{aligned} 5x - 2y + z &= 9 \\ 3x - y &= 9 \\ 3y + 2z &= 15 \end{aligned}$$

7. Invert the following matrices:

   (a) $\begin{pmatrix} 2 & 3 \\ -2 & 1 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$

8. Invert the following matrices:

   (a) $\begin{pmatrix} -4 & 1 \\ 2 & -4 \end{pmatrix}$

   (b) $\begin{pmatrix} 5 & 1 & -1 \\ 0 & 2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$

CHAPTER

7

Systems of equations

Think about the general system of equations

$$Ax = b \tag{7.1}$$

where A is an $n \times n$ matrix and x and b are $n \times 1$ vectors. This expands to the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + ... + a_{nn}x_n &= b_n \end{aligned}$$

The task for this chapter is to determine (1) whether the system has a solution $(x_1, ..., x_n)$, and (2) whether that solution is unique.

We will use three examples to motivate our results. They use $n = 2$ to allow graphical analysis.

Example 3

$$\begin{aligned} 2x + y &= 6 \\ x - y &= -3 \end{aligned}$$

The solution to this one is $(x, y) = (1, 4)$.

Example 4

$$\begin{aligned} x - 2y &= 1 \\ 4y - 2x &= -2 \end{aligned}$$

This one has an infinite number of solutions defined by $(x, y) = (x, \frac{x-1}{2})$.

Example 5

$$\begin{aligned} x - y &= 4 \\ 2y - 2x &= 5 \end{aligned}$$

This one has no solution.

7.1 Identifying the number of solutions

7.1.1 The inverse approach

If A has an inverse A^{-1}, then left-multiplying both sides of (7.1) by A^{-1} yields $x = A^{-1}b$. So, what we really want to know is, *when does an inverse exist*? We already know that an inverse exists if the determinant is nonzero, that is, if $|A| \neq 0$.

7.1.2 Row-echelon decomposition

Write the **augmented matrix**

$$B = (A|b) = \left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & b_n \end{array} \right)$$

which is $n \times (n+1)$. The goal is to transform the matrix B through operations that consist of multiplying one row by a scalar and adding it to another row, and ending up with a matrix in which all the elements below the diagonal are 0. This is the **row-echelon form** of the matrix.

Example 6 (3 continued) *Form the augmented matrix*

$$B = \left(\begin{array}{cc|c} 2 & 1 & 6 \\ 1 & -1 & -3 \end{array} \right)$$

Multiply the first row by $-\frac{1}{2}$ *and add it to the second row to get*

$$R = \left(\begin{array}{cc|c} 2 & 1 & 6 \\ 1-1 & -1-\frac{1}{2} & -3-3 \end{array} \right) = \left(\begin{array}{cc|c} 2 & 1 & 6 \\ 0 & -\frac{3}{2} & -6 \end{array} \right)$$

Example 7 (4 continued) *Form the augmented matrix*

$$B = \left(\begin{array}{cc|c} 1 & -2 & 1 \\ -2 & 4 & -2 \end{array} \right)$$

Multiply the first row by 2 *and add it to the second row to get*

$$R = \left(\begin{array}{cc|c} 1 & -2 & 1 \\ -2+2 & 4-4 & -2+2 \end{array} \right) = \left(\begin{array}{cc|c} 1 & -2 & 1 \\ 0 & 0 & 0 \end{array} \right)$$

Example 8 (5 continued) *Form the augmented matrix*

$$B = \left(\begin{array}{cc|c} 1 & -1 & 4 \\ -2 & 2 & 5 \end{array} \right)$$

Multiply the first row by 2 *and add it to the second row to get*

$$R = \left(\begin{array}{cc|c} 1 & -1 & 4 \\ -2+2 & 2-2 & 5+8 \end{array} \right) = \left(\begin{array}{cc|c} 1 & -1 & 4 \\ 0 & 0 & 13 \end{array} \right)$$

Example 3 has a unique solution, Example 4 has an infinite number of them, and Example 5 has no solution. These results correspond to properties of the row-echelon matrix R. If the row-echelon form of the augmented matrix has only nonzero diagonal elements, there is a unique solution. If it has some rows that are zero, there are infinitely many solutions. If there are rows with zeros everywhere except in the last column, there is no solution.

Definition 1 *The* ***rank*** *of a matrix is the number of nonzero rows in its row-echelon form.*

Proposition 9 *The* $n \times n$ *square matrix* A *has an inverse if and only if its rank is* n.

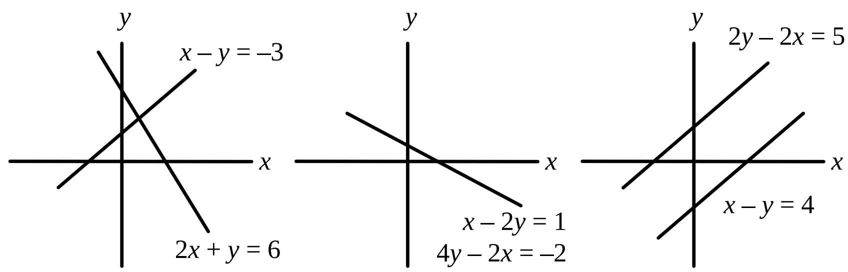


Figure 7.1: Graphing in (x, y) space: One solution when the lines intersect (left graph), infinite number of solutions when the lines coincide (center graph), and no solutions when the lines are parallel (right graph)

7.1.3 Graphing in (x,y) space

This is pretty simple and is shown in Figure 7.1. The equations are lines. In example 3 the equations constitute two different lines that cross at a single point. In example 4 the lines coincide. In example 5 the lines are parallel. We get a unique solution in example 3, an infinite number of them in example 4, and no solution in example 5.

What happens if we move to three dimensions? An equation reduces the dimension by one, so each equation identifies a plane. Two planes intersect in a line. The intersection of a line and a plane is a point. So, we get a unique solution if the planes intersect in a single point, an infinite number of solutions if the three planes intersect in a line or in a plane, and no solution if two or more of the planes are parallel.

7.1.4 Graphing in column space

This approach is completely different but really useful. Each column of A is an $n \times 1$ vector. So is b. The question becomes, is b in the **column space**, that is, the space spanned by the columns of A?

A **linear combination** of vectors $\bar{a}$ and $\bar{b}$ is given by $x\bar{a} + y\bar{b}$, where x and y are scalars.

The **span** of the vectors $\bar{x}$ and $\bar{y}$ is the set of linear combinations of the

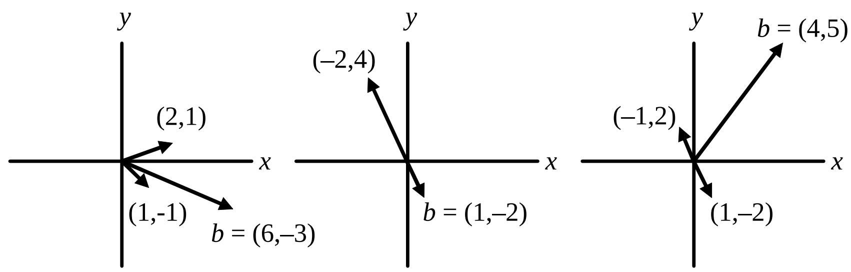


Figure 7.2: Graphing in the column space: When the vector b is in the column space (left graph and center graph) a solution exists, but when the vector b is not in the column space there is no solution (right graph)

two vectors.

Example 9 (3 continued) *The column vectors are* $(2, 1)$ *and* $(1, -1)$. *These span the entire plane. For any vector* (z_1, z_2), *it is possible to find numbers* x *and* y *that solve* $x(2, 1) + y(1, -1) = (z_1, z_2)$. *Written in matrix notation this is*

$$\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}.$$

We already know that the matrix has an inverse, so we can solve this.

Look at left graph in Figure 7.2. The column space is spanned by the two vectors $(2, 1)$ *and* $(1, -1)$. *The vector* $b = (6, -3)$ *lies in the span of the column vectors. This leads to our rule for a solution:* ***If*** b ***is in the span of the columns of*** A***, there is a solution.***

Example 10 (4 continued) *In the center graph in Figure 7.2, the column vectors are* $(1, -2)$ *and* $(-2, 4)$. *They are on the same line, so they only span that line. In this case the vector* $b = (1, -2)$ *is also on that line, so there is a solution.*

Example 11 (5 continued) *In the right panel of Figure 7.2, the column vectors are* $(1, -2)$ *and* $(-1, 2)$, *which span a single line. This time, though, the vector* $b = (4, 5)$ *is not on that line, and there is no solution.*

Two vectors $\bar{a}$ and $\bar{b}$ are **linearly dependent** if there exists a scalar $r \neq 0$ such that $\bar{a} = r\bar{b}$.

Two vectors $\bar{a}$ and $\bar{b}$ are **linearly independent** if there is no scalar $r \neq 0$ such that $\bar{a} = r\bar{b}$. Equivalently, $\bar{a}$ and $\bar{b}$ are linearly independent if there do not exist scalars r_1 and r_2, not both equal to zero, such that

$$r_1\bar{a} + r_2\bar{b} = \bar{0}. \tag{7.2}$$

One more way of writing this is that $\bar{a}$ and $\bar{b}$ are linearly independent if the only solution to the above equation has $r_1 = r_2 = 0$.

This last one has some impact when we write it in matrix form. Suppose that the two vectors are 2-dimensional, and construct the matrix A by using the vectors $\bar{a}$ and $\bar{b}$ as its columns:

$$A = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}.$$

Now we can write (7.2) as

$$A \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

If A has an inverse, there is a unique solution to this equation, given by

$$\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = A^{-1} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

So, invertibility of A and the linear independence of its columns are inextricably linked.

Proposition 10 *The square matrix A has an inverse if its columns are mutually linearly independent.*

7.2 Summary of results

The system of n linear equations in n unknowns given by

$$Ax = b$$

has a *unique* solution if:

1. A^{-1} exists.
2. $|A| \neq 0$.
3. The row-echelon form of A has no rows with all zeros.
4. A has rank n.
5. The columns of A span n-dimensional space.
6. The columns of A are mutually linearly independent.

The system of n linear equations in n unknowns given by $Ax = b$ has an *infinite number* of solutions if:

1. The row-echelon form of the augmented matrix $(A|b)$ has rows with all zeros.
2. The vector b is contained in the span of a *subset* of the columns of A.

The system of n linear equations in n unknowns given by $Ax = b$ has *no* solution if:

1. The row-echelon form of the augmented matrix $(A|b)$ has at least one row with all zeros *except* in the last column.
2. The vector b is *not* contained in the span of the columns of A.

7.3 Problems

1. Determine whether or not the following systems of equations have a unique solution, an infinite number of solutions, or no solution.

   (a)

$$\begin{aligned} 3x + 6y &= 4 \\ 2x - 5z &= 8 \\ x - y - z &= -10 \end{aligned}$$

(b)

$$\begin{aligned} 4x - y + 8z &= 160 \\ 17x - 8y + 10z &= 200 \\ -3x + 2y + 2z &= 40 \end{aligned}$$

(c)

$$\begin{aligned} 2x - 3y &= 6 \\ 3x + 5z &= 15 \\ 2x + 6y + 10z &= 18 \end{aligned}$$

(d)

$$\begin{aligned} 4x - y + 8z &= 30 \\ 3x + 2z &= 20 \\ 5x + y - 2z &= 40 \end{aligned}$$

(e)

$$\begin{aligned} 6x - y - z &= 3 \\ 5x + 2y - 2z &= 10 \\ y - 2z &= 4 \end{aligned}$$

2. Find the values of a for which the following matrices do *not* have an inverse.

(a)

$$\begin{pmatrix} 6 & -1 \\ 2 & a \end{pmatrix}$$

(b)

$$\begin{pmatrix} 5 & a & 0 \\ 4 & 2 & 1 \\ -1 & 3 & 1 \end{pmatrix}$$

(c)

$$\begin{pmatrix} 5 & 3 \\ -3 & a \end{pmatrix}$$

(d)

$$\begin{pmatrix} -1 & 3 & 1 \\ 0 & 5 & a \\ 6 & 2 & 1 \end{pmatrix}$$

CHAPTER

8

Using linear algebra in economics

8.1 IS-LM analysis

Consider the following model of a closed macroeconomy:

$$\begin{aligned} Y &= C + I + G \\ C &= c((1-t)Y) \\ I &= i(R) \\ M &= P \cdot m(Y, R) \end{aligned}$$

with

$$\begin{aligned} 0 &< c' < \frac{1}{1-t} \\ i' &< 0 \\ m_Y &> 0, m_R < 0 \end{aligned}$$

Here Y is GDP, which you can think of as production, income, or spending. The variables C, I, and G are the spending components of GDP, with C

standing for consumption, I for investment (which is spending by businesses on new capital and by consumers on new housing, not what you do in the stock market), and G for government spending. The first equation says that total spending Y is equal to the sum of the components of spending, $C + I + G$. There are no imports or exports, so this is a closed-economy model.

The amount of consumption depends on consumers' after-tax income, and when Y is income and t is the tax rate, after-tax (or disposable) income is $(1 - t)Y$. So the second equation says that consumption is a function $c(\cdot)$ of disposable income, and $c' > 0$ means that it is an increasing function.

Investment is typically spending on large items, and it is often financed through borrowing. Because of this, investment depends on the interest rate R, and when the interest rate increases borrowing becomes more expensive and the amount of investment falls. Consequently the investment function $i(R)$ is decreasing.

M is money supply, and the right-hand side of the fourth equation is money demand. P is the price level, and when things become more expensive it takes more money to purchase the same amount of stuff. When income Y increases people want to buy more stuff, and they need more money to do it with, so money demand increases when income increases. Also, since money is cash and checking account balances, which tend not to earn interest, and so when interest rates rise people tend to move some of their assets into interest-bearing accounts. This means that money demand falls when interest rates rise.

The four equations provide a model of the economy, known as the IS-LM model. The first three equations describe the IS curve from macro courses and the fourth equation describes the LM curve. The model is useful for describing a closed economy in the short run, that is, before the price level has time to adjust.

At this point you should be wondering why we care. The answer is that we want to do some comparative statics analysis. The variables G, t, and M are exogenous policy variables. P is predetermined and assumed constant for the problem. Everything else is endogenous, so everything else is a function of G, t, and M. We are primarily interested in the variables Y and R, and we want to see how they change when the policy variables change.

Let's look for the comparative statics derivatives dY/dG and dR/dG. To find them, first simplify the system of four equations to a system of two equations:

$$\begin{aligned} Y &= c((1-t)Y) + i(R) + G \\ M &= P \cdot m(Y, R) \end{aligned}$$

Implicitly differentiate with respect to G to get

$$\begin{aligned} \frac{dY}{dG} &= (1-t)c'\frac{dY}{dG} + i'\frac{dR}{dG} + 1 \\ 0 &= P \cdot m_Y \frac{dY}{dG} + P \cdot m_R \frac{dR}{dG} \end{aligned}$$

Rearrange as

$$\begin{aligned} \frac{dY}{dG} - (1-t)c'\frac{dY}{dG} - i'\frac{dR}{dG} &= 1 \\ m_Y \frac{dY}{dG} + m_R \frac{dR}{dG} &= 0 \end{aligned}$$

We can write this in matrix form

$$\begin{pmatrix} 1-(1-t)c' & -i' \\ m_Y & m_R \end{pmatrix} \begin{pmatrix} \frac{dY}{dG} \\ \frac{dR}{dG} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Now use Cramer's rule to solve for dY/dG and dR/dG:

$$\frac{dY}{dG} = \frac{\begin{vmatrix} 1 & -i' \\ 0 & m_R \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} = \frac{m_R}{[1-(1-t)c']m_R + m_Y i'}.$$

Both the numerator and denominator are negative, so $dY/dG > 0$. GDP rises when government spending increases. Now for interest rates:

$$\frac{dR}{dG} = \frac{\begin{vmatrix} 1-(1-t)c' & 1 \\ m_Y & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} = \frac{-m_Y}{[1-(1-t)c']m_R + m_Y i'}.$$

The numerator is negative and so is the denominator. Thus, $dR/dG > 0$. An increase in government spending increases both GDP and interest rates in the short run.

It is also possible to find the comparative statics derivatives dY/dt, dR/dt, dY/dM, and dR/dM. You should figure them out yourselves.

8.2 Econometrics

We want to look at the estimation equation

$$\underset{(n\times 1)}{y} = \underset{(n\times k)}{X}\underset{(k\times 1)}{\beta} + \underset{(n\times 1)}{e}. \tag{8.1}$$

The matrix y contains the data on our dependent variable, and the matrix X contains the data on the independent, or explanatory, variables. Each row is an observation, and each column is an explanatory variable. From the equation we see that there are n observations and k explanatory variables. The matrix β is a vector of k coefficients, one for each of the k explanatory variables. The estimates will not be perfect, and so the matrix e contains error terms, one for each of the n observations. The fundamental problem in econometrics is to use data to estimate the coefficients in β in order to make the errors e small. The notion of "small," and the one that is consistent with the standard practice in econometrics, is to make the sum of the squared errors as small as possible.

8.2.1 Least squares analysis

We want to minimize the sum of the squared errors. Rewrite (8.1) as

$$e = y - X\beta$$

and note that

$$e^T e = \sum_{i=1}^{n} e_i^2.$$

Then

$$\begin{aligned} e^T e &= (y - X\beta)^T (y - X\beta) \\ &= y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta \end{aligned}$$

We want to minimize this expression with respect to the parameter vector β. But notice that there is a β and a β^T in the expression. Let's treat these as two separate variables to get two FOCs:

$$\begin{aligned} -X^T y + X^T X\beta &= 0 \\ -y^T X + \beta^T X^T X &= 0 \end{aligned}$$

These are two copies of the same equation, because the first is the transpose of the second. So let's use the first one because it has β instead of β^T. Solving for β yields

$$\begin{aligned} X^T X\beta &= X^T y \\ \hat{\beta} &= (X^T X)^{-1} X^T y \end{aligned}$$

We call $\hat{\beta}$ the OLS estimator. Note that it is determined entirely by the data, that is, by the independent variable matrix X and the dependent variable vector y.

8.2.2 A lame example

Consider a regression with two observations and one independent variable, with the data given in the table below.

Observation number	Dependent variable y	Independent variable X
1	4	3
2	8	4

There is no constant. Our two observations lead to the two equations

$$\begin{aligned} 4 &= 3\beta + e_1 \\ 8 &= 4\beta + e_2 \end{aligned}$$

We want to find the value of β that minimizes $e_1^2 + e_2^2$.

Since $e_1 = 4 - 3\beta$ and $e_2 = 8 - 4\beta$, we get

$$\begin{aligned} e_1^2 + e_2^2 &= (4 - 3\beta)^2 + (8 - 4\beta)^2 \\ &= 16 - 24\beta + 9\beta^2 + 64 - 64\beta + 16\beta^2 \\ &= 80 - 88\beta + 25\beta^2. \end{aligned}$$

Minimize this with respect to β. The FOC is

$$\begin{aligned} -88 + 50\beta &= 0 \\ \beta &= \frac{88}{50} = \frac{44}{25}. \end{aligned}$$

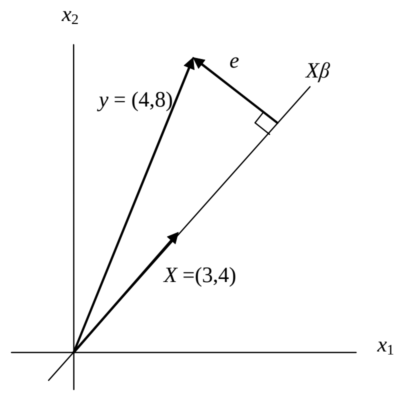


Figure 8.1: Graphing the lame example in column space

Now let's do it with matrices. The two equations can be written

$$\begin{pmatrix} 4 \\ 8 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} (\beta) + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

The OLS estimator is

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T y \\ &= \left(\begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right)^{-1} \begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ 8 \end{pmatrix} \\ &= (25)^{-1}(44) = \frac{44}{25}. \end{aligned}$$

We get the same answer.

8.2.3 Graphing in column space

We want to graph the previous example in column space. The example is lame for precisely this reason – so I can graph it in two dimensions.

The key here is to think about what we are doing when we find the value of β to minimize $e_1^2 + e_2^2$. X is a point, shown by the vector $X = (3, 4)$ in Figure 8.1, and $X\beta$ is the equation of the line through that point. y is another point, shown by the vector $y = (4, 8)$ in the figure. e is the vector connecting some point $X\hat{\beta}$ on the line $X\beta$ to the point y. $e_1^2 + e_2^2$ is the square of the length of e, so we want to minimize the length of e, and we do that by finding the point on the line $X\beta$ that is closest to the point y. Graphically, the closest point is the one that causes the vector e to be at a right angle to the vector X.

Two vectors $\bar{a}$ and $\bar{b}$ are **orthogonal** if $\bar{a} \cdot \bar{b} = 0$. This means we can find our coefficients by having the vector e be orthogonal to the vector X. Remembering that $e = y - X\beta$, we get

$$X \cdot (y - X\hat{\beta}) = 0.$$

Now notice that if we write the two vectors $\bar{a}$ and $\bar{b}$ as column matrices A and B, we have

$$\bar{a} \cdot \bar{b} = A^T B.$$

Thus we can rewrite the above expression as

$$\begin{aligned} X^T(y - X\hat{\beta}) &= 0 \\ X^T y - X^T X\hat{\beta} &= 0 \\ X^T X\hat{\beta} &= X^T y \\ \hat{\beta} &= (X^T X)^{-1} X^T y, \end{aligned}$$

which is exactly the answer we got before.

8.2.4 Interpreting some matrices

We have the estimated parameters given by

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

This tells us that the predicted values of y are

$$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y.$$

The matrix

$$X(X^T X)^{-1} X^T$$

is a **projection matrix**, and it projects the vector y onto the column space of X.

The residuals vector can be written

$$\begin{aligned} e &= y - X\hat{\beta} \\ &= (I - X(X^TX)^{-1}X^T)y. \end{aligned}$$

The two matrices $X(X^TX)^{-1}X^T$ and $(I - X(X^TX)^{-1}X^T)$ have the special property that they are **idempotent**, that is, they satisfy the property that

$$AA = A.$$

Geometrically it is clear why this happens. Suppose we apply the projection matrix $X(X^TX)^{-1}X^T$ to the vector y. That projects y onto the column space of X, so that $X(X^TX)^{-1}X^Ty$ lies in the column space of X. If we apply the same projection matrix a second time, it doesn't do anything because y is already in the column space of X. Similarly, the matrix $(I - X(X^TX)^{-1}X^T)$ projects y onto the space that is *orthogonal* to the column space of X. Doing it a second time does nothing, because it is projecting into the same space a second time.

8.3 Stability of dynamic systems

In macroeconomics and time series econometrics a common theme is the stability of the economy. In this section I show how stability works and relate stability to matrices.

8.3.1 Stability with a single variable

A **dynamic system** takes the form of

$$y_{t+1} = ay_t.$$

The variable t denotes the period number, so period $t+1$ is the period that immediately follows period t. The variable we are interested in is y, and, in particular, we would like to know how y varies over time. The initial period is period 0, and the initial value of y_t is y_0, which for the sake of argument we will assume is positive. The process is not very interesting if $a = 0$, because then $y_1 = y_2 = y_3 = ... = 0$, and it's also not very interesting if $a = 1$, because then $y_1 = y_2 = ... = y_0$. We want some movement in y_t, so let's assume that $a \neq 0$ and $a \neq 1$.

The single-variable system is pretty straightforward (some might even say lame), and it follows the following process:

$$\begin{aligned} y_1 &= ay_0 \\ y_2 &= ay_1 = a^2 y_0 \\ &\vdots \\ y_t &= a^t y_0 \\ &\vdots \end{aligned}$$

The process $y_0, y_1, ...$ is if it eventually converges to a finite value, or, in mathematical terms, if there exists a value y such that

$$\lim_{t \to \infty} y_t = y.$$

If the process is not stable then it explodes, diverging to either $+\infty$ or $-\infty$.

Whether or not the process is stable is determined by the magnitude of the parameter a. To see how, look at

$$\lim_{t \to \infty} y_t = \lim_{t \to \infty} a^t y_0 = y_0 \lim_{t \to \infty} a^t.$$

If $a > 1$ then $a^t \to \infty$, and the process cannot be stationary unless y_0 just happens to be zero, which is unlikely. Similarly, if $a < -1$ the process also diverges, this time cycling between positive and negative values depending on whether t is positive or negative, respectively (because y_0 is positive). On the other hand, if $0 < a < 1$, the value $a^t < 1$ and thus $y_t = a^t y_0 < y_0$. What's more, as $t \to \infty$, the quantity $a^t \to 0$, and therefore $y^t \to 0$. Thus, the process is stable when $a \in [0, 1)$. It also turns out to be stable when $-1 < a < 0$. The reasoning is the same. When $a \in (-1, 0)$ we have $a^t \in (-1, 1)$ and $\lim_{t \to \infty} a^t = 0$.

This reasoning gives us two stability conditions:

$$\begin{aligned} |a| &< 1 \\ \lim_{t \to \infty} y_t &= 0. \end{aligned}$$

8.3.2 Stability with two variables

All of that was pretty simple. Let's look at a dynamic system with two variables:

$$\begin{aligned} y_{t+1} &= ay_t + bz_t \\ z_{t+1} &= cy_t + dz_t \end{aligned}$$

Now both variables depend on the past values of both variables. When is this system stable?

We can write the system in matrix form:

$$\begin{pmatrix} y_{t+1} \\ z_{t+1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \tag{8.2}$$

or, in shorthand notation,

$$\bar{y}_{t+1} = A\bar{y}_t.$$

But this gives us a really complicated system. Sure, we know that

$$\begin{pmatrix} y_t \\ z_t \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^t \begin{pmatrix} y_0 \\ z_0 \end{pmatrix},$$

but the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^t$$

is really complicated. For example,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^3 = \begin{pmatrix} a^3 + 2bca + bcd & b\left(a^2 + ad + d^2 + bc\right) \\ c\left(a^2 + ad + d^2 + bc\right) & d^3 + 2bcd + abc \end{pmatrix}$$

and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{10}$$

is too big to fit on the page.

Things would be easy if the matrix was diagonal, that is, if $b = c = 0$. Then we would have two separate single-variable dynamic processes

$$\begin{aligned} y_{t+1} &= ay_t \\ z_{t+1} &= dz_t \end{aligned}$$

and we already know the stability conditions for these. But the matrix is not diagonal. So let's mess with things to get a system with a diagonal matrix. It will take a while, so just remember the goal when we get there: we want to generate a system that looks like

$$\bar{x}_{t+1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \bar{x}_t \tag{8.3}$$

because then we can treat the stability of the elements of the vector $\bar{x}$ separately.

8.3.3 Eigenvalues and eigenvectors

Begin by remembering that I is the identity matrix. An **eigenvalue** of the square matrix A is a value λ such that

$$\det(A - \lambda I) = \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = 0.$$

Taking the determinant yields

$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = ad - a\lambda - d\lambda + \lambda^2 - bc.$$

So, the eigenvalues are the solutions to the quadratic equation

$$\lambda^2 - (a+d)\lambda + (ad - bc) = 0.$$

In general quadratic equations have two solutions, call them λ_1 and λ_2.

For example, suppose that the matrix A is

$$A = \begin{pmatrix} 3 & 2 \\ 6 & -1 \end{pmatrix}.$$

The eigenvalues satisfy the equation

$$\begin{aligned} \lambda^2 - (3-1)\lambda + (-3-12) &= 0 \\ \lambda^2 - 2\lambda - 15 &= 0 \\ \lambda &= 5, -3. \end{aligned}$$

These are the eigenvalues of A. Look at the two matrices we get from the formula $A - \lambda I$:

$$\begin{pmatrix} 3-5 & 2 \\ 6 & -1-5 \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 6 & -6 \end{pmatrix}$$
$$\begin{pmatrix} 3-(-3) & 2 \\ 6 & -1-(-3) \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ 6 & 2 \end{pmatrix}$$

and both of these matrices are singular, which is what you get when you make the determinant equal to zero.

An **eigenvector** of A is a vector v such that

$$(A - \lambda I)v = 0$$

where λ is an eigenvalue of A. For our example, the two eigenvectors are the solutions to

$$\begin{pmatrix} -2 & 2 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$\begin{pmatrix} 6 & 2 \\ 6 & 2 \end{pmatrix} \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where I made the second subscript denote the number of the eigenvector and the first subscript denote the element of that eigenvector. These two equations have simple solutions. The first equation holds when

$$\begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

because

$$\begin{pmatrix} -2 & 2 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

and the second equation holds when

$$\begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

because

$$\begin{pmatrix} 6 & 2 \\ 6 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The relationship between eigenvalues and eigenvectors is useful for our task. Recall that if λ is an eigenvalue and v is an eigenvector then

$$\begin{aligned} (A - \lambda I)v &= 0 \\ Av - \lambda I v &= 0 \\ Av &= \lambda v. \end{aligned}$$

In particular, we have

$$A \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} = \lambda_1 \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix}$$

and

$$A \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} = \lambda_2 \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix}.$$

Construct the matrix V so that the two eigenvectors are columns:

$$V = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix}.$$

Then we have

$$\begin{aligned} AV &= \left(A \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} \; A \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} \right) \\ &= \left(\lambda_1 \begin{pmatrix} v_{11} \\ v_{21} \end{pmatrix} \; \lambda_2 \begin{pmatrix} v_{12} \\ v_{22} \end{pmatrix} \right) \\ &= \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \\ &= V \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \end{aligned}$$

We are almost there. If V has an inverse, V^{-1}, we can left-multiply both sides by V^{-1} to get

$$V^{-1}AV = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}. \tag{8.4}$$

This is the diagonal matrix we were looking for to make our dynamic system easy.

8.3.4 Back to the dynamic system

Go back to the dynamic system

$$\bar{y}_{t+1} = A\bar{y}_t.$$

Create a different vector $\bar{x}$ according to

$$\bar{x} = V^{-1}\bar{y},$$

where V is the matrix of eigenvectors we constructed above. This implies that

$$\bar{y} = V\bar{x}. \tag{8.5}$$

Then

$$\bar{x}_{t+1} = V^{-1}\bar{y}_{t+1}.$$

Since $\bar{y}_{t+1} = A\bar{y}_t$, we get

$$\begin{aligned} \bar{x}_{t+1} &= V^{-1}(A\bar{y}_t) \\ &= \left(V^{-1}A\right)\bar{y}_t \\ &= (V^{-1}A)(V\bar{x}_t) \\ &= \left(V^{-1}AV\right)\bar{x}_t. \end{aligned}$$

We figured out the formula for $V^{-1}AV$ in equation (8.4), which gives us

$$\bar{x}_{t+1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \bar{x}_t.$$

But this is exactly the diagonal system we wanted in equation (8.3). So we are there. It's about time.

Let's look back at what we have. We began with a matrix A. We found its two eigenvalues λ_1 and λ_2, and we found the two corresponding eigenvectors and combined them to form the matrix V. All of this comes from the matrix A, so we haven't added anything that wasn't in the problem. But our original problem was about the vector $\bar{y}_t$, and our new problem is about the vector $\bar{x}_t$.

Using the intuition we gained from the section with a single-variable dynamic system, we say that the process $\bar{y}_0, \bar{y}_1, ...$ is stable if

$$\lim_{t\to\infty} \bar{y}_t = 0.$$

The dynamic process $\bar{y}_{t+1} = A\bar{y}_t$ was hard to work with, and so it was difficult to determine whether or not it was stable. But the dynamic process

$$\begin{pmatrix} w_{t+1} \\ x_{t+1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} w_t \\ x_t \end{pmatrix}$$

is easy to work with, because multiplying it out yields the two simple, single-variable equations

$$\begin{aligned} w_{t+1} &= \lambda_1 w_t \\ x_{t+1} &= \lambda_2 x_t. \end{aligned}$$

And we know the stability conditions for single-variable equations. The first one is stable if $|\lambda_1| < 1$, and the second one is stable if $|\lambda_2| < 1$. So, if these two equations hold, we have

$$\lim_{t\to\infty} \bar{x}_t = \lim_{t\to\infty} \begin{pmatrix} w_t \\ x_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Finally, remember that we constructed $\bar{x}_t$ according to (see equation (8.5))

$$\bar{y}_t = V\bar{x}_t,$$

and so

$$\lim_{t\to\infty} \bar{y}_t = \lim_{t\to\infty} V\bar{x}_t = V \lim_{t\to\infty} \bar{x}_t = 0$$

and the original system is also stable.

That was a lot of steps, but it all boils down to something simple, and it works for more than two dimensions. *The dynamic system*

$$\bar{y}_{t+1} = A\bar{y}_t$$

is stable if all of the eigenvalues of A have magnitude smaller than one.

8.4 Problems

1. Consider the following IS-LM model:

$$\begin{aligned} Y &= C + I + G \\ C &= c((1-t)Y) \\ I &= i(R) \\ M &= P \cdot m(Y, R) \end{aligned}$$

with

$$\begin{aligned} c' &> 0 \\ i' &< 0 \\ m_Y &> 0, m_R < 0 \end{aligned}$$

The variables G, t, and M are exogenous policy variables. P is predetermined and assumed constant for the problem.

(a) Assume that $(1-t)c' < 1$, so that a \$1 increase in GDP leads to less than a dollar increase in spending. Compute and *interpret* dY/dt and dR/dt.

(b) Compute and *interpret* dY/dM and dR/dM.

(c) Compute and *interpret* dY/dP and dR/dP.

2. Consider a different IS-LM model, this time for an open economy, where X is net exports and T is total tax revenue (as opposed to t which was the marginal tax rate).

$$\begin{aligned} Y &= C + I + G + X \\ C &= c(Y - T) \\ I &= i(R) \\ X &= x(Y, R) \\ M &= P \cdot m(Y, R) \end{aligned}$$

with

$$\begin{aligned} c' &> 0 \\ i' &< 0 \\ x_Y &< 0, x_R < 0 \\ m_Y &> 0, m_R < 0 \end{aligned}$$

The variables G, T, and M are exogenous policy variables. P is predetermined and assumed constant for the problem.

(a) For this problem assume that $c' + x_Y < 1$, so that a \$1 increase in GDP leads to less than a dollar increase in spending. Compute and *interpret* dY/dG and dR/dG.

(b) Compute and *interpret* dY/dT and dR/dT.

(c) Compute and *interpret* dY/dM and dR/dM.

3. The following is a model of the long-run economy:

$$\begin{aligned} Y &= C + I + G + X \\ C &= c((1-t)Y) \\ I &= i(R) \\ X &= x(Y, R) \\ M &= P \cdot m(Y, R) \\ Y &= \bar{Y} \end{aligned}$$

with

$$\begin{aligned} c' &> 0 \\ i' &< 0 \\ x_Y &< 0, x_R < 0 \\ m_Y &> 0, m_R < 0 \end{aligned}$$

The variables G, t, and M are exogenous policy variables, and $\bar{Y}$ is also exogenous but not a policy variable. It is interpreted as potential GDP, or full-employment GDP. The variables Y, C, I, X, P are all endogenous.

(a) Compute and find the signs of dY/dG, dR/dG, and dP/dG.

(b) Compute and find the signs of dY/dM, dR/dM, and dP/dM.

4. Consider the following system of equations:

$$\begin{aligned} q_D &= D(p, I) \\ q_S &= S(p, w) \\ q_D &= q_S \end{aligned}$$

The first equation says that the quantity demanded in the market depends on the price of the good p and household income I. Consistent with this being a normal good, we have $D_p < 0$ and $D_I > 0$. The second equation says that the quantity supplied in the market depends on the price of the good and the wage rate of the work force, w. We have $S_p > 0$ and $S_w < 0$. The third equation says that markets must clear, so that quantity demanded equals quantity supplied. Three variables are endogenous: q_D, q_S, and p. Two variables are exogenous: I and w.

(a) Show that the market price increases when income increases.

(b) Show that the market price increases when the wage rate increases.

5. Find the coefficients for a regression based on the following data table:

Observation number	x_1	x_2	y
1	1	9	6
2	1	4	2
3	1	3	5

6. Consider a regression based on the following data table:

Observation number	x_1	x_2	y
1	2	8	1
2	6	24	0
3	-4	-16	-1

(a) Show that the matrix $X^T X$ is not invertible.

(b) Explain intuitively, using the idea of a column space, why there is no unique coefficient vector for this regression.

7. Find the coefficients for a regression based on the following data table:

Observation number	x_1	x_2	y
1	5	4	10
2	2	1	2
3	3	6	7

8. Suppose that you are faced with the following data table:

Observation number	x_1	x_2	y
1	1	2	10
2	1	3	0
3	1	5	25
4	1	4	15

You are thinking about adding one more explanatory variable, x_3, to the regression. x_3 is given by

$$x_3 = \begin{pmatrix} 24 \\ 36 \\ 60 \\ 48 \end{pmatrix}$$

Explain why this would be a bad idea.

9. Find the eigenvalues and eigenvectors of the following matrices:

   (a) $\begin{pmatrix} 5 & 1 \\ 4 & 2 \end{pmatrix}$

   (b) $\begin{pmatrix} 10 & -1 \\ 12 & 3 \end{pmatrix}$

   (c) $C = \begin{pmatrix} 3 & -6 \\ -2 & 4 \end{pmatrix}$

   (d) $D = \begin{pmatrix} 3 & 4 \\ 0 & 2 \end{pmatrix}$

   (e) $E = \begin{pmatrix} 10 & 4 \\ -3 & -6 \end{pmatrix}$

10. Consider the dynamic system

$$\bar{y}_{t+1} = A\bar{y}_t.$$

   (a) If $A = \begin{pmatrix} 1/3 & 0 \\ 0 & -1/5 \end{pmatrix}$, is the dynamic system stable?

   (b) If $A = \begin{pmatrix} 4/3 & -1/4 \\ 1/3 & 1/4 \end{pmatrix}$, is the dynamic system stable?

   (c) If $A = \begin{pmatrix} 1/4 & 1 \\ 0 & -2/3 \end{pmatrix}$, is the dynamic system stable?

   (d) If $A = \begin{pmatrix} 1/5 & 1 \\ 2 & 8/9 \end{pmatrix}$, is the dynamic system stable?

CHAPTER

9

Second-order conditions

9.1 Taylor approximations for $\mathbb{R} \to \mathbb{R}$

Consider a differentiable function $f : \mathbb{R} \to \mathbb{R}$. Since f is differentiable at point x_0 it has a linear approximation at that point. Note that this is a linear approximation to the function, as opposed to a linear approximation at the point. The linear approximation can be written

$$L(x) = f(x_0) + f'(x_0)(x - x_0).$$

When $x = x_0$ we get the original point, so $L(x_0) = f(x_0)$. When x is not equal to x_0 we multiply the difference $(x - x_0)$ by the slope at point x_0, then add that amount to $f(x_0)$.

All of this is shown in Figure 9.1. The linear approximation $L(x)$ is a straight line that is tangent to the function $f(x)$ at $f(x_0)$. To get the approximation at point x, take the horizontal distance $(x - x_0)$ and multiply it by the slope of the tangent line, which is just $f'(x_0)$. This gives us the quantity $f'(x_0)(x - x_0)$, which must be added to $f(x_0)$ to get the right linear approximation.

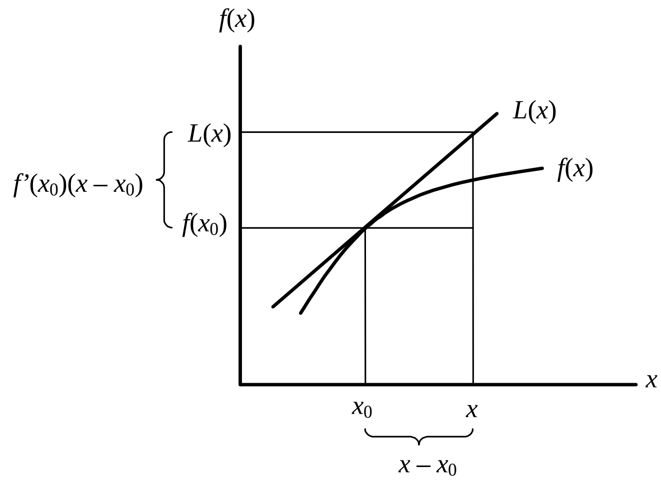


Figure 9.1: A first-order Taylor approximation

The **Taylor approximation** builds on this intuition. Suppose that f is n times differentiable. Then an n-th order approximation is

$$f(x) \approx f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + ... + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n.$$

We care mostly about second-degree approximations, or

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2.$$

The key to understanding the numbers in the denominators of the different terms is noticing that the first derivative of the linear approximation matches the first derivative of the function, the second derivative of the second-order approximation equals the second derivative of the original function, and so on. So, for example, we can write the third-degree approximation of $f(x)$ at $x = x_0$ as

$$g(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \frac{f'''(x_0)}{6}(x - x_0)^3.$$

Differentiate with respect to x to get

$$\begin{aligned} g'(x) &= f'(x_0) + f''(x_0)(x - x_0) + \frac{f'''(x_0)}{2}(x - x_0)^2 \\ g''(x) &= f''(x_0) + f'''(x_0)(x - x_0) \\ g'''(x) &= f'''(x_0) \end{aligned}$$

9.2 Second order conditions for $\mathbb{R} \to \mathbb{R}$

Suppose that $f(x)$ is maximized when $x = x_0$. Take a second-order Taylor approximation:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2.$$

Since f is maximized when $x = x_0$, the first-order condition has $f'(x_0) = 0$. Thus the second term in the Taylor approximation disappears. We are left with

$$f(x) \approx f(x_0) + \frac{f''(x_0)}{2}(x - x_0)^2.$$

If f is maximized, it must mean that any departure from x_0 leads to a decrease in f. In other words,

$$f(x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 \leq f(x_0)$$

for all x. Simplifying gives us

$$\frac{f''(x_0)}{2}(x - x_0)^2 \leq 0$$

and, since $(x - x_0)^2 \geq 0$, it must be the case that $f''(x_0) \leq 0$. This is how we can get the second order condition from the Taylor approximation.

9.3 Taylor approximations for $\mathbb{R}^m \to \mathbb{R}$

This time we are only going to look for a second-degree approximation. We need some notation:

$$\nabla f(\bar{x}) = \begin{pmatrix} f_1(\bar{x}) \\ \vdots \\ f_m(\bar{x}) \end{pmatrix}$$

and is called the **gradient** of f at x. Also

$$H(\bar{x}) = \begin{pmatrix} f_{11}(\bar{x}) & \cdots & f_{1m}(\bar{x}) \\ \vdots & \ddots & \vdots \\ f_{m1}(\bar{x}) & \cdots & f_{mm}(\bar{x}) \end{pmatrix}$$

is called the **Hessian**. It is the matrix of second derivatives.

A second-order Taylor approximation for a function of m variables can be written

$$f(\bar{x}) \approx f(\bar{x}^0) + \sum_{i=1}^{m} f_i(\bar{x}^0)(x_i - x_i^0) + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}(\bar{x}^0)(x_i - x_i^0)(x_j - x_j^0)$$

Let's write this in matrix notation. The first term is simply $f(\bar{x}^0)$. The second term is

$$(\bar{x} - \bar{x}^0) \cdot \nabla f(\bar{x}^0) = (\bar{x} - \bar{x}^0)^T \nabla f(\bar{x}^0).$$

The third term is

$$\frac{1}{2}(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)(\bar{x} - \bar{x}^0).$$

Let's check to make sure this last one works. First check the dimensions, which are $(1 \times m)(m \times m)(m \times 1) = (1 \times 1)$, which is what we want. Then break the problem down. $(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)$ is a $(1 \times m)$ matrix with element j given by

$$\sum_{i=1}^{m} f_{ij}(\bar{x}^0)(x_i - x_i^0).$$

To get $(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)(\bar{x} - \bar{x}^0)$ we multiply each element of $(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)$ by the corresponding element of $(\bar{x} - \bar{x}^0)$ and sum, to get

$$\sum_{i=1}^{m}\sum_{j=1}^{m} f_{ij}(\bar{x}^0)(x_i - x_i^0)(x_j - x_j^0).$$

So, the second degree Taylor approximation of the function $f : \mathbb{R}^m \to \mathbb{R}$ is given by

$$f(\bar{x}) \approx f(\bar{x}^0) + (\bar{x} - \bar{x}^0)^T \nabla f(\bar{x}^0) + \frac{1}{2}(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)(\bar{x} - \bar{x}^0)$$

9.4 Second order conditions for $\mathbb{R}^m \to \mathbb{R}$

Suppose that $f : \mathbb{R}^m \to \mathbb{R}$ is maximized when $\bar{x} = \bar{x}^0$. Then the first-order condition is

$$\nabla f(\bar{x}^0) = \bar{0}$$

and the second term in the Taylor approximation drops out. For f to be maximized when $\bar{x} = \bar{x}^0$ it must be the case that

$$f(\bar{x}^0) + \frac{1}{2}(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)(\bar{x} - \bar{x}^0) \leq f(\bar{x}^0)$$

or

$$(\bar{x} - \bar{x}^0)^T H(\bar{x}^0)(\bar{x} - \bar{x}^0) \leq 0.$$

9.5 Negative semidefinite matrices

The matrix A is **negative semidefinite** if, for every column matrix x, we have

$$x^T A x \leq 0.$$

Obviously, the second order condition for a maximum is that $H(\bar{x}^0)$ is negative semidefinite. In the form it is written in, though, it is a difficult thing to check.

Form a submatrix A_i from the square matrix A by keeping the square matrix formed by the first i rows and first i columns. (Note that this is different from the submatrix we used to find determinants in Section 6.3.) The determinant of A_i is called the **i-th leading principal minor** of A.

Theorem 11 *Let A be a symmetric $m \times m$ matrix. Then A is negative semidefinite if and only if its m leading principle minors alternate in sign so that*

$$(-1)^i |A_i| \geq 0$$

for $i = 1, ..., m$.

There are other corresponding notions:

- A is **negative definite** if $x^T A x < 0$ for all nonzero vectors x, and this occurs if and only if its m leading principle minors alternate in sign so that

  $$(-1)^i |A_i| > 0$$

  for $i = 1, ..., m$.

- A is **positive definite** if $x^T Ax > 0$ for all nonzero vectors x, and this occurs if and only if its m leading principle minors are positive, so that

$$|A_i| > 0$$

for $i = 1, ..., m$.

- A is **positive semidefinite** if $x^T Ax \geq 0$ for all vectors x, and this occurs if and only if its m leading principle minors are nonnegative, so that

$$|A_i| \geq 0$$

for $i = 1, ..., m$.

- A is **indefinite** if none of the other conditions holds.

9.5.1 Application to second-order conditions

Suppose that the problem is to choose x_1 and x_2 to maximize $f(x_1, x_2)$. The FOCs are

$$\begin{aligned} f_1 &= 0 \\ f_2 &= 0 \end{aligned}$$

and the SOC is

$$H = \begin{pmatrix} f_{11} & f_{21} \\ f_{21} & f_{22} \end{pmatrix}$$

is negative semidefinite. Note that f_{21} appears in both off-diagonal elements, which is okay because $f_{12} = f_{21}$. That is, it doesn't matter if you differentiate f first with respect to x_1 and then with respect to x_2 or the other way around. The requirements for H to be negative semidefinite are

$$\begin{aligned} f_{11} &\leq 0 \\ \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} &= f_{11}f_{22} - f_{12}^2 \geq 0 \end{aligned}$$

Note that the two conditions together imply that $f_{22} \leq 0$.

9.5.2 Examples

$A = \begin{pmatrix} -3 & 3 \\ 3 & -4 \end{pmatrix}$ is negative definite because $a_{11} < 0$ and $a_{11}a_{22} - a_{12}a_{21} = 3 > 0$.

$A = \begin{pmatrix} 6 & 1 \\ 1 & 3 \end{pmatrix}$ is positive definite because $a_{11} > 0$ and $a_{11}a_{22} - a_{12}a_{21} = 17 > 0$.

$A = \begin{pmatrix} -5 & -3 \\ -3 & 4 \end{pmatrix}$ is indefinite because $a_{11} < 0$ and $a_{11}a_{22} - a_{12}a_{21} = -29 < 0$.

9.6 Concave and convex functions

All of the second-order conditions considered so far rely on the objective function being twice differentiable. In the single-variable case we require that the second derivative is nonpositive for a maximum and nonnegative for a minimum. In the many-variable case we require that the matrix of second and cross partials (the Hessian) is negative semidefinite for a maximum and positive semidefinite for a minimum. But objective functions are not always differentiable, and we would like to have some second order conditions that work for these cases, too.

Figure 9.2 shows a function with a maximum. It also has the following property. If you choose any two points on the curve, such as points a and b, and draw the line segment connecting them, that line segment always lies below the curve. When this property holds for every pair of points on the curve, we say that the function is **concave**.

It is also possible to characterize a concave function mathematically. Point a in the figure has coordinates $(x_a, f(x_a))$, and point b has coordinates $(x_b, f(x_b))$. Any value x between x_a and x_b can be written as

$$x = tx_a + (1-t)x_b$$

for some value $t \in [0,1]$. Such a point is called **convex combination** of x_a and x_b. When $t = 1$ we get $x = 1 \cdot x_a + 0 \cdot x_b = x_a$, and when $t = 0$ we get $x = 0 \cdot x_a + 1 \cdot x_b = x_b$. When $t = \frac{1}{2}$ we get $x = \frac{1}{2}x_a + \frac{1}{2}x_b$ which is the midpoint between x_a and x_b, as shown in Figure 9.2.

The points on the line segment connecting a and b in the figure have

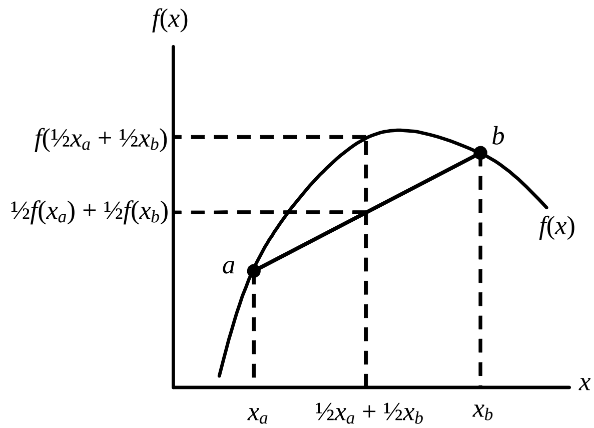


Figure 9.2: A concave function

coordinates

$$(tx_a + (1-t)x_b, tf(x_a) + (1-t)f(x_b))$$

for $t \in [0,1]$. Let's choose one value of t, say $t = \frac{1}{2}$. The point on the line segment is

$$\left(\frac{1}{2}x_a + \frac{1}{2}x_b, \frac{1}{2}f(x_a) + \frac{1}{2}f(x_b)\right)$$

and it is the midpoint between points a and b. But $\frac{1}{2}x_a + \frac{1}{2}x_b$ is just a value of x, and we can evaluate $f(x)$ at $x = \frac{1}{2}x_a + \frac{1}{2}x_b$. According to the figure,

$$f(\frac{1}{2}x_a + \frac{1}{2}x_b) \geq \frac{1}{2}f(x_a) + \frac{1}{2}f(x_b)$$

where the left-hand side is the height of the curve and the right-hand side is the height of the line segment connecting a to b.

Concavity says that this is true for all possible values of t, not just $t = \frac{1}{2}$.

Definition 2 *A function $f(x)$ is **concave** if, for all x_a and x_b,*

$$f(tx_a + (1-t)x_b) \geq tf(x_a) + (1-t)f(x_b)$$

for all $t \in [0,1]$.

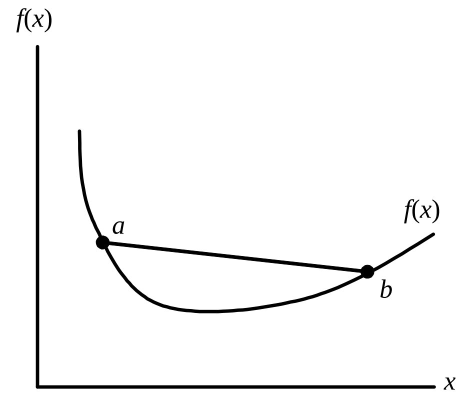


Figure 9.3: A convex function

Concave functions tend to be upward sloping, downward sloping, or have a maximum (that is, upward sloping and then downward sloping). Thus, instead of assuming that a function has the right properties of its second derivative, we can instead assume that it is concave. And notice that nothing in the definition says anything about whether x is single-dimensional or multidimensional. The same definition of concave works for both single-variable and multi-variable optimization problems.

If a concave function is twice differentiable, it has a nonpositive second derivative.

A **convex** function has the opposite property: the line segment connecting any two points on the curve lies above the curve, as in Figure 9.3. We get a corresponding definition.

Definition 3 *A function $f(x)$ is **convex** if, for all x_a and x_b,*

$$f(tx_a + (1-t)x_b) \leq tf(x_a) + (1-t)f(x_b)$$

for all $t \in [0, 1]$.

Convexity is the appropriate assumption for minimization, and if a convex function is twice differentiable its second derivative is nonnegative.

As an example, consider the standard profit-maximization problem, where output is q and profit is given by

$$\pi(q) = r(q) - c(q),$$

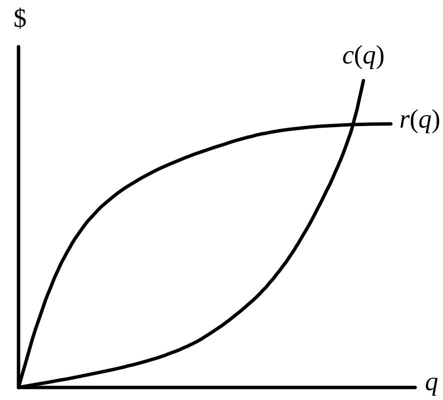


Figure 9.4: Profit maximization with a concave revenue function and a convex cost function

where $r(q)$ is the revenue function and $c(q)$ is the cost function. The standard assumptions are that the revenue function is concave and the cost function is convex. If both functions are twice differentiable, the first-order condition is the familiar

$$r'(q) - c'(q) = 0$$

and the second-order condition is

$$r''(q) - c''(q) \leq 0.$$

This last expression holds if $r''(q) \leq 0$ which occurs when $r(q)$ is concave, and if $c''(q) \geq 0$ which occurs when $c(q)$ is convex. Figure 9.4 shows the standard revenue and cost functions in a profit maximization problem, and in the graph the revenue function is concave and the cost function is convex.

Some functions are neither convex nor concave. More precisely, they have some convex portions and some concave portions. Figure 9.5 provides an example. The function is convex to the left of x_0 and concave to the right of x_0.

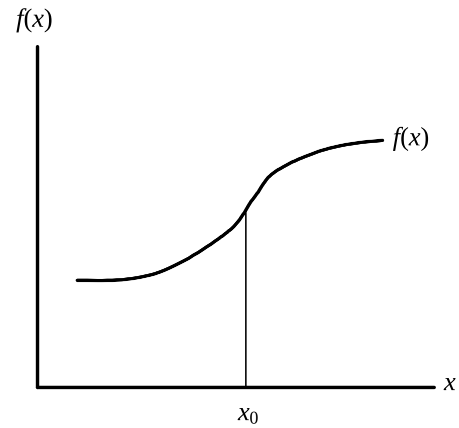


Figure 9.5: A function that is neither concave nor convex

9.7 Quasiconcave and quasiconvex functions

The important property for a function with a maximum, such as the one shown in Figure 9.2, is that it rise and then fall. But the function depicted in Figure 9.6 also rises then falls, and clearly has a maximum. But it is not concave. Instead it is *quasiconcave*, which is the weakest second-order condition we can use. The purpose of this section is to define the terms "quasiconcave" and "quasiconvex," which will take some work.

Before we can define them, we need to define a different term. A *set* S is **convex** if, for any two points $x, y \in S$, the point $\alpha x + (1 - \alpha)y \in S$ for all $\alpha \in [0, 1]$. Let's break this into pieces using Figure 9.7. In the left-hand graph, the set S is the interior of the oval, and choose two points x and y in S. These can be either in the interior of S or on its boundary, but the ones depicted are in the interior. The set $\{z | z = \alpha x + (1 - \alpha)y$ for some $\alpha \in [0, 1]\}$ is just the line segment connecting x to y, as shown in the figure. The set S is convex if the line segment is inside of S, no matter which x and y we choose. Or, using different terminology, the set S is convex if any convex combination of two points in S is also in S.

In contrast, the set S in the right-hand graph in Figure 9.7 is not convex. Even though points x and y are inside of S, the line segment connecting them passes outside of S. In this case the set is **nonconvex** (there is no such thing as a concave set).

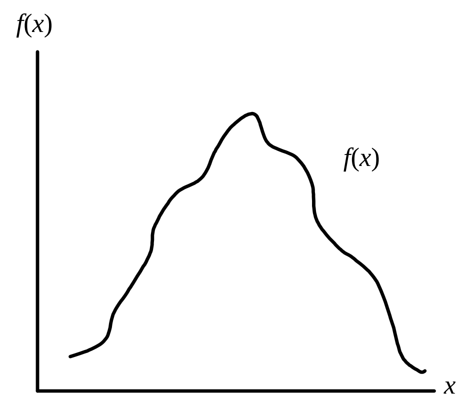


Figure 9.6: A quasiconcave function

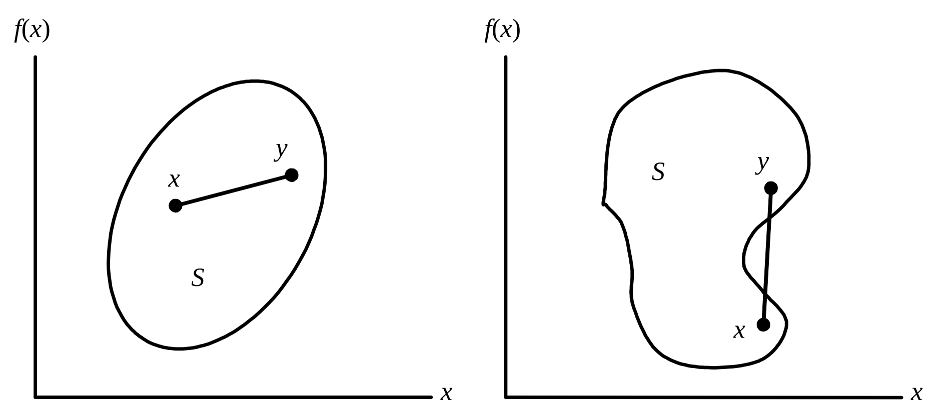


Figure 9.7: Convex sets: The set in the left-hand graph is convex because all line segments connecting two points in the set also lie completely within the set. The set in the right-hand graph is not convex because the line segment drawn does not lie completely within the set.

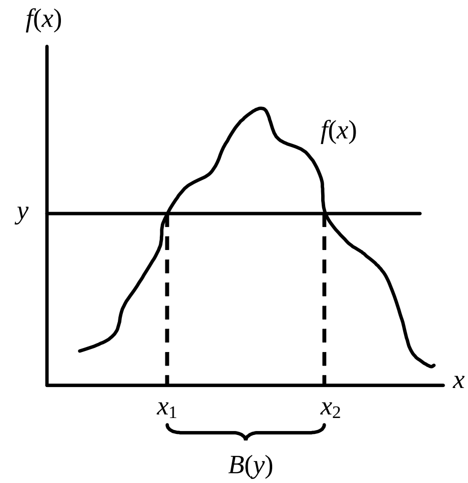


Figure 9.8: Defining quasiconcavity: For any value y, the set $B(y)$ is convex

The graphs in Figure 9.7 are for 2-dimensional sets. The definition of *convex*, though, works in any number of dimensions. In particular, it works for 1-dimensional sets. A 1-dimensional set is a subset of the real line, and it is convex if it is an interval, either open, closed, or half-open/half-closed.

Now look at Figure 9.8, which has the same function as in Figure 9.6. Choose any value y, and look for the set

$$B(y) = \{x|f(x) \geq y\}.$$

This is a *better-than set*, and it contains the values of x that generate a value of $f(x)$ that is at least as high as y. In Figure 9.8 the set $B(y)$ is the closed interval $[x_1, x_2]$, which is a convex set. This gives us our definition of a quasiconcave function.

Definition 4 *The function $f(x)$ is **quasiconcave** if for any value y, the set $B(y) = \{x|f(x) \geq y\}$ is convex.*

To see why quasiconcavity is both important and useful, flip way back to the beginning of the book to look at Figure 1.1 on page 2. That graph depicted either a consumer choice problem, in which case the line is a budget line and the curve is an indifference curve, or it depicted a firm's cost-minimization problem, in which case the line is an isocost line and the curve

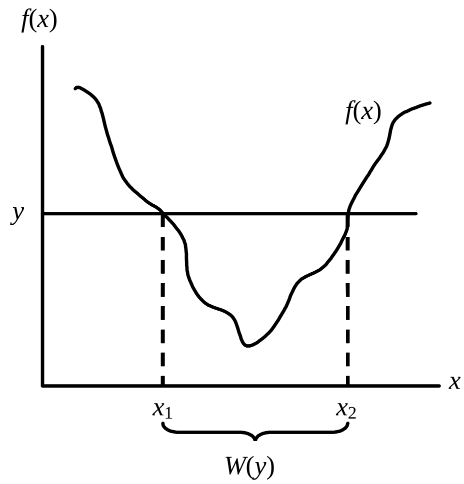


Figure 9.9: A quasiconvex function

is an isoquant. Think about it as a consumer choice problem. The area above the indifference curve is the set of points the consumer prefers to those on the indifference curve. So the set of points above the indifference curve is the better-than set. And it's convex. So the appropriate second-order condition for utility maximization problems is that the utility function is quasiconcave. Similarly, an appropriate second-order condition for cost-minimization is that the production function (the function that gives you the isoquant) is quasiconcave. When you take microeconomics, see where quasiconcavity shows up as an assumption.

Functions can also be quasiconvex.

Definition 5 *The function $f(x)$ is* ***quasiconvex*** *if for any value y, the set $W(y) = \{x|f(x) \leq y\}$ is convex.*

Quasiconvex functions are based on *worse-than sets* $W(y)$. This time the set of points generating values lower than y must be convex. To see why, look at Figure 9.9. This time the points that generate values of $f(x)$ lower than y form an interval, but the better-than set is not an interval.

Concave functions are also quasiconcave, which you can see by looking at Figure 9.2, and convex functions are also quasiconvex, which you can see by looking at Figure 9.3. But a quasiconcave function may not be concave, as in Figure 9.8, and a quasiconvex function may not be convex, as in Figure 9.9.

The easiest way to remember the definition for quasiconcave is to draw a concave function with a maximum. We know that it is also quasiconcave. Choose a value of y and draw the horizontal line, like we did in Figure 9.8. Which set is convex, the better-than set or the worse-than set? As shown in the figure, it's the better-than set that is convex, so we get the right definition. If you draw a convex function with a minimum and follow the same steps you can figure out the right definition for a quasiconvex function.

9.8 Problems

1. Find the gradient of

$$f(x_1, x_2, x_3) = 2x_1x_3^2 + 3x_1x_2^2 - 4x_1^2$$

   and then evaluate it at the point $(5, 2, 0)$.

2. Find the second-degree Taylor approximations of the following functions at $x_0 = 1$:

   (a) $f(x) = -2x^3 - 5x + 9$

   (b) $f(x) = 10x - 40\sqrt{x} + \ln x$

   (c) $f(x) = e^x$

3. Find the second-degree Taylor approximation of the function $f(x) = 3x^3 - 4x^2 - 2x + 12$ at $x_0 = 0$.

4. Find the second-degree Taylor approximation of the function $f(x) = ax^2 + bx + c$ at $x_0 = 0$.

5. Tell whether the following matrices are negative definite, negative semidefinite, positive semidefinite, positive definite, or indefinite.

   (a) $\begin{pmatrix} -3 & 2 \\ 2 & 1 \end{pmatrix}$

   (b) $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$

(c) $\begin{pmatrix} 3 & 4 \\ 4 & -3 \end{pmatrix}$

(d) $\begin{pmatrix} 4 & 0 & 1 \\ 0 & -3 & -2 \\ 1 & -2 & 1 \end{pmatrix}$

(e) $\begin{pmatrix} 6 & 1 \\ 1 & 3 \end{pmatrix}$

(f) $\begin{pmatrix} -4 & 16 \\ 16 & -4 \end{pmatrix}$

(g) $\begin{pmatrix} -2 & 1 \\ 1 & -4 \end{pmatrix}$

(h) $\begin{pmatrix} 3 & 2 & 3 \\ 2 & 4 & 0 \\ 3 & 0 & -1 \end{pmatrix}$

6. State whether the second-order condition is satsified for the following problems.

   (a) $\min_{x,y} 4y^2 - xy$

   (b) $\max_{x,y} 7 + 8x + 6y - x^2 - y^2$

   (c) $\max_{x,y} 5xy - 2y^2$

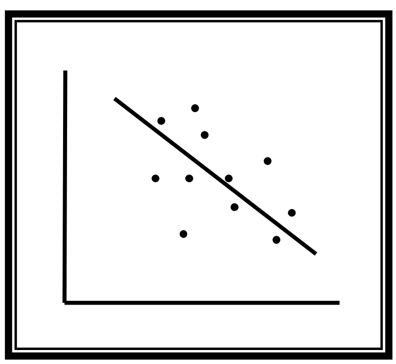
   (d) $\min_{x,y} 6x^2 + 3y^2$

7. Is $(6, 2)$ a convex combination of $(11, 4)$ and $(-1, 0)$? Explain.

8. Use the formula for convexity, and not the second derivative, to show that the function $f(x) = x^2$ is convex.

PART III

ECONOMETRICS



(probability and statistics)

CHAPTER

10

Probability

10.1 Some definitions

An **experiment** is an activity that involves doing something or observing something resulting in an **outcome**. The performance of an experiment is a **trial**. Experiments can be physical, biological, social, or anything else.

The **sample space** for an experiment is the set of possible outcomes of the experiment. The sample space is denoted Ω (the Greek letter omega) and a typical element, or outcome, is denoted ω (lower case omega). The impossible event is the empty set, $\varnothing$. Suppose, for example, that the experiment consists of tossing a coin twice. The sample space is

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$

and each pair is an outcome. Another experiment is an exam. Scores are out of 100, so the sample space is

$$\Omega = \{0,1,...,100\}.$$

We are going to work with subsets, and we are going to work with some mathematical symbols pertaining to subsets. The notation is given in the following table.

In math	In English
$\omega \in A$	omega is an element of A
$A \cap B$	A intersection B
$A \cup B$	A union B
$A \subset B$	A is a strict subset of B
$A \subseteq B$	A is a weak subset of B
A^C	The complement of A

An **event** is a subset of the sample space. If the experiment consists of tossing a coin twice, the event that there is at least one head can be written

$$A = \{(H, H), (H, T), (T, H)\}.$$

Note that the entire sample space is an event, and so is the impossible event $\varnothing$. Sometimes we want to talk about single-element events, and write ω instead of $\{\omega\}$.

It is best to think of outcomes and events as *occurring*. For the latter, an event A occurs if there is some outcome $\omega \in A$ such that ω occurs.

Our eventual goal is to assign probabilities to events. To do this we need notation for the set of all possible events. Call it Σ (for sigma-algebra, which is a concept we will not get into).

Two events A and B are **mutually exclusive** if there is no outcome that is in both events, that is, $A \cap B = \varnothing$. If A and B are mutually exclusive then if event A occurs event B is impossible, and vice versa.

10.2 Defining probability abstractly

A **probability measure** is a mapping $P : \Sigma \to [0, 1]$, that is, a function mapping events into numbers between 0 and 1. The function P has three key properties:

Axiom 1 $P(A) \geq 0$ for any event $A \in \Sigma$.

Axiom 2 $P(\Omega) = 1$

Axiom 3 If $A_1, A_2, ...$ is a (possibly finite) sequence of mutually exclusive events, then

$$P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$$

The first axiom states that probabilities cannot be negative. The second one states that the probability that *something* happens is one. The third axiom states that when events are mutually exclusive, the probability of the union is simply the sum of the probabilities.

These three axioms imply some other properties.

Theorem 12 $P(\varnothing) = 0$.

Proof. The events Ω and $\varnothing$ are mutually exclusive since $\Omega \cap \varnothing = \varnothing$. Since $\Omega \cup \varnothing = \Omega$, axiom 3 implies that

$$1 = P(\Omega \cup \varnothing) = P(\Omega) + P(\varnothing) = 1 + P(\varnothing),$$

and it follows that $P(\varnothing) = 0$. ∎

The next result concerns the relation $\subseteq$, where $A \subseteq B$ means that either A is contained in B or A is equal to B.

Theorem 13 *If $A \subseteq B$ then $P(A) \leq P(B)$.*

Proof. Suppose that A and B are events with $A \subseteq B$. Define $C = \{\omega : \omega \in B \text{ but } \omega \notin A\}$. Then A and C are mutually exclusive with $A \cup C = B$, and axiom 3 implies that

$$P(B) = P(A \cup C) = P(A) + P(C) \geq P(A).$$

∎

For the next theorems, let A^C denote the complement of the event A, that is, $A^C = \{\omega \in \Omega : \omega \notin A\}$.

Theorem 14 $P(A^C) = 1 - P(A)$.

Proof. Note that $A^C \cup A = \Omega$ and $A^C \cap A = \varnothing$. Then $P(A^C \cup A) = P(A^C) + P(A) = 1$, and therefore $P(A^C) = 1 - P(A)$. ■

Note that this theorem implies that $P(A) \leq 1$ for any event A. To see why, first write $P(A) = 1 - P(A^C)$, and by axiom 1 we have $P(A^C) \geq 0$. The result follows.

Theorem 15 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. First note that $A \cup B = A \cup (A^C \cap B)$. You can see this in Figure 10.1. The points in $A \cup B$ are either in A or they are outside of A but in B. The events A and $A^C \cap B$ are mutually exclusive. Axiom 3 says

$$P(A \cup B) = P(A) + P(A^C \cap B).$$

Next note that $B = (A \cap B) \cup (A^C \cap B)$. Once again this is clear in Figure 10.1, but it says that we can separate B into two parts, the part that intersects A and the part that does not. These two parts are mutually exclusive, so

$$P(B) = P(A \cap B) + P(A^C \cap B). \tag{10.1}$$

Rearranging yields

$$P(A^C \cap B) = P(B) - P(A \cap B).$$

Substituting this into equation (10.1) yields

$$\begin{aligned} P(A \cup B) &= P(A) + P(A^C \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

■

10.3 Defining probabilities concretely

The previous section told us what the abstract concept *probability measure* means. Sometimes we want to know actual probabilities. How do we get them? The answer relies on the ability to partition the sample space into equally likely outcomes.

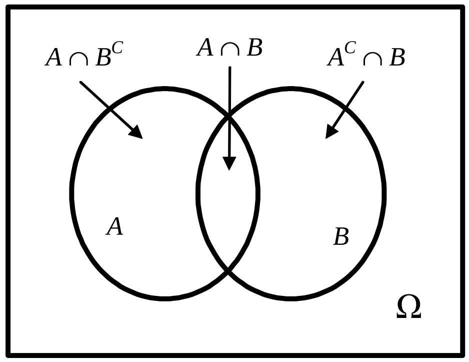


Figure 10.1: Finding the probability of the union of two sets

Theorem 16 *Suppose that every outcome in the sample space* $\Omega = \{\omega_1, ..., \omega_n\}$ *is equally likely. Then the probability of any event* A *is the number of outcomes in* A *divided by* n*.*

Proof. We know that $P(\Omega) = P(\{\omega_1\} \cup \cdots \cup \{\omega_n\}) = 1$. By construction $\{\omega_i\} \cap \{\omega_j\} = \varnothing$ when $i \neq j$, and so the events $\{\omega_1\}, ..., \{\omega_n\}$ are mutually exclusive. By Axiom 3 we have

$$P(\Omega) = P(\{\omega_1\}) + ... + P(\{\omega_n)\} = 1.$$

Since each of the outcomes is equally likely, this implies that

$$P(\{\omega_i\}) = 1/n$$

for $i = 1, ..., n$. If event A contains k of the outcomes in the set $\{\omega_1, ..., \omega_n\}$, it follows that $P(A) = k/n$. ■

This theorem allows us to compute probabilities from experiments like flipping coins, rolling dice, and so on. For example, if a die has six sides, the probability of the outcome 5 is $1/6$. The probability of the event $\{1, 2\}$ is $1/6 + 1/6 = 1/3$, and so on.

For an exercise, find the probability of getting exactly one head in four tosses of a fair coin. Answer: There are 16 possible outcomes. Four of them have one head. So, the probability is $1/4$.

In general events are not equally likely, so we cannot determine probabilities theoretically in this manner. Instead, the probabilities of the events are given directly.

10.4 Conditional probability

Suppose that $P(B) > 0$, so that the event B occurs with positive probability. Then $P(A|B)$ is the **conditional probability** that event A occurs given that event B has occurred. It is given by the formula

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

Think about what this expression means. The numerator is the probability that *both* A and B occur. The denominator is the probability that B occurs. Clearly $A \cap B$ is a subset of B, so the numerator is smaller than the denominator, as required. The ratio can be interpreted as the fraction of the time when B occurs that A also occurs.

Consider the probability distribution given in the table below. Outcomes are two dimensional, based on the values of x and y. The probabilities of the different outcomes are given in the entries of the table. Note that all the entries are nonnegative and that the sum of all the entries is one, as required for a probability measure.

	$y = 1$	$y = 2$	$y = 3$	$y = 4$
$x = 1$	0.02	0.01	0.02	0.10
$x = 2$	0.05	0.00	0.03	0.11
$x = 3$	0.04	0.15	0.02	0.09
$x = 4$	0.10	0.16	0.02	0.08

Find the conditional probability $P(x = 2|y = 4)$. The formula is $P(x = 2$ and $y = 4)/P(y = 4)$. The probability that $x = 2$ and $y = 4$ is just the entry in a single cell, and is 0.11. The probability that $y = 4$ is the sum of the probabilities in the last column, or 0.38. So, $P(x = 2|y = 4) = 0.11/0.38 = 11/38$.

Now find the conditional probability that y is either 1 or 2 given that $x \geq 3$. The probability that $y \leq 2$ and $x \geq 3$ is the sum of the four cells in the lower left, or 0.45. The probability that $x \geq 3$ is 0.66. So, the conditional probability is $45/66 = 15/22$.

Now look at a medical example. A patient can have condition A or not. He takes a test which turns out positive or not. The probabilities are given in the following table:

	Test positive	Test negative
Condition A	0.010	0.002
Healthy	0.001	0.987

Note that condition A is quite rare, with only 12 people in 1000 having it. Also, a positive test is ten times more likely to come from a patient with the condition than from a patient without the condition. We get the following conditional probabilities:

$$\begin{aligned} P(A|\text{positive}) &= 10/11 = 0.909, \\ P(\text{healthy}|\text{negative}) &= 987/989 = 0.998, \\ P(\text{positive}|A) &= 10/12 = 0.833, \\ P(\text{negative}|\text{healthy}) &= 987/988 = 0.999. \end{aligned}$$

10.5 Bayes' rule

Theorem 17 (Bayes' rule) *Assume that $P(B) > 0$. Then*

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

Proof. Note that

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

and

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

Rearranging the second one yields

$$P(A \cap B) = P(B|A)P(A)$$

and the theorem follows from substitution. ■

Let's make sure Bayes' rule works for the medical example given above. We have $P(\text{positive}|A) = 10/12$, $P(\text{positive}) = 11/1000$, and $P(A) = 12/1000$. Using Bayes' rule we get

$$P(A|\text{positive}) = \frac{P(\text{positive}|A) \cdot P(A)}{P(\text{positive})} = \frac{\frac{10}{12} \cdot \frac{12}{1000}}{\frac{11}{1000}} = \frac{10}{11}.$$

The interpretation of Bayes' rule is for responding to updated information. We are interested in the occurrence of event A after we receive some new information. We start with the **prior** $P(A)$. Then we find out that B holds. We should use this new information to update the probability of A occurring. $P(A|B)$ is called the **posterior** probability. Bayes' rule tells us how to do this. We multiply the prior $P(A)$ by the **likelihood**

$$\frac{P(B|A)}{P(B)}.$$

If this ratio is greater than one, the posterior probability is higher than the prior probability. If the ratio is smaller than one, the posterior probability is lower. The ratio is greater than one if A occurring makes B more likely.

Bayes' rule is important in game theory, finance, and macro.

People don't seem to follow it. Here is a famous example (Kahneman and Tversky, 1973 *Psychological Review*). Some subjects are told that a group consists of 70 lawyers and 30 engineers. The rest of the subjects are told that the group has 30 lawyers and 70 engineers. All subjects were then given the following description:

> Dick is a 30 year old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

Subjects were then asked to judge the probability that Dick is an engineer. Subjects in both groups said that it is about 0.5, ignoring the prior information. The new information is uninformative, so $P(B|A)/P(B) = 1$, and according to Bayes' rule the posterior should be the same as the prior.

This example has people overweighting the new information. Psychologists have also come up with studies in which subjects overweight the prior. When subjects overweight the new information it is called *representativeness*, and when they overweight the prior it is called *conservatism.*

10.6 Monty Hall problem

At the end of the game show *Let's Make a Deal* the host, Monty Hall, offers a contestant the choice among three doors, labeled A, B, and C. There is a prize behind one of the doors, and nothing behind the other two. After the contestant chooses a door, to build suspense Monty Hall reveals one of the doors with no prize. He then asks the contestant if she would like to stay with her original door or switch to the other one. What should she do?

The answer is that she should take the other door. To see why, suppose she chooses door A, and that Monty reveals door B. What is the probability that the prize is behind door C given that door B was revealed? Bayes' rule says we use the formula

$$P(\text{prize behind } C|\text{ reveal } B) = \frac{P(\text{reveal } B|\text{ prize } C) \cdot P(\text{prize } C)}{P(\text{reveal } B)}.$$

Before revealing the door, the prize was equally likely to be behind each of the three doors, so $P(\text{prize } A) = P(\text{prize } B) = P(\text{prize } C) = 1/3$. Next find the conditional probability that Monty reveals door B given that the prize is behind door C. Remember that Monty cannot reveal the door with the prize behind it or the door chosen by the contestant. Therefore Monty *must* reveal door B if the prize is behind door C, and the conditional probability $P(\text{reveal } B|\text{ prize } C) = 1$. The remaining piece of the formula is the probability that he reveals B. We can write this as

$$\begin{aligned} P(\text{reveal } B) &= P(\text{reveal } B|\text{ prize } A) \cdot P(A) \\ &\quad + P(\text{reveal } B|\text{ prize } B) \cdot P(B) \\ &\quad + P(\text{reveal } B|\text{ prize } C) \cdot P(C). \end{aligned}$$

The middle term is zero because he cannot reveal the door with the prize behind it. The last term is $1/3$ for the reasons given above. If the prize is behind A he can reveal either B or C and, assuming he does so randomly, the conditional probability $P(\text{reveal } B|\text{ prize } A) = 1/2$. Consequently the first term is $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$. Using all this information, the probability of revealing door B is $P(\text{reveal } B) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$. Plugging this into Bayes' rule yields

$$P(\text{prize } C|\text{ reveal } B) = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

The probability that the prize is behind A given that he revealed door B is $1 - P(\text{prize } C|\text{ reveal } B) = 1/3$. The contestant should switch.

10.7 Statistical independence

Two events A and B are **independent** if and only if $P(A \cap B) = P(A) \cdot P(B)$.

Consider the following table relating accidents to drinking and driving.

	Accident	No accident
Drunk driver	0.03	0.10
Sober driver	0.03	0.84

Notice that from this table that half of the accidents come from sober drivers, but there are many more sober drivers than drunk ones. The question is whether accidents are independent of drunkenness. Compute P(drunk $\cap$ accident) $= 0.03$, P(drunk) $= 0.13$, P(accident) $= 0.06$, and finally

$$P(drunk) \cdot P(accident) = 0.13 \cdot 0.06 = 0.0078 \neq 0.03.$$

So, accidents and drunk driving are not independent events. This is not surprising, as we would expect drunkenness to be a contributing factor to accidents. Note that P(accident|drunk) $= 3/13 = 0.23$ while P(accident|sober) $= 3/87 = 0.03\,4$.

We can prove an easy theorem about independent events.

Theorem 18 *If A and B are independent then $P(A|B) = P(A)$.*

Proof. We have

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

■

According to this theorem, if accidents and drunkenness were independent events, then P(accident|drunk) $=$ P(accident); that is, the probability of getting in an accident when drunk is the same as the overall probability of getting in an accident.

10.8 Problems

1. Answer the questions using the table below.

	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$
$x = 5$	0.01	0.03	0.17	0.00	0.00
$x = 20$	0.03	0.05	0.04	0.20	0.12
$x = 30$	0.11	0.04	0.02	0.07	0.11

   (a) What is the most likely outcome?

   (b) What outcomes are impossible?

   (c) Find the probability that $x = 30$.

   (d) Find the probability that $x \in \{5, 20\}$ and $2 \le y \le 4$.

   (e) Find the probability that $y \le 2$ conditional on $x \ge 20$.

   (f) Verify Bayes' rule for $P(y = 4|x = 20)$.

   (g) Are the events $x \le 20$ and $y \in \{1, 4\}$ statistically independent?

2. Answer the questions from the table below:

	$b = 1$	$b = 2$	$b = 3$	$b = 4$
$a = 1$	0.02	0.02	0.21	0.02
$a = 2$	0.03	0.01	0.05	0.06
$a = 3$	0.01	0.01	0.01	0.06
$a = 4$	0.00	0.05	0.00	0.12
$a = 5$	0.12	0.06	0.00	0.14

   (a) Which event is more likely, $A = \{(a, b) : 3 \le a \le 4\}$ or $B = \{(a, b) : b \le 2 \text{ and } a = 5\}$?

   (b) List the largest impossible event.

   (c) Find the probability that $b \ne 3$.

   (d) Find $P(b = 2|a = 5)$.

   (e) Find $P(a \ge 3|b \in \{1, 4\})$.

   (f) Are the events $a \in \{1, 3\}$ and $b \in \{1, 2, 4\}$ statistically independent?

3. A disease hits 1 in every 20,000 people. A diagnostic test is 95% accurate, that is, the test is positive for 95% of people with the disease, and negative for 95% of the people who do not have the disease. Max just tested positive for the disease. What is the probability he has it?

4. You have data that sorts individuals into occupations and age groups. There are three occupations: doctor, lawyer, and entrpreneur. There are two age categories: below 40 (young) and above 40 (old). You wanted to know the probability that an old person is an entrepreneur. Your grad student misunderstands you, though, and presents you with the following information:

   20% of the sample are doctors and 30% are entrepreneurs

   40% of the doctors are young

   20% of the entrepreneurs are young

   70% of the lawyers are young

   Find the probability that the an old person is an entrepreneur.

CHAPTER

11

Random variables

11.1 Random variables

A **random variable** is a variable whose value is a real number, and that number is determined by the outcome of an experiment. For example, the number of heads in ten coin tosses is a random variable, and the Dow-Jones Industrial Average is a random variable.

The standard notation for a random variable is to place a tilde over the variable. So $\tilde{x}$ is a random variable.

The **realization** of a random variable is based on the outcome of an actual experiment. For example, if I toss a coin ten times and find four heads, the realization of the random variable is 4. When the random variable is denoted $\tilde{x}$ its realization is denoted x.

A random variable is **discrete** if it can take only discrete values (either a finite number or a countable infinity of values). A random variable is **continuous** if it can take any value in an interval.

Random variables have probability measures, and we can use random variables to define events. For example, $P(\tilde{x} = x)$ is the probability that the realization of the random variable $\tilde{x}$ is x, and $P(\tilde{x} \in [2,3])$ is the probability that the realization of the random variable $\tilde{x}$ falls in the interval $[2,3]$. The event in the latter example is the event that the realization of $\tilde{x}$ falls in the interval $[2,3]$.

11.2 Distribution functions

The **distribution function** for the random variable $\tilde{x}$ with probability measure P is given by

$$F(x) = P(\tilde{x} \leq x).$$

The distribution function $F(x)$ tells the probability that the realization of the random variable is no greater than x. Distribution functions are almost always denoted by capital letters.

Theorem 19 *Distribution functions are nondecreasing and take values in the interval* $[0,1]$*.*

Proof. The second part of the statement is obvious. For the first part, suppose $x < y$. Then the event $\tilde{x} \leq x$ is contained in the event $\tilde{x} \leq y$, and by Theorem 13 we have

$$F(x) = P(\tilde{x} \leq x) \leq P(\tilde{x} \leq y) = F(y).$$

■

11.3 Density functions

If the distribution function $F(x)$ is differentiable, the **density function** is

$$f(x) = F'(x).$$

If the distribution function $F(x)$ is discrete, the density function is $P(\tilde{x} = x)$ for each possible value of x. Sometimes distribution functions are neither differentiable nor discrete. This causes headaches that we will not deal with here.

Note that it is possible to go from a density function to a distribution function:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

So, the distribution function is the accumulated value of the density function. This leads to some additional common terminology. The distribution function is often called the *cumulative density function*, or c.d.f. The density function is often called the *probability density function*, or p.d.f.

The **support** of a distribution $F(x)$ is the smallest closed set containing $\{x|f(x) \neq 0\}$, that is, the set of points for which the density is positive. For a discrete distribution this is just the set of outcomes to which the distribution assigns positive probability. For a continuous distribution the support is the smallest closed set containing all of the points that have positive probability. For our purposes there is no real reason for using a closed (as opposed to open) set, but the definition given here is mathematically correct.

11.4 Useful distributions

11.4.1 Binomial (or Bernoulli) distribution

The **binomial distribution** arises when the experiment consists of repeated trials with the same two possible outcomes in each trial. The most obvious example is flipping a coin n times. The outcome is a series of heads and tails, and the probability distribution governing the number of heads in the series of n coin tosses is the binomial distribution.

To get a general formula, label one possible outcome of the trial a *success* and the other a *failure*. These are just labels. In coin tossing, we could count a head as a "success" and a tail as a "failure," or we could do it the other way around. If we are checking lightbulbs to see if they work, we could label a working lightbulb as a "success" and a nonworking bulb as a "failure," or we could do it the other way around. A coauthor (Harold Winter) and I used the binomial distribution to model juror bias, and the two possible outcomes were a juror biased toward conviction and a juror biased toward acquittal. We obviously cared about the total bias of the group. We had to arbitrarily label one form of bias a "success" and the other a "failure."

Suppose that the probability of a success is p in any given trial, which means that the probability of a failure is $q = 1 - p$. Also, assume that the trials are statistically independent, so that a success in trial t has no effect on the probability of success in period $t + 1$.

The question is, what is the probability of x successes in n trials?

Let's work this out for the case of two trials. Let the vector (outcome 1, outcome 2) denote the event in which outcome 1 is realized in the first trial and outcome 2 in the second trial. Using P as the probability measure, we have

$$\begin{aligned} P(\text{success, success}) &= p^2 \\ P(\text{success, failure}) &= pq \\ P(\text{failure, success}) &= pq \\ P(\text{failure, failure}) &= q^2 \end{aligned}$$

The probability of two successes in two trials is p^2, the probability of one success in two trials is $2pq$, and the probability of two failures is q^2.

With three trials, letting s denote a success and f a failure, we have

$$\begin{aligned} P(s,s,s) &= p^3 \\ P(s,s,f) &= P(s,f,s) = P(f,s,s) = p^2q \\ P(s,f,f) &= P(f,s,f) = P(f,f,s) = pq^2 \\ P(f,f,f) &= q^3 \end{aligned}$$

Thus the probability of three successes is p^3, the probability of two successes is $3p^2q$, the probability of one success is $3pq^2$, and the probability of no successes is q^3.

In general, the rule for x successes in n trials when the probability of success in a single trial is p is

$$b(x,n,p) = \binom{n}{x} p^x q^{n-x}$$

There are two pieces of the formula. The probability of a single, particular configuration of x successes and $n-x$ failures is $p^x q^{n-x}$. For example, the probability that the first x trials are successes and the last $n-x$ trials are failures is $p^x q^{n-x}$. The number of possible configurations with x successes and $n-x$ failures is $\binom{n}{x}$, which is given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \frac{1 \cdot 2 \cdot \ldots \cdot n}{[1 \cdot \ldots \cdot x][1 \cdot \ldots \cdot (n-x)]}.$$

Note that the function $b(x,n,p)$ is a *density* function. The binomial distribution is a discrete distribution, so $b(x,n,p) = P(\tilde{x} = x)$, where $\tilde{x}$ is the random variable measuring the number of successes in n trials.

The binomial distribution function is

$$B(x,n,p) = \sum_{i=0}^{x} b(x,n,p),$$

so that it is the probability of getting x or fewer successes in n trials.

Microsoft Excel, and probably similar programs, make it easy to compute the binomial density and distribution. The formula for $b(x,n,p)$ is

$$=\text{BINOMDIST}(x,n,p,0)$$

and the formula for $B(x,n,p)$ is

$$=\text{BINOMDIST}(x,n,p,1)$$

The last argument in the function just tells the program whether to compute the density or the distribution.

Let's go back to my jury example. Suppose we want to draw a pool of 12 jurors from the population, and that 20% of them are biased toward acquittal, with the rest biased toward conviction. The probability of drawing 2 jurors biased toward acquittal and 10 biased toward conviction is

$$b(2,12,0.2) = 0.283$$

The probability of getting *at most* two jurors biased toward acquittal is

$$B(2,12,0.2) = .558$$

The probability of getting a jury in which every member is biased toward conviction (and no member is biased toward acquittal) is

$$b(0,12,0.2) = 0.069$$

11.4.2 Uniform distribution

The great appeal of the uniform distribution is that it is easy to work with mathematically. It's density function is given by

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

The corresponding distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} \quad \text{if} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Okay, so these look complicated. But, if x is in the support interval $[a, b]$, then the density function is the constant function $f(x) = 1/(b-a)$ and the distribution function is the linear function $F(x) = (x-a)/(b-a)$.

Graphically, the uniform density is a horizontal line. Intuitively, it spreads the probability evenly (or uniformly) throughout the support $[a, b]$, which is why it has its name. The distribution function is just a line with slope $1/(b-a)$ in the support $[a, b]$.

11.4.3 Normal (or Gaussian) distribution

The normal distribution is the one that gives us the familiar bell-shaped density function, as shown in Figure 11.1. It is also central to statistical analysis, as we will see later in the course. We begin with the **standard normal distribution**. For now, its density function is

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

and its support is the entire real line: $(-\infty, \infty)$. The distribution function is

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

and we write it as an integral because there is no simple functional form.

Later on we will find the mean and standard deviation for different distributions. The standard normal has mean 0 and standard deviation 1. A more general normal distribution with mean μ and standard deviation σ has density function

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

and distribution function

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-\mu)^2/2\sigma^2} dt.$$

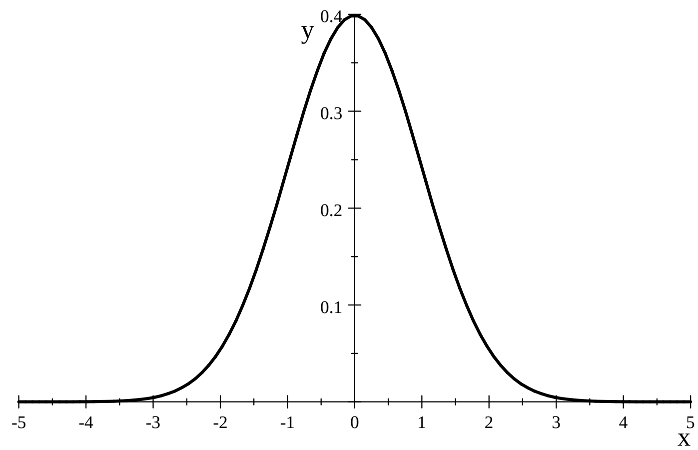


Figure 11.1: Density function for the standard normal distribution

The effects of changing the mean μ can be seen in Figure 11.2. The peak of the normal distribution is at the mean, so the standard normal peaks at $x = 0$, which is the thick curve in the figure. The thin curve in the figure has a mean of 2.5.

Changing the standard deviation σ has a different effect, as shown in Figure 11.3. The thick curve is the standard normal with $\sigma = 1$, and the thin curve has $\sigma = 2$. As you can see, increasing the standard deviation lowers the peak and spreads the density out, moving probability away from the mean and into the tails.

11.4.4 Exponential distribution

The density function for the exponential distribution is

$$f(x) = \frac{1}{\theta} e^{-x/\theta}$$

and the distribution function is

$$F(x) = 1 - e^{-x/\theta}.$$

It is defined for $x > 0$. The exponential distribution is often used for the failure rate of equipment: the probability that a piece of equipment will fail

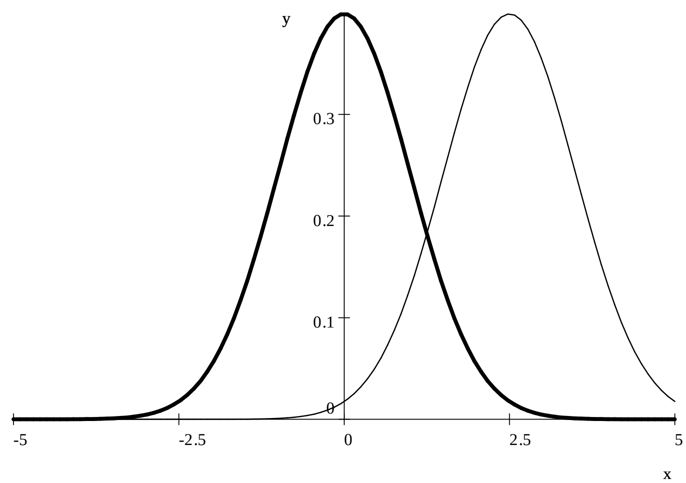


Figure 11.2: Changing the mean μ of the normal density

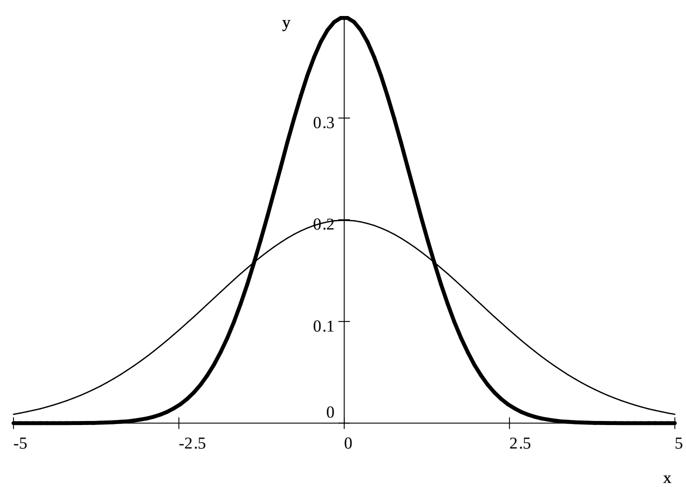


Figure 11.3: Increasing the standard deviation σ of the normal density

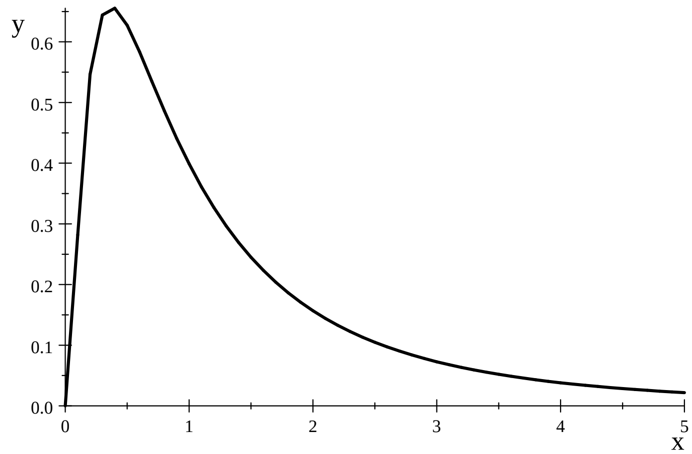


Figure 11.4: Density function for the lognormal distribution

by time x is $F(x)$. Accordingly, $f(x)$ is the probability of a failure at time x.

11.4.5 Lognormal distribution

The random variable $\tilde{x}$ has the **lognormal distribution** if the random variable $\tilde{y} = \ln \tilde{x}$ has the normal distribution. The density function is

$$f(x) = \frac{1}{\sqrt{2\pi}} \left[e^{-(\ln x)^2/2} \right] \cdot \frac{1}{x}$$

and it is defined for $x \geq 0$. It is shown in Figure 11.4.

11.4.6 Logistic distribution

The **logistic distribution** function is

$$F(x) = \frac{1}{1 + e^{-x}}.$$

Its density function is

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

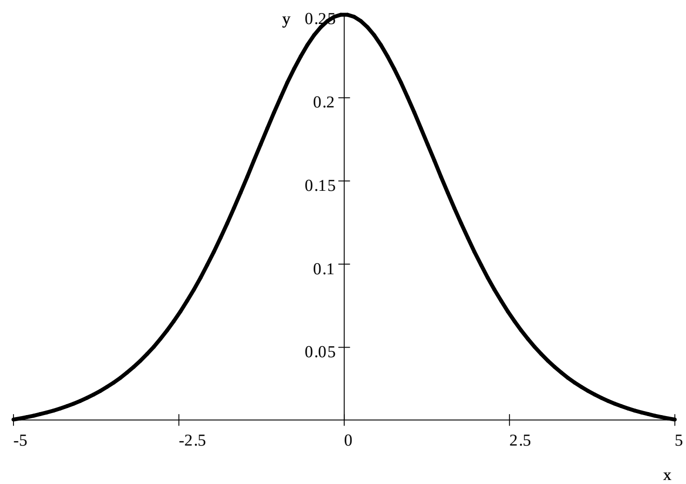


Figure 11.5: Density function for the logistic distribution

It is defined over the entire real line and gives the bell-shaped density function shown in Figure 11.5.

CHAPTER

12

Integration

If you took a calculus class from a mathematician, you probably learned two things about integration: (1) integration is the opposite of differentiation, and (2) integration finds the area under a curve. Both of these are correct. Unfortunately, in economics we rarely talk about the area under a curve. There are exceptions, of course. We sometimes think about profit as the area between the price line and the marginal cost curve, and we sometimes compute consumer surplus as the area between the demand curve and the price line. But this is not the primary reason for using integration in economics.

Before we get into the interpretation, we should first deal with the mechanics. As already stated, integration is the opposite of differentiation. To make this explicit, suppose that the function $F(x)$ has derivative $f(x)$. The following two statements provide the fundamental relationship between derivatives and integrals:

$$\int_a^b f(x)dx = F(b) - F(a), \tag{12.1}$$

and

$$\int f(x)dx = F(x) + c, \tag{12.2}$$

where c is a constant. The integral in (12.1) is a **definite integral**, and its distinguishing feature is that the integral is taken over a finite interval. The integral in (12.2) is an **indefinite integral**, and it has no endpoints. The reason for the names is that the solution in (12.1) is unique, or definite, while the solution in (12.2) is not unique. This occurs because when we integrate the function $f(x)$, all we know is the slope of the function $F(x)$, and we do not know anything about its height. If we choose one function that has slope $f(x)$, call it $F^*(x)$, and we shift it upward by one unit, its slope is still $f(x)$. The role of the constant c in (12.2), then, is to account for the indeterminacy of the height of the curve when we take an integral.

The two equations (12.1) and (12.2) are consistent with each other. To see why, notice that

$$\int f(x)dx = \int_{-\infty}^{\infty} f(x)dx,$$

so an indefinite integral is really just an integral over the entire real line $(-\infty, \infty)$. Furthermore,

$$\begin{aligned} \int_a^b f(x)dx &= \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx \\ &= [F(b) + c] - [F(a) + c] \\ &= F(b) - F(a). \end{aligned}$$

Some integrals are straightforward, but others require some work. From our point of view the most important ones follow, and they can be checked by differentiating the right-hand side.

$$\begin{aligned} \int x^n dx &= \frac{x^{n+1}}{n+1} + c \text{ for } n \neq -1 \\ \int \frac{1}{x} dx &= \ln x + c \\ \int e^{rx} dx &= \frac{e^{rx}}{r} + c \\ \int \ln x dx &= x \ln x - x + c \end{aligned}$$

There are also two useful rules for more complicated integrals:

$$\int af(x)dx = a\int f(x)dx$$
$$\int [f(x)+g(x)]\,dx = \int f(x)dx + \int g(x)dx.$$

The first of these says that a constant inside of the integral can be moved outside of the integral. The second one says that the integral of the sum of two functions is the sum of the two integrals. Together they say that integration is a linear operation.

12.1 Interpreting integrals

Repeat after me (okay, way after me, because I wrote this in April 2008):

Integrals are used for adding.

They can also be used to find the area under a curve, but in economics the primary use is for addition.

To see why, suppose we wanted to do something strange like measure the amount of water flowing in a particular river during a year. We have not figured out how to measure the total volume, but we can measure the flow at any instant in time using our Acme Hydroflowometer$^{\text{TM}}$. At time t the volume of water flowing past a particular point as measured by the Hydroflowometer$^{\text{TM}}$ is $h(t)$.

Suppose that we break the year into n intervals of length T/n each, where T is the amount of time in a year. We measure the flow once per time interval, and our measurement times are $t_1, ..., t_n$. We use the measured flow at time t_i to calculate the total flow for period i according to the formula

$$h(t_i)\frac{T}{n}$$

where $h(t_i)$ is our measure of the instantaneous flow and T/n is the length of time in the interval. The total estimated volume for the year is then

$$V(n) = \sum_{t=t_1}^{t_n} h(t_i)\frac{T}{n}. \tag{12.3}$$

We can make our estimate of the volume more accurate by taking more measurements of the flow. As we do this n becomes larger and T/n becomes smaller.

Now suppose that we could measure the flow at every instant of time. Then $T/n \to 0$, and if we tried to do the summation in equation (12.3) we would add up a whole bunch of numbers, each of which is multiplied by zero. But the total volume is not zero, so this cannot be the right approach. It's not. The right approach uses integrals. The idea behind an integral is adding an infinite number of zeroes together to get something that is not zero. Our correct formula for the volume would be

$$V = \int_0^T h(t)dt.$$

The expression dt takes the place of the expression T/n in the sum, and it is the length of each measurement interval, which is zero.

We can use this approach in a variety of economic applications. One major use is for taking expectations of continuous random variables, which is the topic of the next chapter. Before going on, though, there are two useful tricks involving integrals that deserve further attention.

12.2 Integration by parts

Integration by parts is a very handy trick that is often used in economics. It is also what separates us from the lower animals. The nice thing about integration by parts is that it is simple to reconstruct. Start with the product rule for derivatives:

$$\frac{d}{dx}[f(x)g(x)] = f'(x)g(x) + f(x)g'(x).$$

Integrate both sides of this with respect to x and over the interval $[a, b]$:

$$\int_a^b \frac{d}{dx}[f(x)g(x)]dx = \int_a^b f'(x)g(x)dx + \int_a^b f(x)g'(x)dx. \tag{12.4}$$

Note that the left-hand term is just the integral (with respect to x) of a derivative (with respect to x), and combining those two operations leaves the function unchanged:

$$\begin{aligned}\int_a^b \frac{d}{dx}[f(x)g(x)]dx &= f(x)g(x)|_a^b \\ &= f(b)g(b) - f(a)g(a).\end{aligned}$$

Plugging this into (12.4) yields

$$f(x)g(x)|_a^b = \int_a^b f'(x)g(x)dx + \int_a^b f(x)g'(x)dx.$$

Now rearrange this to get the rule for **integration by parts**:

$$\int_a^b f'(x)g(x)dx = f(x)g(x)|_a^b - \int_a^b f(x)g'(x)dx. \tag{12.5}$$

When you took calculus, integration by parts was this mysterious thing. Now you know the secret – it's just the product rule for derivatives.

12.2.1 Application: Choice between lotteries

Here is my favorite use of integration by parts. You may not understand the economics yet, but you will someday. Suppose that an individual is choosing between two lotteries. Lotteries are just probability distributions, and the individual's objective function is

$$\int_a^b u(x)F'(x)dx \tag{12.6}$$

where a is the lowest possible payoff from the lottery, b is the highest possible payoff, u is a utility function defined over amounts of money, and $F'(x)$ is the density function corresponding to the probability distribution function $F(x)$. (I use the derivative notation $F'(x)$ instead of the density notation $f(x)$ to make the use of integration by parts more transparent.)

The individual can choose between lottery $F(x)$ and lottery $G(x)$, and we would like to find properties of $F(x)$ and $G(x)$ that guarantee that the individual likes $F(x)$ better. How can we do this? We must start with what we know about the individual. The individual likes money, and $u(x)$ is the utility of money. If she likes money, $u(x)$ must be nondecreasing, so

$$u'(x) \geq 0.$$

And that's all we know about the individual.

Now look back at expression (12.6). It is the integral of the product of $u(x)$ and $F'(x)$. But the only thing we know about the individual is that $u'(x) \geq 0$, and expression (12.6) does not have $u'(x)$ in it. So let's integrate by parts.

$$\int_a^b u(x)F'(x)dx = u(x)F(x)|_a^b - \int_a^b u'(x)F(x)dx.$$

To simplify this we need to know a little more about probability distribution functions. Since a is the lowest possible payoff from the lottery, the distribution function must satisfy $F(a) = 0$ (this comes from Theorem 19). Since b is the highest possible payoff from the lottery, the distribution function must satisfy $F(b) = 1$. So, the above expression reduces to

$$\int_a^b u(x)F'(x)dx = u(b) - \int_a^b u'(x)F(x)dx. \tag{12.7}$$

We can go through the same analysis for the other lottery, $G(x)$, and find

$$\int_a^b u(x)G'(x)dx = u(b) - \int_a^b u'(x)G(x)dx. \tag{12.8}$$

The individual chooses the lottery $F(x)$ over the lottery $G(x)$ if

$$\int_a^b u(x)F'(x)dx \geq \int_a^b u(x)G'(x)dx,$$

that is, if the lottery $F(x)$ generates a higher value of the objective function than the lottery $G(x)$ does. Or written differently, she chooses $F(x)$ over $G(x)$ if

$$\int_a^b u(x)F'(x)dx - \int_a^b u(x)G'(x)dx \geq 0.$$

Subtracting (12.8) from (12.7) yields

$$\begin{aligned}
& \int_a^b u(x)F'(x)dx - \int_a^b u(x)G'(x)dx \\
= & \left[u(b) - \int_a^b u'(x)F(x)dx\right] - \left[u(b) - \int_a^b u'(x)G(x)dx\right] \\
= & \int_a^b u'(x)G(x)dx - \int_a^b u'(x)F(x)dx \\
= & \int_a^b u'(x)\left[G(x) - F(x)\right]dx.
\end{aligned}$$

The difference depends on something we know about: $u'(x)$, which we know is nonnegative. The individual chooses $F(x)$ over $G(x)$ if the above expression is nonnegative, that is, if

$$\int_a^b u'(x)\left[G(x) - F(x)\right]dx \geq 0.$$

We know that $u'(x) \geq 0$. We can guarantee that the product $u'(x)\left[G(x) - F(x)\right]$ is nonnegative if the other term, $[G(x) - F(x)]$, is also nonnegative. So, we are certain that she will choose $F(x)$ over $G(x)$ if

$$G(x) - F(x) \geq 0 \text{ for all } x.$$

This turns out to be the answer to our question. Any individual who likes money will prefer lottery $F(x)$ to lottery $G(x)$ if $G(x) - F(x) \geq 0$ for all x. There is even a name for this condition – *first-order stochastic dominance.* But the goal here was not to teach you about choice over lotteries. The goal was to show the usefulness of integration by parts. So let's look back and see exactly what it did for us. The objective function was an integral of the product of two terms, $u(x)$ and $F'(x)$. We could not assume anything about $u(x)$, but we could assume something about $u'(x)$. So we used integration by parts to get an expression that involved the term we knew something about. And that is its beauty.

12.3 Differentiating integrals

As you have probably noticed, in economics we differentiate a lot. Sometimes, though, the objective function has an integral, as with expression

(12.6) above. Often we want to differentiate an objective function to find an optimum, and when the objective function has an integral we need to know how to differentiate it. There is a rule for doing so, called *Leibniz's rule*, named after the 17th-century German mathematician who was one of the two independent inventors of calculus (along with Newton).

We want to find

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(x,t)dx.$$

Note that we are differentiating with respect to t, and we are integrating with respect to x. Nevertheless, t shows up three times in the expression, once in the upper limit of the integral, $b(t)$, once in the lower limit of the integral, $a(t)$, and once in the integrand, $f(x,t)$. We need to figure out what to do with these three terms.

A picture helps. Look at Figure 12.1. The integral is the area underneath the curve $f(x,t)$ between the endpoints $a(t)$ and $b(t)$. Three things happen when t changes. First, the function $f(x,t)$ shifts, and the graph shows an upward shift, which makes the integral larger because the area under a higher curve is larger. Second, the right endpoint $b(t)$ changes, and the graph shows it getting larger. This again increases the integral because now we are integrating over a larger interval. Third, the left endpoint $a(t)$ changes, and again the graph shows it getting larger. This time, though, it makes the integral smaller because moving the left endpoint rightward shrinks the interval over which we are integrating.

Leibniz's rule accounts for all three of these shifts. Leibniz's rule says

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(x,t)dx = \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}dx + b'(t)f(b(t),t) - a'(t)f(a(t),t).$$

Each of the three terms corresponds to one of the shifts in Figure 12.1. The first term accounts for the upward shift of the curve $f(x,t)$. The term $\partial f(x,t)/\partial t$ tells how far upward the curve shifts at point x, and the integral

$$\int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}dx$$

tells how much the area changes because of the upward shift in $f(x,t)$.

The second term accounts for the movement in the right endpoint, $b(t)$. Using the graph, the amount added to the integral is the area of a rectangle

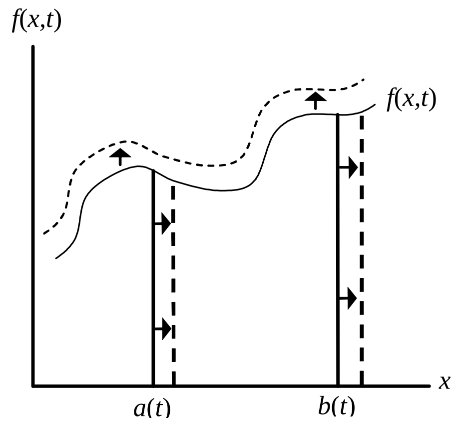


Figure 12.1: Leibniz's rule

that has height $f(b(t), t)$, that is, $f(x, t)$ evaluated at $x = b(t)$, and width $b'(t)$, which accounts for how far $b(t)$ moves when t changes. Since area is just length times width, we get $b'(t)f(b(t), t)$, which is exactly the second term.

The third term accounts for the movement in the left endpoint, $a(t)$. Using the graph again, the change in the integral is the area of a rectangle that has height $f(a(t), t)$ and width $a'(t)$. This time, though, if $a(t)$ increases we are *reducing* the size of the integral, so we must *subtract* the area of the rectangle. Consequently, the third term is $-a'(t)f(a(t), t)$.

Putting these three terms together gives us Leibniz's rule, which looks complicated but hopefully makes sense.

12.3.1 Application: Second-price auctions

A simple application of Leibniz's rule comes from auction theory. A first-price sealed bid auction has bidders submit bids simultaneously to an auctioneer who awards the item to the highest bidder who then pays his bid. This is a very common auction form. A *second*-price sealed bid auction has bidders submit bids simultaneously to an auctioneer who awards the item to the highest bidder, just like before, but this time the winning bidder pays the *second-highest* price.

To model the second-price auction, suppose that there are n bidders and that bidder i values the item being auctioned at v_i, which is independent of how much everyone else values the item. Bidders do not know their opponents' valuations, but they do know the probability distribution of the opponents' valuations. Bidder i must choose his bid b_i.

Let $F_i(b)$ be the probability that the highest *other* bid faced by i, that is, the highest bid except for b_i, is no larger than b. Then $F_i(b)$ is a probability distribution function, and its density function is $f_i(b)$. Bidder i's expected payoff is

$$V_i(b_i) = \int_0^{b_i} (v_i - b) f_i(b) db.$$

Let's interpret this function. Bidder i wins if his is the highest bid, which occurs if the highest other bid is between 0 (the lowest possible bid) and his own bid b_i. If the highest other bid is above b_i bidder i loses and gets a payoff of zero. This is why the integral is taken over the interval $[0, b_i]$. If bidder i wins he pays the highest other bid b, which is distributed according to the density function $f_i(b)$. His surplus if he wins is $v_i - b$, his value minus how much he pays.

Bidder i chooses the bid b_i to maximize his expected payoff $V_i(b_i)$. Since this is a maximization problem we should find the first-order condition:

$$V_i'(b_i) = \frac{d}{db_i} \int_0^{b_i} (v_i - b) f_i(b) db = 0.$$

Notice that we are differentiating with respect to b_i, which shows up only as the upper endpoint of the integral. Using Leibniz's rule we can evaluate this first-order condition:

$$\begin{aligned} 0 &= \frac{d}{db_i} \int_0^{b_i} (v_i - b) f_i(b) db \\ &= \int_0^{b_i} \frac{\partial}{\partial b_i} \left[(v_i - b) f_i(b) \right] db + \frac{db_i}{db_i} \cdot (v_i - b_i) f_i(b_i) - \frac{d0}{db_i} \cdot (v_i - 0) f_i(0). \end{aligned}$$

The first term is zero because $(v_i - b) f_i(b)$ is not a function of b_i, and so the partial derivative is zero. The second term reduces to $(v_i - b_i) f_i(b_i)$ because db_i/db_i is simply one. The third term is zero because the derivative $d0/db_i = 0$. This leaves us with the first-order condition

$$0 = (v_i - b_i) f_i(b_i).$$

Since density functions take on only nonnegative values, the first-order condition holds when $v_i - b_i = 0$, or $b_i = v_i$. In a second-price auction the bidder should bid his value.

This result makes sense intuitively. Let b_i be bidder i's bid, and let b denote the highest other bid. Suppose first that bidder i bids more than his value, so that $b_i > v_i$. If the highest other bid is in between these, so that $v_i < b < b_i$, bidder i wins the auction but pays $b - v_i$ more than his valuation. He could have avoided this by bidding his valuation, v_i. Now suppose that bidder i bids less than his value, so that $b_i < v_i$. If the highest other bid is between these two, so that $b_i < b < v_i$, bidder i loses the auction and gets nothing. But if he had bid his value he would have won the auction and paid $b < v_i$, and so he would have been better off. Thus, the best thing for him to do is bid his value.

12.4 Problems

1. Suppose that $f(x)$ is the density function for a random variable distributed uniformly over the interval $[2, 8]$.

   (a) Compute
   $$\int_2^8 xf(x)dx$$
   (b) Compute
   $$\int_2^8 x^2 f(x)dx$$

2. Compute the following derivative:
   $$\frac{d}{dt}\int_{-t^2}^{t^2} tx^2 dx$$

3. Find the following derivative:
   $$\frac{d}{dt}\int_{-3t}^{4t^2} t^2 x^3 dx$$

4. Let $U(a, b)$ denote the uniform distribution over the interval $[a, b]$. Find conditions on a and b that guarantee that $U(a, b)$ first-order stochastically dominates $U(0, 1)$.

CHAPTER

13

Moments

13.1 Mathematical expectation

Let $\tilde{x}$ be a random variable with density function $f(x)$ and let $u(x)$ be a real-valued function. The **expected value** of $u(\tilde{x})$ is denoted $E[u(\tilde{x})]$ and it is found by the following rules. If $\tilde{x}$ is discrete taking on value x_i with probability $f(x_i)$ then

$$E[u(\tilde{x})] = \sum_i u(x_i)f(x_i).$$

If $\tilde{x}$ is continuous the expected value of $u(\tilde{x})$ is given by

$$E[u(\tilde{x})] = \int_{-\infty}^{\infty} u(x)f(x)dx.$$

Since integrals are for adding, as we learned in the last chapter, these formulas really do make sense and go together.

The **expectation operator** $E[\cdot]$ is linear, which means that

$$\begin{aligned} E[au(\tilde{x})] &= aE[u(\tilde{x})] \\ E[u(\tilde{x}) + v(\tilde{x})] &= E[u(\tilde{x})] + E[v(\tilde{x})] \end{aligned}$$

13.2 The mean

The **mean** of a random variable $\tilde{x}$ is $\mu = E[\tilde{x}]$, that is, it is the expected value of the function $u(x) = x$.

Consider the discrete distribution with outcomes $(4, 10, 12, 20)$ and corresponding probabilities $(0.1, 0.2, 0.3, 0.4)$. The mean is

$$E[\tilde{x}] = (4)(0.1) + (10)(0.2) + (12)(0.3) + (20)(0.4) = 14$$

13.2.1 Uniform distribution

The mean of the uniform distribution over the interval $[a, b]$ is $(a + b)/2$. If you don't believe me, draw it. To figure it out from the formula, compute

$$\begin{aligned} E[\tilde{x}] &= \int_a^b x \cdot \frac{1}{b-a} dx \\ &= \frac{1}{b-a} \int_a^b x dx \\ &= \frac{1}{b-a} \cdot \frac{1}{2} x^2 \Big|_a^b \\ &= \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} \\ &= \frac{b+a}{2}. \end{aligned}$$

13.2.2 Normal distribution

The mean of the general normal distribution is the parameter μ. Recall that the normal density function is

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

The mean is

$$E[\tilde{x}] = \int \frac{x}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx.$$

Use the change-of-variables formula $y = \frac{x-\mu}{\sigma}$ so that $x = \mu+\sigma y$, $(x-\mu)^2/\sigma^2 = y^2$, and $dx = \sigma dy$. Then we can rewrite

$$\begin{aligned} E[\tilde{x}] &= \int \frac{x}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \int_{-\infty}^{\infty} \frac{\mu+\sigma y}{\sigma\sqrt{2\pi}} e^{-y^2/2} \sigma dy \\ &= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy. \end{aligned}$$

The first integral is the integral of the standard normal density, and like all densities its integral is 1. The second integral can be split into two parts:

$$\int_{-\infty}^{\infty} y e^{-y^2/2} dy = \int_{-\infty}^{0} y e^{-y^2/2} dy + \int_{0}^{\infty} y e^{-y^2/2} dy.$$

Use the change of variables $y = -z$ in the first integral on the right-hand side. Then $y^2 = z^2$ and $dy = -dz$, so

$$\int_{-\infty}^{0} y e^{-y^2/2} dy = -\int_{0}^{\infty} z e^{-z^2/2} dz$$

Plugging this back into the expression above it yields

$$\int_{-\infty}^{\infty} y e^{-y^2/2} dy = -\int_{0}^{\infty} z e^{-z^2/2} dz + \int_{0}^{\infty} y e^{-y^2/2} dy.$$

But both integrals on the right-hand side are the same, so the expression is zero. Thus, we get $E[\tilde{x}] = \mu$.

13.3 Variance

The **variance** of the random variable $\tilde{x}$ is $E[(\tilde{x}-\mu)^2]$, where $\mu = E[\tilde{x}]$ is the mean of the random variable. The variance is denoted σ^2. Note that

$$\begin{aligned} E[(\tilde{x}-\mu)^2] &= E[\tilde{x}^2 - 2\mu\tilde{x} + \mu^2] \\ &= E[\tilde{x}^2] - 2\mu E[\tilde{x}] + \mu^2 \\ &= E[\tilde{x}^2] - 2\mu^2 + \mu^2 \\ &= E[\tilde{x}^2] - \mu^2. \end{aligned}$$

We can find the variance of the discrete random variable used in the preceding section. The outcomes were $(4, 10, 12, 20)$ and the corresponding probabilities were $(0.1, 0.2, 0.3, 0.4)$. The mean was 14. The variance is

$$\begin{aligned} E[(\tilde{x}-\mu)^2] &= (0.1)(4-14)^2 + (0.2)(10-14)^2 + \\ & \quad (0.3)(12-14)^2 + (0.4)(20-14)^2 \\ &= (0.1)(100) + (0.2)(16) + (0.3)(4) + (0.4)(36) \\ &= 28.8 \end{aligned}$$

We can also find it using the alternative formula:

$$E[\tilde{x}^2] - \mu^2 = (0.1)(4^2) + (0.2)(10^2) + (0.3)(12)^2 + (0.4)(20^2) - 14^2.$$

You should be able to show that

$$Var(a\tilde{x}) = a^2 Var(\tilde{x}).$$

The **standard deviation** of the random variable $\tilde{x}$ is $\sqrt{E[(\tilde{x}-\mu)^2]}$, which means that the standard deviation is simply σ. It is the square root of the variance.

13.3.1 Uniform distribution

The variance of the uniform distribution can be found from

$$\begin{aligned} E[\tilde{x}^2] &= \int_a^b \frac{x^2}{b-a} dx \\ &= \frac{1}{b-a} \cdot \frac{1}{3} x^3 \bigg|_a^b \\ &= \frac{1}{b-a} \cdot \frac{b^3-a^3}{3} \end{aligned}$$

and note that $b^3 - a^3 = (b-a)(b^2+ab+a^2)$. Consequently,

$$\begin{aligned}
\sigma^2 &= E[\tilde{x}^2] - \mu^2 \\
&= \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4} \\
&= \frac{4b^2+4ab+4a^2-3b^2-6ab-3a^2}{12} \\
&= \frac{b^2-2ab+a^2}{12} \\
&= \frac{(b-a)^2}{12}.
\end{aligned}$$

13.3.2 Normal distribution

The variance of the standard normal distribution is 1. Let's take that on faith. The variance of the general normal distribution had better be the parameter σ^2. To make sure, compute

$$E[(\tilde{x}-\mu)^2] = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx$$

Using the same change-of-variables trick as before, we get

$$\begin{aligned}
E[(\tilde{x}-\mu)^2] &= \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \\
&= \int_{-\infty}^{\infty} \frac{(\mu+\sigma y-\mu)^2}{\sigma\sqrt{2\pi}} e^{-y^2/2} \sigma dy \\
&= \int_{-\infty}^{\infty} \frac{\sigma y^2}{\sqrt{2\pi}} e^{-y^2/2} dy \\
&= \sigma^2 \int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2} dy.
\end{aligned}$$

The integral is the variance of the standard normal, which we already said was 1.

13.4 Application: Order statistics

Suppose that you make n independent draws from the random variable $\tilde{x}$ with distribution function $F(x)$ and density $f(x)$. The value of the highest of them is a random variable, the value of the second highest is a random variable, and so on, for n random variables. The **n-th order statistic** is the expected value of the n-th highest draw. So, the first order statistic is the expected value of the highest of the n draws, the second order statistic is the expected value of the second highest of the n draws, and so on.

We use order statistics in a variety of settings, but the most straightforward one is auctions. Think about a first-price sealed bid auction in which n bidders submit their bids simultaneously and then the highest bidder wins and pays her bid. The seller's expected revenue, then, is the expected value of the highest of the n bids, which is the first order statistic. Now think about the second-price sealed-bid auction. In this auction n bidders submit their bids simultaneously, the highest bid wins, and the winner pays the second-highest bid. The seller's expected revenue in this auction is the expected value of the second-highest of the n bids, which is the second order statistic.

As a side note, order statistics have also played a role in cosmology, the study of the cosmos, and in particular they were used by Edwin Hubble. Hubble was clearly an overachiever. In set the Illinois state high school record for high jump. He was a Rhodes scholar. He was the first astronomer to use the giant 200-inch Hale telescope at Mount Palomar. He was honored with a 41 cent postage stamp. He has the Hubble Space Telescope named after him. Importantly for this story, though, he established that the universe extends beyond our galaxy, the Milky Way. This was a problem because we know that stars that are farther away are dimmer, but not all stars have the same brightness. So, we can't tell whether a particular star is dim because it's far away or because it's just not very bright (econometricians would call this an identification problem). Astronomers before Hubble made the heroic (that is, unreasonable) assumption that all starts were the same brightness and worked from there. Hubble used the milder assumption that the *brightest* star in every nebula (or galaxy, but they didn't know the difference at the time) is equally bright. In other words, he assumed that the first order statistic is the same for every nebula.

We want to find the order statistics, and, in particular, the first and second order statistics. To do this we have to find some distributions. Think about the first order statistic. It is the expected value of the highest of the n draws, and the highest of the n draws is a random variable with a distribution. But what is the distribution? We must construct it from the underlying distribution F.

Let $G^{(1)}(x)$ denote the distribution for the highest of the n values drawn independently from $F(x)$. We want to derive $G^{(1)}(x)$. Remember that $G^{(1)}(x)$ is the probability that the highest draw is less than or equal to x. For the highest draw to be less than or equal to x, it must be the case that every draw is less than or equal to x. When $n = 1$ the probability that the one draw is less than or equal to x is $F(x)$. When $n = 2$ the probability that both draws are less than or equal to x is $(F(x))^2$. And so on. When there are n draws the probability that all of them are less than or equal to x is $(F(x))^n$, and so

$$G^{(1)}(x) = F^n(x).$$

From this we can get the density function by differentiating $G^{(1)}$ with respect to x:

$$g^{(1)}(x) = nF^{n-1}(x)f(x).$$

Note the use of the chain rule.

This makes it possible to compute the *first order statistic*, since we know the distribution and density functions for the highest of n draws. We just take the expected value in the usual way:

$$s^{(1)} = \int xnF^{n-1}(x)f(x)dx.$$

Example 12 *Uniform distribution over* $(0, 1)$. *We have* $F(x) = x$ *on* $[0, 1]$, *and* $f(x) = 1$ *on* $[0, 1]$. *The first order statistic is*

$$\begin{aligned} s^{(1)} &= \int xnF^{n-1}(x)f(x)dx \\ &= \int_0^1 xnx^{n-1}dx \\ &= n\int_0^1 x^n dx \\ &= n\frac{x^{n+1}}{n+1}\bigg|_0^1 \\ &= \frac{n}{n+1}. \end{aligned}$$

This answer makes some sense. If $n = 1$ *the first order statistic is just the mean, which is* $1/2$. *If* $n = 2$ *and the distribution is uniform so that the* *draws are expected to be evenly spaced, then the highest draw should be about* $2/3$ *and the lowest should be about* $1/3$. *If* $n = 3$ *the highest draw should be about* $3/4$, *and so on.*

We also care about the second order statistic. To find it we follow the same steps, beginning with identifying the distribution of the second-highest draw. To make this exercise precise, we are looking for the probability that the second-highest draw is no greater than some number, call it y. There are a total of $n+1$ ways that we can get the second-highest draw to be below y, and they are listed below:

Event	Probability
Draw 1 is above y and the rest are below y	$(1-F(y))F^{n-1}(y)$
Draw 2 is above y and the rest are below y	$(1-F(y))F^{n-1}(y)$
$\vdots$	$\vdots$
Draw n is above y and the rest are below y	$(1-F(y))F^{n-1}(y)$
All the draws are below y	$F^n(y)$

Let's figure out these probabilities one at a time. Regarding the first line, the probability that draws 2 through n are below y is the probability of getting $n-1$ draws below y, which is $F^{n-1}(y)$. The probability that draw 1 is above y is $1-F(y)$. Multiplying these together yields the probability of getting draw 1 above y and the rest below. The probability of getting draw 2 above y and the rest below is the same, and so on for the first n rows of the table. In the last row all of the draws are below y, in which case both the highest and the second highest draws are below y. The probability of all n draws being below y is just $F^n(y)$, the same as when we looked at the first order statistic. Summing the probabilities yields the distribution of the second-highest draw:

$$G^{(2)}(y) = n(1-F(y))F^{n-1}(y) + F^n(y).$$

Multiplying this out and simplifying yields

$$G^{(2)}(y) = nF^{n-1}(y) - (n-1)F^n(y).$$

The density function is found by differentiating $G^{(2)}$ with respect to y:

$$g^{(2)}(y) = n(n-1)F^{n-2}(y)f(y) - n(n-1)F^{n-1}(y)f(y).$$

It can be rearranged to get

$$g^{(2)}(y) = n(n-1)(1-F(y))F^{n-2}(y)f(y).$$

The *second order statistic* is the expected value of the second-highest draw, which is

$$\begin{aligned} s^{(2)} &= \int y g^{(2)}(y)dy \\ &= \int yn(n-1)(1-F(y))F^{n-2}(y)f(y)dy. \end{aligned}$$

Example 13 *Uniform distribution over* $[0,1]$*.*

$$\begin{aligned} s^{(2)} &= \int yn(n-1)(1-F(y))F^{n-2}(y)f(y)dy \\ &= \int_0^1 yn(n-1)(1-y)y^{n-2}dy \\ &= n(n-1)\int_0^1 \left(y^{n-1}-y^n\right)dy \\ &= n(n-1)\frac{y^n}{n}\Big|_0^1 - n(n-1)\frac{y^{n+1}}{n+1}\Big|_0^1 \\ &= (n-1) - \frac{n(n-1)}{n+1} \\ &= \frac{n-1}{n+1}. \end{aligned}$$

If there are four draws uniformly dispersed between 0 *and* 1*, the highest draw is expected to be at* $3/4$*, the second highest at* $2/4$*, and the lowest at* $1/4$*. If there are five draws, the highest is expected to be at* $4/5$ *and the second highest is expected to be at* $3/5$*, and so on.*

13.5 Problems

1. Suppose that the random variable $\tilde{x}$ takes on the following values with the corresponding probabilities:

Value	Probability
7	.10
4	.23
2	.40
-2	.15
-6	.10
-14	.02

   (a) Compute the mean.

   (b) Compute the variance.

2. The following table shows the probabilities for two random variable, one with density function $f(x)$, and one with density function $g(x)$.

x	$f(x)$	$g(x)$
10	0.15	0.20
15	0.5	0.30
20	0.05	0.1
30	0.1	0.1
100	0.2	0.3

   (a) Compute the means of the two variables.

   (b) Compute the variances of the two variables.

   (c) Compute the standard deviations of the two variables.

3. Consider the triangular density given by $f(x) = 2x$ on the interval $[0, 1]$.

   (a) Find its distribution function F.

   (b) Verify that it is a distribution function, that is, and specifically for this case, that F is increasing, $F(0) = 0$, and $F(1) = 1$.

   (c) Find the mean.

(d) Find the variance.

4. Consider the triangular density given by $f(x) = \frac{1}{8}x$ on the interval $[0, 4]$.

   (a) Find its distribution function F.

   (b) Verify that it satisfies the properties of a distribution function, that is, $F(0) = 0$, $F(4) = 1$, and F increasing.

   (c) Find the mean.

   (d) Find the variance.

5. Show that if the variance of $\tilde{x}$ is σ^2 then the variance of $a\tilde{x}$ is $a^2\sigma^2$, where a is a scalar.

6. Show that if the variance of $\tilde{x}$ is σ_x^2 and if $\tilde{y} = 3\tilde{x} - 1$, then the variance of $\tilde{y}$ is $9\sigma_x^2$.

7. Suppose that the random variable $\tilde{x}$ takes the value 6 with probability $\frac{1}{2}$ and takes the value y with probability $\frac{1}{2}$. Find the derivative $d\sigma^2/dy$, where σ^2 is the variance of $\tilde{x}$.

8. Let $G^{(1)}$ and $G^{(2)}$ be the distribution functions for the highest and second highest draws, respectively. Show that $G^{(1)}$ first-order stochastically dominates $G^{(2)}$.

CHAPTER

14

Multivariate distributions

Multivariate distributions arise when there are multiple random variables. For example, what we normally refer to as "the weather" is comprised of several random variables: temperature, humidity, rainfall, etc. A multivariate distribution function is defined over a vector of random variables. A bivariate distribution function is defined over two random variables. In this chapter I restrict attention to bivariate distributions. Everything can be extended to multivariate distributions by adding more random variables.

14.1 Bivariate distributions

Let $\tilde{x}$ and $\tilde{y}$ be two random variables. The distribution function $F(x,y)$ is given by

$$F(x,y) = P(\tilde{x} \leq x \text{ and } \tilde{y} \leq y).$$

It is called the **joint distribution function**. The function $F(x,\infty)$ is the probability that $\tilde{x} \leq x$ and $\tilde{y} \leq \infty$. The latter is sure to hold, and so $F(x,\infty)$ is the *univariate* distribution function for the random variable $\tilde{x}$.

Similarly, the function $F(\infty, y)$ is the univariate distribution function for the random variable $\tilde{y}$.

The density function depends on whether the random variables are continuous or discrete. If they are both discrete then the density is given by $f(x, y) = P(\tilde{x} = x \text{ and } \tilde{y} = y)$. If they are both continuous the density is given by

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

This means that the distribution function can be recovered from the density using the formula

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

14.2 Marginal and conditional densities

Consider the following example with two random variables:

	$\tilde{y} = 1$	$\tilde{y} = 2$
$\tilde{x} = 1$	0.1	0.3
$\tilde{x} = 2$	0.2	0.1
$\tilde{x} = 3$	0.1	0.2

The random variable $\tilde{x}$ can take on three possible values, and the random variable $\tilde{y}$ can take on two possible values. The probabilities in the table are the values of the joint density function $f(x, y)$.

Now add a total row and a total column to the table:

	$\tilde{y} = 1$	$\tilde{y} = 2$	$f_{\tilde{x}}(x)$
$\tilde{x} = 1$	0.1	0.3	0.4
$\tilde{x} = 2$	0.2	0.1	0.3
$\tilde{x} = 3$	0.1	0.2	0.3
$f_{\tilde{y}}(y)$	0.4	0.6	1

The sum of the first column is the total probability that $\tilde{y} = 1$, and the sum of the second column is the total probability that $\tilde{y} = 2$. These are the **marginal densities**. For a discrete bivariate random variable $(\tilde{x}, \tilde{y})$ we define the marginal density of $\tilde{x}$ by

$$f_{\tilde{x}}(x) = \sum_{i=1}^{n} f(x, y_i)$$

where the possible values of $\tilde{y}$ are $y_1, ..., y_n$. For the continuous bivariate random variable we define the marginal density of $\tilde{x}$ by

$$f_{\tilde{x}}(x) = \int_{-\infty}^{\infty} f(x, y) dy.$$

From this we can recover the marginal distribution function of $\tilde{x}$ by integrating with respect to x:

$$F_{\tilde{x}}(x) = \int_{-\infty}^{x} f_{\tilde{x}}(t) dt = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(t, y) dy dt.$$

We have already discussed conditional probabilities. We would like to have conditional densities. From the table above, it is apparent that the conditional density of $\tilde{x}$ given the realization of $\tilde{y}$ is $f(x|y) = f(x, y)/f(y)$. To see that this is true, look for the probability that $\tilde{x} = 3$ given $\tilde{y} = 2$. The probability that $\tilde{y} = 2$ is $f_{\tilde{y}}(2) = 0.6$. The probability that $\tilde{x} = 3$ and $\tilde{y} = 2$ is $f(3, 2) = 0.2$. The conditional probability is $f(\tilde{x} = 3|\tilde{y} = 2) = f(3, 2)/f_{\tilde{y}}(2) = 0.2/0.6 = 1/3$. So, the rule is just what we would expect in the discrete case.

What about the continuous case? The same formula works:

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

So, it doesn't matter in this case whether the random variables are discrete or continuous for us to figure out what to do. Both of these formulas require conditioning on a single realization of $\tilde{y}$. It is possible, though, to define the conditional density much more generally. Let A be an event, and let P be the probability measure over events. Then we can write the conditional density

$$f(x|A) = \frac{f(x, A)}{P(A)}$$

where $f(x, A)$ denotes the probability that both x and A occur, written as a density function. For example, if $A = \{x : \tilde{x} \leq x_0\}$, so that A is the event that the realization of $\tilde{x}$ is no greater than x_0, we know that $P(A) = F(x_0)$. Therefore

$$f(x|\tilde{x} \leq x_0) = \begin{cases} \frac{f(x)}{F(x_0)} & \text{if } x \leq x_0 \\ 0 & \text{if } x > x_0 \end{cases}$$

At this point we have too many functions floating around. Here is a table to help with notation and terminology.

Function	Notation	Formula
Density	$f(x, y)$ or $f_{\tilde{x},\tilde{y}}(x, y)$	
Distribution	$F(x, y)$ or $F_{\tilde{x},\tilde{y}}(x, y)$	Discrete: $\sum_{\tilde{y} \leq y} \sum_{\tilde{x} \leq x} f(x, y)$ Continuous: $\int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)dsdt$
Univariate dist.	$F(x)$	$F(x, \infty)$
Marginal density	$f_{\tilde{x}}(x)$	Discrete: $\sum_y f(x, y)$ Continuous: $\int_{-\infty}^{\infty} f(x, y)dy$
Conditional density	$f(x\|y)$ or $f_{\tilde{x}\|y}(x\|y)$	$f(x, y)/f(y)$

The random variables $\tilde{x}$ and $\tilde{y}$ are **independent** if $f_{\tilde{x},\tilde{y}}(x, y) = f_{\tilde{x}}(x)f_{\tilde{y}}(y)$. In other words, the random variables are independent if the bivariate density is the product of the marginal densities. Independence implies that $f(x|y) = f_{\tilde{x}}(x)$ and $f(y|x) = f_{\tilde{y}}(y)$, so that the conditional densities and the marginal densities coincide.

14.3 Expectations

Suppose we have a bivariate random variable $(\tilde{x}, \tilde{y})$. Let $u(x, y)$ be a real-valued function, in which case $u(\tilde{x}, \tilde{y})$ is a univariate random variable. Then the expected value of $u(\tilde{x}, \tilde{y})$ is

$$E[u(\tilde{x}, \tilde{y})] = \sum_y \sum_x u(x, y)f(x, y)$$

in the discrete case and

$$E[u(\tilde{x},\tilde{y})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y)f(x,y)dxdy$$

in the continuous case.

It is still possible to compute the means of the random variables $\tilde{x}$ and $\tilde{y}$ separately. We can do this using the marginal densities. So, for example, in the table above the mean of $\tilde{y}$ is $(0.4)(1) + (0.6)(2) = 1.6$.

A particularly important case is where $u(x, y) = (x - \mu_x)(y - \mu_y)$, where μ_x is the mean of $\tilde{x}$ and μ_y is the mean of $\tilde{y}$. The resulting expectation is called the **covariance** of $\tilde{x}$ and $\tilde{y}$, and it is denoted

$$\sigma_{xy} = Cov(\tilde{x}, \tilde{y}) = E[(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)].$$

Note that σ_{xx} is just the variance of $\tilde{x}$. Also, it is easy to show that $\sigma_{xy} =$ $E[\tilde{x}\tilde{y}] - \mu_x\mu_y$.

The quantity

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

is called the **correlation coefficient** between $\tilde{x}$ and $\tilde{y}$. The following theorems apply to correlation coefficients.

Theorem 20 *If $\tilde{x}$ and $\tilde{y}$ are independent then $\rho_{xy} = \sigma_{xy} = 0$.*

Proof. When the random variables are independent, $f(x,y) = f_{\tilde{x}}(x)f_{\tilde{y}}(y)$. Consequently we can write

$$\sigma_{xy} = E[(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)] = E[\tilde{x} - \mu_x] \cdot E[\tilde{y} - \mu_y)].$$

But each of the expectations on the right-hand side are zero, and the result follows. ■

It is important to remember that the converse is not true: sometimes two variables are not independent but still happen to have a zero covariance. An example is given in the table below. One can compute that $\sigma_{xy} = 0$ but note that $f(2, 6) = 0$ while $f_{\tilde{x}}(2) \cdot f_{\tilde{y}}(6) = (0.2)(0.4) \neq 0$.

	$\tilde{y}=6$	$\tilde{y}=8$	$\tilde{y}=10$
$\tilde{x}=1$	0.2	0	0.2
$\tilde{x}=2$	0	0.2	0
$\tilde{x}=3$	0.2	0	0.2

Theorem 21 $\left|\rho_{xy}\right| \leq 1$.

Proof. Consider the random variable $\tilde{x} - t\tilde{y}$, where t is a scalar. Because variances cannot be negative, we have

$$\begin{aligned} 0 &\leq \sigma^2_{x-ty} \\ &= E[\tilde{x}^2 - 2t\tilde{x}\tilde{y} + t^2\tilde{y}^2] - (\mu_x^2 - 2t\mu_x\mu_y + t^2\mu_y^2) \\ &= \left(E[\tilde{x}^2] - \mu_x^2\right) + t^2\left(E[\tilde{y}^2] - \mu_y^2\right) - 2t\left(E[\tilde{x}\tilde{y}] - \mu_x\mu_y\right) \\ &= \sigma_x^2 + t^2\sigma_y^2 - 2t\sigma_{xy}. \end{aligned}$$

Since this is true for any scalar t, choose

$$t = \frac{\sigma_{xy}}{\sigma_y^2}.$$

Substituting gives us

$$\begin{aligned} 0 &\leq \sigma_x^2 + \left(\frac{\sigma_{xy}}{\sigma_y^2}\right)^2 \sigma_y^2 - 2\frac{\sigma_{xy}}{\sigma_y^2}\sigma_{xy} \\ 0 &\leq \sigma_x^2 - \frac{\sigma_{xy}^2}{\sigma_y^2} \\ \frac{\sigma_{xy}^2}{\sigma_x^2\sigma_y^2} &\leq 1 \\ \left|\frac{\sigma_{xy}}{\sigma_x\sigma_y}\right| &\leq 1. \end{aligned}$$

■

The theorem says that the correlation coefficient is bounded between -1 and 1. If $\rho_{xy} = 1$ it means that the two random variables are perfectly correlated, and once you know the value of one of them you know the value of the other. If $\rho_{xy} = -1$ the random variables are perfectly negatively correlated. This contains just as much information as perfect correlation. If you know that $\tilde{x}$ has attained its highest possible value and $\tilde{x}$ and $\tilde{y}$ are perfectly negatively correlated, then $\tilde{y}$ must have attained its lowest value. Finally, if $\rho_{xy} = 0$ the two variables are perfectly uncorrelated (and possibly independent).

14.4 Conditional expectations

When there are two random variables, $\tilde{x}$ and $\tilde{y}$, one might want to find the expected value of $\tilde{x}$ given that $\tilde{y}$ has attained a particular value or set of values. This would be the conditional mean. We can use the above table for an example. What is the expected value of $\tilde{x}$ given that $\tilde{y} = 8$? $\tilde{x}$ can only take one value when $\tilde{y} = 8$, and that value is 2. So, the conditional mean of $\tilde{x}$ given that $\tilde{y} = 8$ is 2. The conditional mean of $\tilde{x}$ given that $\tilde{y} = 10$ is also 2, but for different reasons this time.

To make this as general as possible, let $u(x)$ be a function of x but not of y. I will only consider the continuous case here; the discrete case is similar. The **conditional expectation** of $u(\tilde{x})$ given that $\tilde{y} = y$ is given by

$$E[u(\tilde{x})|\tilde{y} = y] = \int_{-\infty}^{\infty} u(x)f(x|y)dx = \frac{1}{f_{\tilde{y}}(y)} \int_{-\infty}^{\infty} u(x)f(x,y)dx.$$

Note that this expectation is a function of y but not a function of x. The reason is that x is integrated out on the right-hand side, but y is still there.

14.4.1 Using conditional expectations - calculating the benefit of search

Consider the following search process. A consumer, Max, wants to buy a particular digital camera. He goes to a store and looks at the price. At that point he has three choices: (i) but the camera at that store, (ii) go to another store to check its price, or (iii) go back to a previous store and buy the camera there. Stores draw their prices independently from the distribution

$F(p)$ given by

$$\begin{aligned} p &= 200 \text{ with probability } 0.2 \\ p &= 190 \text{ with probability } 0.3 \\ p &= 180 \text{ with probability } 0.4 \\ p &= 170 \text{ with probability } 0.1 \end{aligned}$$

We want to answer the following question: If the lowest price so far is q, what is the expected benefit from checking one more store?

Let's begin by answering this in the most straightforward way possible. Suppose that $q = 200$, so that the lowest price found so far is the worst possible price. If Max searches one more time there is a 10% chance of finding a price of \$170 and saving \$30, a 40% chance of finding a price of \$180 and saving \$20, a 30% chance of fincing a price of \$190 and saving only \$10, and a 20% chance of finding another store that charges the highest possible price of \$200, in which case the savings are zero. The expected saving is $(.1)(30) + (.4)(20) + (.3)(10) + (.2)(0) = 14$. When $q = 200$, the expected benefit of search is \$14.

Now suppose that $q = 190$, so that the best price found so far is \$190. Max has a 10% chance of finding a price of \$170 and saving \$20, a 40% chance of finding a price of \$180 and saving \$10, a 30% chance of finding the same price and saving nothing, and a 20% chance of finding a higher price of \$200, in which case he also saves nothing. The expected saving is $(.1)(20) + (.4)(10) + (.3)(0) + (.2)(0) = 6$. When the best price found so far is $q = 190$, the expected benefit of search is \$6.

Finally, suppose that $q = 180$. Now there is only one way to improve, which comes by finding a store that charges a price of \$170, leading to a \$10 saving. The probabiliyt of finding such a store is 10%, and the expected saving from search is \$1.

So now we know the answers, and let's use these answers to figure out a general formula, specifically one involving conditional expectations. Note that when Max finds a price of p and the best price so far is q, his benefit is $q - p$ **if** the new price p is lower than the old price q. Otherwise the benefit is zero because he would be better off buying the item at a store he's already found. This "if" statement lends itself to a conditional expectation. In particular, the "if" statement pertains to the conditional expectation $E[q - \tilde{p}|\tilde{p} < q]$, where the expectation is taken over the random variable p. This epxression tells us what the average benefit is *provided* that the benefit is nonnegative. The actual expected benefit is

$$\Pr\{\tilde{p} < q\}E[q - \tilde{p}|\tilde{p} < q],$$

which is the probability that the benefit is positive times the expected benefit conditional on the benefit being positive.

Let's make sure this works using the above example. In particular, let's look at $q = 190$. The conditional expectation is

$$\begin{aligned} E[190 - \tilde{p}|\tilde{p} &< 190] = (.4)(190 - 180) + (.1)(190 - 170) \\ &= 6, \end{aligned}$$

which is exactly what we found before.

The conditional expectation lets us work with more complicated distributions. Suppose that prices are drawn independently from the uniform distribution over the interval $[150, 200]$. Let the corresponding distribution function be $F(p)$ and the density function be $f(p)$. The expected benefit from searching at another store when the lowest price so far is q is

$$\begin{aligned} \Pr\{\tilde{p} &< q\}E[q - \tilde{p}|\tilde{p} < q] = F(q) \int_{150}^{q} [q - p] \frac{f(p)}{F(q)} dp \\ &= \int_{150}^{q} [q - p] f(p) dp. \end{aligned}$$

To see why this works, look at the top line. The probability that $\tilde{p} < q$ is simply $F(q)$, because that is the definition of the distribution function. That gives us the first term on the right-hand side. For the second term, note that we are taking the expectation of $q - p$, so that term is in brackets. To find the conditional expectation, we multiply by the conditional density which is the density of the random variable p divided by the probability that the conditioning event ($\tilde{p} < q$) occurs. We take the integral over the interval $[150, q]$ because outside of this interval the value of the benefit is zero. When we multiply the two terms on the right-hand side of the top line together, we find that the $F(q)$ term cancels out, leaving us with the very simple bottom line. Using it we can find the net benefit of searching at one more store when the best price so far is \$182.99:

$$\int_{150}^{q} [q - p] f(p) dp = \int_{150}^{182.99} [182.99 - p] \frac{1}{50} dp = 10.883.$$

14.4.2 The Law of Iterated Expectations

There is an important result concerning conditional expectations. It is called the **Law of Iterated Expectations**, and it goes like this.

$$E_y[E_x[u(\tilde{x})|\tilde{y} = y]] = E_x[u(\tilde{x})].$$

It's a complicated statement, so let's look at what it means. The inside of the left-hand side is the conditional expectation of $u(\tilde{x})$ given that $\tilde{y}$ takes some value y. As we have already learned, this is a function of y but not a function of x. Let's call it $v(y)$, and $v(\tilde{y})$ is a random variable. So now let's take the expectation of $v(\tilde{y})$. The Law of Iterated Expectations says that $E[v(\tilde{y})] = E_x[u(\tilde{x})]$.

Another way of looking at it is taking the expectation of a conditional expectation. Doing that removes the conditional part.

The best thing to do here is to look at an example to see what's going on. Let's use one of our previous examples:

	$\tilde{y} = 1$	$\tilde{y} = 2$	$f_{\tilde{x}}(x)$
$\tilde{x} = 1$	0.1	0.3	0.4
$\tilde{x} = 2$	0.2	0.1	0.3
$\tilde{x} = 3$	0.1	0.2	0.3
$f_{\tilde{y}}(y)$	0.4	0.6	1

Begin by finding the conditional expectations $E[\tilde{x}|\tilde{y} = 1]$ and $E[\tilde{x}|\tilde{y} = 2]$. We get $E[\tilde{x}|\tilde{y} = 1] = 2$ and $E[\tilde{x}|\tilde{y} = 2] = 11/6$. Now take the expectation over y to get

$$\begin{aligned} E_y[E_x[u(\tilde{x})|\tilde{y} &= y]] = f_{\tilde{y}}(1) \cdot E[\tilde{x}|\tilde{y} = 1] + f_{\tilde{y}}(2) \cdot E[\tilde{x}|\tilde{y} = 2] \\ &= (0.4)(2) + (0.6)(11/6) = 1.9. \end{aligned}$$

Now find the unconditional expectation of $\tilde{x}$. It is

$$\begin{aligned} E_x[\tilde{x}] &= f_{\tilde{x}}(1) \cdot 1 + f_{\tilde{x}}(2) \cdot 2 + f_{\tilde{x}}(3) \cdot 3 \\ &= (0.4)(1) + (0.3)(2) + (0.3)(3) \\ &= 1.9. \end{aligned}$$

It works.

Now let's look at it generally using the continuous case. Begin with

$$\begin{aligned} E_x[u(\tilde{x})] &= \int_{-\infty}^{\infty} u(x) f_{\tilde{x}}(x) dx \\ &= \int_{-\infty}^{\infty} u(x) \left(\int_{-\infty}^{\infty} f(x,y) dy \right) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x) f(x,y) dy dx. \end{aligned}$$

Note that $f(x,y) = f(x|y) f_{\tilde{y}}(y)$, so we can rewrite the above expression

$$\begin{aligned} E_x[u(\tilde{x})] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x) f(x|y) f_{\tilde{y}}(y) dy dx \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} u(x) f(x|y) dx \right] f_{\tilde{y}}(y) dy \\ &= \int_{-\infty}^{\infty} E_x[u(\tilde{x}) | \tilde{y} = y] f_{\tilde{y}}(y) dy \\ &= E_y[E_x[u(\tilde{x}) | \tilde{y} = y]]. \end{aligned}$$

14.5 Problems

1. There are two random variables, $\tilde{x}$ and $\tilde{y}$, with joint density $f(x,y)$ given by the following table.

$f(x,y)$	$\tilde{y} = 10$	$\tilde{y} = 20$	$\tilde{y} = 30$
$\tilde{x} = 1$	.04	0	.20
$\tilde{x} = 2$	.07	0	.18
$\tilde{x} = 3$	.02	.11	.07
$\tilde{x} = 4$	.01	.12	.18

(a) Construct a table showing the distribution function $F(x,y)$.

(b) Find the univariate distributions $F_{\tilde{x}}(x)$ and $F_{\tilde{y}}(y)$.

(c) Find the marginal densities $f_{\tilde{x}}(x)$ and $f_{\tilde{y}}(y)$.

(d) Find the conditional density $f(x|\tilde{y} = 20)$.

(e) Find the mean of $\tilde{y}$.

(f) Find the mean of $\tilde{x}$ conditional on $\tilde{y} = 20$.

(g) Are $\tilde{x}$ and $\tilde{y}$ independent?

(h) Verify that the Law of Iterated Expectations works.

2. There are two random variables, $\tilde{x}$ and $\tilde{y}$, with joint density given by the following table:

$f(x,y)$	$\tilde{y}=3$	$\tilde{y}=8$	$\tilde{y}=10$
$\tilde{x}=1$	0.03	0.02	0.20
$\tilde{x}=2$	0.02	0.12	0.05
$\tilde{x}=3$	0.05	0.01	0.21
$\tilde{x}=4$	0.07	0.11	0.11

(a) Construct a table showing the distribution function $F(x,y)$.

(b) Find the univariate distributions $F_{\tilde{x}}(x)$ and $F_{\tilde{y}}(y)$.

(c) Find the marginal densities $f_{\tilde{x}}(x)$ and $f_{\tilde{y}}(y)$.

(d) Find the conditional density $f(y|\tilde{x}=3)$.

(e) Find the means of $\tilde{x}$ and $\tilde{y}$.

(f) Find the mean of $\tilde{x}$ conditional on $\tilde{y}=3$.

(g) Are $\tilde{x}$ and $\tilde{y}$ independent?

(h) Find $Var(\tilde{x})$ and $Var(\tilde{y})$.

(i) Find $Cov(\tilde{x},\tilde{y})$.

(j) Find the correlation coefficient between $\tilde{x}$ and $\tilde{y}$.

(k) Verify the Law of Iterated Expectations for finding $E_x[\tilde{x}] =.E_y[E_x[\tilde{x}|y]]$.

3. Let $F(x)$ be the uniform distribution over the interval $[a,b]$, and suppose that $c \in (a,b)$. Show that $F(x|x \leq c)$ is the uniform distribution over $[a,c]$.

4. Consider the table of probabilities below:

$f(x,y)$	$\tilde{y}=10$	$\tilde{y}=20$
$\tilde{x}=-1$	0.1	a
$\tilde{x}=+1$	0.3	b

What values must a and b take for $\tilde{x}$ and $\tilde{y}$ to be independent?

CHAPTER

15

Statistics

15.1 Some definitions

The set of all of the elements about which some information is desired is called the **population**. Examples might be the height of all people in Knoxville, or the ACT scores of all students in the state, or the opinions about Congress of all people in the US. Different members of the population have different values for the variable, so we can treat the population variable as a random variable $\tilde{x}$. So far everything we have done in probability theory is about the population random variable. In particular, its mean is μ and its variance is σ^2.

A **random sample** from a population random variable $\tilde{x}$ is a set of independent, identically distributed (IID) random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$, each of which has the same distribution as the parent random variable $\tilde{x}$.

The reason for random sampling is that sometimes it is too costly to measure all of the elements of a population. Instead, we want to infer properties of the entire population from the random sample. This is statistics.

Let $x_1, ..., x_n$ be the outcomes of the random sample. A **statistic** is a function of the outcomes of the random sample which does not contain any unknown parameters. Examples include the sample mean and the sample variance.

15.2 Sample mean

The **sample mean** is given by

$$\bar{x} = \frac{x_1 + ... + x_n}{n}.$$

Note that we use different notation for the sample mean ($\bar{x}$) and the population mean (μ).

The expected value of the sample mean can be found as follows:

$$\begin{aligned} E[\bar{x}] &= \frac{1}{n}\sum_{i=1}^{n} E[\tilde{x}_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu \\ &= \frac{1}{n}(n\mu) = \mu. \end{aligned}$$

So, the expected value of the sample mean is the population mean.

The variance of the sample mean can also be found. To do this, though, let's figure out the variance of the sum of two independent random variables $\tilde{x}$ and $\tilde{y}$.

Theorem 22 *Suppose that $\tilde{x}$ and $\tilde{y}$ are independent. Then $Var(\tilde{x} + \tilde{y}) = Var(\tilde{x}) + Var(\tilde{y})$.*

Proof. Note that

$$(x - \mu_x + y - \mu_y)^2 = (x - \mu_x)^2 + (y - \mu_y)^2 + 2(x - \mu_x)(y - \mu_y).$$

Take the expectations of both sides to get

$$\begin{aligned} Var(\tilde{x} + \tilde{y}) &= E[(\tilde{x} - \mu_x + \tilde{y} - \mu_y)^2] \\ &= E[(\tilde{x} - \mu_x)^2] + E[(\tilde{y} - \mu_y)^2] + 2E[(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)] \\ &= Var(\tilde{x}) + Var(\tilde{y}) + 2Cov(\tilde{x}, \tilde{y}). \end{aligned}$$

But, as shown in Theorem 20, since $\tilde{x}$ and $\tilde{y}$ are independent, $Cov(\tilde{x}, \tilde{y}) = 0$, and the result follows. ■

We can use this theorem to find the variance of the sample mean. Since a random sample is a set of IID random variables, the theorem applies. Also, recall that $Var(a\tilde{x}_i) = a^2 Var(\tilde{x}_i)$. So,

$$\begin{aligned} Var(\bar{x}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{x}_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(\tilde{x}_i) \\ &= \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

This is a really useful result. It says that the variance of the sample mean around the population mean shrinks as the sample size becomes larger. So, bigger samples imply better fits, which we all knew already but we didn't know why.

15.3 Sample variance

We are going to use two different pieces of notation here. One is

$$m^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and the other is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Both of these can be interpreted as the estimates of variance, and many scientific calculators compute both of them. What you should remember is to use m^2 as the computed variance when the random sample coincides with the entire population, and to use s^2 when the random sample is a subset of the population. In other words, you will almost always use s^2, and we refer to s^2 as the **sample variance.** But, m^2 is useful for what we are about to do.

We want to find the expected value of the sample variance s^2. It is easier to find the expected value of m^2 and note that

$$s^2 = \frac{n}{n-1} m^2.$$

We get

$$\begin{aligned} E[m^2] &= \frac{1}{n} E\left[\sum_{i=1}^{n} (\tilde{x}_i - \bar{x})^2\right] \\ &= \frac{1}{n} E\left[\sum_{i=1}^{n} \tilde{x}_i{}^2\right] - E[\bar{x}^2] \end{aligned}$$

which follows from a previous manipulation of variance: $Var(\tilde{x}) = E[(\tilde{x} - \mu)^2] = E[\tilde{x}^2] - \mu^2$. Rearranging that formula and applying it to the sample mean tells us that

$$E[\bar{x}^2] = E[\bar{x}]^2 + Var(\bar{x}),$$

so

$$E[m^2] = \left(\frac{1}{n} \sum_{i=1}^{n} E[\tilde{x}_i^2]\right) - E[\bar{x}]^2 - Var(\bar{x}).$$

But we already know some of these values. We know that $E[\bar{x}] = \mu$ and $Var(\bar{x}) = \sigma^2/n$. Finally, note that since $\tilde{x}_i$ has mean μ and variance σ^2, we have

$$E[\tilde{x}_i^2] = \mu^2 + \sigma^2.$$

Plugging this all in yields

$$\begin{aligned} E[m^2] &= \left(\frac{1}{n} \sum_{i=1}^{n} (\mu^2 + \sigma^2)\right) - \mu^2 - \frac{\sigma^2}{n} \\ &= (\mu^2 + \sigma^2) - \mu^2 - \frac{\sigma^2}{n} \\ &= \frac{n-1}{n} \sigma^2 \end{aligned}$$

Now we can get the mean of the sample variance s^2:

$$\begin{aligned} E[s^2] &= E\left[\frac{n}{n-1}m^2\right] \\ &= \frac{n}{n-1}E[m^2] \\ &= \frac{n}{n-1}\cdot\frac{n-1}{n}\sigma^2 \\ &= \sigma^2. \end{aligned}$$

The reason for using s^2 as the sample variance instead of m^2 is that s^2 has the right expected value, that is, the expected value of the sample variance is equal to the population variance.

We have found that $E[\bar{x}] = \mu$ and $E[s^2] = \sigma^2$. Both $\bar{x}$ and s^2 are statistics, because they depend only on the observed values of the random sample and they have no unknown parameters. They are also **unbiased** because their expected values are equal to the population parameters. Unbiasedness is an important and valuable property. Since we use random samples to learn about the characteristics of the entire population, we want statistics that match, in expectation, the parameters of the population distribution. We want the sample mean to match the population mean in expectation, and the sample variance to match the population variance in expectation.

Is there any intuition behind dividing by $n-1$ in the sample variance instead of dividing by m? Here is how I think about it. The random sample has n observations in it. We only need one observation to compute a sample mean. It may not be a very good or precise estimate, but it is still an estimate. Since we can use the first observation to compute a sample mean, we can use all of the data to compute all of the data to compute a sample mean. This may seem cryptic and obvious, but now think about what we need in order to compute a sample variance. Before we can compute the sample variance, we need to compute the sample mean, and we need at least one observation to do this. That leaves us with $n-1$ observations to compute the sample variance. The terminology used in statistics is *degrees of freedom*. With n observations we have n degrees of freedom when we compute the sample mean, but we only have $n-1$ degrees of freedom when we compute the sample variance because one degree of freedom was used to compute the sample mean. In both calculations (sample mean and sample variance) we divide by the number of degrees of freedom, n for the sample mean $\bar{x}$ and $n-1$ for the sample variance s^2.

15.4 Convergence of random variables

In this section we look at what happens when the sample size becomes infinitely large. The results are often referred to as **asymptotic properties**. We have two main results, both concerning the sample mean. One is called the Law of Large Numbers, and it says that as the sample size grows without bound, the sample mean converges to the population mean. The second is the Central Limit Theorem, and it says that the distribution of the sample mean converges to a normal distribution, regardless of whether the population is normally distributed or not.

15.4.1 Law of Large Numbers

Let $\bar{x}_n$ be the sample mean from a sample of size n. The basic law of large numbers is

$$\bar{x}_n \to \mu \text{ when } n \to \infty.$$

The only remaining issue is what that convergence arrow means.

The **Weak Law of Large Numbers** states that for any $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|\bar{x}_n - \mu| < \varepsilon\right) = 1.$$

To understand this, take any small positive number ε. What is the probability that the sample mean $\bar{x}_n$ is within ε of the population mean? As the sample size grows, the sample mean should get closer and closer to the population mean. And, if the sample mean truly converges to the population mean, the probability that the sample mean is within ε of the population mean should get closer and closer to 1. The Weak Law says that this is true no matter how small ε is.

This type of convergence is called **convergence in probability**, and it is written

$$\bar{x}_n \xrightarrow{P} \mu \text{ when } n \to \infty.$$

The **Strong Law of Large Numbers** states that

$$P\left(\lim_{n\to\infty} \bar{x}_n = \mu\right) = 1.$$

This one is a bit harder to understand. It says that the sample mean is almost sure to converge to the population mean. In fact, this type of convergence is called **almost sure convergence** and it is written

$$\bar{x}_n \xrightarrow{a.e.} \mu \text{ when } n \to \infty.$$

15.4.2 Central Limit Theorem

This is the most important theorem in asymptotic theory, and it is the reason why the normal distribution is so important to statistics.

Let $N(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2. Let $\Phi(z)$ be the distribution function (or cdf) of the normal distribution with mean 0 and variance 1 (or the standard normal $N(0,1)$). To state it, compute the **standardized mean:**

$$Z_n = \frac{\bar{x}_n - E[\bar{x}_n]}{\sqrt{Var(\bar{x}_n)}}.$$

We know some of these values: $E[\bar{x}_n] = \mu$ and $Var(\bar{x}_n) = \sigma^2/n$. Thus we get

$$Z_n = \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}.$$

The **Central Limit Theorem** states that if $Var(\tilde{x}_i) = \sigma^2 < \infty$, that is, if the population random variable has finite variance, then

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z).$$

In words, the distribution of the standardized sample mean converges to the standard normal distribution. This kind of convergence is called **convergence in distribution**.

15.5 Problems

1. You collect the following sample of size $n = 12$:

$$10, 4, -1, 3, -2, 8, 6, 8, 6, 1, -5, 10$$

   Find the sample mean and sample variance.

CHAPTER

16

Sampling distributions

Remember that statistics, like the mean and the variance of a random variable, are themselves random variables. So, they have probability distributions. We know from the Central Limit Theorem that the distribution of the sample mean converges to the normal distribution as the sample size grows without bound. The purpose of this chapter is to find the distributions for the mean and other statistics when the sample size is finite.

16.1 Chi-square distribution

The chi-square distribution turns out to be fundamental for doing statistics because it is closely related to the normal distribution. Chi-square random variables can only have nonnegative values. It turns out to be the distribution you get when you square a normally distributed random variable.

The density function for the chi-square distribution with n degrees of freedom is

$$f(x) = \frac{(x/2)^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2\Gamma(n/2)}$$

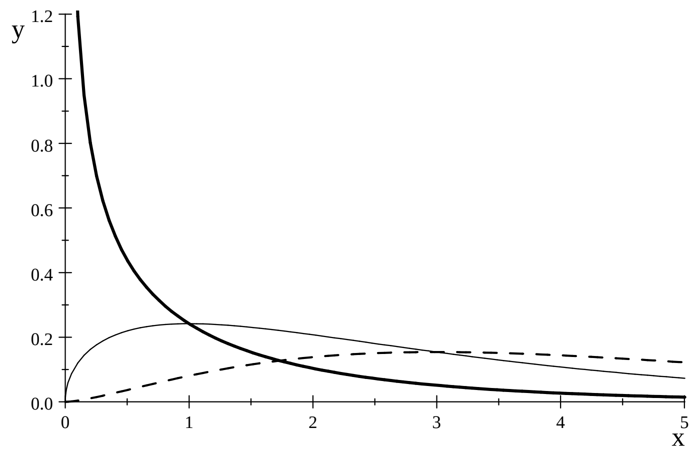


Figure 16.1: Density for the chi-square distribution. The thick line has 1 degree of freedom, the thin line has 3, and the dashed line has 5.

where $\Gamma(a)$ is the Gamma function defined as $\Gamma(a) = \int_0^\infty y^{a-1}e^{-y}dy$ for $a > 0$. We use the notation $\tilde{y} \sim \chi_n^2$ to denote a chi-square random variable with n degrees of freedom. The density for the chi-square distribution with different degrees of freedom is shown in Figure 16.1. The thick line is the density with 1 degree of freedom, the thin line has 3 degrees of freedom, and the dashed line has 5. Changing the degrees of freedom radically changes the shape of the density function.

One strange thing about the chi-square distribution is its mean and variance. The mean of a χ_n^2 random variable is n, the number of degrees of freedom, and the variance is $2n$.

The relationship between the standard normal and the chi-square distribution is given by the following theorem.

Theorem 23 *If $\tilde{x}$ has the standard normal distribution, then the random variable $\tilde{x}^2$ has the chi-square distribution with 1 degree of freedom.*

Proof. The distribution function for the variable $\tilde{y} = \tilde{x}^2$ is

$$\begin{aligned} F_{\tilde{y}}(y) &= P(\tilde{y} \leq y) \\ &= P(\tilde{x}^2 \leq y) \\ &= P(\sqrt{y} \leq x \leq \sqrt{y}) \\ &= 2P(0 \leq x \leq \sqrt{y}) \end{aligned}$$

where the last equality follows from the fact that the standard normal distribution is symmetric around its mean of zero. From here we can compute

$$F_{\tilde{y}}(y) = 2\int_0^{\sqrt{y}} f_{\tilde{x}}(x)dx$$

where $f_{\tilde{x}}(x)$ is the standard normal density. Using Leibniz's' rule, differentiate this with respect to y to get the density of $\tilde{y}$:

$$f_{\tilde{y}}(y) = 2f_{\tilde{x}}(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$$

where the last term is the derivative of $\sqrt{y}$ with respect to y. Plug in the formula for the normal density to get

$$f_{\tilde{y}}(y) = \frac{1}{\sqrt{2\pi}}e^{-y/2} \cdot y^{-1/2} = \frac{(y/2)^{-1/2}e^{-y/2}}{2\sqrt{\pi}}.$$

This looks exactly like the formula for the density of χ_1^2 except for the denominator. But, $\Gamma(1/2) = \sqrt{\pi}$, and the formula is complete. ■

Both the normal distribution and the chi-square distribution have the property that they are additive. That is, if $\tilde{x}$ and $\tilde{y}$ are independent normally distributed random variables, then $\tilde{z} = \tilde{x} + \tilde{y}$ is also normally distributed. If $\tilde{x}$ and $\tilde{y}$ are independent chi-square random variables with n_x and n_y degrees of freedom, respectively, then $\tilde{z} = \tilde{x} + \tilde{y}$ is has a chi-square distribution with $n_x + n_y$ degrees of freedom.

16.2 Sampling from the normal distribution

We use the notation $\tilde{x} \sim N(\mu, \sigma^2)$ to denote a random variable that is distributed normally with mean μ and variance σ^2. Similarly, we use the notation

$\tilde{x} \sim \chi_n^2$ when $\tilde{x}$ has the chi-square distribution with n degrees of freedom. The next theorem describes the distribution of the sample statistics of the standard normal distribution.

Theorem 24 *Let $x_1, ..., x_n$ be an IID random sample from the standard normal distribution $N(0, 1)$. Then*

(a) $\bar{x} \sim N(0, \frac{1}{n})$.

(b) $\sum(x_i - \bar{x})^2 \sim \chi_{n-1}^2$.

(c) *The random variables $\bar{x}$ and $\sum(x_i - \bar{x})^2$ are statistically independent.*

The above theorem relates to random sampling from a *standard* normal distribution. If $x_1, ..., x_n$ are an IID random sample from the general normal distribution $N(\mu, \sigma^2)$, then

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}).$$

Also,

$$\frac{\sum(x_i - \bar{x})^2}{\sigma^2} \sim \chi_{n-1}^2.$$

To figure out the distribution of the sample variance, first note that since the sample variance is

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

and $E[s^2] = \sigma^2$, we get that

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

This looks more useful than it really is. In order to get it you need to know the population variance, σ^2, in which case you don't really need the sample variance, s^2. The next section discusses the sample distribution of s^2 when σ^2 is unknown.

The properties to take away from this section are that the sample mean of a normal distribution has a normal distribution, and the sample variance of a normal distribution has a chi-square distribution (after multiplying by $(n-1)/\sigma^2$).

16.3 t and F distributions

In econometrics the two most frequently encountered distributions are the t distribution and the F distribution. To briefly say where they arise, in econometrics one often runs a linear regression of the form

$$y_t = b_0 + b_1 x_{1,t} + ... + b_k x_{k,t} + \tilde{\varepsilon}_t$$

where t is the observation number, $x_{i,t}$ is an observation of the i-th explanatory variable, and $\tilde{\varepsilon}_t \sim N(0, \sigma^2)$. One then estimates the coefficients $b_0, ..., b_k$ in ways we have described earlier in this book. One uses a t distribution to test whether the individual coefficients $b_0, ..., b_k$ are equal to zero. If the test rejects the hypothesis, that explanatory variable has a statistically significant impact on the dependent variable. The F test is used to test if linear combinations of the coefficients are equal to zero.

Let's start with the **t distribution**. Graphically, its density looks like the normal density but with fatter tails, as in Figure 16.2. To get a t distribution, let $\tilde{x}$ have a standard normal distribution and let $\tilde{y}$ have a chi-square distribution with n degrees of freedom. Assume that $\tilde{x}$ and $\tilde{y}$ are independent. Construct the random variable $\tilde{t}$ according to the formula

$$t = \frac{x}{(y/n)^{1/2}}.$$

Then $\tilde{t}$ has a t distribution with n degrees of freedom. In shorthand,

$$\tilde{t}_n \sim \frac{N(0,1)}{\sqrt{\chi^2_n/n}}.$$

Now look back at Theorem 24. If we sample from a standard normal, the sample mean has an $N(0,1)$ distribution and the sum of the squared deviations has a χ^2_n distribution. So, we get a t distribution when we use the sample mean in the numerator and something like the sample variance in the denominator. The trick is to figure out exactly what we need.

If $\tilde{x} \sim N(\mu, \sigma^2)$, then we know that $\bar{x} \sim N(\mu, \sigma^2/n)$, in which case

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

We also know that $\left(\sum(x_i - \bar{x})^2\right)/\sigma^2 \sim \chi^2_{n-1}$. Putting this all together yields

$$\tilde{t} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\left(\frac{\sum(x_i-\bar{x})^2}{\sigma^2}\right)/(n-1)}} = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}. \tag{16.1}$$

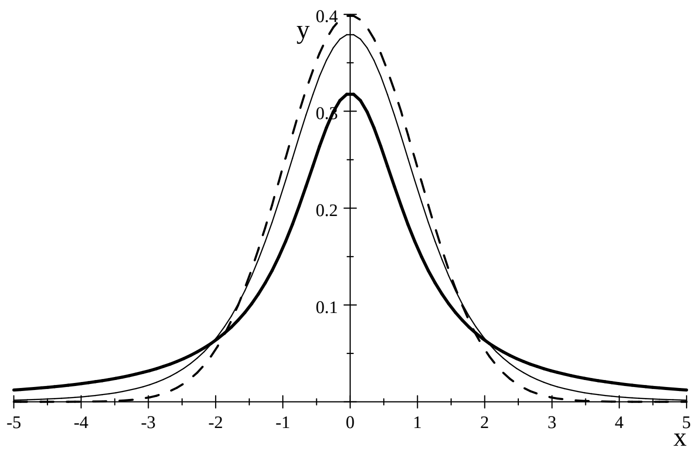


Figure 16.2: Density for the t distribution. The thick curve has 1 degree of freedom and the thin curve has 5. The dashed curve is the standard normal density.

The random variable $\tilde{t}$ has a t distribution with $n-1$ degrees of freedom. Also notice that it does not depend on σ^2. It does depend on μ, but we will discuss the meaning of this more in the next chapter. The statistic computed in expression (16.1) is commonly referred to as a **t-statistic**.

Like the t distribution, the **F distribution** is a ratio of two other distributions. In this case it is the ratio of two chi-square distributions. The formula is

$$F_{m,n} \sim \frac{\chi_m^2/m}{\chi_n^2/n},$$

and the density function is shown in Figure 16.3. Because chi-square distributions assign positive probability only to non-negative outcomes, F distributions also assign positive probability only to non-negative outcomes.

The F distribution and the t distribution are related. If $\tilde{t}_n$ has the t distribution with n degrees of freedom, then

$$(\tilde{t}_n)^2 \sim F_{1,n}.$$

Applying this to expression (16.1) tells us that the following sample statistic

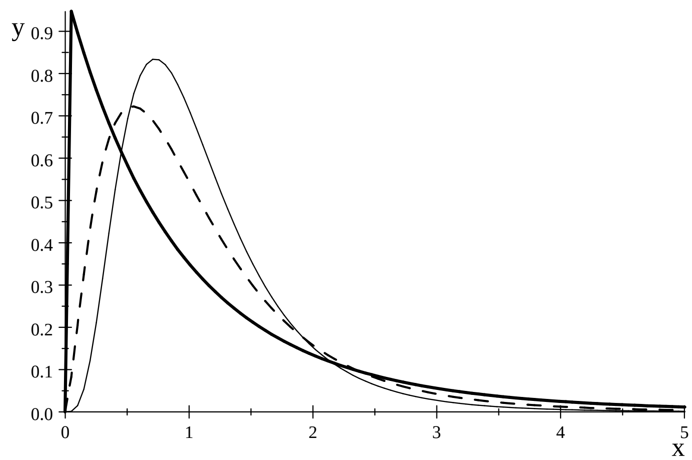


Figure 16.3: Density function for the F distribution. The thick curve is $F_{2,20}$, the dashed curve is $F_{5,20}$, and the thin curve is $F_{10,20}$.

has an F distribution:

$$\tilde{F} = \frac{(\bar{x} - \mu)^2}{s^2/n} \sim F_{1,n-1}.$$

16.4 Sampling from the binomial distribution

Recall that the binomial distribution is used for computing the probability of getting no more than s successes from n independent trials when the probability of success in any trial is p.

A series of random trials is going to result in a sequence of successes or failures, and we can use the random variable $\tilde{x}$ to capture this. Let $x_i = 1$ if there was a success in trial i and $x_i = 0$ if there was a failure in trial i. Then $\sum_{i=1}^{n} x_i$ is the number of successes in n trials, and $\bar{x} = \left(\sum x_i\right)/n$ is the average number of successes. Notice that $\bar{x}$ is also the **sample frequency**, that is, the fraction of successes in the sample. The sample frequency has the following properties:

$$E[\bar{x}] = p$$

$$Var(\bar{x}) = \frac{p(1-p)}{n}.$$

CHAPTER

17

Hypothesis testing

The tool of statistical analysis has two primary uses. One is to describe data, and we do this using such things as the sample mean and the sample variance. The other is to test hypotheses. Suppose that you have, for example, a sample of UT graduates and a sample of Vanderbilt graduates, both from the class of 2002. You may want to know whether or not the average UT grad makes more than the national average income, which was about $35,700 in 2006. You also want to know if the two classes have the same income. You would perform hypothesis tests to either support or reject the hypotheses that UT grads have higher average earnings than the national average and that both UT grads and Vanderbilt grads have the same average income.

In general, hypothesis testing involves the value of some parameter θ that is determined by the data. There are two types of tests. One is to determine if the realized value of θ is in some set Ω_0. The other is to compute two different values of θ from two different samples and determine if they are the same or if one is larger than the other.

17.1 Structure of hypothesis tests

The first part of testing a hypothesis is forming one. In general a hypothesis takes the form of $\theta \in \Omega_0$, where Ω_0 is a nonempty set of values for θ. It could be a single value, or it could be a range of values. The statement $\theta \in \Omega_0$ is called the **null hypothesis**. The **alternative hypothesis** is that $\theta \notin \Omega_0$. We typically write these as

$$\boxed{\begin{array}{c} H_0 \text{ (Null hypothesis): } \theta \in \Omega_0 \\ \text{vs.} \\ H_1 \text{ (Alternative hypothesis): } \theta \notin \Omega_0. \end{array}}$$

The form of the alternative hypothesis is determined completely by the form of the null hypothesis. So, if H_0 is $\theta = \theta_0$, H_1 is $\theta \neq \theta_0$. If H_0 is $\theta \leq \theta_0$, H_1 is $\theta > \theta_0$. And so on.

Another issue in hypothesis testing is that hypotheses can be rejected but they cannot be accepted. So, you can establish that something is false, but not that something is true. Because of this, empirical economists often make the null hypothesis something they would like to be false. If they can reject the null hypothesis, that lends support to the alternative hypothesis. For example, if one thinks that the variance of returns to the Dow-Jones Industrial Average is smaller than the variance of returns to the S&P 500 index, one would form the null hypothesis that the variance is at least as great for the Dow and then try to reject it. When running linear regressions, one tests the null hypothesis that the coefficients are zero.

One uses a **statistical test** to either reject or support the null hypothesis. The nature of the test is as follows. First we compute a **test statistic** for θ. Let's call it T. For example, if θ is the mean of the population distribution, T would be the sample mean. As we know, the value of T is governed by a random process. The statistical test identifies a range of values A for the random variable T such that if $T \in A$ the null hypothesis is "accepted" and if $T \notin A$ the null hypothesis is rejected. The set A is called the **critical region**, and it is important to note that A and Ω_0 are two completely different things. For example, a common null hypothesis is H_0: $\theta = 0$. In that case $\Omega_0 = \{0\}$. But, we do not reject the null hypothesis if T is anything but zero, because then we would reject the hypothesis with probability 1. Instead, we reject the hypothesis if T is sufficiently far from zero, or, in our new terminology, if T is outside of the critical region A.

Statistical tests can have errors because of the inherent randomness. It might be the case that $\theta \in \Omega_0$, so that the null hypothesis is really true, but $T \notin A$ so we reject the null hypothesis. Or, it might be the case that $\theta \notin \Omega_0$ so that the null hypothesis is really false, but $T \in A$ and we "accept" it anyway. The possible outcomes of the test are given in the table below.

		Value of test statistic T	
		$T \in A$	$T \notin A$
True value of parameter θ	$\theta \in \Omega_0$	Correctly "accept" null	Incorrectly reject null Type I error
	$\theta \notin \Omega_0$	Incorrectly "accept" null Type II error	Correctly reject null

A **type I error** occurs when one rejects a true null hypothesis. A **type II error** occurs when a false null hypothesis is not rejected. A problem arises because reducing the probability of a type I error generally increases the probability of a type II error. After all, reducing the probability of a type I error means rejecting the hypothesis less often, whether it is true or not.

Let $F(z|\theta)$ be the distribution of the test statistic z conditional on the value of the parameter θ. The entire previous chapter was about these distributions. If the null hypothesis is really true, the probability that the null is "accepted" is $F(A|\theta \in \Omega_0)$. This is called the **confidence level**. This probability of a type I error is $1 - F(A|\theta \in \Omega_0)$. This probability is called the **significance level**. The standard is to use a 5% significance level, but 10% and 1% significance levels are also reported. The 5% significance level corresponds to a 95% confidence level. I usually interpret the confidence level as the level of certainty with which the null hypothesis is false when it is rejected. So, if I reject the null hypothesis with a 95% confidence level, I am 95% sure that the null hypothesis is really false.

Here, then, is the "method of proof" entailed in statistical analysis. Think of a statement you want to be true. Make this the alternative hypothesis. The null hypothesis is therefore a statement you would like to reject. Construct a test statistic related to the null hypothesis. Reject the null hypothesis if you are 95% sure that it is false given the value of the test statistic.

For example, suppose that you want to test the null hypothesis that the mean of a normal distribution is 0. This makes the hypothesis

$$\begin{aligned} H_0 &: \mu = 0 \\ &\text{vs.} \\ H_1 &: \mu \neq 0 \end{aligned}$$

You take a sample $x_1, ..., x_n$ and compute the sample mean $\bar{x}$ and sample variance s^2. Construct the test statistic

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} \tag{17.1}$$

which we know from Chapter 16 has a t distribution with $n-1$ degrees of freedom. In this case μ is the hypothesized mean, which is equal to zero. Now we must construct a critical range A for the test statistic t. Our critical range will be an interval around zero so that we reject the null hypothesis if t is too far from zero. We call this interval the **95% confidence interval**, and it takes the form (t_L, t_H). Let $T_{n-1}(t)$ be the distribution function for the t distribution with $n-1$ degrees of freedom. Then the endpoints of the confidence interval satisfy

$$\begin{aligned} T_{n-1}(t_L) &= 0.025 \\ T_{n-1}(t_H) &= 0.975 \end{aligned}$$

The first line says that the probability that t is lower than t_L is 2.5%, and the second says that the probability that t is higher than t_H is 2.5%. Combining these means that the probability that the test statistic t is outside of the interval (t_L, t_H) is 5%.

All of this is shown in Figure 17.1. The probability that $t \leq t_L$ is 2.5%, as shown by the shaded area. The probability that $t \geq t_H$ is also 2.5%. The probability that t is between these two points is 95%, and so the interval (t_L, t_H) is the 95% confidence interval. The hypothesis is rejected if the value of t lies outside of the confidence interval.

Another way to perform the same test is to compute $T_{n-1}(t)$, where t is the test statistic given in (17.1) above. Reject the hypothesis if

$$T_{n-1}(t) < 0.025 \text{ or } T_{n-1}(t) > 0.975.$$

If the first of these inequalities hold, the value of the test statistic is outside of the confidence interval and to the left, and if the one on the right holds

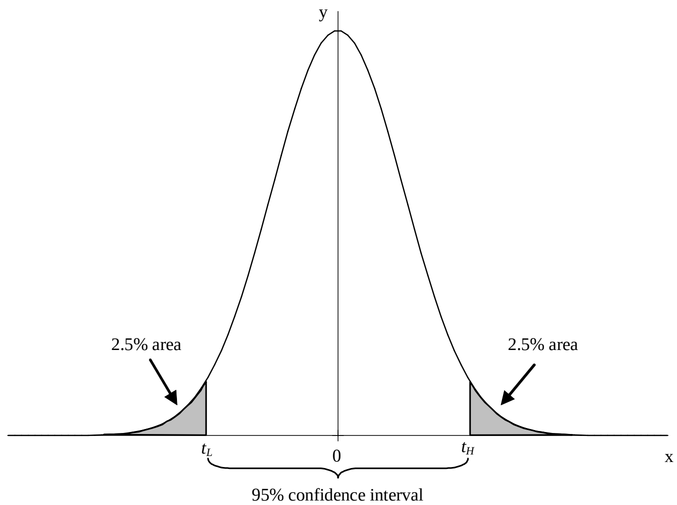


Figure 17.1: Two-tailed hypothesis testing

the test statistic is outside of the confidence interval and to the right. The null hypothesis cannot be rejected if $0.025 \leq T_{n-1}(t) \leq 0.975$.

17.2 One-tailed and two-tailed tests

The tests described above are **two-tailed tests**, because rejection is based on the test statistic lying in one of the two tails of the distribution. Two-tailed tests are used when the null hypothesis is an equality hypothesis, so that it is violated if the test statistic is either too high or too low.

A **one-tailed test** is used for inequality-based hypotheses, such as the one below:

$$\begin{aligned} H_0 &: \mu \geq 0 \\ &\quad \text{vs.} \\ H_1 &: \mu < 0 \end{aligned}$$

In this case the null hypothesis is rejected if the test statistic is both far from zero and negative. Large test statistics are compatible with the null hypothesis as long as they are positive. This contrasts with the two-sided tests where large test statistics led to rejection regardless of the sign.

To test the null hypothesis that the mean of a normal distribution is nonnegative, compute the test statistic given in (17.1). We know from Chapter 16 that it has the t distribution with $n-1$ degrees of freedom. Letting $T_{n-1}(t)$ be the cdf of the t distribution with $n-1$ degrees of freedom, we reject the null hypothesis at the 95% confidence level (or 5% significance level) if

$$T_{n-1}(t) < 0.05.$$

There are two differences between the one-tailed criterion for rejection and the two-tailed criterion in the previous section. One difference is that the one-tailed criterion can only be satisfied one way, with $T_{n-1}(t)$ small, while the two-tailed criterion can be satisfied two ways, with $T_{n-1}(t)$ either close to zero or close to one. The second difference is that the one-tailed criterion has a cutoff point of 0.05, while the two-tailed criterion has a lower cutoff point half as big at 0.025. The reason for this is that the two-tailed test splits the 5% probability mass equally between the two tails, while the one-tailed criterion puts the whole 5% in the lower tail.

The following table gives the rules for the one-tailed and two-tailed tests with significance level α and confidence level $1-\alpha$. The test statistic is z with distribution function $G(z)$, and the hypotheses concern some parameter θ.

<table>
<tr><th>Type of test</th><th>Hypothesis</th><th>Reject H_0 if</th><th>p-value</th></tr>
<tr><td>Two-tailed</td><td>$H_0: \theta = \theta_0$
vs.
$H_1: \theta \neq \theta_0$</td><td>$G(z) < \frac{\alpha}{2}$
or
$G(z) > 1 - \frac{\alpha}{2}$</td><td>$2[1 - G(|z|)]$</td></tr>
<tr><td>Upper one-tailed</td><td>$H_0: \theta \leq \theta_0$
vs.
$H_1: \theta > \theta_0$</td><td>$G(z) > 1 - \alpha$</td><td>$1 - G(z)$</td></tr>
<tr><td>Lower one-tailed</td><td>$H_0: \theta \geq \theta_0$
vs.
$H_1: \theta < \theta_0$</td><td>$G(z) < \alpha$</td><td>$G(z)$</td></tr>
</table>

The **p-value** can be thought of as the exact significance level for the test. The null hypothesis is rejected if the p-value is smaller than α, the desired significance level.

17.3 Examples

17.3.1 Example 1

The following sequence is a random sample from the distribution $N(\mu, \sigma^2)$. The task is to test hypotheses about μ. The sequence is: 56, 74, 55, 66, 51, 61, 55, 48, 48, 47, 56, 57, 54, 75, 49, 51, 79, 59, 68, 72, 64, 56, 64, 62, 42.

Test the null hypothesis that $\mu = 65$. This is a two-tailed test based on the statistic in equation (17.1). We compute $\bar{x} = 58.73$, $s = 9.59$, and $n = 25$. We get

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{58.73 - 65}{9.59/5} = -3.27.$$

The next task is to find the p-value using $T_{24}(t)$, the t distribution with $n - 1 = 24$ degrees of freedom. Excel allows you to do this, but it only allows positive values of t. So, use the command

$$\begin{aligned} &= \text{TDIST}(\,|t|\,, \text{degrees of freedom, number of tails}) \\ &= \text{TDIST}(3.27, 24, 2) = 0.00324 \end{aligned}$$

Thus, we can reject the null hypothesis at the 5% significance level. In fact, we are 99.7% sure that the null hypothesis is false. Maple allows for both positive and negative values of t. Using the table above, the p-value can be found using the formula

$$2\,[1 - \text{TDist}(\,|t|\,, \text{degrees of freedom})]$$

$$\text{TDist}(3.27, 24) = 0.99838.$$

The p-value is $2(1 - 0.99838) = 0.00324$, which is the same answer we got from Excel.

Now test the null hypothesis that $\mu = 60$. This time the test statistic is $t = -0.663$ which yields a p-value of 0.514. We cannot reject this null hypothesis.

What about the hypothesis that $\mu \geq 65$? The sample mean is 58.73, which is less than 65, so we should still do the test. (If the sample mean had been above 65, there is no way we could reject the hypothesis.) This is a one-tailed test based on the same test statistic which we have already computed, $t = -3.27$. We have to change the Excel command to reduce the number of tails:

=TDIST(3.27, 24, 1) = 0.00162

Once again we reject the hypothesis at the 5% level. Notice, however, that the p-value is twice what it was for the two-tailed test. This is as it should be, as you can figure out by looking at the above table.

17.3.2 Example 2

You draw a sample of 100 numbers drawn from a normal distribution with mean μ and variance σ^2. You compute the sample mean and sample variance, and they are $\bar{x} = 10$ and $s^2 = 16$. The null hypothesis is

$$H_0 : \mu = 9$$

Do the data support or reject the hypothesis?

Compute the t-statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{10 - 9}{\sqrt{16}/\sqrt{100}} = 2.5$$

This by itself does not tell us anything. We must plug it into the appropriate t distribution. The t-statistic has $100 - 1 = 99$ degrees of freedom, and we can find

$$\text{TDist}(2.5, 99) = 0.992\,97$$

and we reject if this number is either less than 0.025 or greater than 0.975. It is greater than 0.975, so we can reject the hypothesis. Another way to see it is by computing the p-value

$$p = 2(1 - \text{TDist}(2.5, 99)) = 0.014$$

which is much smaller than the 0.05 required for rejection.

17.3.3 Example 3

Use the same information as example 2, but instead test the null hypothesis

$$H_0 : \mu = 10.4.$$

Do the data support or reject this hypothesis?

Compute the t-statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{10 - 10.4}{\sqrt{16}/\sqrt{100}} = -1.0$$

We can find

$$\text{TDist}(-1, 99) = 0.16$$

:and we reject if this number is either less than 0.025 or greater than 0.975. It is not, so we cannot reject the hypothesis. Another way to see it is by computing the p-value

$$p = 2(1 - \text{TDist}(1, 99)) = 0.32$$

which is much larger than the 0.05 required for rejection.

17.4 Problems

1. Consider the following random sample from a normal distribution with mean μ and standard deviation σ^2:

$$134, 99, 21, 38, 98, 19, 53, -52, 115, 30,$$
$$65, 149, 4, 55, 43, 26, 122, 47, 54, 97,$$
$$87, 34, 114, 44, 26, 98, 38, 24, 30, 86.$$

   (a) Test the hypothesis that $\mu = 0$. Do the data support or reject the hypothesis?

   (b) Test the hypothesis that $\mu = 30$.

   (c) Test the hypothesis that $\mu \geq 65$.

   (d) Test the hypothesis that $\mu \leq 100$.

2. Answer the following questions based on this random sample generated from a normal distribution with mean μ and variance σ^2.

$$89, 51, 12, 17, 71$$
$$39, 47, 37, 42, 75$$
$$78, 67, 20, 9, 9$$
$$44, 71, 32, 13, 61$$

(a) What are the best estimates of μ and σ^2?

(b) Test the hypothesis that $\mu = 40$. Do the data support or reject the hypothesis?

(c) Test the hypothesis that $\mu = 60$. Do the data support or reject the hypothesis?

CHAPTER

18

Solutions to end-of-chapter problems

Solutions for Chapter 2

1. **(a)** $f'(x) = 72x^2\left(x^3+1\right) + \frac{3}{x} + \frac{20}{x^5}$

   (b) $f'(x) = -\frac{20}{(4x-2)^6}$

   (c) $f'(x) = -e^{2x-14x^3}\left(42x^2-2\right)$

   (d) $f'(x) = \frac{9}{x^{1.3}} - \frac{2.7}{x^{1.3}}\ln x$

   (e) $f'(x) = \frac{c}{(d-cx)^2}\left(b-ax^2\right) - 2a\frac{x}{d-cx}$

2. **(a)** $f'(x) = 24x - 24$

   (b) $g'(x) = \frac{1}{4x^3} - \frac{1}{2x^3}\ln 3x = (1-2\ln 3x)/(4x^3)$

   (c) $h'(x) = -4(6x-2)/(3x^2-2x+1)^5$

   (d) $f'(x) = (1-x)e^{-x}$

(e)

$$g'(x) = \frac{9}{(9x-8)^2}\left(2x^2-3\right)\sqrt{5x^3+6} - 4\frac{x}{9x-8}\sqrt{5x^3+6} - \frac{15}{2}\frac{x^2}{9x-8}\frac{2x^2-3}{\sqrt{5x^3+6}}$$

3. Let $f(x) = x^2$. Then

$$\begin{aligned} f'(x) &= \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} \\ &= \lim_{h\to 0}\frac{(x+h)^2-x^2}{h} \\ &= \lim_{h\to 0}\frac{x^2+2xh+h^2-x^2}{h} \\ &= \lim_{h\to 0}\frac{2xh+h^2}{h} \\ &= \lim_{h\to 0}(2x+h) \\ &= 2x. \end{aligned}$$

4. Use the formula

$$\begin{aligned} \lim_{h\to 0}\frac{f(x+h)-f(x)}{h} &= \lim_{h\to 0}\frac{\frac{1}{x+h}-\frac{1}{x}}{h} \\ &= \lim_{h\to 0}\frac{\frac{x}{x(x+h)}-\frac{x+h}{x(x+h)}}{h} \\ &= \lim_{h\to 0}\frac{\frac{-h}{x(x+h)}}{h} \\ &= \lim_{h\to 0}\frac{-h}{xh(x+h)} \\ &= \lim_{h\to 0}\frac{-1}{x(x+h)} = -\frac{1}{x^2} \end{aligned}$$

5. **(a)** Compute $f'(3) = -18 < 0$, so it is decreasing.

(b) Compute $f'(13) = \frac{1}{13} > 0$, so it is increasing.

(c) Compute $f'(4) = -5.0e^{-4} < 0$, so it is decreasing.

(d) Compute $f'(2) = \frac{9}{16} > 0$, so it is increasing.

6. **(a)** $f'(x) = \frac{3}{x^2+4x} - (3x-2)\frac{2x+4}{(x^2+4x)^2}$ and $f'(-1) = \frac{1}{9} > 0$. It is increasing.

   (b) $f'(x) = -\frac{1}{x\ln^2 x}$ and $f'(e) = -\frac{1}{e} < 0$. It is decreasing.

   (c) $f'(x) = 10x + 16$ and $f'(-6) = -44 < 0$. It is decreasing.

7. **(a)** The FOC is $f'(x) = 10 - 8x = 0$ so $x^* = 5/4$. Compute $f''(5/4) = -8 < 0$, so this is a maximum.

   (b) The FOC is $f'(x) = \frac{84}{x^{0.3}} - 6 = 0$ so $x^* = 14^{1/0.3}$. Compute $f''(14^{1/0.3}) = -2.721\,7 \times 10^{-4} < 0$, so this is a maximum.

   (c) The FOC is $f'(x) = 4 - \frac{3}{x} = 0$ so $x^* = 3/4$. Compute $f''(3/4) = \frac{16}{3} > 0$, so this is a minimum.

8. **(a)** The foc is $f'(x) = 8x - 24 = 0$, which is solved when $x = 3$. Also, $f''(x) = 8 > 0$, so it is a minimum.

   (b) The foc is $f'(x) = 20/x - 4 = 0$, which is solved when $x = 5$. Also, $f''(x) = -20/x^2 < 0$, so it is a maximum.

   (c) The foc is

$$f'(x) = \frac{x+1}{(x+2)^2} - \frac{1}{x+2} + 6 = 0,$$

   which has two solutions: $x = -\frac{13}{6}$ and $x = -\frac{11}{6}$. The second derivative is $f''(x) = \frac{2}{(x+2)^2} - 2\frac{x+1}{(x+2)^3}$, and $f''(-\frac{13}{6}) = -432$ while $f''(-\frac{11}{6}) = 432$. The function has a local maximum when $x = -\frac{13}{6}$ and a local minimum when $x = -\frac{11}{6}$.

9. **(a)** Compute $f''(x) = 2a$. Need $a < 0$.

   (b) Need $a > 0$.

10. **(a)** The problem is

$$\max_m b(m) - c(m)$$

   The FOC is

$$b'(m) = c'(m)$$

   Interpretation is that marginal benefit equals marginal cost.

(b) We need $b''(m) - c''(m) \leq 0$. A better way is to assume $b''(m) \leq 0$ and $c''(m) \geq 0$. This is diminishing marginal benefit and increasing marginal cost.

(c) The problem is

$$\max_m wm - c(m)$$

The FOC is

$$w = c'(m)$$

Interpretation is that marginal effort cost equals the wage.

(d) $c''(m) \geq 0$, or increasing marginal effort cost.

11. The first-order condition is

$$\pi'(L) = \frac{90}{\sqrt{L}} - 90 = 0.$$

Solving for L gives us $L = 1$. Bilco devotes 1 unit of labor to widget production and the other 59 to gookey production. It produces $W = 20(1)^{1/2} = 20$ widgets and $G = 30(59) = 1770$ gookeys.

Solutions for Chapter 3

1. **(a)** $(26, -3, 33, 25)$

(b) $\bar{x} \leq \bar{y}$ and $\bar{x} < \bar{y}$ and $\bar{x} \ll \bar{y}$

(c) 77

(d) Yes. $\bar{x} \cdot \bar{x} = 65$, $\bar{y} \cdot \bar{y} = 135$, and $(\bar{x} + \bar{y}) \cdot (\bar{x} + \bar{y}) = 354$. We have

$$\sqrt{\bar{x} \cdot \bar{x}} + \sqrt{\bar{y} \cdot \bar{y}} = \sqrt{65} + \sqrt{135} = 19.681$$

and

$$\sqrt{(\bar{x} + \bar{y}) \cdot (\bar{x} + \bar{y})} = \sqrt{354} = 18.815.$$

2. **(a)** $(18, -8, -48, -20)$

(b) -7

(c) $\sqrt{\bar{x} \cdot \bar{x}} = \sqrt{65} = 8.0623$, $\sqrt{\bar{y} \cdot \bar{y}} = \sqrt{26} = 5.099$, and $\sqrt{(\bar{x} + \bar{y}) \cdot (\bar{x} + \bar{y})} = \sqrt{77} = 8.7750$, which is smaller than $\sqrt{65} + \sqrt{26} = 13.161$.

3. **(a)** $f_x(x,y) = 8x - 12y + 18$

   (b) $f_y(x,y) = 6y - 12x$

   (c) $\left(\frac{9}{8}, \frac{9}{4}\right)$

4. **(a)** $f_x(x,y) = 16y - 4$.

   (b) $f_y(x,y) = 16x - 2/y^2$.

   (c) The two foc's are $16y - 4 = 0$ and $16x - 2/y^2 = 0$. The first one implies that $y = \frac{1}{4}$. Plugging this into the second expressions yields

$$\begin{aligned} 16x - \frac{2}{y^2} &= 0 \\ 16x &= \frac{2}{(\frac{1}{4})^2} = 32 \\ x &= 2 \end{aligned}$$

   The critical point is $(x,y) = (2, \frac{1}{4})$.

5. **(a)** $3 \ln x + 2 \ln y = k$

   (b) Implicitly differentiate $3 \ln x + 2 \ln y(x) = k$ with respect to x to get

$$\frac{3}{x} + \frac{2}{y}\frac{dy}{dx} = 0$$

$$\frac{dy}{dx} = -\frac{3/x}{2/y} = -\frac{3y}{2x}$$

6. **(a)** $\pi(q) = p(q)q - cq = 120q - 4q^2 - cq$

   (b) The FOC is

$$\begin{aligned} 120 - 8q^* - c &= 0 \\ q^* &= 15 - \frac{c}{8} \end{aligned}$$

   (c) Using the answer to (b), we have $dq^*/dc = -1/8 < 0$

(d) Plug $q^* = 15 - \frac{c}{8}$ into $\pi(q)$ to get

$$\begin{aligned}\pi(q) &= 120\left(15 - \frac{c}{8}\right) - 4\left(15 - \frac{c}{8}\right)^2 - c\left(15 - \frac{c}{8}\right) \\ &= 1800 - 15c - 900 + 15c - \frac{c^2}{16} - 15c + \frac{c^2}{8} \\ &= 900 - 15c + \frac{c^2}{16}\end{aligned}$$

(e) Differentiating yields

$$\pi'(c) = -15 + \frac{c}{8}$$

(f) Compare the answers to (b) and (e). Note that $-q$ is also the partial derivative of $\pi(q) = p(q)q - cq$ with respect to c, which is why this works.

7. **(a)** Implicitly differentiate to get

$$30x\frac{dx}{da} + 3a\frac{dx}{da} + 3x - \frac{5}{a} \cdot \frac{dx}{da} + \frac{5x}{a^2} = 0.$$

Solving for dx/da yields

$$\begin{aligned}\left(30x + 3a - \frac{5}{a}\right)\frac{dx}{da} &= -\left(3x + \frac{5x}{a^2}\right) \\ \frac{dx}{da} &= -\frac{3x + \frac{5x}{a^2}}{30x + 3a - \frac{5}{a}} = -\frac{x}{a}\frac{3a^2 + 5}{3a^2 + 30xa - 5}\end{aligned}$$

(b) Implicitly differentiate to get

$$12xa\frac{dx}{da} + 6x^2 = 5 - 5a^2\frac{dx}{da} - 10xa.$$

Solving for dx/da yields

$$\begin{aligned}\left(12xa + 5a^2\right)\frac{dx}{da} &= 5 - 10xa - 6x^2 \\ \frac{dx}{da} &= \frac{5 - 10xa - 6x^2}{12xa + 5a^2}\end{aligned}$$

8. **(a)** The foc is

$$\frac{30}{\sqrt{L}} - w = 0$$

and solving it for L gets

$$\begin{aligned} 30 &= w\sqrt{L} \\ \sqrt{L} &= \frac{30}{w} \\ L^* &= \frac{900}{w^2} \end{aligned}$$

(b) Since

$$L^* = \frac{900}{w^2}$$

we have

$$\frac{dL^*}{dw} = -\frac{1800}{w^3} < 0$$

The firm uses fewer workers when the wage rises.

(c) Plugging L^* into the profit function yields

$$\pi^* = 30\sqrt{4 \cdot \frac{900}{w^2}} - w \cdot \frac{900}{w^2} = \frac{900}{w}$$

and from there we find

$$\frac{d\pi^*}{dw} = -\frac{900}{w^2} < 0.$$

Profit falls when the wage rises. This happens for two reasons. One is that the firm must pay workers more, and the other is that it uses fewer workers (see part b) and produces less output.

9. **(a)** Implicitly differentiate to find dK/dL:

$$\begin{aligned} F_K(K,L)\frac{dK}{dL} + F_L(K,L) &= 0 \\ \frac{dK}{dL} &= -\frac{F_L(K,L)}{F_K(K,L)} \end{aligned}$$

(b) Both F_L and F_K are positive, so $dK/dL = -F_L/F_K < 0$ and the isoquant slopes downward.

Solutions for Chapter 4

1. Set up the Lagrangian

$$\mathcal{L}(x, y, \lambda) = 12x^2y^4 + \lambda(120 - 2x - 4y).$$

The foc's are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 24xy^4 - 2\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 48x^2y^3 - 4\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 120 - 2x - 4y = 0 \end{aligned}$$

This is three equations in three unknowns, so now we solve for the values of x, y, and λ. There are many ways to do this, and one of them can be found on page 39. Here is another. Solve the third equation for x:

$$\begin{aligned} 120 - 2x - 4y &= 0 \\ x &= 60 - 2y \end{aligned}$$

Substitute this into the first two equations

$$\begin{aligned} 24xy^4 - 2\lambda &= 0 \\ 48x^2y^3 - 4\lambda &= 0 \end{aligned}$$

to get

$$\begin{aligned} 24(60 - 2y)y^4 - 2\lambda &= 0 \\ 48(60 - 2y)^2y^3 - 4\lambda &= 0 \end{aligned}$$

Multiply the top equation by -2 and add the result to the second equation to get

$$48(60 - 2y)^2y^3 - 4\lambda - \left[48(60 - 2y)y^4 - 4\lambda\right] = 0$$

The terms with λ in them cancel out, and we are left with

$$48(60 - 2y)^2y^3 - 48(60 - 2y)y^4 = 0$$

Divide both sides by $48(60-2y)y^3$ to get

$$\begin{aligned}(60-2y)-y &= 0\\ 60-3y &= 0\\ y &= 20\end{aligned}$$

Substitute this back into things we know to get

$$x = 60 - 2y = 20$$

and

$$\lambda = 12(60-2y)y^4 = 12(20)(20^4) = 38,400,000.$$

2. The Lagrangian is

$$\mathcal{L}(a,b,\lambda) = 3\ln a + 2\ln b + \lambda(400 - 12a - 14b)$$

The FOCs are

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial a} &= \frac{3}{a} - 12\lambda = 0\\ \frac{\partial \mathcal{L}}{\partial b} &= \frac{2}{b} - 14\lambda = 0\\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 400 - 12a - 14b = 0\end{aligned}$$

Solving the first two yields $a = 3/12\lambda$ and $b = 2/14\lambda$. Substituting into the third equation gives us

$$\begin{aligned}400 - 12\left(\frac{3}{12\lambda}\right) - 14\left(\frac{2}{14\lambda}\right) &= 0\\ 400 - \frac{3}{\lambda} - \frac{2}{\lambda} &= 0\\ 400 &= \frac{5}{\lambda}\\ \lambda &= \frac{5}{400} = \frac{1}{80}\end{aligned}$$

Plugging into the earlier expressions,

$$a = \frac{3}{12\lambda} = \frac{3}{12/80} = \frac{240}{12} = 20$$

and

$$b = \frac{2}{14\lambda} = \frac{2}{14/80} = \frac{160}{14} = \frac{80}{7}.$$

3. The Lagrangian is

$$\mathcal{L}(x, y, \lambda) = 16x + y + \lambda(1 - x^{1/4}y^{3/4}).$$

The FOCs are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial x} &= 16 - \frac{1}{4}\lambda\left(\frac{y}{x}\right)^{3/4} = 16 - \frac{\lambda}{4}\frac{x^{1/4}y^{3/4}}{x} = 0 \\
\frac{\partial \mathcal{L}}{\partial y} &= 1 - \frac{3}{4}\lambda\left(\frac{x}{y}\right)^{1/4} = 1 - \frac{3\lambda}{4}\frac{x^{1/4}y^{3/4}}{y} = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda} &= 1 - x^{1/4}y^{3/4} = 0
\end{aligned}$$

From the third FOC we know that

$$x^{1/4}y^{3/4} = 1,$$

so the other two FOCs simplify to

$$\lambda = 64x$$

and

$$\lambda = \frac{4}{3}y.$$

Setting these equal to each other gives us

$$\begin{aligned}
\frac{4}{3}y &= 64x \\
y &= 48x.
\end{aligned}$$

Plugging this into the third FOC yields

$$\begin{aligned}
x^{1/4}y^{3/4} &= 1 \\
x^{1/4}(48x)^{3/4} &= 1 \\
x &= \frac{1}{48^{3/4}} = \frac{3^{1/4}}{24}.
\end{aligned}$$

We can then solve for

$$y = 48x = 2 \cdot 3^{1/4}$$

and

$$\lambda = \frac{4y}{3} = \frac{8 \cdot 3^{1/4}}{3}.$$

4. Set up the Lagrangian

$$\mathcal{L}(x, y, \lambda) = 3xy + 4x + \lambda(80 - 4x - 12y).$$

The foc's are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 3y + 4 - 4\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 3x - 12\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 80 - 4x - 12y = 0 \end{aligned}$$

The first equation reduces to $y = 4(\lambda - 1)/3$ and the second equation tells us that $x = 4\lambda$. Substituting these into the third equation yields

$$\begin{aligned} 80 - 4x - 12y &= 0 \\ 80 - 4(4\lambda) - 12(4)(\lambda - 1)/3 &= 0 \\ 96 - 32\lambda &= 0 \\ \lambda &= 3 \end{aligned}$$

Plugging this into the equations we already derived gives us the rest of the solution:

$$\begin{aligned} x &= 4\lambda = 12 \\ y &= 4(\lambda - 1)/3 = 8/3. \end{aligned}$$

5. Set up the Lagrangian

$$\mathcal{L}(x, y, \lambda) = 5x + 2y + \lambda(80 - 3x - 2xy)$$

The foc's are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 5 - 3\lambda - 2y\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 2 - 2x\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 80 - 3x - 2xy = 0 \end{aligned}$$

$$\begin{aligned} 5 - 3\lambda - 2y\lambda &= 0 \\ 2 - 2x\lambda &= 0 \\ 80 - 3x - 2xy &= 0 \end{aligned}$$

Now we solve these equations. The third one reduces to

$$\begin{aligned} 80 - 3x - 2xy &= 0 \\ 2xy &= 80 - 3x \\ y &= \frac{80 - 3x}{2x} \end{aligned}$$

and the second one reduces to

$$\begin{aligned} 2 - 2x\lambda &= 0 \\ \lambda &= \frac{1}{x}. \end{aligned}$$

Substitute these into the first one to get

$$\begin{aligned} 5 - 3\lambda - 2y\lambda &= 0 \\ 5 - 3\left(\frac{1}{x}\right) - 2\left(\frac{80 - 3x}{2x}\right)\left(\frac{1}{x}\right) &= 0 \end{aligned}$$

Multiplying through by x^2 yields

$$\begin{aligned} 5x^2 - 3x - 80 + 3x &= 0 \\ 5x^2 &= 80 \\ x^2 &= 16 \\ x &= \pm 4 \end{aligned}$$

Note that we only use the positive root in economics, so $x = 4$. Substituting into the other two equations yields

$$y = \frac{80 - 3x}{2x} = \frac{17}{2}$$

and

$$\lambda = \frac{1}{x} = \frac{1}{4}.$$

6. **(a)** $\pi'(x) = 400 + 4x > 0$. It says that increasing the size of the farm leads to increased profit, which seems sensible when the farm starts off small.

   (b) $\pi''(x) = 4$. This is questionable. But, it could arise because of increasing returns to scale or because of fixed inputs.

   (c) The Lagrangian is

   $$\mathcal{L}(x, \lambda) = 400x + 2x^2 + \lambda(10 - x).$$

   The FOCs are

   $$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 400 + 4x - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= 10 - x = 0 \end{aligned}$$

   The second one tells us that $x = 10$ and the first one tells us that $\lambda = 400 + 4x = 440$.

   (d) It is the marginal value of land.

   (e) That would be $\pi'(10) = 440$. This, by the way, is why the lame problem is useful. Note that the answers to (d) and (e) are the same.

   (f) No. remember that $\pi'(x) > 0$, so π is increasing and more land is better. Profit is maximized subject to the constraint. Obviously, constrained optimization will require a different set of second order conditions than unconstrained optimization does.

7. **(a)** $\pi'(L) = 30/\sqrt{L} - 10$ which is positive when $L < 9$. We would hope for an upward-sloping profit function, so this works, especially since L is only equal to 4.

   (b) $\pi''(L) = -15/L^{3/2}$ which is negative. Profit grows at a decreasing rate, which makes sense.

   (c) The Lagrangian is

   $$\mathcal{L}(L, \lambda) = 30\sqrt{4L} - 10L + \lambda(4 - L)$$

The foc's are

$$\frac{\partial \mathcal{L}}{\partial L} = \frac{30}{\sqrt{L}} - 10 - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 4 - L = 0$$

The second equation can be solved to get $L = 4$. Plugging $L = 4$ into the first equation yields

$$\begin{aligned} \frac{30}{\sqrt{L}} - 10 - \lambda &= 0 \\ \frac{30}{\sqrt{L}} - 10 - \lambda &= 0 \\ 5 &= \lambda \end{aligned}$$

(d) The Lagrange multiplier is always the marginal value of relaxing the constraint, where the value comes from whatever the objective function measures. In this case the objective function is the profit function, and the constraint is on the number of workers the firm can use at one time, so the Lagrange multiplier measures the marginal profit from adding workers.

(e) This is

$$\pi'(4) = 30/\sqrt{4} - 10 = 5.$$

Note that this matches the answer from (c).

(f) No. The first derivative of the profit function is positive (and equal to 5) when $L = 4$, which means that profit is increasing when L is 4. The second derivative does not tell us whether we are at a maximum or minimum when there is a constraint.

8. **(a)** The Lagrangian is

$$\mathcal{L}(x, y, \lambda) = x^{\alpha} y^{1-\alpha} + \lambda(M - p_x x - p_y y).$$

The FOCs are

$$\frac{\partial \mathcal{L}}{\partial x} = \alpha x^{\alpha-1} y^{1-\alpha} - \lambda p_x = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = (1-\alpha) x^{\alpha} y^{-\alpha} - \lambda p_y = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = M - p_x x - p_y y = 0.$$

Rearrange the first two to get

$$\begin{aligned}
\frac{\alpha\left(\frac{y}{x}\right)^{1-\alpha}}{p_x} &= \lambda \\
\frac{(1-\alpha)\left(\frac{x}{y}\right)^{\alpha}}{p_y} &= \lambda.
\end{aligned}$$

Set them equal to each other to get

$$\begin{aligned}
\frac{\alpha\left(\frac{y}{x}\right)^{1-\alpha}}{p_x} &= \frac{(1-\alpha)\left(\frac{x}{y}\right)^{\alpha}}{p_y} \\
\left(\frac{y}{x}\right)^{1-\alpha}\left(\frac{y}{x}\right)^{\alpha} &= \frac{(1-\alpha)}{\alpha}\frac{p_x}{p_y} \\
\frac{y}{x} &= \frac{(1-\alpha)}{\alpha}\frac{p_x}{p_y} \\
y &= \frac{(1-\alpha)}{\alpha}\frac{p_x}{p_y}x.
\end{aligned}$$

Now substitute this into the budget constraint to get

$$\begin{aligned}
p_x x + p_y y &= M \\
p_x x + p_y\frac{(1-\alpha)}{\alpha}\frac{p_x}{p_y}x &= M \\
p_x x + \frac{(1-\alpha)}{\alpha}p_x x &= M \\
p_x x &= \frac{M}{1+\frac{(1-\alpha)}{\alpha}} = \alpha M \\
x &= \frac{\alpha M}{p_x}.
\end{aligned}$$

Substituting this back into what we found for y yields

$$\begin{aligned}
y &= \frac{(1-\alpha)}{\alpha}\frac{p_x}{p_y}x \\
&= \frac{(1-\alpha)}{\alpha}\frac{p_x}{p_y}\frac{\alpha M}{p_x} \\
&= \frac{(1-\alpha)M}{p_y}.
\end{aligned}$$

(b) These are easy.

$$\frac{\partial x^*}{\partial M} = \frac{\alpha}{p_x} > 0$$
$$\frac{\partial y^*}{\partial M} = \frac{1-\alpha}{p_y} > 0.$$

(c) Again, these are easy.

$$\frac{\partial x^*}{\partial p_x} = -\frac{\alpha M}{p_x^2} < 0$$
$$\frac{\partial y^*}{\partial p_x} = 0.$$

The demand curve for good x is downward-sloping, and it is independent of the price of the other good.

9. **(a)** Denote labor devoted to widget production by w and labor devoted to gookey production by g. The Lagrangian is

$$\mathcal{L}(w, g, \lambda) = (9)(20w^{1/2}) + (3)(30g) - (11)(w + g) + \lambda(60 - w - g).$$

The foc's are

$$\frac{\partial \mathcal{L}}{\partial w} = 90w^{-\frac{1}{2}} - 11 - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial g} = 90 - 11 - \lambda = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda} = 60 - w - g = 0$$

The second equation says that $\lambda = 79$. Plugging this into the first equation yields

$$\frac{90}{\sqrt{w}} - 11 - 79 = 0$$
$$90 = 90\sqrt{w}$$
$$w = 1$$

The third equation then implies that $g = 60 - w = 59$. These are the same as the answers to the question 4 on the first homework.

(b) The Lagrange multiplier is the marginal value of adding workers.

10. **(a)** The farmer's problem is

$$\max_{L,W} LW$$

$$\begin{aligned} \text{s.t. } 2L + 2W &= F \\ W &= S \end{aligned}$$

(b) The Lagrangian is

$$\mathcal{L}(L, W, \lambda, \mu) = LW + \lambda(F - 2L - 2W) + \mu(S - W).$$

The first-order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial L} &= W - 2\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial W} &= L - 2\lambda - \mu = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= F - 2L - 2W = 0 \\ \frac{\partial \mathcal{L}}{\partial \mu} &= S - W = 0 \end{aligned}$$

We must solve this set of equations:

$$\begin{aligned} W &= S \text{ (from fourth equation)} \\ L &= F/2 - S \text{ (from third equation)} \\ \lambda &= S/2 \text{ (from second equation)} \\ \mu &= F/2 - 2S \text{ (from first equation)} \end{aligned}$$

(c) It depends. The marginal impact on area comes directly from the Lagrange multipliers. λ is the marginal impact of having a longer fence while keeping the shortest side fixed, and μ is the marginal impact of lengthening the shortest side while keeping the total fence length constant. We want to know which is greater, $S/2$ or $F/2 - 2S$. We can find

$$\begin{aligned} \lambda &\geq \mu \\ S/2 &\geq F/2 - 2S \\ 5S/2 &\geq F/2 \\ S &\geq F/5. \end{aligned}$$

When the shortest side is more than one-fifth of the total amount of fencing, the farmer would rather lengthen the fence than lengthen the shortest side. When the shortest side is smaller than a fifth of the fence lenght, she would rather lengthen that side, keeping the total fence length fixed.

Solutions for Chapter 5

1. **(a)** The solution to the alternative problem is $(x, y) = (8, \frac{8}{3})$. Note that $4 \cdot 8 + \frac{8}{3} = 34\frac{2}{3} > 20$, so the second constraint does not hold.

   (b) The solution to the alternative problem is $(x, y) = (\frac{10}{3}, \frac{20}{3})$. Note that $2 \cdot \frac{10}{3} + 3 \cdot \frac{20}{3} = \frac{80}{3} > 24$, so the first constraint does not hold.

   (c) If the solution to the alternative problem in (a) had satisfied the second constraint, the second constraint would have been non-binding and its Lagrange multiplier would have been zero. This is not what happened, though, so the second constraint must bind, in which case $\lambda_2 > 0$. Similarly, part (b) shows us that the first constraint must also bind, and so $\lambda_1 > 0$.

   (d) Because both constraints bind, the problem becomes

$$\begin{gathered} \max_{x,y} x^2 y \\ \text{s.t. } 2x + 3y = 24 \\ 4x + y = 20 \end{gathered}$$

   This is easy to solve because there is only one point that satisfies both constraints: $(x, y) = (\frac{18}{5}, \frac{28}{5})$. Now find the Lagrange multipliers. The FOCs for the equality-constrained problem are

$$\begin{aligned} 2xy - 2\lambda_1 - 4\lambda_2 &= 0 \\ x^2 - 3\lambda_1 - \lambda_2 &= 0 \\ 24 - 2x - 3y &= 0 \\ 20 - 4x - y &= 0 \end{aligned}$$

We already used the last two to find x and y. Plug those values into the first two to get two equations in two unknowns:

$$\begin{aligned} 2\lambda_1 - 4\lambda_2 &= \frac{1008}{25} \\ 3\lambda_1 + \lambda_2 &= \frac{324}{25} \end{aligned}$$

The solution to this is $(\lambda_1, \lambda_2) = (\frac{144}{125}, \frac{1188}{125})$.

2. **(a)** The solution to the alternative problem is $(x, y) = (8, \frac{8}{3})$. Note that $4 \cdot 8 + \frac{8}{3} = 34\frac{2}{3} < 36$, so the second constraint does hold this time.

(b) The solution to the alternative problem is $(x, y) = (6, 12)$. Note that $2 \cdot 6 + 3 \cdot 12 = 48 > 24$, so the first constraint does not hold.

(c) The solution to the alternative problem in (a) satisfies the second constraint, so the second contrainst is nonbinding. Therefore $\lambda_1 > 0$ and $\lambda_2 = 0$.

(d) Because only the first constraint binds, the problem becomes

$$\max_{x,y} x^2 y$$
$$\text{s.t. } 2x + 3y = 24$$

We know from part (a) that $(x, y) = (8, \frac{8}{3})$. We also know that $\lambda_2 = 0$. To find λ_1 use the FOCs for the equality-constrained problem:

$$\begin{aligned} 2xy - 2\lambda_1 &= 0 \\ x^2 - 3\lambda_1 &= 0 \\ 24 - 2x - 3y &= 0 \end{aligned}$$

Plug $x = 8$ into the second equation to get $\lambda_1 = \frac{64}{3}$. Or, plug $x = 8$ and $y = \frac{8}{3}$ into the first equation to get the same thing.

3. **(a)** Setting $\lambda_2 = 0$ in the original Lagrangian we get the first-order conditions

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 4y - 6x - \lambda_1 = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 4x - 4\lambda_1 = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_1} &= 36 - x - 4y = 0 \end{aligned}$$

We solve these for x, y, and λ_1 and get

$$x = \frac{9}{2},\ y = \frac{63}{8},\ \lambda_1 = \frac{9}{2}$$

We then have $5x + 2y = \frac{45}{2} + \frac{63}{4} = \frac{153}{4} < 45$ and the second constraint is satisfied.

(b) Setting $\lambda_1 = 0$ in the original Lagrangian, the foc's are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x} &= 4y - 6x - 5\lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 4x - 2\lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= 45 - 5x - 2y = 0 \end{aligned}$$

The solution is

$$x = \frac{45}{13},\ y = \frac{180}{13},\ \lambda_2 = \frac{90}{13}$$

We then have $x + 4y = \frac{45}{13} + \frac{720}{13} = \frac{765}{13} > 36$ and the first constraint is not satisfied.

(c) Part (a) shows that we can get a solution when the first constraint binds and the second doesn't, and part (b) shows that we cannot get a solution when the second constraint binds but the first does not. So, the answer comes from part (a), with

$$x = \frac{9}{2},\ y = \frac{63}{8},\ \lambda_1 = \frac{9}{2},\ \lambda_2 = 0.$$

4. **(a)** The foc's are

$$\frac{\partial \mathcal{L}}{\partial x} = 3y - 8 - \lambda_1 = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 3x - 4\lambda_1 = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = 24 - x - 4y = 0$$

The solution is

$$x = \frac{20}{3},\ y = \frac{13}{3},\ \lambda_1 = 5$$

and since $5x + 2y = \frac{100}{3} + \frac{26}{3} = 42 > 30$ the second constraint is not satisfied.

(b) The foc's are

$$\frac{\partial \mathcal{L}}{\partial x} = 3y - 8 - 5\lambda_2 = 0$$
$$\frac{\partial \mathcal{L}}{\partial y} = 3x - 2\lambda_2 = 0$$
$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = 30 - 5x - 2y = 0$$

$$3y - 8 - 5\lambda_2 = 0$$
$$3x - 2\lambda_2 = 0$$
$$30 - 5x - 2y = 0$$

The solution is

$$x = \frac{37}{15},\ y = \frac{53}{6},\ \lambda_2 = \frac{37}{10}$$

and since $x + 4y = \frac{189}{5} > 24$ the first constraint is not satisfied.

(c) Since when one constraint binds the other fails, they must both bind. The place where they both bind is the intersection of the two "budget lines," or where the following system is solved:

$$x + 4y = 24$$
$$5x + 2y = 30$$

The solution is $x = 4$, $y = 5$. Now we have to find the values of λ_1 and λ_2. To do this, go back to the foc's for the entire original Lagrangian:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x} &= 3y - 8 - \lambda_1 - 5\lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 3x - 4\lambda_1 - 2\lambda_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_1} &= 24 - x - 4y = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= 30 - 5x - 2y = 0\end{aligned}$$

Plug the values for x and y into the first two equations to get

$$\begin{aligned}15 - 8 - \lambda_1 - 5\lambda_2 &= 0 \\ 12 - 4\lambda_1 - 2\lambda_2 &= 0\end{aligned}$$

and solve for λ_1 and λ_2. The solution is $\lambda_1 = \frac{23}{9}$ and $\lambda_2 = \frac{8}{9}$.

5. **(a)**

$$K(x, y, \lambda) = x^2y + \lambda[42 - 4x - 2y].$$

(b)

$$\begin{aligned}x\frac{\partial K}{\partial x} &= x(2xy - 4\lambda) = 0 \\ y\frac{\partial K}{\partial y} &= y(x^2 - 2\lambda) = 0 \\ \lambda\frac{\partial K}{\partial \lambda} &= \lambda(42 - 4x - 2y) = 0 \\ & x, y, \lambda \geq 0\end{aligned}$$

(c) First notice that the objective function is x^2y, which is zero if either x or y is zero. Consequently, neither $x \geq 0$ nor $y \geq 0$ can be binding. The other, budget-like constraint is binding because x^2y is increasing in both arguments, and so $\lambda > 0$. The Kuhn-Tucker conditions reduce to

$$\begin{aligned}2xy - 4\lambda &= 0 \\ x^2 - 2\lambda &= 0 \\ 42 - 4x - 2y &= 0\end{aligned}$$

Solving yields $(x, y, \lambda) = (7, 7, \frac{49}{2})$.

6. **(a)**

$$K(x, y, \lambda) = xy + 40x + 60y + \lambda(12 - x - y)$$

(b)

$$\begin{aligned} x\frac{\partial K}{\partial x} &= x(y + 40 - \lambda) = 0 \\ y\frac{\partial K}{\partial y} &= y(x + 60 - \lambda) = 0 \\ \lambda\frac{\partial K}{\partial \lambda} &= \lambda(12 - x - y) = 0 \end{aligned}$$

$$x, y, \lambda \geq 0$$

(c) This one is complicated, because we can identify three potential solutions: (i) x is zero and y is positive, (ii) x is positive and y is zero, and (iii) both x and y are positive. The only thing to do is try them one at a time.

Case (i): $x = 0$. Then $y = 12$ from the third equation, and $\lambda = 60$ from the second equation. The value of the objective function is $xy + 40x + 60y = 720$.

Case (ii): $y = 0$. Then $x = 12$ from the third equation, and $\lambda = 40$ from the first equation. The value of the objective function is 480. This case is not as good as case (i), so it cannot be the answer.

Case (iii): $x, y > 0$. Divide both sides of the first K-T condition by x, which is legal since $x > 0$, divide the second by y, and divide the third by λ. We get

$$\begin{aligned} y + 40 - \lambda &= 0 \\ x + 60 - \lambda &= 0 \\ 12 - x - y &= 0 \end{aligned}$$

The solution to this system of equations is $x = -4$, $y = 16$, $\lambda = 56$. This is not allowed, though, because $x < 0$.

The final solution is case (i): $x = 0$, $y = 12$, $\lambda = 60$.

Solutions for Chapter 6

1. **(a)** $\begin{pmatrix} 19 & -16 & -22 \\ 2 & -3 & 61 \end{pmatrix}$

   (b) $\begin{pmatrix} 10 & 0 & -1 \\ 13 & 18 & 5 \end{pmatrix}$

   (c) $\begin{pmatrix} 7 & 1 & 2 \\ 19 & 2 & 2 \\ -23 & -36 & 1 \end{pmatrix}$

   (d) 39

2. **(a)** $\begin{pmatrix} 14 & -23 \\ -21 & 32 \\ 9 & 11 \end{pmatrix}$

   (b) $\begin{pmatrix} 7 & 0 \\ 12 & 14 \\ 0 & 9 \end{pmatrix}$

   (c) $\begin{pmatrix} 33 & -13 & 32 \\ 7 & -11 & 36 \end{pmatrix}$

   (d) 84

3. **(a)** 14

   (b) 134

4. **(a)** 21

   (b) -12

5. In matrix form the system of equations is

$$\begin{pmatrix} 6 & -2 & -3 \\ 2 & 4 & 1 \\ 3 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 8 \end{pmatrix}.$$

Using Cramer's rule we get

$$x = \frac{\det\begin{pmatrix} 1 & -2 & -3 \\ -2 & 4 & 1 \\ 8 & 0 & -1 \end{pmatrix}}{\det\begin{pmatrix} 6 & -2 & -3 \\ 2 & 4 & 1 \\ 3 & 0 & -1 \end{pmatrix}} = \frac{80}{2} = 40$$

$$y = \frac{\det\begin{pmatrix} 6 & 1 & -3 \\ 2 & -2 & 1 \\ 3 & 8 & -1 \end{pmatrix}}{\det\begin{pmatrix} 6 & -2 & -3 \\ 2 & 4 & 1 \\ 3 & 0 & -1 \end{pmatrix}} = \frac{-97}{2}$$

$$z = \frac{\det\begin{pmatrix} 6 & -2 & 1 \\ 2 & 4 & -2 \\ 3 & 0 & 8 \end{pmatrix}}{\det\begin{pmatrix} 6 & -2 & -3 \\ 2 & 4 & 1 \\ 3 & 0 & -1 \end{pmatrix}} = \frac{224}{2} = 112$$

6. In matrix form the system of equations is

$$\begin{pmatrix} 5 & -2 & 1 \\ 3 & -1 & 0 \\ 0 & 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 9 \\ 9 \\ 15 \end{pmatrix}.$$

Using Cramer's rule we get

$$x = \frac{\det\begin{pmatrix} 9 & -2 & 1 \\ 9 & -1 & 0 \\ 15 & 3 & 2 \end{pmatrix}}{\det\begin{pmatrix} 5 & -2 & 1 \\ 3 & -1 & 0 \\ 0 & 3 & 2 \end{pmatrix}} = \frac{60}{11}$$

$$y = \frac{\det\begin{pmatrix} 5 & 9 & 1 \\ 3 & 9 & 0 \\ 0 & 15 & 2 \end{pmatrix}}{\det\begin{pmatrix} 5 & -2 & 1 \\ 3 & -1 & 0 \\ 0 & 3 & 2 \end{pmatrix}} = \frac{81}{11}$$

$$z = \frac{\det\begin{pmatrix} 5 & -2 & 9 \\ 3 & -1 & 9 \\ 0 & 3 & 15 \end{pmatrix}}{\det\begin{pmatrix} 5 & -2 & 1 \\ 3 & -1 & 0 \\ 0 & 3 & 2 \end{pmatrix}} = \frac{-39}{11}$$

7. **(a)** $\frac{1}{8}\begin{pmatrix} 1 & -3 \\ 2 & 2 \end{pmatrix}$

   (b) $\frac{1}{2}\begin{pmatrix} 1 & -2 & 1 \\ -1 & 2 & 1 \\ 1 & 0 & -1 \end{pmatrix}$

8. **(a)** $\frac{1}{14}\begin{pmatrix} -4 & -1 \\ -2 & -4 \end{pmatrix}$

   (b) $\frac{1}{25}\begin{pmatrix} 5 & -4 & 3 \\ 0 & 15 & -5 \\ 0 & -5 & 10 \end{pmatrix}$

Solutions for Chapter 7

1. **(a)** The determinant of the matrix

$$\begin{pmatrix} 3 & 6 & 0 \\ 2 & 0 & -5 \\ 1 & -1 & -1 \end{pmatrix}$$

   is -33, and so there is a unique solution.

(b) The determinant of the matrix

$$\begin{pmatrix} 4 & -1 & 8 \\ 17 & -8 & 10 \\ -3 & 2 & 2 \end{pmatrix}$$

is 0, and so there is not a unique solution. To find out whether there is no solution or an infinite number, get the augmented matrix in row-echelon form.

$$\left(\begin{array}{ccc|c} 4 & -1 & 8 & 160 \\ 17 & -8 & 10 & 200 \\ -3 & 2 & 2 & 40 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 4 & -1 & 8 & 160 \\ 0 & -\frac{15}{4} & -24 & -480 \\ 0 & \frac{5}{4} & 8 & 160 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 4 & -1 & 8 & 160 \\ 0 & -\frac{15}{4} & -24 & -480 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Since the bottom row is zeros all the way across, there are infinitely many solutions.

(c) The determinant of the matrix

$$\begin{pmatrix} 2 & -3 & 0 \\ 3 & 0 & 5 \\ 2 & 6 & 10 \end{pmatrix}$$

is 0, and so there is not a unique solution. To find out whether there is no solution or an infinite number, get the augmented matrix in row-echelon form.

$$\left(\begin{array}{ccc|c} 2 & -3 & 0 & 6 \\ 3 & 0 & 5 & 15 \\ 2 & 6 & 10 & 18 \end{array}\right)$$

Multiply the first row by 2 and add it to the third row:

$$\left(\begin{array}{ccc|c} 2 & -3 & 0 & 6 \\ 3 & 0 & 5 & 15 \\ 6 & 0 & 10 & 30 \end{array}\right)$$

Multiply the second row by 2 and subtract it from the third row:

$$\left(\begin{array}{ccc|c} 2 & -3 & 0 & 6 \\ 3 & 0 & 5 & 15 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

There is a row of all zeros, so there is an infinite number of solutions.

(d) The determinant of the matrix

$$\begin{pmatrix} 4 & -1 & 8 \\ 3 & 0 & 2 \\ 5 & 1 & -2 \end{pmatrix}$$

is 0, and so there is not a unique solution. To find out whether there is no solution or an infinite number, get the augmented matrix in row-echelon form.

$$\left(\begin{array}{ccc|c} 4 & -1 & 8 & 30 \\ 3 & 0 & 2 & 20 \\ 5 & 1 & -2 & 40 \end{array}\right)$$

Add the top row to the bottom row:

$$\left(\begin{array}{ccc|c} 4 & -1 & 8 & 30 \\ 3 & 0 & 2 & 20 \\ 9 & 0 & 6 & 70 \end{array}\right)$$

Multiply the middle row by 3 and subtract it from the bottom row:

$$\left(\begin{array}{ccc|c} 4 & -1 & 8 & 30 \\ 3 & 0 & 2 & 20 \\ 0 & 0 & 0 & 10 \end{array}\right)$$

Since the bottom row is zeros all the way across except for the last column, there is no solution.

(e) The determinant of the matrix

$$\begin{pmatrix} 6 & -1 & -1 \\ 5 & 2 & -2 \\ 0 & 1 & -2 \end{pmatrix}$$

is -27, there is a unique solution.

2. **(a)** There is no inverse if the determinant is zero, which leads to the equation

$$\begin{aligned} 6a+2 &= 0 \\ a &= -\frac{1}{3}. \end{aligned}$$

(b) Setting the determinant equal to zero and solving for a yields

$$\begin{aligned} -5a-5 &= 0 \\ a &= -1 \end{aligned}$$

(c) There is no inverse if the determinant is zero, which leads to the equation

$$\begin{aligned} 5a+9 &= 0 \\ a &= -\frac{9}{5}. \end{aligned}$$

(d) Setting the determinant equal to zero and solving for a yields

$$\begin{aligned} 20a-35 &= 0 \\ a &= \frac{7}{4} \end{aligned}$$

Solutions for Chapter 8

1. **(a)** Rewrite the system as

$$\begin{aligned} Y &= c((1-t)Y)+i(R)+G \\ M &= P\cdot m(Y,R) \end{aligned}$$

Implicitly differentiate with respect to t to get

$$\begin{aligned} \frac{dY}{dt} &= -c'Y+(1-t)c'\frac{dY}{dt}+i'\frac{dR}{dt} \\ 0 &= Pm_Y\frac{dY}{dt}+Pm_R\frac{dR}{dt} \end{aligned}$$

Write in matrix form:

$$\begin{pmatrix} 1-(1-t)c' & -i' \\ m_Y & m_R \end{pmatrix}\begin{pmatrix} \frac{dY}{dt} \\ \frac{dR}{dt} \end{pmatrix}=\begin{pmatrix} -c'Y \\ 0 \end{pmatrix}$$

Use Cramer's rule to get:

$$\begin{aligned}
\frac{dY}{dt} &= \frac{\begin{vmatrix} -c'Y & -i' \\ 0 & m_R \end{vmatrix}}{\begin{vmatrix} (1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} \\
&= -c'Y \cdot \frac{m_R}{(1-(1-t)c')m_R + m_Y i'}
\end{aligned}$$

which is $-c'Y$ times the derivative from the lecture. It is negative, so an increase in the tax rate reduces GDP.

$$\begin{aligned}
\frac{dR}{dt} &= \frac{\begin{vmatrix} 1-(1-t)c' & -c'Y \\ m_Y & 0 \end{vmatrix}}{\begin{vmatrix} (1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} \\
&= -c'Y \cdot \frac{m_Y}{(1-(1-t)c')m_R + m_Y i'}
\end{aligned}$$

which is $-c'Y$ times the derivative from the lecture. It is negative, so an increase in the tax rate reduces the interest rate.

(b) Implicitly differentiate the system with respect to M to get

$$\begin{aligned}
\frac{dY}{dM} &= (1-t)c'\frac{dY}{dM} + i'\frac{dR}{dM} \\
1 &= Pm_Y\frac{dY}{dM} + Pm_R\frac{dR}{dM}
\end{aligned}$$

Write in matrix form:

$$\begin{pmatrix} (1-t)c' & -i' \\ m_Y & m_R \end{pmatrix} \begin{pmatrix} \frac{dY}{dM} \\ \frac{dR}{dM} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Use Cramer's rule to get:

$$\begin{aligned}
\frac{dY}{dM} &= \frac{\begin{vmatrix} 0 & -i' \\ 1 & m_R \end{vmatrix}}{\begin{vmatrix} (1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} \\
&= \frac{i'}{(1-(1-t)c')m_R + m_Y i'}
\end{aligned}$$

Both the numerator and denominator are negative, making the derivative positive, and so an increase in money supply increases GDP.

$$\begin{aligned} \frac{dR}{dt} &= \frac{\begin{vmatrix} (1-t)c' & 0 \\ m_Y & 1 \end{vmatrix}}{\begin{vmatrix} (1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} \\ &= \frac{(1-t)c'}{(1-(1-t)c')m_R + m_Y i'} \end{aligned}$$

The numerator is positive, making the derivative negative, and so an increase in money supply reduces the interest rate.

(c) Implicitly differentiate the system with respect to P to get

$$\begin{aligned} \frac{dY}{dP} &= (1-t)c'\frac{dY}{dP} + i'\frac{dR}{dP} \\ 0 &= m + Pm_Y\frac{dY}{dP} + Pm_R\frac{dR}{dP} \end{aligned}$$

Write in matrix form:

$$\begin{pmatrix} 1-(1-t)c' & -i' \\ m_Y & m_R \end{pmatrix} \begin{pmatrix} \frac{dY}{dP} \\ \frac{dR}{dP} \end{pmatrix} = \begin{pmatrix} 0 \\ -m \end{pmatrix}$$

The derivatives are $-m$ times the derivatives from part (b), and so an increase in the price level reduces GDP and increases the interest rate.

2. **(a)** First simplify to two equations:

$$\begin{aligned} Y &= c(Y-T) + i(R) + G + x(Y,R) \\ M &= P \cdot m(Y,R) \end{aligned}$$

Implicitly differentiate with respect to G to get

$$\begin{aligned} \frac{dY}{dG} &= c'\frac{dY}{dG} + i'\frac{dR}{dG} + 1 + x_Y\frac{dY}{dG} + x_R\frac{dR}{dG} \\ 0 &= P \cdot m_Y\frac{dY}{dG} + P \cdot m_R\frac{dR}{dG} \end{aligned}$$

Rearrange as

$$\begin{aligned}\frac{dY}{dG} - c'\frac{dY}{dG} - i'\frac{dR}{dG} - x_Y\frac{dY}{dG} - x_R\frac{dR}{dG} &= 1 \\ m_Y\frac{dY}{dG} + m_R\frac{dR}{dG} &= 0\end{aligned}$$

We can write this in matrix form

$$\begin{pmatrix} 1-c'-x_Y & -i'-x_R \\ m_Y & m_R \end{pmatrix}\begin{pmatrix} \frac{dY}{dG} \\ \frac{dR}{dG} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Now use Cramer's rule to solve for dY/dG and dR/dG:

$$\begin{aligned}\frac{dY}{dG} &= \frac{\begin{vmatrix} 1 & -i'-x_R \\ 0 & m_R \end{vmatrix}}{\begin{vmatrix} 1-c'-x_Y & -i'-x_R \\ m_Y & m_R \end{vmatrix}} \\ &= \frac{m_R}{(1-c'-x_Y)m_R + m_Y(i'+x_R)}\end{aligned}$$

The numerator is negative. The denominator is negative. So, $dY/dG > 0$.

$$\begin{aligned}\frac{dR}{dG} &= \frac{\begin{vmatrix} 1-c'-x_Y & 1 \\ m_Y & 0 \end{vmatrix}}{\begin{vmatrix} 1-c'-x_Y & -i'-x_R \\ m_Y & m_R \end{vmatrix}} \\ &= \frac{-m_Y}{(1-c'-x_Y)m_R + m_Y(i'+x_R)}\end{aligned}$$

The numerator is negative and so is the denominator. Thus, $dR/dG > 0$. An increase in government spending increases both GDP and interest rates in the short run.

(b) In matrix form we get

$$\begin{pmatrix} 1-c'-x_Y & -i'-x_R \\ m_Y & m_R \end{pmatrix}\begin{pmatrix} \frac{dY}{dG} \\ \frac{dR}{dG} \end{pmatrix} = \begin{pmatrix} -c' \\ 0 \end{pmatrix}$$

Thus, the derivatives are $-c'$ times those in part (a), so an increase in tax revenue reduces both GDP and interest rates.

(c)

$$\frac{dY}{dM} = \frac{\begin{vmatrix} 1 & -i' - x_R \\ 0 & m_R \end{vmatrix}}{\begin{vmatrix} 1 - c' - x_Y & -i' - x_R \\ m_Y & m_R \end{vmatrix}} = \frac{i' + x_R}{(1 - c' - x_Y)m_R + m_Y(i' + x_R)}$$

Both the numerator and denominator are negative, making the derivative positive, and so an increase in money supply increases GDP.

$$\frac{dR}{dM} = \frac{\begin{vmatrix} 1 - c' - x_Y & 0 \\ m_Y & 1 \end{vmatrix}}{\begin{vmatrix} (1-t)c' & -i' \\ m_Y & m_R \end{vmatrix}} = \frac{1 - c' - x_Y}{(1 - c' - x_Y)m_R + m_Y(i' + x_R)}$$

The numerator is positive, making the derivative negative, and so an increase in money supply reduces the interest rate.

3. **(a)** Rewrite the system as

$$\begin{aligned} Y &= c((1-t)Y) + i(R) + x(Y,R) + G \\ M &= P \cdot m(Y,R) \\ Y &= \bar{Y} \end{aligned}$$

Implicitly differentiate with respect to G to get

$$\begin{aligned} \frac{dY}{dG} &= (1-t)c'\frac{dY}{dG} + i'\frac{dR}{dG} + x_Y\frac{dY}{dG} + x_R\frac{dR}{dG} + 1 \\ 0 &= m(Y,R)\frac{dP}{dG} + Pm_Y\frac{dY}{dG} + Pm_R\frac{dR}{dG} \\ \frac{dY}{dG} &= 0 \end{aligned}$$

The last line implies, obviously, that $dY/dG = 0$. This makes sense because Y is fixed at the exogenous level $\bar{Y}$. Even so, let's go through the effort of writing the equation in matrix notation:

$$\begin{pmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} dY/dG \\ dR/dG \\ dP/dG \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Use Cramer's rule to get

$$\frac{dY}{dG} = \frac{\begin{vmatrix} 1 & -i'-x_R & 0 \\ 0 & Pm_R & m \\ 0 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{vmatrix}} = 0$$

where the result follows immediately from the row with all zeroes. The increase in government spending has no long-run impact on GDP. As for interest rates,

$$\frac{dR}{dG} = \frac{\begin{vmatrix} 1-(1-t)c'-x_Y & 1 & 0 \\ Pm_Y & 0 & m \\ 1 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{vmatrix}} = \frac{m}{-mi'-mx_R} = -\frac{1}{i'+x_R} > 0.$$

Increased government spending leads to an increase in interest rates in the long run. Finally,

$$\frac{dP}{dG} = \frac{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 1 \\ Pm_Y & Pm_R & 0 \\ 1 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{vmatrix}} = \frac{-Pm_R}{-mi'-mx_R} > 0.$$

An increase in government spending leads to an increase in the price level.

(b) This time implicitly differentiate with respect to M to get

$$\begin{aligned}\frac{dY}{dM} &= (1-t)c'\frac{dY}{dM} + i'\frac{dR}{dM} + x_Y\frac{dY}{dM} + x_R\frac{dR}{dM} \\ 1 &= m(Y,R)\frac{dP}{dM} + Pm_Y\frac{dY}{dM} + Pm_R\frac{dR}{dM} \\ \frac{dY}{dM} &= 0\end{aligned}$$

Write the equation in matrix notation:

$$\begin{pmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{pmatrix}\begin{pmatrix} dY/dM \\ dR/dM \\ dP/dM \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Use Cramer's rule to get

$$\frac{dY}{dM} = \frac{\begin{vmatrix} 0 & -i'-x_R & 0 \\ 1 & Pm_R & m \\ 0 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{vmatrix}} = 0$$

The increase in money supply has no long-run impact on GDP. As for interest rates,

$$\frac{dR}{dM} = \frac{\begin{vmatrix} 1-(1-t)c'-x_Y & 0 & 0 \\ Pm_Y & 1 & m \\ 1 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{vmatrix}} = \frac{0}{-mi'-mx_R} = 0.$$

Increasing the money supply has no long-run impact on interest rates, either. Finally,

$$\frac{dP}{dM} = \frac{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & 1 \\ 1 & 0 & 0 \end{vmatrix}}{\begin{vmatrix} 1-(1-t)c'-x_Y & -i'-x_R & 0 \\ Pm_Y & Pm_R & m \\ 1 & 0 & 0 \end{vmatrix}} = \frac{-i'-x_R}{-mi'-mx_R} = \frac{1}{m} > 0.$$

An increase the money supply leads to an increase in the price level. That's the *only* long-run impact of an increase in money supply.

4. **(a)** Implicitly differentiate the system with respect to I:

$$\begin{aligned}\frac{dq_D}{dI} &= D_p\frac{dp}{dI} + D_I\\ \frac{dq_S}{dI} &= S_p\frac{dp}{dI}\\ \frac{dq_D}{dI} &= \frac{dq_S}{dI}\end{aligned}$$

Write it in matrix form:

$$\begin{pmatrix} 1 & 0 & -D_p \\ 0 & 1 & -S_p \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} dq_D/dI \\ dq_S/dI \\ dp/dI \end{pmatrix} = \begin{pmatrix} D_I \\ 0 \\ 0 \end{pmatrix}$$

Solve for dp/dI using Cramer's rule:

$$\frac{dp}{dI} = \frac{\begin{vmatrix} 1 & 0 & D_I \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & -D_p \\ 0 & 1 & -S_p \\ 1 & -1 & 0 \end{vmatrix}} = \frac{-D_I}{D_P - S_P} > 0$$

where the result follows because $D_I > 0$, $D_p < 0$, and $S_p > 0$.

(b) Implicitly differentiate the system with respect to w:

$$\begin{aligned}\frac{dq_D}{dw} &= D_p\frac{dp}{dw}\\ \frac{dq_S}{dw} &= S_p\frac{dp}{dw} + S_w\\ \frac{dq_D}{dw} &= \frac{dq_S}{dw}\end{aligned}$$

Write it in matrix form:

$$\begin{pmatrix} 1 & 0 & -D_p \\ 0 & 1 & -S_p \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} dq_D/dw \\ dq_S/dw \\ dp/dw \end{pmatrix} = \begin{pmatrix} 0 \\ S_w \\ 0 \end{pmatrix}$$

Solve for dp/dw using Cramer's rule:

$$\frac{dp}{dw} = \frac{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & S_w \\ 1 & -1 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & -D_p \\ 0 & 1 & -S_p \\ 1 & -1 & 0 \end{vmatrix}} = \frac{S_w}{D_P - S_P} > 0,$$

where the result follows because $S_w < 0$, $D_p < 0$, and $S_p > 0$.

5. We can write the regression as $y = X\beta + e$ where

$$y = \begin{pmatrix} 6 \\ 2 \\ 5 \end{pmatrix} \text{ and } X = \begin{pmatrix} 1 & 9 \\ 1 & 4 \\ 1 & 3 \end{pmatrix}.$$

The estimated coefficients are given by

$$\hat{\beta} = (X^T X)^{-1} X^T y = \frac{1}{62} \begin{pmatrix} 146 \\ 23 \end{pmatrix}.$$

6. **(a)** The matrix is

$$X^T X = \begin{pmatrix} 2 & 6 & -4 \\ 8 & 24 & -16 \end{pmatrix} \begin{pmatrix} 2 & 8 \\ 6 & 24 \\ -4 & -16 \end{pmatrix} = \begin{pmatrix} 56 & 224 \\ 224 & 896 \end{pmatrix}$$

and its determinant is 0.

(b) The second column of x is a scalar multiple of the first, and so the two vectors span the same column space. The regression projects the y vector onto this column space, but there are infinitely-many ways to write the resulting projection as a combination of the two column vectors.

7.

$$\hat{\beta} = (X^T X)^{-1} X^T y = \frac{1}{138} \begin{pmatrix} 205 \\ 64 \end{pmatrix}$$

8. $x_3 = 12x_2$, and so the new variable does not expand the space spanned by the columns of the data matrix. All it does is make the solution indeterminant, and the matrix X^TX will not be invertible. To see this, note that if we add the column

$$X = \begin{pmatrix} 1 & 2 & 24 \\ 1 & 3 & 36 \\ 1 & 5 & 60 \\ 1 & 4 & 48 \end{pmatrix} \text{ and } X^TX = \begin{pmatrix} 4 & 14 & 168 \\ 14 & 54 & 648 \\ 168 & 648 & 7776 \end{pmatrix}$$

The determinant of X^TX is 0. Also, the third column is 12 times the second column.

9. **(a)** The eigenvalues are given by the solution to the problem

$$\begin{vmatrix} 5-\lambda & 1 \\ 4 & 2-\lambda \end{vmatrix} = 0.$$

Taking the determinant yields

$$\begin{aligned} (5-\lambda)(2-\lambda) - 4 &= 0 \\ 6 - 7\lambda + \lambda^2 &= 0 \\ \lambda &= 6, 1 \end{aligned}$$

Eigenvectors satisfy

$$\begin{pmatrix} 5-\lambda & 1 \\ 4 & 2-\lambda \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

When $\lambda = 1$, this is

$$\begin{pmatrix} 4 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

There are many solutions, but one of them is

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -4 \end{pmatrix}.$$

When $\lambda = 6$, the equation is

$$\begin{pmatrix} -1 & 1 \\ 4 & -4 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Again there are many solutions, but one of them is

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

(b) Use the same steps as before. The eigenvalues are $\lambda = 7$ and $\lambda = 6$. When $\lambda = 7$ an eigenvector is $(1, 3)$, and when $\lambda = 6$ an eigenvector is $(1, 4)$.

(c) The eigenvalues are $\lambda = 7$ and $\lambda = 0$. When $\lambda = 7$ an eigenvector is $(3, -2)$ and when $\lambda = 0$ an eigenvector is $(2, 1)$.

(d) The eigenvalues are $\lambda = 3$ and $\lambda = 2$. When $\lambda = 3$ an eigenvector is $(1, 0)$ and when $\lambda = 2$ an eigenvector is $(-4, 1)$.

(e) The eigenvalues are $\lambda = 2 + 2\sqrt{13}$ and $\lambda = 2 - 2\sqrt{13}$. When $\lambda = 2 + 2\sqrt{13}$ an eigenvector is $(-2\sqrt{13} - 8, 3)$ and when $\lambda = 2 - 2\sqrt{13}$ an eigenvector is $(2\sqrt{13} - 8, 3)$.

10. **(a)** Yes. The eigenvalues are $\lambda = 1/3$ and $\lambda = -1/5$, both of which are less than one.

(b) No. The eigenvalues are $\lambda = 5/4$ and $\lambda = 1/3$. The first one is larger than 1, so the system is unstable.

(c) Yes. The eigenvalues are $\lambda = 1/4$ and $\lambda = -2/3$, both of which have magnitude less than one.

(d) No. The eigenvalues are $\lambda = 2$ and $\lambda = -41/45$. The first one is larger than 1, so the system is unstable.

Solutions for Chapter 9

1.

$$\nabla f(x_1, x_2, x_3) = \begin{pmatrix} 2x_3^2 + 3x_2^2 - 8x_1 \\ 6x_1x_2 \\ 4x_1x_3 \end{pmatrix}$$

$$\nabla f(5, 2, 0) = \begin{pmatrix} -28 \\ 60 \\ 0 \end{pmatrix}$$

2. **(a)** The second-order Taylor approximation is

$$f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$$

We have

$$\begin{aligned} f(1) &= 2 \\ f'(x) &= -6x^2 - 5 \\ f'(1) &= -11 \\ f''(x) &= -12x \\ f''(1) &= -12 \end{aligned}$$

and so the Taylor approximation at 1 is

$$2 - 11(x - 1) - \frac{12}{2}(x - 1)^2 = -6x^2 + x + 7$$

(b) We have

$$\begin{aligned} f(1) &= -30 \\ f'(x) &= 10 - \frac{20}{\sqrt{x}} + \frac{1}{x} \\ f'(1) &= -9 \\ f''(x) &= \frac{10}{x^{\frac{3}{2}}} - \frac{1}{x^2} \\ f''(1) &= 9 \end{aligned}$$

and the Taylor approximation at 1 is

$$-30 - 9(x - 1) + \frac{9}{2}(x - 1)^2 = \frac{9}{2}x^2 - 18x - \frac{33}{2}$$

(c) We have

$$\begin{aligned} f(x) &= f'(x) = f''(x) = e^x \\ f(1) &= f'(1) = f''(1) = e \end{aligned}$$

and the Taylor approximation is

$$e + e(x - 1) + \frac{e}{2}(x - 1)^2 = \frac{1}{2}e\left(x^2 + 1\right)$$

3.

$$\begin{aligned} f(x) &\approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 \\ &= 12 - 2x - 4x^2 \end{aligned}$$

4.

$$\begin{aligned} f(x) &\approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 \\ &= c + bx + ax^2 \end{aligned}$$

The second-degree Taylor approximation gives you a second-order polynomial, and if you begin with a second-order polynomial you get a perfect approximation.

5. **(a)** Negative definite because $a_{11} < 0$ and $a_{11}a_{22} - a_{12}a_{21} = -7$.

(b) Positive semidefinite because $a_{11} > 0$ but $a_{11}a_{22} - a_{12}a_{21} = 0$.

(c) Indefinite because $a_{11} > 0$ but $a_{11}a_{22} - a_{12}a_{21} = -25$.

(d) Indefinite because

$$|a_{11}| > 0, \begin{vmatrix} 4 & 0 \\ 0 & -3 \end{vmatrix} = -12, \text{ and } \begin{vmatrix} 4 & 0 & 1 \\ 0 & -3 & -2 \\ 1 & -2 & 1 \end{vmatrix} - 25.$$

(e) Positive definite because $|A_1| = 6 > 0$ and $|A_2| = 17 > 0$.

(f) Indefinite because $|A_1| = -4 < 0$ but $|A_2| = -240 < 0$.

(g) Negative definite because $|A_1| = -2 < 0$ and $|A_2| = 7 > 0$.

(h) Indefinite because $|A_1| = 3 > 0$, $|A_2| = 8 > 0$, and $|A_3| = -44 < 0$.

6. **(a)** Letting $f(x, y)$ denote the objective function, the first partials are

$$\begin{aligned} f_x &= -y \\ f_y &= 8y - x \end{aligned}$$

and the matrix of second partials is

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ -1 & 8 \end{pmatrix}.$$

This matrix is indefinite and so the second-order conditions for a minimum are not satisfied.

(b) Letting $f(x, y)$ denote the objective function, the first partials are

$$\begin{aligned} f_x &= 8 - 2x \\ f_y &= 6 - 2y \end{aligned}$$

and the matrix of second partials is

$$\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

which is negative definite. The second-order conditions are satisfied.

(c) We have

$$\nabla f = \begin{pmatrix} 5y \\ 5x - 4y \end{pmatrix}$$

and

$$H = \begin{pmatrix} 0 & 5 \\ 5 & -4 \end{pmatrix}$$

This matrix is indefinite because $|H_1| = 0$ but $|H_2| = -25 < 0$. The second-order condition is not satisfied.

(d) We have

$$\nabla f = \begin{pmatrix} 12x \\ 6y \end{pmatrix}$$

and

$$H = \begin{pmatrix} 12 & 0 \\ 0 & 6 \end{pmatrix}$$

This matrix is positive definite because $|H_1| = 12 > 0$ and $|H_2| = 72 > 0$. The second-order condition is satisfied.

7. If it is a convex combination there must be some number $t \in [0, 1]$ such that

$$\begin{pmatrix} 6 \\ 2 \end{pmatrix} = t \begin{pmatrix} 11 \\ 4 \end{pmatrix} + (1 - t) \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Writing these out as two equations gives us

$$6 = 11t + (-1)(1 - t)$$

and

$$2 = 4t + 0(1 - t).$$

Solving the first one yields $t = 7/12$ and solving the second one yields $t = 1/2$. These are not the same so it is not a convex combination.

8. Let t be a scalar between 0 and 1. Given x_a and x_b, we want to show that

$$f(tx_a + (1-t)x_b) \leq tf(x_a) + (1-t)f(x_b)$$

Looking at the left-hand side,

$$\begin{aligned} f(tx_a + (1-t)x_b) &= (tx_a + (1-t)x_b)^2 \\ &= (x_b + tx_a - tx_b)^2 \end{aligned}$$

Looking at the right-hand side,

$$\begin{aligned} tf(x_a) + (1-t)f(x_b) &= tx_a^2 + (1-t)x_b^2 \\ &= x_b^2 + tx_a^2 - tx_b^2 \end{aligned}$$

Subtracting the left-hand side from the right-hand side gives us

$$x_b^2 + tx_a^2 - tx_b^2 - (x_b + tx_a - tx_b)^2 = t(1-t)(x_a - x_b)^2$$

which has to be nonnegative because t, $1-t$, and anything squared are all nonnegative.

Solutions for Chapter 10

1. **(a)** $x = 20$ and $y = 4$.

   (b) There are two of them: $x = 15$ and $y = 4$, and $x = 15$ and $y = 5$.

   (c) Sum the probabilities along the row to get 0.35.

   (d) $0.03 + 0.17 + 0.00 + 0.05 + 0.04 + 0.20 = 0.49$.

   (e)

$$P(y \leq 2 | x \geq 20) = \frac{P(y \leq 2 \text{ and } x \geq 20)}{P(x \geq 20)} = \frac{0.23}{0.79} = \frac{23}{79}$$

(f) Bayes' rule says

$$P(y=4|x=20)=\frac{P(x=20|y=4)\cdot P(y=4)}{P(x=20)}.$$

We have $P(y=4|x=20)=0.20/0.44=5/11$. Also,

$$\frac{P(x=20|y=4)\cdot P(y=4)}{P(x=20)}=\frac{(0.20/0.27)\cdot(0.27)}{0.44}=\frac{20}{44}=\frac{5}{11}.$$

(g) Two events are statistically independent if the probability of their intersection equals the product of their probabilities. We have

$$\begin{aligned}
P(x &\leq 20)=0.65\\
P(y &\in \{1,4\})=0.42\\
P(x &\leq 20)\cdot P(y\in\{1,4\})=(0.65)(0.42)=0.273\\
P(x &\leq 20 \text{ and } y\in\{1,4\})=0.24
\end{aligned}$$

They are not statistically independent.

2. **(a)** $P(A)=0.26$ and $P(B)=.18$, so A is more likely.

(b) The numbers in parentheses are (a,b) pairs: $\{(4,1),(4,3),(5,3)\}$.

(c) 0.73.

(d) $P(b=2|a=5)=P(b=2 \text{ and } a=5)/P(a=5)=0.06/0.32=3/16=0.1875$.

(e)

$$P(a\geq 3 \text{ and } b\in\{1,4\})=0.45$$

and

$$P(b\in\{1,4\})=0.58$$

so

$$P(a\geq 3|b\in\{1,4\})=\frac{0.45}{0.58}=0.77586$$

(f) $P(a\in\{1,3\} \text{ and } b\in\{1,2,4\})=0.14$, but $P(a\in\{1,3\})=0.36$ and $P(b\in\{1,2,4\})=0.73$. We have

$$P((a\in\{1,3\})P(b\in\{1,2,4\})=0.36*0.73=0.2628\neq 0.14.$$

They are not statistically independent.

3. We want $P(\text{disease} \mid \text{positive})$, which is

$$P(\text{disease} \mid \text{positive}) = \frac{P(\text{disease and positive})}{P(\text{positive})}.$$

Note that

$$\begin{aligned} P(\text{disease and positive}) &= P(\text{positive} \mid \text{disease}) \cdot P(\text{disease}) \\ &= 0.95 \cdot \frac{1}{20,000} \\ &= 0.0000475 \end{aligned}$$

and

$$\begin{aligned} P(\text{positive}) &= P(\text{positive} \mid \text{disease}) \cdot P(\text{disease}) + P(\text{positive} \mid \text{healthy}) \cdot P(\text{healthy}) \\ &= 0.95 \cdot \frac{1}{20,000} + 0.05 \cdot \frac{19,999}{20,000} \\ &= 0.0000475 + 0.0499975 \\ &= 0.050045 \end{aligned}$$

Now we get

$$\begin{aligned} P(\text{disease} \mid \text{positive}) &= \frac{P(\text{disease and positive})}{P(\text{positive})} \\ &= \frac{0.0000475}{0.050045} \\ &= 0.000949 \end{aligned}$$

In spite of the positive test, it is still very unlikely that Max has the disease.

4. Use Bayes' rule:

$$P(\text{entrepreneur} \mid \text{old}) = \frac{P(\text{old} \mid \text{entrpreneur})P(\text{entrepreneur})}{P(\text{old})}$$

Your grad assistant told you that $P(\text{old} \mid \text{entrepreneur}) = 0.8$ and that $P(\text{entrepreneur}) = 0.3$. But she didn't tell you $P(\text{old})$, so you must calculate it:

$$\begin{aligned} P(\text{old}) &= P(\text{old} \mid \text{doctor})P(\text{doctor}) \\ &\quad +P(\text{old} \mid \text{lawyer})P(\text{lawyer}) \\ &\quad +P(\text{old} \mid \text{entrpreneur})P(\text{entrpreneur}) \\ &= (0.6)(0.2) + (0.3)(0.5) + (0.8)(0.3) \\ &= 0.51 \end{aligned}$$

Plugging this into Bayes' rule yields

$$P(\text{entrepreneur} \mid \text{old}) = \frac{(0.8)(0.3)}{0.51} = 0.47$$

47% of old people are entrepreneurs.

Solutions for Chapter 12

1. **(a)** Plugging in $f(x) = 1/6$ gives us

$$\begin{aligned} \int_2^8 xf(x)dx &= \frac{1}{6}\int_2^8 xdx \\ &= \left.\frac{1}{12}x^2\right|_2^8 \\ &= \frac{64}{12} - \frac{4}{12} = 5 \end{aligned}$$

(b) Following the same strategy,

$$\begin{aligned} \int_2^8 x^2 f(x)dx &= \frac{1}{6}\int_2^8 x^2 dx \\ &= \left.\frac{1}{18}x^3\right|_2^8 \\ &= \frac{512}{18} - \frac{8}{18} = 28 \end{aligned}$$

2. Leibniz' rule says

$$\frac{d}{dt}\int_{a(t)}^{b(t)} f(x,t)dx = \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}dx + b'(t)f(b(t),t) - a'(t)f(a(t),t).$$

Here

$$f(x,t) = tx^2,\ b(t) = t^2, \text{ and } a(t) = -t^2$$

so

$$\frac{\partial f(x,t)}{\partial t} = x^2,\ f(b(t),t) = t\cdot(t^2)^2 = t^5, \text{ and } f(a(t),t) = t\cdot(-t^2)^2 = t^5.$$

Leibniz' rule then becomes

$$\begin{aligned}\frac{d}{dt}\int_{-t^2}^{t^2} tx^2 dx &= \int_{-t^2}^{t^2} x^2 dx + (2t)(t^5) - (-2t)(t^5) \\ &= \left.\frac{x^3}{3}\right|_{-t^2}^{t^2} + 2t^6 + 2t^6 \\ &= \frac{t^6}{3} - \frac{-t^6}{3} + 4t^6 \\ &= \frac{14}{3}t^6.\end{aligned}$$

3. Use Leibniz' rule:

$$\begin{aligned}\frac{d}{dt}\int_{a(t)}^{b(t)} f(x,t)dx &= \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}dx + b'(t)f(b(t),t) - a'(t)f(a(t),t) \\ \frac{d}{dt}\int_{-3t}^{4t^2} t^2x^3dx &= 2\int_{-3t}^{4t^2} tx^3dx + (8t)\cdot t^2\cdot(4t^2)^3 - (-3)\cdot t^2\cdot(-3t)^3 \\ &= \left.\frac{2}{4}tx^4\right|_{-3t}^{4t^2} + 512t^9 - 81t^5 \\ &= 128t^9 - \frac{81}{2}t^5 + 512t^9 - 81t^5 \\ &= 640t^9 - \frac{243}{2}t^5\end{aligned}$$

4. Let $F(x)$ denote the distribution function for $U(a,b)$, and let $G(x)$ denote the distribution function for $U(0,1)$. Then

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} \quad \text{for} & a \le x < b \\ 1 & x \ge b \end{cases}$$

and

$$G(x) = \begin{cases} 0 & x < 0 \\ x \quad \text{for} & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

First-order stochastic dominance requires $F(x) \le G(x)$. The requirements on a and b are

$$\begin{aligned} a &\ge 0 \\ b &\ge 1 \end{aligned}$$

The easiest way to see this is by graphing it. But, from looking at the equations, if $0 \le a \le 1 \le b$ we can write

$$G(x) - F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \le x < a \\ x - \frac{x-a}{b-a} \quad \text{for} & a \le x < 1 \\ 1 - \frac{x-a}{b-a} & 1 \le x < b \\ 0 & x \ge b \end{cases}$$

Note that

$$x - \frac{x-a}{b-a} = \frac{bx - ax - x + a}{b-a} = \frac{a(1-x)}{b-a} + \frac{(b-1)x}{b-a}$$

which is positive when $b \ge 1 \ge x \ge a \ge 0$, and

$$1 - \frac{x-a}{b-a} = \frac{b-a-x+a}{b-a} = \frac{b-x}{b-a}$$

which is positive when $b \ge x$. So $G(x) \ge F(x)$ as desired.

Solutions for Chapter 13

1. **(a)** $\mu = (.10)(7)+(.23)(4)+(.40)(2)+(.15)(-2)+(.10)(-6)+(.02)(-14) = 1.24$

   (b) $\sigma^2 = (.10)(7-1.24)^2 + (.23)(4-1.24)^2 + (.40)(2-1.24)^2 + (.15)(-2-1.24)^2 + (.10)(-6-1.24)^2 + (.02)(-14-1.24)^2 = 16.762.$

2. **(a)** The means are

$$\mu_f = (10)(.15) + (15)(.5) + (20)(.05) + (30)(.1) + (100)(.2) = 33$$

and

$$\mu_g = (10)(.2) + (15)(.3) + (20)(.1) + (30)(.1) + (100)(.3) = 41.5.$$

(b) The variances are

$$\begin{aligned} \sigma_f^2 &= (10-33)^2(.15) + (15-33)^2(.5) + (20-33)^2(.05) \\ &\quad +(30-33)^2(.1) + (100-33)^2(.2) \\ &= 1148.5 \end{aligned}$$

and

$$\begin{aligned} \sigma_g^2 &= (10-41.5)^2(.2) + (15-41.5)^2(.3) + (20-41.5)^2(.1) \\ &\quad +(30-41.5)^2(.1) + (100-41.5)^2(.3) \\ &= 1495.3 \end{aligned}$$

(c) The standard deviations are

$$\sigma_f = \sqrt{1148.5} = 33.890$$

and

$$\sigma_g = \sqrt{1495.3} = 38.669$$

3. **(a)**

$$\begin{aligned} F(x) &= \int_0^x f(t)dt \\ &= \int_0^x 2tdt \\ &= t^2\Big|_0^x \\ &= x^2. \end{aligned}$$

(b) All those things hold.

(c)

$$\begin{aligned}
\mu &= \int_0^1 x \cdot 2x dx \\
&= 2\int_0^1 x^2 dx \\
&= \frac{2}{3}x^3\Big|_0^1 \\
&= \frac{2}{3}.
\end{aligned}$$

(d) First find

$$\begin{aligned}
E[\tilde{x}^2] &= \int_0^1 x^2 \cdot 2x dx \\
&= 2\int_0^1 x^3 dx \\
&= \frac{2}{4}x^4\Big|_0^1 \\
&= \frac{1}{2}.
\end{aligned}$$

Then note that

$$\sigma^2 = E[\tilde{x}^2] - \mu^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

4. **(a)** For $x \in [0, 4]$ we have

$$\begin{aligned}
F(x) &= \int_0^x \frac{1}{8} t dt \\
&= \frac{1}{16}t^2\Big|_0^x \\
&= \frac{1}{16}x^2
\end{aligned}$$

Outside of this interval we have $F(x) = 0$ when $x < 0$ and $F(x) = 1$ when $x > 1$.

(b) Using $F(x) = x^2/16$, we get $F(0) = 0$, $F(4) = 1$, and $F'(x) = x/8 \geq 0$.

(c)

$$\mu = \int_0^4 x \left(\frac{1}{8}x\right) dx = \frac{8}{3}$$

(d)

$$\sigma^2 = \int_0^4 \left(x - \frac{8}{3}\right)^2 \left(\frac{1}{8}x\right) dx = \frac{8}{9}$$

5. The mean of the random variable $\alpha\tilde{x}$ is $a\mu$, where μ is the mean of $\tilde{x}$. The variance is

$$\begin{aligned} Var(a\tilde{x}) &= E[(a\tilde{x} - a\mu)^2] \\ &= E[a^2(\tilde{x} - \mu)^2] \\ & a^2E[\tilde{x} - \mu]^2 \\ &= a^2\sigma^2. \end{aligned}$$

The first line is the definition of variance, the second factors out the a, the third works because the expectations operator is a linear operator, and the third is the definition of σ^2.

6. We know that

$$E[(x - \mu_x)^2] = \sigma_x^2$$

We want to find

$$\sigma_y^2 = E[(y - \mu_y)^2]$$

Note that $\mu_y = 3\mu_x - 1$, and that $y = 3x - 1$. Substituting these in yields

$$\begin{aligned} \sigma_y^2 &= E[(y - \mu_y)^2] \\ &= E[(3x - 1 - (3\mu_x - 1))^2] \\ &= E[(3x - 3\mu_x)^2] \\ &= E[9(x - \mu_x)^2] \\ &= 9E[(x - \mu_x)^2] \\ &= 9\sigma_x^2. \end{aligned}$$

7. The mean is $\mu = 3 + \frac{1}{2}y$. The variance is therefore

$$\begin{aligned}\sigma^2 &= \frac{1}{2}[6-(3+\frac{1}{2}y)]^2 + \frac{1}{2}[y-(3+\frac{1}{2}y)]^2 \\ &= \frac{1}{4}y^2 - 3y + 9.\end{aligned}$$

The derivative with respect to y is

$$\frac{d\sigma^2}{dy} = \frac{1}{2}y - 3.$$

8. All we have to do is show that $G^{(2)}(x) \geq G^{(1)}(x)$ for all x. We have

$$\begin{aligned}G^{(2)}(x) - G^{(1)}(x) &= \left[nF^{n-1}(x)(1-F(x)) + F^n(x)\right] - [F^n(x)] \\ &= nF^{n-1}(x)(1-F(x)) \geq 0.\end{aligned}$$

Solutions for Chapter 14

1. **(a)**

$F(x,y)$	$\tilde{y}=10$	$\tilde{y}=20$	$\tilde{y}=30$
$\tilde{x}=1$	.04	.04	.24
$\tilde{x}=2$	.11	.11	.49
$\tilde{x}=3$	.13	.24	.69
$\tilde{x}=4$	.14	.37	1.00

(b) $F_{\tilde{x}}$ is given by the last column of part (a), and $F_{\tilde{y}}$ is given by the bottom row of part (a).

(c) $f_{\tilde{x}}(1) = .24$, $f_{\tilde{x}}(2) = .25$, $f_{\tilde{x}}(3) = .20$, $f_{\tilde{x}}(4) = .31$. Similarly, $f_{\tilde{y}}(10) = .14$, $f_{\tilde{y}}(20) = .23$, $f_{\tilde{y}}(30) = .63$.

(d) The formula for conditional density is $f(x|\tilde{y}=20) = f(x,20)/f_{\tilde{y}}(20)$, which gives us $f(1|\tilde{y}=20) = 0/.23 = 0$, $f(2|\tilde{y}=20) = 0$, $f(3|\tilde{y}=20) = 11/23$, and $f(4|\tilde{y}=20) = 12/23$.

(e) Using the marginal density from part (c), the mean is

$$\mu_y = (.14)(10) + (.23)(20) + (.63)(30) = 24.9$$

(f) Using part (d),

$$\mu_{x|y=20} = (0)(1) + (0)(2) + (11/23)(3) + (12/23)(4) = 81/23$$

(g) No. For the two to be independent we need $f(x,y) = f_{\tilde{x}}(x)f_y(y)$. This does not hold. For example, we have $f(3,20) = .11$, $f_{\tilde{x}}(3) = .20$, and $f_{\tilde{y}}(20) = .23$, which makes $f_{\tilde{x}}(3)f_{\tilde{y}}(20) = .046 \neq .11$.

(h) We have

$$\begin{aligned} E_x[\tilde{x}|\tilde{y} &= 10] = \frac{.04(1) + .07(2) + .02(3) + .01(4)}{.14} = 2.0 \\ E_x[\tilde{x}|\tilde{y} &= 20] = \frac{0(1) + 0(2) + .11(3) + .12(4)}{.23} = 3.52 \\ E_x[\tilde{x}|\tilde{y} &= 30] = \frac{.2(1) + .18(2) + .07(3) + .18(4)}{.63} = 2.36 \\ E_y[E_x[\tilde{x}|y] &= (.14)(2.0) + (.23)(3.52) + (.63)(2.36) = 2.58 \end{aligned}$$

Finally, using the marginal density from part (c) yields

$$E_x[\tilde{x}] = (.24)(1) + (.25)(2) + (.20)(3) + (.31)(4) = 2.58.$$

It works.

2. **(a)**

$f(x,y)$	$\tilde{y}=3$	$\tilde{y}=8$	$\tilde{y}=10$
$\tilde{x}=1$	0.03	0.05	0.25
$\tilde{x}=2$	0.05	0.19	0.44
$\tilde{x}=3$	0.10	0.25	0.71
$\tilde{x}=4$	0.17	0.43	1.00

(b) $F_{\tilde{x}}(1) = 0.25$, $F_{\tilde{x}}(2) = 0.44$, $F_{\tilde{x}}(3) = 0.71$, $F_{\tilde{x}}(4) = 1.00$ and $F_{\tilde{y}}(3) = 0.17$, $F_{\tilde{y}}(8) = 0.43$, $F_{\tilde{y}}(10) = 1.00$.

(c) $f_{\tilde{x}}(1) = 0.25$, $f_{\tilde{x}}(2) = 0.19$, $f_{\tilde{x}}(3) = 0.27$, $f_{\tilde{x}}(4) = 0.29$ and $f_{\tilde{y}}(3) = 0.17$, $f_{\tilde{y}}(8) = 0.26$, $f_{\tilde{y}}(10) = 0.57$.

(d) $f(\tilde{y} = 3|\tilde{x} = 1) = 0.12$, $f(\tilde{y} = 8|\tilde{x} = 1) = 0.08$, and $f(\tilde{y} = 10|\tilde{x} = 1) = 0.80$.

(e) $\mu_x = 2.6$ and $\mu_y = 8.29$.

(f) $E[\tilde{x}|\tilde{y}=3] = [(0.03)(1) + (0.02)(2) + (0.05)(3) + (0.07)(4)]/0.17 = 2.94$.

(g) No. The following table shows the entries for $f_{\tilde{x}}(x)f_{\tilde{y}}(y)$:

$f_{\tilde{x}}(x)f_{\tilde{y}}(y)$	$\tilde{y}=3$	$\tilde{y}=8$	$\tilde{y}=10$
$\tilde{x}=1$	0.043	0.065	0.143
$\tilde{x}=2$	0.032	0.049	0.108
$\tilde{x}=3$	0.046	0.070	0.154
$\tilde{x}=4$	0.049	0.075	0.165

None of the entries are the same as those in the $f(x,y)$ table.

(h) $\sigma_x^2 = 1.32$ and $\sigma_y^2 = 6.45$.

(i) $Cov(\tilde{x},\tilde{y}) = -0.514$

(j) $\sigma_{xy} = -.176$.

(k) We have

$$E_x[\tilde{x}|\tilde{y} = 3] = \frac{(.03)(1)+(.02)(2)+(.05)(3)+(.07)(4)}{.17} = 2.94$$
$$E_x[\tilde{x}|\tilde{y} = 8] = \frac{(.02)(1)+(.12)(2)+(.01)(3)+(.11)(4)}{.26} = 2.81$$
$$E_x[\tilde{x}|\tilde{y} = 10] = \frac{(.2)(1)+(.05)(2)+(.21)(3)+(.11)(4)}{.57} = 2.40$$

$$E_y[E_x[\tilde{x}|y]] = (.17)(2.94) + (.26)(2.81) + (.57)(2.40) = 2.6$$

which is the same as the mean of $\tilde{x}$ found above.

3. The uniform distribution over $[a,b]$ is

$$F(x) = \frac{x-a}{b-a}$$

when $x \in [a,b]$, it is 1 for $x > b$, and 0 for $x < a$. The conditional distribution is

$$F(x|x \le c) = \frac{F(x)}{F(c)} = \frac{\frac{x-a}{b-a}}{\frac{c-a}{b-a}} = \frac{x-a}{c-a}$$

for $x \in [a,c]$, it is 1 for $x > c$, and 0 for $x < a$. But this is just the uniform distribution over $[a,c]$.

4. $\tilde{x}$ and $\tilde{y}$ are independent if $f(x,y) = f_{\tilde{x}}(x)f_{\tilde{y}}(y)$ or, equivalently, if $f(x|y) = f(x)$. This answer uses the latter formulation. We can see that $f(\tilde{x} = -1|\tilde{y} = 10) = 1/4$, and for $\tilde{x}$ and $\tilde{y}$ to be independent it must also be the case that $f(\tilde{x} = -1|\tilde{y} = 20) = 1/4$. But

$$f(\tilde{x} = -1|\tilde{y} = 20) = \frac{a}{a+b}.$$

We also know that $a + b$ must equal 0.6 so that the probabilities sum to one. Thus,

$$\begin{aligned} \frac{a}{a+b} &= \frac{a}{0.6} = \frac{1}{4} \\ a &= \frac{.6}{4} = 0.15 \\ b &= 0.6 - a = 0.45. \end{aligned}$$

18.1 Solutions for Chapter 15

1. $\bar{x} = 4$ and $s^2 = 24$.

Solutions for Chapter 17

1. **(a)** Compute the t-statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{60.02 - 0}{44.37/\sqrt{30}} = 7.41$$

which has 29 degrees of freedom. Use the Excel formula

=TDIST(7.41, 29, 2)

to get the p-value of 0.0000000365. The data reject the hypothesis.

(b) The t-statistic is 3.706, the p-value is 0.000882, and the hypothesis is rejected.

(c) The t-statistic is -0.614, the p-value for the one-tailed test is 0.27, and the hypothesis is supported.

(d) The sample mean is 60.02 which is smaller than 100, so the hypothesis is supported.

2. **(a)** The best estimate of μ is the sample mean $\bar{x}$ and the best estimate of σ^2 is the sample variance s^2.

$$\begin{aligned}\bar{x} &= 44.2\\ s^2 &= 653.8\end{aligned}$$

(b) Compute the t-statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{44.2 - 40}{25.6/\sqrt{20}} = 0.73$$

and the t-statistic has 19 degrees of freedom. From here compute the p-value of $2(1 - \text{TDist}(0.73, 19)) = 0.47 > 0.05$ and the data support the hypothesis.

(c) Compute the t-statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{44.2 - 60}{25.6/\sqrt{20}} = -2.76$$

and again the t-statistic has 19 degrees of freedom. From here compute the p-value of $2(1 - \text{TDist}(2.76, 19)) = .01\,25 < 0.05$ and the data reject the hypothesis.

INDEX