Introduction to the
Economics of Uncertainty and Information

Timothy Van Zandt
INSEAD

November 2004

Author's address:

INSEAD	Voice:	+33 1 6072 4981
Boulevard de Constance	Fax:	+33 1 6074 6192
77305 Fontainebleau CEDEX	Email:	tvz@insead.edu
FRANCE	WWW:	faculty.insead.edu/vanzandt

Table of Contents

1 Choosing among Uncertain Prospects **1**

1.1 Introduction to decision theory 1
1.2 Lotteries and objective expected utility 4
1.3 States of nature and subjective expected utility 24

2 Choosing when there is new information **33**

2.1 Representing information 33
2.2 Bayes' Theorem 37
2.3 Independence and Exchangeability 48

3 Risk Preferences **57**

3.1 Money outcomes and risk aversion 57
3.2 Comparing the risk aversion of people 65
3.3 Comparing the riskiness of acts 68

4 Market Decisions in the Presence of Risk **79**

4.1 Demand for a state-contingent contract 79
4.2 Techniques for characterizing market decisions 85
4.3 Comparative statics 96
4.4 State-dependent utility 103

5 Markets for state-contingent contracts **109**

5.1 Gains from sharing risks 110
5.2 Insurance market 117

6 Asset Markets **121**

6.1 The nature of asset markets 121
6.2 Market equilibrium 125
6.3 Complete versus incomplete markets 139
6.4 Capital Asset Pricing Model 148

7 Contracting with Hidden Actions **153**

7.1 Efficient contracts with moral hazard 153
7.2 Contracts that give one party the gains from trade 154

8 Monopolistic screening with hidden valuations **157**

8.1 Nonlinear pricing 157
8.2 Differentiated products 157
8.3 Bundling 158

9 Screening with adverse selection **159**

9.1 The nature of adverse selection 159
9.2 Screening 161
9.3 Efficient contracts with adverse selection 168
9.4 The efficient contracts with zero expected profits 170

9.5 Monopolistic screening 178

10 Signaling 181

10.1 The difference between screening and signaling 181

10.2 Signaling in labor markets 181

11 Long-term versus short-term contracting 183

11.1 Ex-ante versus ex-post efficiency 183

11.2 Long-term and short-term insurance contracts 185

Preface

Our subject

This book is about uncertainty and information in economics. Uncertainty and information are inexorably linked. You face uncertainty when you do not know something about the world, such as the consequences of your actions. This means that you lack complete information. Obtaining information then resolves this uncertainty.

As topics in economics, we distinguish between the economics of uncertainty and the economics of information. Both topics involve uncertainty and information. Here is how they differ:

- The *economics of uncertainty* is about economic situations in which there is uncertainty but all involved parties have the *same* information.
- The *economics of information* is about situations in which the parties have *different* or *asymmetric* information.

Here is a scenario that would fall under the economics of uncertainty. When you buy life insurance for a specific flight, it is because you face uncertainty about whether the airplane will crash. The life insurance policy transfers some of this risk to the insurance company, which then also faces uncertainty about the insurance losses it will pay. It is a rough approximation that both you and the insurance company have the same information about the likelihood that the plane will crash. Both you and the insurance company can easily determine whether the plane crashes. With respect to this transaction, you and the insurance company have the same information. When studying this kind of transaction, we will be interested in the premiums offered by insurance companies and the optimal amount of insurance the flyer should buy.

Here is a scenario that would fall under the economics of information. Suppose you are buying health insurance. As with the flight insurance, your goal is to transfer risk to an insurance company. This time, you and the insurance company have different information about several aspects of the transaction:

- Before buying the insurance, you know more about whether you have a high risk of heart attack or other diseases.
- After buying the insurance, you can take actions that increase the chances of health problems, such as smoking, eating a high-fat diet, or engaging in high-risk sex, and the insurance company cannot observe whether you take these actions.
- When you become unhealthy, you demand reimbursements or health care. You—and your doctor—have better information about the state of your health and the value of the medical procedures than the insurance company does.

When contracting with you, the insurance companies have to take into account that you may lie about your existing health condition before contracting, you may secretly engage in high-risk activities after contracting, and you may exaggerate health problems in order to receive more medical attention or reimbursements. When studying such an insurance transaction with asymmetric information, we will be interested in how the insurance contracts contain special provisions, such as deductibles, that are meant to compensate for the informational asymmetries.

Uncertainty and information are not specific to a particular set of economic activities. Rather, they are a potential aspect of any economic activity. Therefore, the economics of uncertainty and information is not a separate subdiscipline, but rather it is part of all the subdisciplines:

- It is part of labor economics because, for example, employees know more about their capabilities within a firm than the employers do.
- It is important in finance because most financial instruments have uncertain returns, and traders have different information about these returns; furthermore, most financial instruments serve the role of sharing risks, the way insurance contracts do.
- It is part of health economics, as illustrated by the health insurance example given above.
- It is part of industrial organization and market structure because sellers have better information about the value of their goods than buyers do, and because each firm has better information about its costs than its competitors do.

Our methods

This list could go on. In this book, we will cover many of these topics. However, rather than studying a series of seemingly unrelated anecdotes about uncertainty and asymmetric information, the goal is to develop unifying models and theories.

A model is a simplified representation of a phenomenon, that emphasizes only the most important aspects, that is specified by using a formal language, and that allows us to derive consequences of the parameters of the model.

Simplification is a goal of modeling, not an unintended negative consequence. The world is too complex to be understood by considering all its detail at the same time. The purpose of models to identify certain important cause-and-effect relationships between just a few components of the world. If the model is good, then the components we ignore might introduce new relationships, but will not obliterate those we have identified. Of course, all conclusions drawn from models have to be taken with a grain of salt and not applied dogmatically to real-world situations; but this should be a lesson you learned ago about all the education you have received.

The formal language we use for the models in this book is mathematics, although limited to simple algebra, single-variable calculus, and a little probability theory. The use of mathematics does not mean that our theories are meant to generative quantitative answers from the specific values of the data in the model. Instead, much of our theory is qualitative, meaning that the specific data may be difficult to observe and so we attempt to say as much as possible using only a few qualitative properties of the data.

Chapter 1

Choosing among Uncertain Prospects

1.1 Introduction to decision theory

1.1.1 Why individual decision theory in economics?

In microeconomics, we build models by first identifying the individual agents, who may be individuals (such as workers and managers) or organizations (such as households, firms and countries). Then we specify the motives of the individual agents and derive the behavior of these agents from their motives. Hence, the foundation of microeconomics is individual decision theory.

1.1.2 Descriptive, prescriptive and normative theories

Decision theory has two goals: To *describe* how agents *do* make decisions (descriptive decision theory) and to *prescribe* how agents *should* make decisions (prescriptive decision theory). As in any theoretical modeling, decision theory balances accuracy and simplicity. A prescriptive decision theory that is too complicated to explain or learn is hardly useful for a decision maker. A descriptive theory should be simple because it is meant to be a framework that organizes and unifies a variety of situations and behavior, because it should be tractable enough to derive conclusions from it, and because we may have to estimate the parameters of the theory from a limited amount of data.

There is a third branch of decision theory, called normative decision theory, whose goal is to describe how a hypothetical, infinitely intelligent being would make decisions. This exercise is simpler than descriptive or prescriptive decision theory, because we do not have to worry about complications such as errors or forgetting and we do not have to worry about the heterogeneity of the intelligence and experience of decision makers. There are only a few ways to be perfect, but many ways to be imperfect! This simplicity has made normative decision theory a good source of models for both descriptive and prescriptive decision theory. In this book, all our models will be drawn from normative theories.

Since people are far from infinitely smart, it is easy to come up with empirical violations of these theories and I will do so on several occasions. The purpose of these examples is *not* to convince you that the models are bad; a model is supposed to be a simplification—rather than replication—of reality. Instead, the purpose is to give you a healthy appreciation of the limitations of the theories. In particular, I want to avoid some extreme and unreasonable views that have polarized decision theory in the past. Economists, who have to make many concessions to simplicity at the level of decision theory in order to build up complex models involving many agents, have predominantly used normative decision theory; in the process, some economists have forgotten that the theory is just an approximation, and not the Truth. For example, some economists have claimed that even if a person violated a normative theory, he would not do so after the inconsistency was explained to him; this ignores the fact that implementing a normative theory requires computational power beyond that of real humans. On the other hand, many decision theorists outside of economics—especially in psychology and philosophy, whose focus is on the behavior of individuals and who therefore can develop more realistic descriptive models—have sometimes insinuated that the accumulated evidence of violations of normative theories invalidates the economic models built on these theories. This view ignores the fact that, even if there are benefits to more complex models of decision making,[1] the simpler models have been useful approximations in the study of many economic phenomena.

1.1.3 A review of choice, preferences and utility

As an introduction to the kind of decision-theory exercises in which we will engage in subsequent sections, I will review a decision theory for choice without uncertainty.

First, we have to specify the objects of choice and the decision process. Let's take a basic situation in which an agent can choose from a set A of feasible alternatives, which belong to a large set X containing all potential alternatives. For example, X is the set of all consumption bundles, and A is the budget set, which depends on prices and the agent's wealth. Alternatively, X is the set of potential presidential candidates and A is the set of candidates who appear on the ballot. We want a model that tells us what the agent would choose from each set of feasible alternatives. We do not want to determine exactly what choices would be made; this would be a list of data for a single individual rather than a model that would apply to many individuals whose choices would differ. Instead, we want to find some consistency conditions that allow us to come up with a simple representation of choice that can be used to derive conclusions without knowing the actual choices, and that has a few parameters that could be estimated from limited data.

Assume that X is finite and every non-empty subset A of X is a potential feasible set.[2] For each set $A \subset X$ of feasible alternatives, let $C(A)$ be the elements of A that the decision maker might choose from A. The decision maker always has to choose something, which means that $C(A)$ is non-empty, but $C(A)$ may contain more than one item because of indifference. $C(\cdot)$ is called the decision maker's *choice rule*.

Let x and y belong to X. If $x \in C(A)$ for some $A \subset X$ containing x and y (that is, y is available but x may be chosen), then we say that x is revealed weakly preferred to y. If also $y \notin C(A)$, then we say that x is revealed preferred to y or, if we want to emphasize that the preference is not weak, that x is revealed strictly preferred to y.

We now impose a consistency condition, called the *Weak Axiom of Revealed Preference* (WARP). It says that if x is revealed weakly preferred to y, then y is not revealed preferred to x:

1. This is an area of current research in economics
2. The discussion here requires a few extra technical details for the case where X is infinite.

Assumption 1. Let x and y belong to X, and let A and B be subsets of X containing x and y. If $x \in C(A)$ and $y \in C(B)$, then $x \in C(B)$.

You can see why this is a natural axiom of a normative theory, but at best an approximation for descriptive or prescriptive theories. Especially when choice sets are large and complex, achieving such consistency is difficult.

This one consistency condition gets us far. First, it implies that choices can be represented by *preferences*, which are defined by binary choices. For each x and y in X, write $x \succeq y$ if $x \in C(\{x, y\})$. This is read "x is weakly preferred to y." If also $y \notin C(\{x, y\})$, then write also $x \succ y$ ("x is (strictly) preferred to y"); otherwise, write also $x \sim y$ ("x is indifferent to y"). The symbols $\succeq$, $\succ$ and $\sim$ are called the (weak) preference, strict preference, and indifference relations, respectively.

Definition 1. The choice rule $C(\cdot)$ satisfies preference maximization if, for every $A \subset X$ and $x \in A$:

$$x \in C(A) \iff x \succeq y \text{ for all } y \in A.$$

In words, choices from large sets are consistent with binary choices. If we know binary choices (preferences), then we know $C(\cdot)$. Preference maximization implies a considerable savings in the amount of information we need in order to know $C(\cdot)$.

It is also useful that preferences satisfy some consistency conditions themselves.

Definition 2. The preference relation $\succeq$ is said to be *rational* if it satisfies:

1. (Completeness) For all $x, y \in X$, we have $x \succeq y$ or $y \succeq x$ (or both).
2. (Transitivity) For all $x, y, z, \in X$, if $x \succeq y$ and $y \succeq z$, then $x \succeq z$.

You can probably guess what comes next.

Proposition 1. *The choice rule $C(\cdot)$ satisfies WARP if and only if it satisfies preference maximization and the preference relation is rational.*

One advantage of rational preferences is that they can be represented by a utility function.

Proposition 2. *If the preference relation $\succeq$ is rational, then there is a utility function $U\colon X \to \mathbb{R}$ such that, for all $x, y, \in X$,*

$$x \succeq y \iff U(x) \geq U(y).$$

The utility representation of $\succeq$ is not unique.[3] For example, if x contains just money values and the decision maker prefers more money to less, then any strictly increasing function $U\colon X \to \mathbb{R}$ is a representation of the preferences.

3. If $U\colon X \to \mathbb{R}$ is a utility representation, and if $f : \mathbb{R} \to R$ is strictly increasing, then the composition $V : X \to \mathbb{R}$ of U and f, defined by $V(x) = f(U(x))$, is also a utility representation. V is called a *monotonic transformation* of U.

If I just want to say that x is weakly preferred to y, then it is more straightforward to write $x \succeq y$ rather than $U(x) \geq U(y)$. Utility functions have more important uses. If I want to present an example of preferences and if X is large, it can be easier to write down a utility function rather than a list of preferences. For example, if there are two consumption goods, it is easy to say that a consumer's utility function for quantities x_1 and x_2 of the two goods is

$$U(x_1, x_2) = \log x_1 + \log x_2,$$

but impossible to list the preferences for all x_1 and x_2. Furthermore, if I do not know preferences, I can posit a utility function with a small number of unknown parameters and then estimate these parameters using econometrics.

You might be having the following thought.

> *OK, so we have a model in which a decision maker is a utility maximizer. Why didn't we just assume this in the first place? Why bother with WARP and all that other stuff in between?*

We could have done so. However, the intermediate steps are important because they tell us in more intuitive terms what lies behind the model. If we just assumed that agents are utility maximizers, we would not have a good rebuttal to the complaint that utility maximization is a crazy idea since no one really walks around with a utility function in her head. Instead, we have the following response.

> *Utility maximization is just a modeler's tool for representing choices. WARP is a compelling consistency condition even for decision makers who are not explicitly maximizing utility, and it implies that the utility representation is valid. Furthermore, although it it is difficult to test whether a consumer has a particular utility function, it is easy to test for violations of WARP from observed behavior, and hence check the empirical validity of the existence of a utility representation.*

1.2 Lotteries and objective expected utility

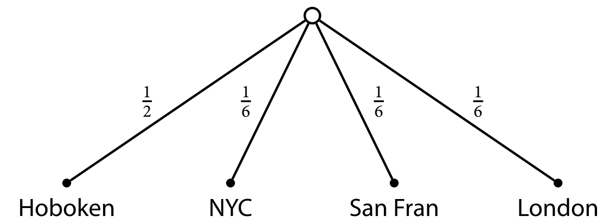
1.2.1 Simple and compound lotteries

Now we turn to decision making under uncertainty. We begin with the simplest case, where *outcomes* (also called *consequences* or *prizes*) are determined by some randomization device such that the probability of each outcome can be objectively determined. Leading examples are lotteries and other gambles. It is assumed that the decision maker cares only about the outcome or the probability distribution over outcomes—not about the random process that generates the outcome. For example, the decision maker is indifferent between the following three choices because they yield the same probability distributions over outcomes:

1. A die is thrown; the DM gets \$1 if the die is odd and \$0 otherwise.
2. A die is thrown; the DM gets \$1 if the die is at least four and \$0 otherwise.
3. A coin is tossed; the DM gets \$1 if the coin comes up heads and \$0 otherwise.

Figure 1.1

Outcome	Prob.
Hoboken	1/2
NYC	1/6
San Fran	1/6
London	1/6



Two ways to present a simple lottery.

This means that when describing the DM's choices, we need specify only the probability distribution over outcome, rather than the random process by which outcomes are selected. (We *could* also specify the random process, but this would be irrelevant detail.)

Formally, let X be the set of outcomes. We assume, for now at least, that X is finite. Each uncertain prospect is represented by a *probability measure* $P: X \to \mathbb{R}$ on X, where $P(x)$ is the probability of outcome x. What makes a function $P: X \to \mathbb{R}$ a probability measure are the following two properties:

1. $P(x) \geq 0$ for every $x \in X$.
2. $\sum_{x \in X} P(x) = 1$.

That is, (1) the probability of each outcome is non-negative and (2) the probability that some outcome occurs is 1. We call these uncertain prospects or probability measures *lotteries*. A simple lottery can be visualized as a tree, with two levels. We let the outcomes be the terminal nodes and write the probabilities on the branches. An example is given in Figure 1.1.

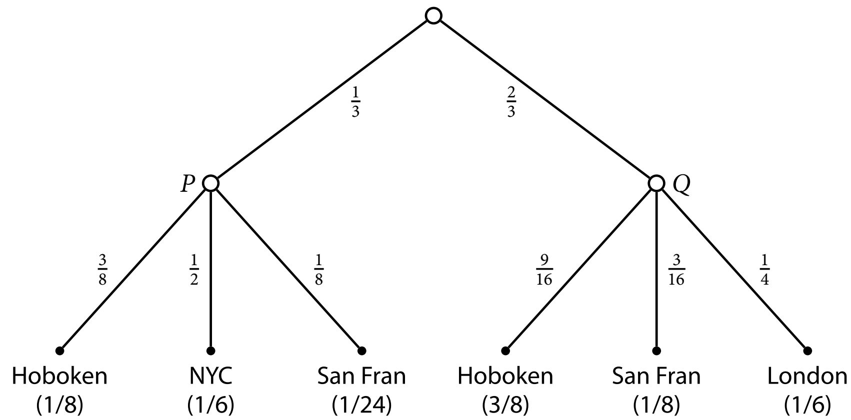
Let $\mathscr{L}$ be the set of lotteries, i.e., the set of probability measures on X. Our task is to characterize how a decision maker chooses from any subset of feasible lotteries. We can start by applying the results of Section 1.1.3. Although I described that section as a review of choice under certainty, the set X of alternatives was arbitrary and could have been the set $\mathscr{L}$ of lotteries. Thus, under WARP,[4] the decision maker's choices maximize a complete, transitive preference relation $\succeq$, which can be represented by a utility function $U: \mathscr{L} \to \mathbb{R}$.

Are we finished? Is there no difference between choice under certainty and choice under uncertainty? We could stop here, but we have not taken advantage of the fact that lotteries have special structures. Using the special properties of choice under uncertainty, we can obtain stronger results. We do this by imposing some restrictions on preferences over lotteries that are natural for a normative theory.

For this purpose, we introduce compound lotteries. In a compound lottery, first some uncertainty is resolved and then the DM faces a new lottery. For example, consider a housing lottery where first you are assigned randomly to a dorm and then you are assigned randomly to a roommate. Formally, given a set X of outcomes: (i) a *simple* lottery is a lottery as we defined above; (ii) a *compound* lottery is a lottery in which the outcomes are simple lotteries. A compound lottery can be represented by a tree with

4. With a technical modification, because the set of lotteries is not finite.

Figure 1.2



A compound lottery. It has the same overall distribution over outcomes as the simple lottery in Figure 1.1.

three levels, in which each subtree is a simple lottery and the probability of each simple lottery is written on the branches of the top level. An example is given in Figure 1.2.

We can find the probability of each terminal node in a compound lottery by multiplying the probabilities of the branches leading to the terminal nodes. For example, the probability of the left-most terminal node in Figure 1.2 is the probability of facing the lottery P times the probability of drawing Hoboken in lottery P, or $1/3 \times 3/8 = 1/8$. In Figure 1.2, the probability of each terminal node is written below the node.

Each outcome can appear in more than one terminal node of a compound lottery. That is, it can be a possible outcome of more than one of the possible second-stage lotteries. To find the overall probability of each outcome, we have to sum the probability of each terminal node containing the outcome. If $\alpha_1, \ldots, \alpha_n$ are the probabilities of each of the possible second-stage lotteries $P_1, \ldots, P_n$, then the overall probability $R(x)$ of outcome x in the compound lottery is

$$R(x) \;=\; \alpha_1 P_1(x) + \alpha_2 P_2(x) + \cdots + \alpha_n P_n(x). \tag{1.1}$$

This overall probability measure R on X is a simple lottery that we call the *reduced lottery* of the compound lottery. Performing these calculations for the compound lottery in Figure 1.2 reveals that its reduced lottery is the simple lottery in Figure 1.1.

The formula in equation (1.1) for the probabilities in the reduced lottery suggests the following notation for denoting the reduced lottery:

$$R \;=\; \alpha_1 P_1 + \alpha_2 P_2 + \cdots + \alpha_n P_n. \tag{1.2}$$

That is, if $P_1, \ldots, P_n$ are simple lotteries and $\alpha_1, \ldots, \alpha_n$ are positive numbers that sum to 1, then equation (1.2) shows the simple lottery for which the probability of outcome x is given by equation (1.1). Hence, it is the reduced lottery of the compound lottery in which lottery P_i occurs in the second stage with probability α_i.

In keeping with our assumption that the decision maker cares only about outcomes or consequences—not about the random process that determines the outcomes—we assume that the decision maker is indifferent between a compound lottery and its reduced lottery. (This is an example of what is called *consequentialism*.) If each compound lottery is equivalent to its reduced lottery, then compound lotteries become what we said we should avoid: irrelevant detail. However, they are useful for stating some assumptions about the decision maker's preferences in an intuitive and compelling way.

1.2.2 The Independence Axiom

Your friend Akbar has received a free international round-trip ticket and is planning to use it for his winter vacation. Unfortunately, he is making his reservation late. His preferred destinations, Hawaii and Madrid, are sold out. So he makes a reservation for Cancun. He can also choose to be wait-listed for Hawaii or Madrid, but not both. For either destination, the probability that he can ultimately get a reservation is 1/2. Depending on whether he chooses to be wait-listed for Hawaii or Madrid, he faces one of the following lotteries:

Figure 1.3

Akbar, not having read this book and bewildered by the uncertainty in this decision, asks for your advice.

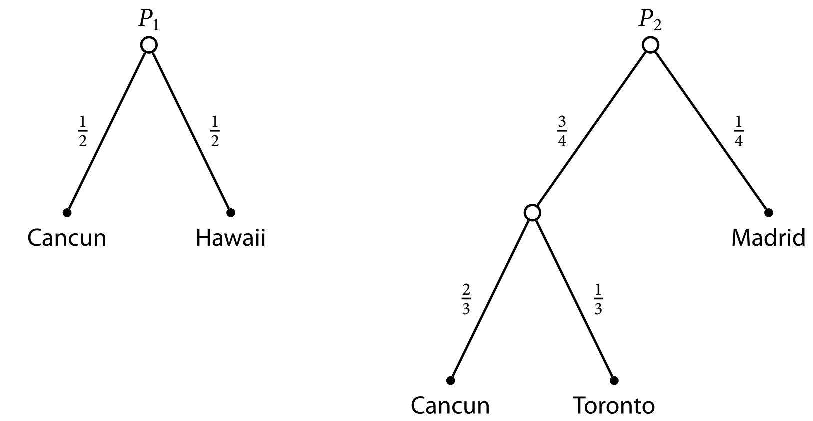
"Simple!" you tell him. "Which destination do you prefer, Hawaii or Madrid?"

"Madrid, by a long shot!" he replies.

You advise him to get on the waiting list for Madrid and you give him the following reason: In either lottery, the probability of not getting a reservation is the same (1/2) and the consequence when he doesn't get a reservation is the same (ending up in Cancun). Therefore, all that should matter is what happens when he does get a reservation. If he would prefer Madrid for sure over Hawaii for sure, then he should get on the waiting list for Madrid.

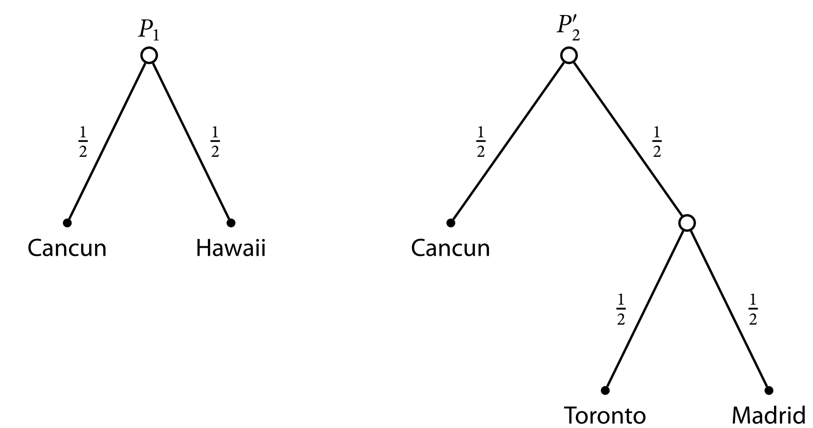
Akbar goes off to make his reservation, but comes back to you the next day looking gloomy and in need of more help. It turns out that he had previously misunderstood the travel agent. If he decides to get on the waiting list for Hawaii, then the situation is as before. However, if he decides to get on the waiting list for Madrid, then the situation is completely different. First, the probability of getting a reservation for Madrid is only 1/4 (rather than 1/2). Second, to get on the waiting list, he has to drop his reservation for Cancun. If he then doesn't get a reservation for Madrid, there is a 2/3 chance he can get back his reservation for Cancun, but there is a 1/3 chance he will only be able to get a reservation for Toronto. Thus, he really has to choose from the lotteries in Figure 1.4.

Figure 1.4



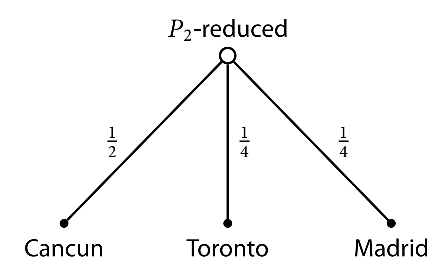
To help, you suggest that Akbar instead compare the lotteries in Figure 1.5.

Figure 1.5



You explain that P_2 and P_2' have the same reduced lottery, shown in Figure 1.6.

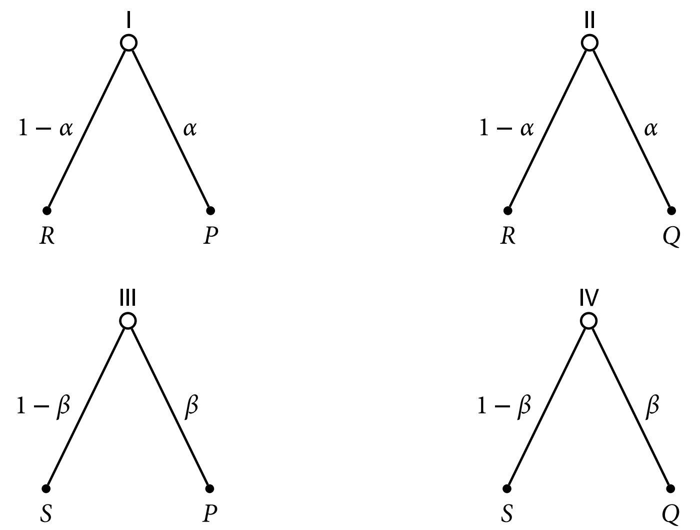
Figure 1.6



Hence, the ranking of P_1 and P_2 should be the same as the ranking of P_1 and P_2'.

Akbar resists this last idea. He claims that this rewriting of P_2 has changed the decision problem, because P_2' does not reflect the way uncertainty is resolved. In the tree, it looks as if first Akbar learns whether he goes to Cancun and then, if not, he faces a 50–50 chance of Toronto or Madrid. But you explain to Akbar that all that should matter is what simple lottery the compound lottery reduces to. If, for example, the destination is chosen by flipping a coin twice, should Akbar's preferences depend on whether he closes his eyes during the coin tosses?

Figure 1.7



P, Q, R and S are lotteries, and α and β are probabilities. The Independence Axiom states that lottery I is preferred to lottery II if and only if lottery P is preferred to lottery Q. An implication of the Independence Axiom is that lottery III is preferred to lottery IV if and only if lottery I is preferred to lottery II.

Then you explain that you rewrote P_2 because P_2' is easier to compare with P_1. You ask Akbar: "What would you prefer, to go to Hawaii for sure, or to have a 50-50 chance between going to Toronto and Madrid?" (If he can't answer this question, then there isn't much you can do to help him!)

He replies, "To go to Hawaii for sure."

Then you advise him to get on the waiting list for Hawaii. He is not quite convinced, but you explain: "Compare P_1 and P_2'. In either case, there is a 50-50 chance of ending up in Cancun. The difference is that—conditional on not ending up in Cancun—you get to go to Hawaii for sure if you get wait-listed for Hawaii, whereas you face a 50-50 chance between going to Toronto or Madrid if you get wait-listed for Madrid. Only this difference should matter, not the left branches of the trees."

Here is a general statement of the principle. Let P, Q and R be simple lotteries. Let I be a compound lottery that, in the first stage, yields lottery P with probability α and lottery R with probability $1-\alpha$. Let II be a compound lottery that yields lottery Q with probability α and lottery R with probability $1-\alpha$. These compound lotteries are shown in the top row of Figure 1.7, with the simple lotteries P, Q and R drawn as terminal nodes rather than subtrees (for brevity). It is plausible that, normatively, your choices between I and II should not depend on R or α. Instead, you choose I over II if and only if you would choose P over Q. This is called the *Independence Axiom*.

Recall that the reduced lotteries for lotteries I and II are denoted by $\alpha P + (1-\alpha)R$ and $\alpha Q + (1-\alpha)R$, respectively. Then the Independence Axiom can be stated formally as follows:

Assumption 2. (*Independence Axiom*) For all lotteries $P, Q, R \in \mathscr{L}$ and all α such that $0 < \alpha < 1$,

$$P \succeq Q \iff \alpha P + (1-\alpha)R \succeq \alpha Q + (1-\alpha)R. \tag{1.3}$$

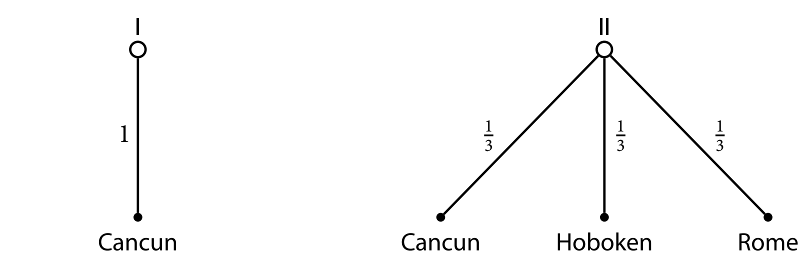
The Independence Axiom does not tell us how a DM should rank a particular pair of lotteries. Instead, it is a consistency condition on multiple rankings. By successive applications, we can find consistency conditions for a wider range of choices than is stated directly in the axiom. Let S be a simple lottery and let β be such that $0 < \beta < 1$. The following compound lotteries

$$\begin{aligned} \text{III} &= \beta P + (1-\beta)S \\ \text{IV} &= \beta Q + (1-\beta)S \end{aligned}$$

are shown in bottom row of Figure 1.7. Take a moment to convince yourself that, for example, the strict preferences I $\succ$ II and IV $\succ$ III violate the Independence Axiom.

Let's return to Akbar's travel plans to see this consistency condition in use. The next year, Akbar again receives free tickets and again makes his reservations a little bit late. But this time, Akbar comes to you proud and smiling; he says that he was able to make his choices without your help this year. Here is a summary of the choices he faced and the decisions he made: When he first went to the travel agent, he had to choose between the lotteries in Figure 1.8.

Figure 1.8



Akbar chose II over I. Once again, the travel agent had made a mistake and Akbar had to choose again, this time between the lotteries in Figure 1.9.

Figure 1.9

This time, Akbar chose III over IV.

"Akbar," you say, "Your choices are inconsistent!"

"What!?!" he replies. "You are just imposing your own preferences over vacation spots on my decisions; these may not be the choices you would have made, but they are not inconsistent."

Just looking at the simple lotteries as drawn above, it is hard to explain to Akbar why the choices are inconsistent. But if we can rewrite the four lotteries in the form shown in Figure 1.7, then we have shown that Akbar's choices violate the Independence Axiom.

To do this rewriting, we need to find lotteries P, Q, R and S and probabilities α and β such that the compound lotteries in Figure 1.7 reduce to the corresponding simple lotteries in Akbar's decision problems. That is, for each possible outcome x, we must have

$$\begin{aligned}
\mathrm{I}(x) &= \alpha P(x) + (1-\alpha)R(x) \\
\mathrm{II}(x) &= \alpha Q(x) + (1-\alpha)R(x) \\
\mathrm{III}(x) &= \beta P(x) + (1-\beta)S(x) \\
\mathrm{IV}(x) &= \beta Q(x) + (1-\beta)S(x).
\end{aligned}$$

The first thing to do is restrict the possible outcomes of each lottery. For example, P can place positive probability only on those outcomes that both I and III place positive probability on; R can place positive probability only on those outcomes that both I and II place positive probability on. We therefore know that possible outcomes for each of the simple lotteries are at most:

Lottery	Possible Outcomes
P	Cancun
Q	Hoboken, Rome
R	Cancun
S	Hoboken

This example ends up being quite simple. P, R and S have only one possible outcome. Q has two possible outcomes; hence we only have to find 1 probability number for Q. Together with the probabilities α and β, we have three unknowns (and many equations). We can solve the system of equations by inspection:

$$\left.\begin{aligned}
Q(\text{Cancun}) &= 0 \\
R(\text{Cancun}) &= 1 \\
\mathrm{II}(\text{Cancun}) &= 1/3
\end{aligned}\right\} \implies \alpha = 1 - 1/3 = 2/3$$

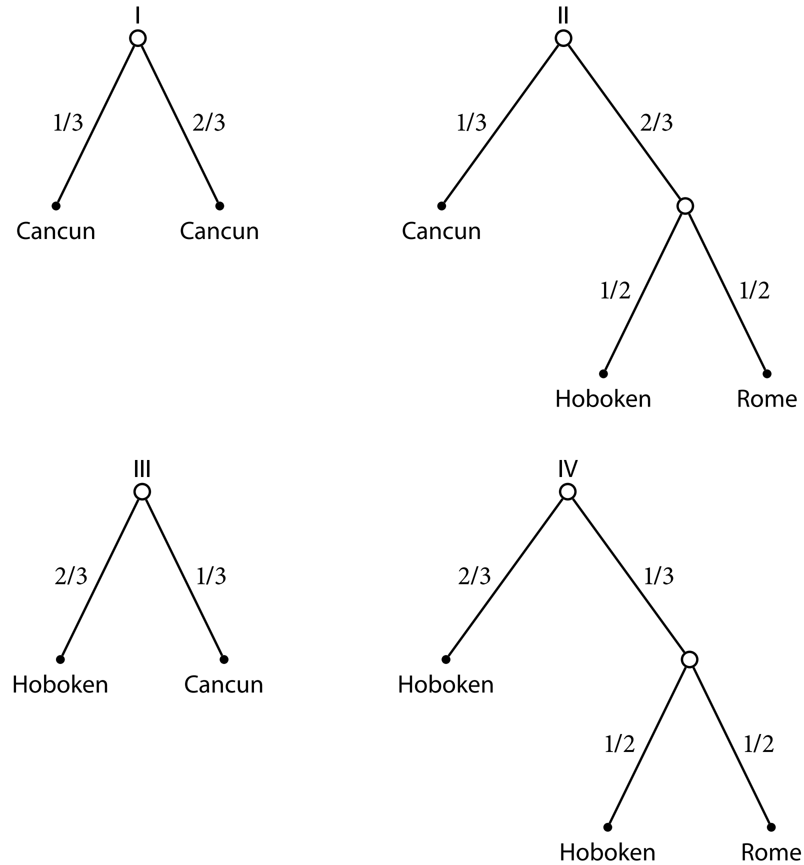
$$\left.\begin{aligned}
R(\text{Hoboken}) &= 0 \\
\alpha &= 2/3 \\
\mathrm{II}(\text{Hoboken}) &= 1/3
\end{aligned}\right\} \implies Q(\text{Hoboken}) = 1/2$$

$$\left.\begin{aligned}
P(\text{Cancun}) &= 1 \\
S(\text{Cancun}) &= 0 \\
\mathrm{III}(\text{Cancun}) &= 1/3
\end{aligned}\right\} \implies \beta = 1/3$$

This rewriting of Akbar's choices as compound lotteries is shown in Figure 1.10. You can check that each compound lottery in Figure 1.10 reduces to the corresponding simple lottery in the Akbar's travel problem.

Although the Independence Axiom may be a compelling normative principal, it is not always trivial to see when it applies to particular choices. The consistency of

Figure 1.10



Akbar's choices from Figures 1.8 and 1.9, rewritten as compound lotteries to match the general scheme shown in Figure 1.7. Akbar's choices of II over I and of III over IV violate the Independence Axiom.

the axiom is difficult for real humans to achieve. Consider the four lotteries shown on the left of Figure 1.11. The Independence Axiom implies that for the choices in Figure 1.11, I should be preferred to II if and only if III is preferred to IV. However, I $\succ$ II and IV $\succ$ III are commonly observed.[5]

Exercise 1.1. Consider the pairs of lotteries in Figures E1.1 and E1.2.

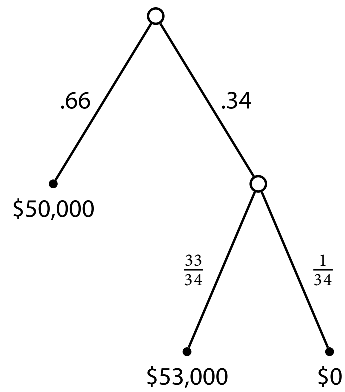
5. This example is called the Allais paradox because Maurice Allais, a French economist, presented an example like this one in debates about expected utility theory in the 1950's.

Figure 1.11

Lottery: **Simple form:** **Decomposed:**

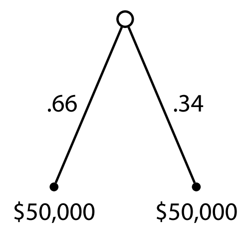
I

Prob.	Prize
.66	$50,000
.33	$53,000
.01	$0



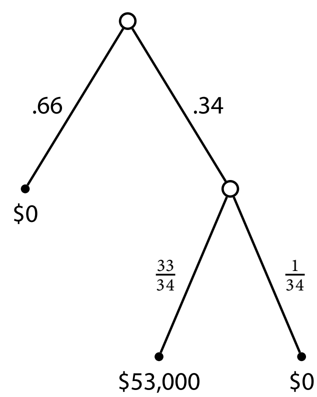
II

Prob.	Prize
1	$50,000



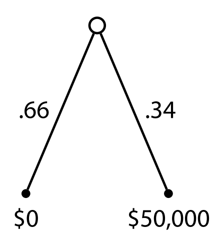
III

Prob.	Prize
.67	$0
.33	$53,000



IV

Prob.	Prize
.66	$0
.34	$50,000



The simple lotteries on the left are the reduced lotteries of the compound lotteries on the right. The preferences I $\succsim$ II and IV $\succsim$ III violate the Independence Axiom, but are common for subjects in decision experiments.

Figure E1.1

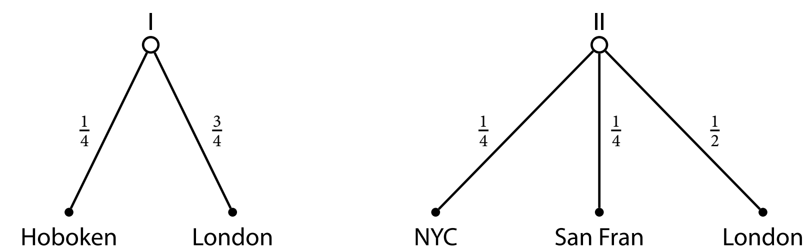
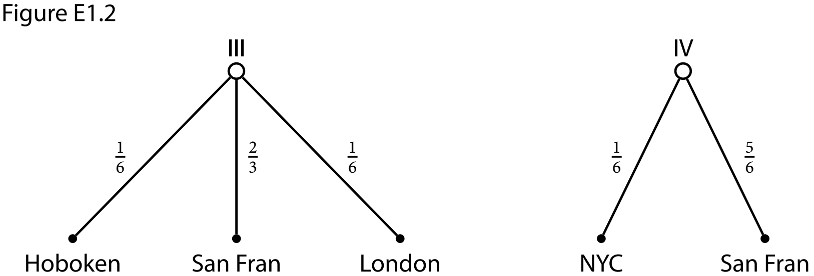


Figure E1.2



Show that—to be consistent with the Independence Axiom—if I is chosen over II then III should be chosen over IV.

Exercise 1.2. A decision maker has maximin preferences over lotteries if, for some ranking of outcomes, the decision maker chooses the lottery whose worst possible outcome is the best.

This is not a complete definition, because it does not say how the decision maker ranks two lotteries when indifferent between their worst possible outcomes. There are various ways to complete the definition, but a simple one that will suffice for the purpose of this exercise is to assume that the decision maker is then indifferent between the two lotteries. (The alternative is to describe more complicated rules for breaking this indifference, such as looking at the second-worst outcome or looking at the probability placed on the common worst outcome.)

a. Let $\succsim$ be a rational preference ordering on the set $\mathscr{L}$ of lotteries on a set X. Let each element x of X also denote the "lottery" that puts probability 1 on x, so that $\succsim$ also is an ordering on X. With this notation in mind, state formally what it means for $\succsim$ to be maximin preferences.

b. Show that maximin preferences violate the Independence Axiom. (You will need a minor auxiliary assumption.)

Exercise 1.3. Let $\succ$ be a strict preference relation over a set P of lotteries. Suppose that $\succ$ satisfies the following:

(Axiom 1) If $p \succ q$, then for all $a \in (0, 1)$ and $r \in P$ it follows that

$$ap + (1-a)r \succ aq + (1-a)r. \tag{E1.1}$$

Show that $\succ$ also satisfies the following:

(Axiom 2) If $p \succ q$ and $a, b \in (0, 1)$ are such that $a > b$, then

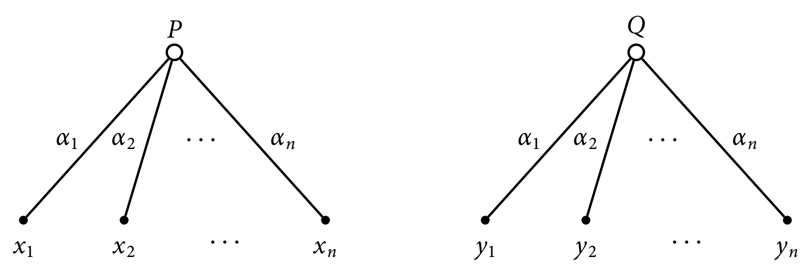
$$ap + (1-a)q \succ bp + (1-b)q. \tag{E1.2}$$

1.2.3 First-order stochastic dominance

Let $\succsim$ be the preference relation on the set $\mathcal{L}$ of lotteries. For two outcomes $x, y \in X$, we write $x \succsim y$ if the lottery that puts probability 1 on x is weakly preferred to the lottery that puts probability 1 on y.

Let P and Q be two lotteries with outcomes $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, respectively, such that for each $i \in \{1, \ldots, n\}$, $P(x_i) = Q(y_i)$. Let α_i be this common probability. We assume that $\alpha_i > 0$ for all i, but allow that $x_i = x_j$ or $y_i = y_j$ for some $i \neq j$. Lotteries P and Q are shown in tree form in Figure 1.12.

Figure 1.12



Definition 3. Suppose P and Q can be written as above, and $x_i \succsim y_i$ for all i. Then P is said to *weakly first-order stochastically dominate* Q. If also $x_i \succ y_i$ for some i, then P is said to *(strictly) first-order stochastically dominate* Q. (We abbreviate this P f.o.s.d. Q.)

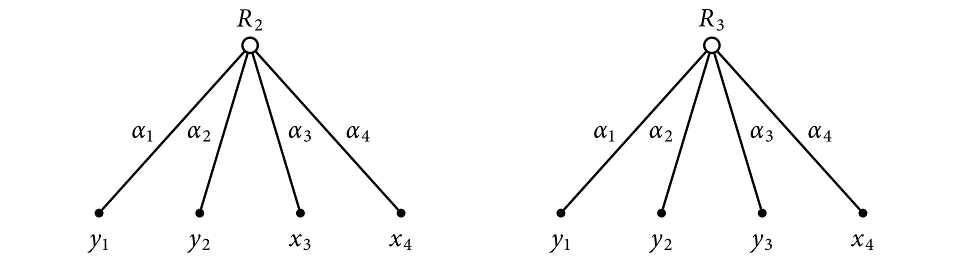
It is fairly intuitive that P should be preferred to Q if P f.o.s.d. Q. For example, suppose that the lotteries are based on the roll of a die with n faces, such that the probability of face i is α_i and the outcome when face i comes up is x_i or y_i. Then, for any roll of the die, the outcome of lottery P is preferred to the outcome of lottery Q.

Here is a formal statement and proof:

Proposition 3. *Suppose that $\succsim$ satisfies the Independence Axiom. If P weakly (resp., strictly) first-order stochastically dominates Q, then $P \succsim Q$ (resp., $P \succ Q$).*

Proof. Let R_i (for $i = 0, \ldots, n$) be the lottery with outcomes $y_1, \ldots, y_i, x_{i+1}, \ldots, x_n$ that occur with probabilities $\alpha_1, \ldots, \alpha_n$, respectively. For example, if $n = 4$, then R_2 and R_3 are lotteries in Figure 1.13.

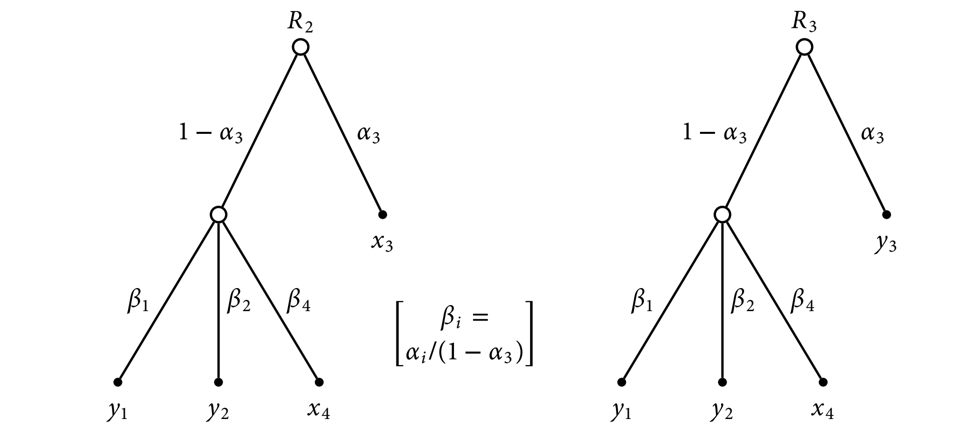
Figure 1.13



Observe that $R_0 = P$ and $R_n = Q$.

Let $i \in \{1, \ldots, n\}$. Lotteries R_{i-1} and R_i are the same except that R_{i-1} places probability α_i on x_i and R_i places probability α_i on y_i. For example, if $n = 4$, then R_2 and R_3 can be decomposed as in Figure 1.14.

Figure 1.14



The Independence Axiom implies that, if $x_3 \succsim y_3$, then $R_2 \succsim R_3$, and if $x_3 \succ y_3$, then $R_2 \succ R_3$. In general, if $x_i \succsim y_i$ (resp., $x_i \succ y_i$), then $R_{i-1} \succsim R_i$ (resp., $R_{i-1} \succ R_i$).

Therefore, we have shown that if $x_i \succsim y_i$ for all i, then

$$P = R_0 \succsim R_1 \succsim \cdots \succsim R_{n-1} \succsim R_n = Q. \tag{1.4}$$

From the transitivity of $\succsim$, it follows that $P \succsim Q$. Furthermore, if $x_i \succ y_i$ for some i, then one of the preferences equation (1.4) is strict, and it follows that $P \succ Q$. □

Not all lotteries can be ranked by f.o.s.d. However, when they can, we can determine preferences over the ranked lotteries using only information about the ranking of outcomes, and not the strength of the preferences over outcomes. For example, suppose that

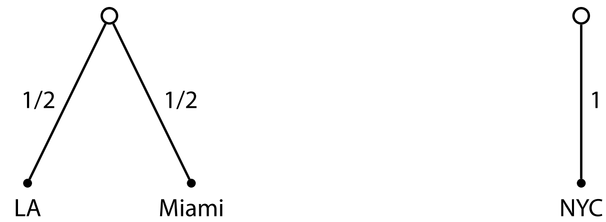
$$X = \{\text{LA, NYC, Miami}\}$$

and we know only that $\succsim$ satisfies IA and that

$$\text{LA} \succ \text{NYC} \succ \text{Miami}.$$

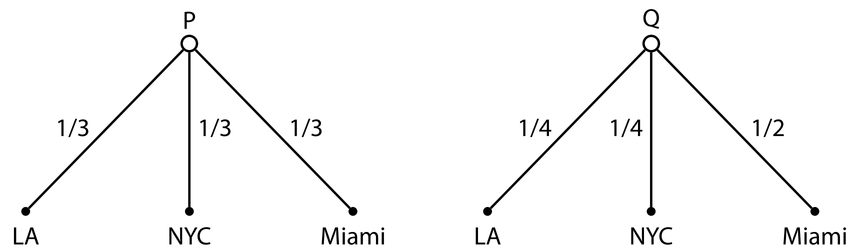
Then we cannot determine the DM's ranking of the two lotteries in Figure 1.15

Figure 1.15



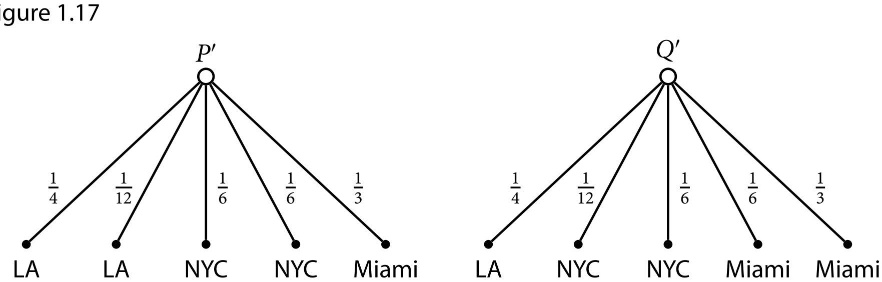
because it depends on how strongly the DM prefers LA to NYC and NYC to Miami. However, we can conclude that lottery P is preferred to lottery Q, in Figure 1.16, because P f.o.s.d. Q.

Figure 1.16



For this example, it probably was not obvious to you that P and Q could be made to match the definition of first-order stochastic dominance, even though you found it intuitive that P would be preferred to Q. Observe, however, that P and Q are equivalent to P' and Q' in Figure 1.17.

Figure 1.17



I will explain how I found P' and Q', but you will only understand the procedure if you try it out yourself. I started filling in branches of the tree from left to right. On the left-most branch, I put the best outcome for each lottery, which happened to be the same (LA) in this case. I gave the branch as much probability as I could without exceeding the probability of the outcome for either lottery. This used up all the probability of LA for lottery Q but not for lottery P. Therefore, on the next branch I put LA for lottery P' and the next best outcome (NYC) for lottery Q'. I again gave this branch as much probability as possible, without exceeding the probability of the respective outcome for each lottery. In this case, since the first branch put probability 1/4 on LA in P' and the total probability of LA in P is 1/3, I could only put probability 1/12 on the second branch. And so on. I could also have reversed the procedure, going from the worst to best outcomes.

This procedure suggests the following criterion for determining f.o.s.d., which can be checked mechanically.

PROPOSITION 4. *Let $X = \{x_1, \ldots, x_n\}$, and (without loss of generality) assume that $x_1 \succsim x_2 \succsim \cdots \succsim x_n$. P weakly f.o.s.d. Q if and only if, for every $k \in \{1, \ldots, n\}$:*

$$P(x_1) + \cdots + P(x_k) \;\geq\; Q(x_1) + \cdots + Q(x_k). \tag{1.5}$$

If equation (1.5) holds with strict inequality for k such that $x_k \succ x_{k+1}$, then P strictly f.o.s.d. Q.

Note that, since the probabilities sum to 1, equation (1.5) is equivalent to[6]

$$P(x_k) + \cdots + P(x_n) \;\leq\; Q(x_k) + \cdots + Q(x_n). \tag{1.6}$$

In other words, we can either compare the cumulative probabilities of the worst outcomes or of the best outcomes.

Let's apply this criterion to the following two lotteries:

P:

Roll of die:	1,6	5	2,3	4
One week in:	Hoboken	DC	San Fran	London

Q:

Roll of die:	Odd	2	4	6
One week in:	Hoboken	NYC	San Fran	London

Assume that

$$\text{London} \succ \text{San Fran} \succ \text{DC} \succ \text{NYC} \succ \text{Hoboken}.$$

Here are the values for equation (1.5):

x	$P\{x' \in X \mid x' \succsim x\}$	$Q\{x' \in X \mid x' \succsim x\}$
London	1/6	1/6
San Fran	1/2	1/3
DC	2/3	1/3
NYC	2/3	1/2
Hoboken	1	1

Since the $P(\cdot)$ values are as high as the $Q(\cdot)$ values in each case, with strict inequality in some places, P f.o.s.d. Q.

6. $P(x_k) + \cdots + P(x_n)$ can also be written $P\{x \in X \mid x_k \succsim x\}$, and $P(x_1) + \cdots + P(x_k)$ can also be written $P\{x \in X \mid x \succsim x_k\}$. Note that $P\{x \in X \mid x_k \succsim x\}$ is like a cumulative distribution function for the ordering on X induced by preferences. You are more likely to have seen the definition of first-order stochastic dominance for the special case in which X is an interval of real numbers representing monetary payoffs and the DM prefers more money over less. Suppose that lotteries P and Q have cumulative distribution functions F and G, respectively. Then the criterion in equation (1.6) becomes $F(x) \leq G(x)$ for all $x \in X$. If X instead represents costs, so that lower costs are preferred to higher costs, then the inequality is reversed: P weakly f.o.s.d. Q if $F(x) \geq G(x)$ for all $x \in X$.

Exercise 1.4. You are going to have pizza for dinner, and are trying to decide whether to have to pizza delivered, or whether to pick it up yourself. In the end, all that matters to you is how much the pizza costs and whether the pizza is hot or cold (e.g., the trip to the parlor is irrelevant). The pizza costs \$10. If it is delivered, you pay a \$2 delivery charge, unless the pizza is cold when it arrives, in which case the pizza and the delivery are free. The pizza parlor delivers cold pizza 1 out of 50 times. If you decide to pick the pizza up, there is no delivery charge. However, there is a 1 in 10 chance that you will be late and the pizza will be cold. There is also a 1 in 100 chance (independent of whether you are late) that you will be the 200th customer to go into the pizzeria today, in which case the pizza is free.

a. Write your decision as a choice between two lotteries.

b. If I only know that you like hot pizza more than cold pizza (other things equal) and cheap pizza more than expensive pizza (other things equal), can I determine your ranking of the possible outcomes in the two lotteries? (Explain.)

c. Is there any ranking of the outcomes consistent with the above (hot better than cold, etc) such that having the pizza delivered first-order stochastically dominates picking up the pizza? (Explain.)

d. Give a ranking of the outcomes consistent with the above such that picking the pizza up first-order stochastically dominates having the pizza delivered.

e. Give a ranking of the outcomes consistent with the above such that picking the pizza up does not first-order stochastically dominate having the pizza delivered.

Exercise 1.5. Let Z be a finite set of money values, and let p and q be lotteries on Z. Show that the following are equivalent:

1. For all $\bar{z} \in Z$: $p(z \leq \bar{z}) \leq q(z \leq \bar{z})$.
2. For all $\bar{z} \in Z$: $p(z \geq \bar{z}) \geq q(z \geq \bar{z})$.

Exercise 1.6. Suppose that you are considering insuring a piece of luggage. Given the risks and the insurance premium quoted, you decide that you are indifferent between getting and not getting the insurance. Then the airline offers you a "probabilistic" insurance policy. You pay the premium, as usual. If the luggage is lost, then with probability 1/2 you receive the value of the luggage, and with probability 1/2 your premium is instead returned to you.

Suppose that your preferences satisfy the Independence Axiom. How do you rank this probabilistic insurance compared to getting full insurance?

Notes and hints:

1. You should assume the premium is such that, if you know you have lost your luggage then you prefer to be insured (e.g., the premium is lower than the value of the luggage).

2. You should answer this question drawing trees and applying the IA directly, rather than using a utility representation.
3. As you should expect, you need to start by setting up the set of outcomes.

1.2.4 Continuity Axiom

Our next normative axiom is called the Continuity or Archimedean Axiom. In words:

> There is nothing so good (or so bad) that it does not become insignificant if it occurs with small enough probability.

Here is the formal statement:

Assumption 3. (*Continuity Axiom*) If $P, Q \in \mathscr{L}$ are such that $P \succ Q$, then for all $R \in \mathscr{L}$, there is α such that $0 < \alpha < 1$ and

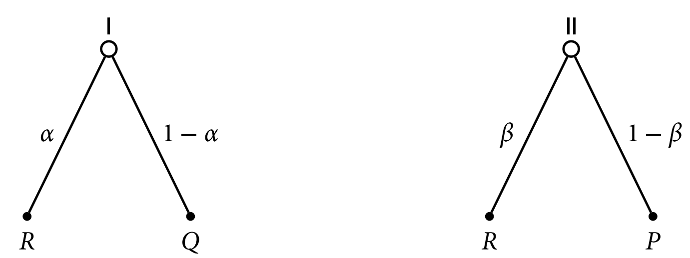
$$P \succ (1-\alpha)Q + \alpha R$$

and there is β such that $0 < \beta < 1$ and

$$(1-\beta)P + \beta R \succ Q.$$

Finally, here it is in tree form. If $P \succ Q$ and if lottery R is reached with small enough probability (small α and β) in the compound lotteries in Figure 1.18, then $P \succ$ I and II $\succ Q$.

Figure 1.18



To find a violation of the Continuity Axiom, we need to imagine a decision maker for whom something *is* that bad or that good. For example, suppose X contains "death" and some monetary outcomes:

$$X = \{\text{death}, \$100, \$1\text{B}\}.$$

Suppose also that the decision maker always chooses the lottery with the lowest probability of death, but if two lotteries have the same probability of death, she chooses the one with the highest expected monetary payoff.

To construct our formal counterexample to the Continuity Axiom, we need to find lotteries P, Q and R such that $P \succ Q$ and *for every* β such that $0 < \beta < 1$, we have

$$Q \succeq (1-\alpha)P + \alpha R. \tag{1.7}$$

R should be the "bad" thing, and so we let R be the lottery that yields death for sure. Let P and Q be the lotteries that yield \$1B and \$100, respectively, for sure. Then $P \succ Q$ since neither puts any probability on death, but equation (1.7) holds because the lotter $(1-\alpha)P + \alpha R$ puts positive probability on death whereas lottery Q does not.

You might be thinking that, with this counterexample, I am describing an often-observed violation of the Continuity Axiom. Not true—in fact, the Continuity Axiom, as technical as it may sounds, is quite sound empirically. Although people may talk about how death is worse than anything, or about how they would pay any amount of money to reduce the chance that children die in car accidents, in practice people always accept a small chance of a terrible outcome in return for a high chance of greater enjoyment or monetary return. For example, we risk our lives every day to go to work, and we risk our children's lives every day when we take them to school.

Exercise 1.7. Recall the maximin preferences defined in Exercise 1.2. Show that these preferences violate the Continuity Axiom. (You will need a minor auxiliary assumption.)

1.2.5 Expected Utility

Like WARP, the Independence and Continuity Axioms are consistency conditions that help us analyze decision problems. Also like WARP, these axioms have the added benefit of a representation of preferences that is very useful to modelers.

We already noted that with WARP, preferences $\succeq$ over lotteries have a utility representation $U\colon \mathscr{L} \to \mathbb{R}$, which is such that[7]

$$P \succeq Q \iff U(P) \geq U(Q).$$

We now use the special structure of lotteries to decompose this utility representation into probabilities and utility over outcomes.

Definition 4. Preference $\succeq$ over lotteries $\mathscr{L}$ satisfy *expected-utility maximization* if there is a function $u\colon Z \to \mathbb{R}$ such that, for any lotteries P and Q, $P \succeq Q$ if and only if

$$\sum_{z\in Z} P(z)u(z) \;\geq\; \sum_{z\in Z} Q(z)u(z).$$

If $Z = \{\text{death}, \$100\}$, one's notion of the "expected value of a lottery" does not make sense. What is the average of death and \$100? However, given a utility function $u\colon Z \to \mathbb{R}$, which is a random variable, each outcome gets a utility value and we can find the expected value of the utility for each lottery. For a lottery P, this expected value is $\sum_{z\in Z} P(z)u(z)$, and is called the expected utility of P. A DM is an expected-utility maximizer if, for some utility function $u\colon Z \to \mathbb{R}$, she always prefers the lottery with the highest expected utility.

And now the anticipated punch-line:

7. WARP is sufficient for such a utility representation on any finite subset of lotteries.

Theorem 1. (von Neumann–Morgenstern) *If $\succeq$ satisfies the Independence and Continuity Axioms, then it satisfies expected-utility maximization.*

This theorem is due to John von Neumann (who needs no introduction) and Oskar Morgenstern (a Princeton economist). Von Neumann and Morgenstern did not invent expected utility maximization, but they were the first to derive it as a consequence of consistency conditions on preferences over lotteries (while developing their theory of games). Expected utility in this framework is often called "von Neumann–Morgenstern (VNM) expected utility", and the utility function $u: Z \to \mathbb{R}$ is often called a "von Neumann–Morgenstern utility function".

The utility function in decision making without uncertainty only *ranks* alternatives, whereas the VNM utility function also measures the *strength* of preferences over outcomes. For example, suppose that the set of alternatives without uncertainty is

$$X = \{\text{LA, NYC, Miami}\},$$

and preferences over X are given by

$$\text{LA} \succ \text{NYC} \succ \text{Miami} .$$

Then here are two equally valid and equivalent utility representations of these preferences:

x	$U(x)$	$V(x)$
LA	1,000	1,000
NYC	2	999
Miami	1	1

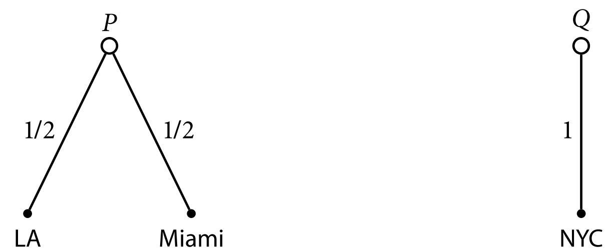
The utility function U suggests that the preference of LA over NYC is much stronger than the preference of NYC over Miami, whereas the utility function V suggest the opposite. However, both of these statements are meaningless, because only the ranking, and not the actual utility values, are important. In fact, any strictly increasing (monotonic) transformation of U or V is also a utility representation of $\succeq$. For these reasons, U and V are sometimes called *ordinal* utility functions.

Now suppose that the alternatives are lotteries over a set

$$X = \{\text{LA, NYC, Miami}\}$$

of outcomes. The functions U and V, as VNM utility functions, represent different preferences over lotteries. For example, if an expected-utility maximizer's VNM utility function is U, then she prefers P over Q in Figure 1.19, but if it is V, then she prefers Q over P.

Figure 1.19



Because a VNM utility function measures strength of preference over outcomes, it is sometimes called a *cardinal* utility function.

It is still true that VNM preferences over lotteries can be represented by many different VNM utility functions, but there is not as much freedom as with ordinal utility functions. Instead, only positive affine transformations preserve the ranking of lotteries.[8] If $u\colon X \to \mathbb{R}$ is a VNM utility function, then $v\colon X \to \mathbb{R}$ is a positive affine transformation of u if there are numbers $a > 0$ and b such that $v(x) = au(x) + b$ for all $x \in X$. As an exercise, check that if v is a positive affine transformation of u, then u and v induce the same preferences over lotteries.

Exercise 1.8. Let $\succeq$ be VNM preferences over lotteries $\mathscr{L}$, represented by a VNM utility function $u : Z \to \mathbb{R}$. Suppose $v : Z \to \mathbb{R}$ is a positive affine transformation of u. Show that v also represents the preferences $\succeq$.

Exercise 1.9. A person you know (with an odd view about fun) points a revolver at your head. It has six chambers and n bullets. He is going to spin the chambers and pull the trigger for sure, but first he makes you an offer. If you give him a certain amount of money, he will first remove one of the bullets.

Most people say that in such a situation, they would pay more if initially there were a single bullet than if there were four bullets. That is, there is some number x of dollars such that they would agree to pay x dollars to remove the bullet if $n = 1$, but they would refuse to pay x to remove a bullet if $n = 4$.

The purpose of this problem is to show that such choices are inconsistent with expected-utility maximization, assuming that (i) if you survive, you prefer more money over less money (ii) if you die, you don't care how much money you have. Don't confuse things by reading too much into the problem.

a. Within the VNM framework, what exactly are the two choice problems (involving a total of four alternatives)? (Be explicit, which doesn't mean verbose.)

b. Show directly that the choices violate the Independence Axiom.

c. Now show that the choices are inconsistent with expected utility maximization by stating what the decisions mean for the utility function, and deriving a contradiction.

d. What is the intuition? Use the extreme case, where $n = 6$, as a way to illustrate the

8. This is often stated: "u is unique up to a positive affine transformation."

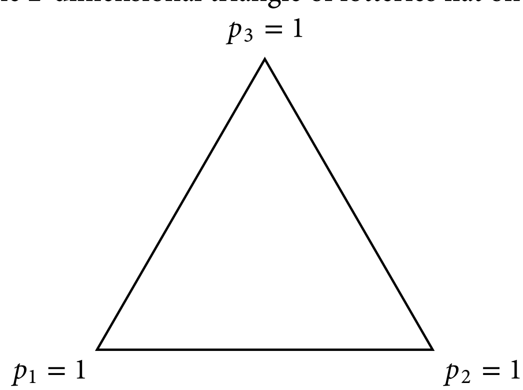
intuition.

Exercise 1.10. It is said of preferences over lotteries that satisfy expected utility maximization that they are "linear in probabilities". That is, if $U: \mathscr{L} \to \mathbb{R}$ is a utility function over lotteries that has an expected utility representation, then U is a linear function of the probabilities of the outcomes. For concreteness, assume that the set of outcomes Z has three outcomes, z_1, z_2 and z_3. Each lottery can be specified by three numbers: the probabilities p_1, p_2 and p_3 of the three outcomes.

a. Write down the mathematical definition of the set of lotteries, as a subset of $\mathbb{R}^3$. Given an example of an expected utility representation, and use it to explain that the utility function over lotteries is linear in probabilities.

b. Draw the set of lotteries in $\mathbb{R}^3$ the best that your 3D-drawing skills allow. This set should be a 2-dimensional triangle, even though it is sitting in $\mathbb{R}^3$. It is called a *simplex*.

We can redraw the 2-dimensional triangle of lotteries flat on the page:



For example, at the $p_2 = 1$ vertex, the probability of z_2 is 1 and the probability of the other outcomes is zero. Along the side opposite this vertex, the probability of z_2 is zero. For all points in a given line parallel to this size, the probability of z_2 is the same.

Specify two lotteries P and Q, and plot them on the simplex. Indicate the position of the lottery $(1/3)P + (2/3)Q$, as defined in class.

For the utility representation you gave in Problem 1.10, draw two indifference curves on the simplex.

1.3 States of nature and subjective expected utility

1.3.1 States, actions and outcomes

Reducing a decision problem under uncertainty to selecting among lotteries over outcomes is a useful simplification if we are not interested in the source of uncertainty and how a decision maker generates different lotteries. However, it is not a rich enough model in many other cases.

A richer model, which we call the "states-of-nature" model, has the following components:

$$\begin{aligned} X &= \text{set of outcomes (what the DM cares about)} \\ S &= \text{set of states (uncertain factors beyond the control of the DM)} \\ A &= \text{set of actions (what the DM controls).} \end{aligned}$$

The outcome is determined jointly by the action and the state. (States are also called states of nature or states of the world.) We can summarize this relation by a function $F\colon A \times S \to X$, where $F(a, s)$ is the outcome when the action is a and the state is s.

The specification of X, S and A in this richer model requires great care. This is the subject of Sections 1.3.2 and 1.3.3.

1.3.2 From large worlds to small worlds

What do we mean by the state? A grand interpretation is that a state is a complete specification of the past, present and future configuration of the world, except for those details that are part of the DM's actions. There is uncertainty because we do not know which is the true configuration, and consider there to be many possible configurations. S is the set of all such possible configurations.

A problem with this notion of state is that it is *too* grand. It contains more detail than is manageable and necessary in applications. There are too many possible configurations, and each one requires an infinite amount of information to specify. Therefore, we prefer a coarser description of uncertainty that does not distinguish between all possible states.

Let S^* be the "grand" set of states. (Hopefully we can conceptualize S^* even if we cannot even mentally specify a particular grand state.) Although we cannot write down a particular grand state, it is easy to specify some sets of grand states, such as the set of all states in which it rains. A set E of states—that is, a subset E of S^*—is called an event.

A set $\{E_1, \ldots, E_n\}$ of disjoint events whose union is S^* is called a partition of S^*. That is, $\{E_1, \ldots, E_n\}$ is a partition of S^* if

1. $E_i \subset S^*$ for all i,
2. $E_i \cap E_j = \varnothing$ for all $i \neq j$, and
3. $E_1 \cup E_2 \cup \cdots \cup E_n = S^*$.

Stated yet another way, conditions (2) and (3) mean that each state belongs to one and only one event in the partition.

We replace S^* by a partition $\{E_1, \ldots, E_n\}$ of S^* that is *fine* enough that, for each event in the partition, it is not necessary to distinguish between the different states in that event (for the decision problem being studied). That is, for every action, any two states in the same event should result in the same outcome. The partition should otherwise be as *coarse* as possible for the sake of simplicity. Events in this partition are called *elementary events*.

Let's consider the selection of a suitable partition of S^* into elementary events for modeling a particular decision problem. I have to decide whether to take a nap in my office and, if so, for what time to set my alarm clock. Suppose that the alarm clock can only be set for multiples of 5 minutes, and it is now 3:15. I want to sleep for at most 20 minutes, but I am trying to decide whether to set the clock for an earlier time because my boss might walk into the office. I care about how long I sleep, except if my boss walks in, in which case I feel equally lousy however long I have slept.

Let's start by writing down the sets of actions and of outcomes:

$$A = \left\{ \begin{array}{lcl} a_1 & = & \text{do not sleep} \\ a_2 & = & \text{set alarm for 5 minutes} \\ a_3 & = & \text{set alarm for 10 minutes} \\ a_4 & = & \text{set alarm for 15 minutes} \\ a_5 & = & \text{set alarm for 20 minutes} \end{array} \right\}$$

$$X = \left\{ \begin{array}{lcl} x_1 & = & \text{do not sleep} \\ x_2 & = & \text{sleep 5 minutes} \\ x_3 & = & \text{sleep 10 minutes} \\ x_4 & = & \text{sleep 15 minutes} \\ x_5 & = & \text{sleep 20 minutes} \\ x_6 & = & \text{awakened by boss} \end{array} \right\}$$

I will now propose a collection $\mathscr{E}$ of elementary events that is flawed, and then we will see how to fix it. $\mathscr{E}$ consists of these five events:

$$\begin{aligned} E_1 &= \left\{ s \in S^* \mid \text{boss enters in next 5 minutes} \right\} \\ E_2 &= \left\{ s \in S^* \mid \text{boss enters in next 10 minutes} \right\} \\ E_3 &= \left\{ s \in S^* \mid \text{boss enters 10 to 15 minutes from now} \right\} \\ E_4 &= \{ s \in S^* \mid \text{boss does not enter in next 15 minutes} \\ &\qquad \text{and is eating a bagel now} \} \\ E_5 &= \{ s \in S^* \mid \text{boss does not enter in next 20 minutes} \\ &\qquad \text{and is not eating a bagel now} \} \end{aligned}$$

Before reading further, carefully examine these five events and look for reasons why they are not a good set of elementary events for this problem. You should be able to find at least four reasons.

Two of the reasons have to do with the fact that the events do not form a partition. The other two have to do with the coarseness/fineness of the events. Let's start with the partition problems, since if the events do not form a partition, the model will be mathematical nonsense.

Problem 1. $\mathscr{E}$ is not a partition because events E_1 and E_2 intersect. They both contain the states in which the boss enters in the next 5 minutes.

Solution. We can modify E_2 to eliminate the overlap:

$$E_2' = \left\{ s \in S^* \mid \text{boss enters 5 to 10 minutes from now} \right\}.$$

Problem 2. $\mathscr{E}$ is not a partition because the union of the events does not include all the states. States in which my boss enters 15 to 20 minutes from now and is not eating a bagel do not lie in any of the events.

Solution. We can expand E_5 to include the missing states:

$$E_5' = \{s \in S^* \mid \text{boss does not enter in next 15 minutes and is not eating a bagel now}\}.$$

Problem 3. The partition is too fine. For the decision problem, there is no need to distinguish between states in which my boss is eating a bagel and states in which he is not. (This is the only one of the four problems that is not critical. We fix this problem for the sake of simplicity only.)

Solution. We can combine events E_4 and E_5', replacing them by

$$E_4' = \{s \in S^* \mid \text{boss does not enter in next 15 minutes}\}.$$

Problem 4. The partition is too coarse. For the decision problem, states in which my boss enters 15 minutes from now and states in which she enters 22 minutes from now all lie in E_4', but lead to different outcomes if I set my alarm for 20 minutes.

Solution. We can divide E_4' into two events:

$$\begin{aligned} E_4'' &= \{s \in S^* \mid \text{boss enters 15 to 20 minutes from now}\} \\ E_5'' &= \{s \in S^* \mid \text{boss does not enter in next 20 minutes}\}. \end{aligned}$$

In summary, here is a proper set of elementary events for this decision problem:

$$\begin{aligned} E_1 &= \{s \in S^* \mid \text{boss enters in next 5 minutes}\} \\ E_2' &= \{s \in S^* \mid \text{boss enters 5 to 10 minutes from now}\} \\ E_3 &= \{s \in S^* \mid \text{boss enters 10 to 15 minutes from now}\} \\ E_4'' &= \{s \in S^* \mid \text{boss enters 15 to 20 minutes from now}\} \\ E_5'' &= \{s \in S^* \mid \text{boss does not enter in next 20 minutes}\}. \end{aligned}$$

Since we will not be "peeking inside" elementary events, we can simplify terminology and notation by thinking of each event as a "point" rather than a set, calling each elementary event a state, and letting S be the set of these "small-worlds" states.

We will be using probability measures on S. We should make sure that this construction of S does not screw up the use of probability. To simplify the mathematics of this discussion assume (here only) that S^* is finite. Let $\mathcal{E} = \{E_1, \dots, E_n\}$ be a partition of S^* into elementary events, and let $S = \{s_1, \dots, s_n\}$ be the corresponding set of states. A subset $E \subset S$ is also called an event, and it corresponds to an event in S^*. For example, the event $\{s_1, s_3, s_4\} \subset S$ corresponds to the event $E_1 \cup E_3 \cup E_4$.

Let the set S^* of grand states have a probability measure $\pi^* \colon S^* \to \mathbb{R}$ that specifies the probability of each state. The probability of an event $E \subset S^*$ is $\pi^*(E) = \sum_{s \in E} \pi^*(s)$. Recall that π^* has the following properties:

1. $0 \le \pi^*(s) \le 1$ for all $s \in S^*$.
2. $\sum_{s \in S^*} \pi^*(s) = 1$.

I will now show that there is a probability measure $\pi\colon S \to \mathbb{R}$ on S that is consistent with π^*. Define $\pi(s_i) = \pi^*(E_i)$. For $E \subset S$, define $\pi(E) = \sum_{s\in E} \pi(s)$. You can verify that $\pi(E)$ is equal to the probability of the corresponding event in S^*. Furthermore,

$$\sum_{s\in S} \pi(s) \;=\; \pi^*(E_1) + \ldots + \pi^*(E_n) \;=\; 1,$$

since $\{E_1, \ldots, E_n\}$ is a partition. Hence, π is also a probability measure.

From now on, the story of the grand state need only lurk in the background. When writing down a model, or thinking about a decision problem, we simply say "The set S of states is …."

1.3.3 Outcomes and state-independent preferences

For defining the set X of outcomes, we have two options. One is to adopt the definition of outcome from our model of lotteries: an outcome is *everything* the decision maker cares about and that is potentially affected by his action in the decision problem being studied. Sometimes this involves repeating information about the state that affects the DM's preferences. In this case, it is sometimes simpler to omit this information from the outcomes.

For example, suppose the decision problem is to choose a contract for provision of heart surgery, which might be contingent on whether you have a heart attack. The set of states is

$$S \;=\; \{\text{heart attack, no heart attack}\}.$$

Consider the following two sets of outcomes:

$$X \;=\; \left\{ \begin{array}{lcl} x_1 & = & \text{heart attack, heart surgery} \\ x_2 & = & \text{heart attack, no heart surgery} \\ x_3 & = & \text{no heart attack, heart surgery} \\ x_4 & = & \text{no heart attack, no heart surgery} \end{array} \right\}$$

$$X^* \;=\; \left\{ \begin{array}{lcl} x_1^* & = & \text{heart surgery} \\ x_2^* & = & \text{no heart surgery} \end{array} \right\}$$

Outcomes in X include everything you care about. Outcomes in X^* omit whether you have a heart attack, but this information is not lost because it is part of the states.

Preferences over outcomes in X are *state independent*. Although having a heart attack may affect the likelihood of outcome x_1, how you feel about having a heart and heart surgery does not depend on whether you have a heart attack. If you find the last sentence nonsensical, it will help to imagine the following hypothetical scenario. Suppose that there is a special device that can "undo" a heart attack (including all the pain and bad memories) if you do have one, and can induce a heart attack if you do not. Then any outcome in X is feasible for either state in S. You can now ask yourself the following question: Does the way feel about having a heart attack and heart surgery depend on whether or not you naturally had a heart attack? I think it is a plausible approximation that it does not. In this hypothetical scenario, the uncertainty about whether you have a heart attack is simply a randomization device that generates lotteries.

In contrast, preferences over outcomes as defined in X^* are *state dependent*. That is, how you feel about having heart surgery depend on whether you have had a heart attack. (This requires no explanation.)

Which is the best way to set up a model? Sometimes, preferences are naturally state independent. For example, if you are speculating on the Zimbabwean dollar, and you do not live in Zimbabwe or have other connections with the Zimbabwean economy, then presumably you only care about the Zimbabwean exchange rate in so far as it affects the return on your transactions. We can let the set of outcomes be your net profit, and preferences are state independent.

On the other hand, in the heart-attack problem, preferences are naturally state dependent. We can still define outcomes rich enough so that preferences are state independent. This requires adding and repeating information about the state in the description of the outcomes, but allows us to distinguish between the state of nature as purely a source of uncertainty and the state of nature as something that directly affects our lives. We will use both state-dependent and state-independent preferences, depending on which leads to a simpler and more intuitive model. A common case in which we will have state-dependent preferences is in insurance models, or other models in which it is convenient to think of outcomes as monetary returns or commodity trades, over which preferences are state dependent.

1.3.4 Objective expected utility

The decision maker's problem is to choose an action from A. Every action $a \in A$ leads to a mapping $\tilde{x}_a : S \to X$ from states to outcomes, which is defined by

$$\tilde{x}_a(s) = F(a, s).$$

Such a mapping is called an *act*. Our first assumption is that the decision maker cares about the act induced by an action, rather than the action itself. Rather than a behavior assumption, this is requirement on how we set up the model. We are supposed to have defined the outcomes so that they include any aspects of the action that the decision maker cares about.

Suppose that there is an objectively ascertained probability measure $\pi : S \to \mathbb{R}$ on the set of states. That is, π can be determined by some frequentist approach to statistics and all parties agree that π measures the probability of each state. For example, this is true if the state is the number that comes up on a roll of a die, or is this week's winning lottery number.

Each act $\tilde{x} : S \to X$ then induces a probability distribution $P : X \to \mathbb{R}$ on the set of outcomes, defined by

$$P(x) = \sum_{s \in \tilde{x}^{-1}(x)} \pi(s).$$

That is, $\tilde{x}^{-1}(x)$ is the set of states that leads to outcomes x, and $P(x)$ is the probability that the true state lies in this set. In statistics, $\tilde{x}$ is called a *random object* and P is called the *distribution* of $\tilde{x}$.

Suppose that preferences over money are state independent. Then a distribution P of an act $\tilde{x}$ is a lottery over outcomes and you rank acts the way you would rank their distributions as lotteries. In particular, you are indifferent between any two acts with the same distributions. Preferences over acts thereby reduce to preferences over lotteries, and we can turn to VNM expected utility. Let $u\colon X \to \mathbb{R}$ be the decision maker's VNM utility function. The expected utility of an act $\tilde{x}\colon S \to X$ is

$$\sum_{s \in S} \pi(s) u(\tilde{x}(s)).$$

In this expression, $\tilde{x}(s)$ is the outcome in state s, and $u(\tilde{x}(s))$ is the utility of this outcome.

Suppose instead that preferences are state dependent. We can still use VNM expected utility theory, but we have to construct a set $\hat{X}$ of outcomes that fit the state independence assumption of VNM expected utility. The state and outcome together describe everything the DM cares about, and so we let $\hat{X} = X \times S$. A typical "outcome" is $\langle s, x\rangle$. Each act $\tilde{x}\colon S \to X$ induces a probability measure $P\colon \hat{X} \to \mathbb{R}$ on $\hat{X}$, defined by

$$P(x,s) = \begin{cases} \pi(s) & \text{if } \tilde{x}(s) = x \\ 0 & \text{otherwise.} \end{cases} \tag{1.8}$$

We again assume that the DM's preferences over lotteries over $\hat{X}$ satisfy the VNM expected utility maximization, and let $u\colon \hat{X} \to \mathbb{R}$ be the VNM utility function. The expected utility of an act $\tilde{x}\colon S \to X$ is

$$\sum_{s \in S} \pi(s) u(\tilde{x}(s), s).$$

In this expression, $\tilde{x}(s)$ is the outcome in state s, and $u(\tilde{x}(s), s)$ is the utility of this outcome in state s.

This derivation of expected utility for state-dependent preferences makes some people uncomfortable. Many of the lotteries over $\hat{X}$ may be impossible. For example, it is difficult to generate casino-style gambles in which the prizes include earthquakes in California. Hence, the choices from a subset of hypothetical lotteries are themselves hypothetical. We have to assume that the DM can reasonably consider such hypothetical choices.

1.3.5 Subjective expected utility

Suppose that there is no objective probability measure on the set of states. For example, what is the probability that Candidate Zorg will win the next election? There is much room for disagreement, and we will not be able to convince each other about the correct answer to this question. How do we represent preferences over acts in this case?

The short answer is that we assume that the DM has subjective beliefs, which are given by a probability measure $\pi\colon S \to \mathbb{R}$, and then behaves like an expected utility maximizer as described in the previous sections. Such behavior is called subjective expected utility (SEU) maximization.

If this book were just about decision theory, I would instead derive subjective expected utility maximization from consistency conditions on preferences over acts. This derivation was first accomplished by Leonard Savage, who used decision theory to provide a foundation for statistics.[9] However, we must press on if we are to reach multi-agent economic models.

9. Rather than the other way around!

Subjective expected utility is even more contentious then VNM expected utility. First, any violation of VNM expected utility maximization is a violation of SEU, since objective expected utility is a special case of SEU. In addition, the formulation of subjective beliefs clearly complicates the decision problem and makes our normative model further removed from actual behavior. There are even very simple situations in which people do not act as if they had probabilistic beliefs over the states.

Exercise 1.11. Maximin preferences are appealing in the states-of-the-world framework because beliefs about the relative likelihood of events is not important. A decision model that does involve the likelihood of events but does not involve "arbitrary" subjective beliefs is the "principle of insufficient reason". According to this model, in situations in which the probability of events cannot be objectively determined, the decision maker assigns each state equal probability, and then acts as an expected utility maximizer. For example, if there are three states, "sunny", "cloudy" and "rainy", then each state has probability 1/3. Show that the principle of insufficient reason can lead to different decisions for the exact same decision problem depending on how the modeler or decision maker chooses to specify the set of states. This requires a simple example and a clear explanation.

1.3.6 Statewise dominance

In the state-of-nature model, there is a criterion that is related to but stronger than first-order stochastic dominance. We call it statewise dominance, or simply dominance.

An act $\tilde{x}$ dominates another act $\tilde{y}$ if it results in a better outcome in every state. We again write $x \succeq y$ for $x, y \in X$ if the sure outcome x is preferred to the sure outcome y. Then $\tilde{x}$ weakly dominates $\tilde{y}$ if $\tilde{x}(s) \succeq \tilde{y}(s)$ for every $s \in S$. If also $\tilde{x}(s) \succ \tilde{y}(s)$ for some $s \in S$, then $\tilde{x}$ (strictly) dominates $\tilde{y}$.

If $\tilde{x}$ dominates $\tilde{y}$, then the lottery induced by $\tilde{x}$ first-order stochastically dominates $\tilde{y}$ no matter what is the probability measure on S (as long as all states get positive probability—otherwise we may only get weak f.o.s.d.). Hence, dominance is an especially useful criterion when uncertainty is subjective, because the ranking of $\tilde{x}$ and $\tilde{y}$ does not depend on beliefs. However, because it is stronger, it is also less frequently applicable than f.o.s.d. For a particular probability measure on S, an act $\tilde{x}$ may f.o.s.d. $\tilde{y}$ even though it does not statewise dominate $\tilde{y}$.

Exercise 1.12. Here are some decision theories for the Savage setup (states of the world without objective uncertainty) that differ from Subjective Expected Utility theory:

Maximin For each action, there is a worst-case (worst over all possible states). Choose the action whose worst-case is the best.

Maximax For each action, there is a best-case (over all possible states). Choose the action whose best-case is the best.

Minimax regret For each action you choose, and each state that occurs, there might be some other action you wish you had chosen. The difference between the utility you would have gotten in the state if you had chosen the best action, and the utility you actually got given the action you chose, is called the *regret* for that state and action. Now, for each action, there is a worse-case in terms of regret, i.e., a maximum regret over all possible states. Choose the action whose maximum regret is the lowest.

Insufficient reason If you don't know the objective probabilities of the states, then you should simply place the same probability on each state. Then choose the action that maximizes expected utility given this probability distribution.

a. Consider the following payoff matrix, where the numbers are "utility" payoffs:

$$
\begin{array}{cc|ccc|}
 & & & \text{States} & \\
 & & s_1 & s_2 & s_3 \\
\cline{3-5}
 & a_1 & 0 & 10 & 2 \\
\text{Acts} & a_2 & 3 & 4 & 0 \\
 & a_3 & 2 & 0 & 9 \\
\cline{3-5}
\end{array}
$$

What decision (choice of an action) does each decision rule listed above lead to? Explain in each case.

b. Replace the entries in the payoff matrix by arbitrary prizes. Suppose you only know the person's ordinal preferences over (relative ranking of) prizes. For which of the decision rules given above is this enough information to deduce the person's choice? Explain.

c. Explain why one might say that maximin is a pessimistic decision rule, and that maximax is an optimistic decision rule.

d. Explain why the insufficient reason decision rule is sensitive to the specification of the states (e.g., to whether you consider "rain" to be a singe state of the world, or distinguish between "rain in the morning only" and "rain all day").

e. Pick one of the decision rules, and compare it to SEU, including your own subjective view on which is better.

Chapter 2

Choosing when there is new information

2.1 Representing information

2.1.1 Information as an event

In the Introduction, I stated that uncertainty and information are inexorably linked. We have modeled decisions under uncertainty in which the decision maker's information is constant in the course of the decision problem. Now we consider decision problems during which new information is revealed. This is one case where we need the states-of-nature model. That model allows a surprisingly general—though by no means universal—representation of information.

Consider a piece of information, such as "it rained yesterday in Milwaukee". How does it fit into our states-of-nature model? A first answer is that it causes our beliefs to change. This is true, but we want a representation of this information that provides a simple rule for changing beliefs.

Let us return, for a moment, to the "grand" set of states S^*. Imagine that we list every possible statement about the world. In a particular state, each of these statements is either true or false. (In fact, we can think of a state S^* as just a list of the truth values of each of these statements.) For example, in each state the statement "It rained yesterday in Milwaukee" is either true or false. Let E be the set of states in which it is true. By learning the statement, what we have learned—and all we have learned—is that the true state is in E and not in the complement of E. Therefore, we can represent the contents of a piece of information by an event $E \subset S^*$.[1]

2.1.2 Setting up the small-worlds model correctly

Even if you believe every sentence that I wrote in the previous subsection, you may not find such a representation of information intuitive and general. Does not information usually make you change your beliefs about states in a smooth way, rather than simply ruling out the possibility of certain states?

Our theory can represent such partial information as long as we set up the right small-worlds state space for the situation we have in mind. We must choose elementary events that are fine enough that our information does not allow us to distinguish between two large-worlds states that are in the same elementary event. I illustrate this with an example.

1. Remember that the definition of an event is a subset of the set of states. Only the context in which the term is used can impute further meaning to the term.

Consider the following decision problem, told from the point of view of a woman. You are thinking about inviting a man named Homer to the movies. Homer will accept the invitation if and only if he likes you, but you are not sure about this. Getting turned down is a worse outcome for you than not inviting Homer at all. Here is how, according to Section 1.3, we would define the actions, states and outcomes in a small-worlds model of the decision problem:

$$\text{(actions)}\quad A = \left\{ \begin{array}{lcl} a_1 & = & \text{invite Homer} \\ a_2 & = & \text{not invite Homer} \end{array} \right\},$$

$$\text{(states)}\quad S = \left\{ \begin{array}{lcl} s_1 & = & \text{he likes me} \\ s_2 & = & \text{he likes me not} \end{array} \right\},$$

$$\text{(outcomes)}\quad X = \left\{ \begin{array}{lcl} x_1 & = & \text{invitation declined} \\ x_2 & = & \text{invitation accepted.} \\ x_3 & = & \text{no invitation} \end{array} \right\}$$

Now suppose that you passed Homer in the street this morning and he smiled at you. This might tell you something about whether Homer likes you, even though Homer sometimes smiles at people he does not like and sometimes does not smile at people he likes. How can we represent this information?

If it can be represented as an event in our model, we just have to check which event corresponds to this information. Because there are only two states, there are only four events (only four subsets of S):

$$\varnothing\,, \quad \{s_1\}\,, \quad \{s_2\}\,, \quad \{s_1, s_2\}.$$

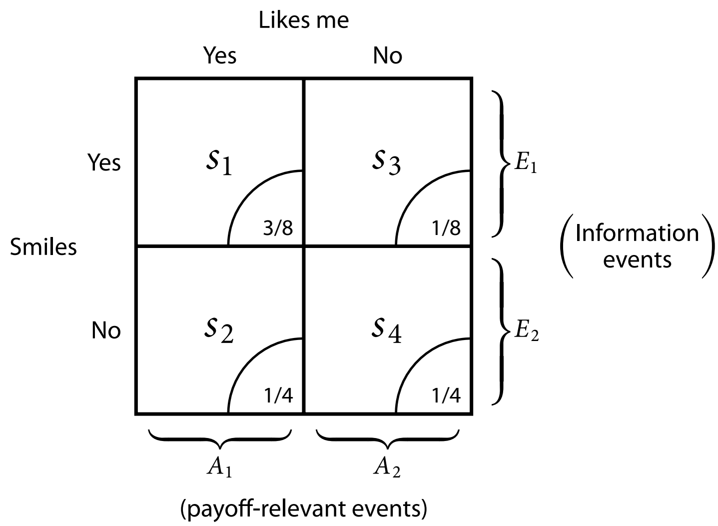
Remember that the interpretation of an event E as information is that "all I have learned is that the true state is in E". The null event $\varnothing$ does not make sense as information—there is something wrong with the model if you learn that the true state is in $\varnothing$, and hence does not exist! The event $\{s_1\}$ does not correspond to your information, because you still consider s_2 possible. We rule out $\{s_2\}$ in the same way. The event $\{s_1, s_2\}$ seems like the best pick: you still consider s_1 and s_2 possible, and so it is true that you know that the true state is in $\{s_1, s_2\}$. However, you knew this before you saw Homer smile. If this is truly all that the smile tells you, then you have learned nothing and have no basis for changing your beliefs about s_1 and s_2.

The problem is that the elementary events we chose for our small-worlds model are not fine enough to allow the representation of this information. Your information allows you to distinguish between grand states in which Homer likes you and smiles and grand states in which Homer likes you and does not smile, but these two states lie in the same elementary event.

We solve this problem by enriching the set of states:

$$S = \left\{ \begin{array}{lcl} s_1 & = & \text{he likes me and smiles} \\ s_2 & = & \text{he likes me and does not smile} \\ s_3 & = & \text{he does not like me and smiles} \\ s_4 & = & \text{he does not like me and does not smile} \end{array} \right\}.$$

Figure 2.1



Enriched state space for example. Shows partition $\{A_1, A_2\}$ of S into payoff-relevant events and partition $\{E_1, E_2\}$ of S into information events. Prior probability of each state is written in lower-right corner of cell.

See also the taxonomy in Figure 2.1. Homer's smile then provides you with the information $\{s_1, s_3\}$, which we denote by E_1.

2.1.3 Conditional beliefs

We now assume that you—as an unboundedly rational decision maker—react to this information by updating your *prior* beliefs, and then maximizing expected utility using your *posterior* beliefs. Your state-independent or state-dependent utility function over *outcomes* does not change. However, the change in beliefs may cause your preferences over *acts* and *actions* to change.

Let $\pi\colon S \to \mathbb{R}$ (a probability measure) be your prior beliefs. For $s \in S$, denote by $\pi(s \mid E)$ your posterior beliefs given that you have learned that the true state is in E. $\pi(\cdot \mid E)\colon S \to \mathbb{R}$ is also a probability measure.

Figure 2.1 shows prior beliefs for the example. According to these prior beliefs you consider s_1 to be 3 times more likely than s_3 (since $\pi(s_1) = 3/8$ and $\pi(s_3) = 1/8$). After you learn only that the true state lies in $E_1 = \{s_1, s_3\}$, you should still consider s_1 to be 3 times more likely than s_3. However, you now consider states s_2 and s_4 to be impossible. The probabilities in your posterior beliefs should still sum to 1. Hence, we need to rescale the probabilities of s_1 and s_3 by dividing them by the prior probability of $\{s_1, s_3\}$, so that $\pi(s_1 \mid E_1) = 3/4$ and $\pi(s_3 \mid E_1) = 1/4$.

The general rule is

$$\pi(s \mid E) = \begin{cases} \dfrac{\pi(s)}{\pi(E)} & \text{if } s \in E, \\ 0 & \text{otherwise.} \end{cases}$$

For an event $A \subset S$, the conditional probability is

$$\pi(A \mid E) = \sum_{s \in A} \pi(s \mid E) = \sum_{s \in A \cap E} \pi(s \mid E)$$
$$= \sum_{s \in A \cap E} \frac{\pi(s)}{\pi(E)} = \frac{\sum_{s \in A \cap E} \pi(s)}{\pi(E)} = \frac{\pi(A \cap E)}{\pi(E)}.$$

This formula,

$$\pi(A \mid E) = \frac{\pi(A \cap E)}{\pi(E)},$$

is called *Bayes' Rule*.

Suppose that

$$u(\text{invitation declined}) = 0,$$
$$u(\text{invitation accepted}) = 4,$$
$$u(\text{no invitation}) = 2\tfrac{3}{4}.$$

The action "invite Homer" leads to the act $\tilde{x}(s_1) = \tilde{x}(s_2) = x_1$ and $\tilde{x}(s_3) = \tilde{x}(s_4) = x_2$. The action "not invite Homer" leads to the act $\tilde{y}(s_1) = \tilde{y}(s_2) = \tilde{y}(s_3) = \tilde{y}(s_4) = x_3$. The expected utility of $\tilde{y}$ is always $2\frac{3}{4}$. The expected utility of $\tilde{x}$ for the prior beliefs is

$$\left(\frac{5}{8}\right) 4 + \left(\frac{3}{8}\right) 0 = 2\frac{1}{2}.$$

Hence, "not invite" is chosen over "invite" if you do not have any information. However, if Homer smiles at you, then the expected utility of f for your posterior beliefs is

$$\left(\frac{3}{4}\right) 4 + \left(\frac{1}{4}\right) 0 = 3.$$

Hence, you prefer the action "invite".

2.1.4 Information as a partition

Compare the statements "Homer smiled at me this morning" and "this morning I will see whether Homer smiles at me". The latter statement describes, from a point of view prior to observing your information, the possible things you will learn. This is called an *information structure*. The former describes, from a point of view posterior to observing your information, what you actually learned. This is called a *realization* of your information; it is also what I have called a "piece of information". (I will frequently refer to both as just "information", when no confusion will arise.)

Let's consider how we would represent an information structure. For each state s, let $I(s)$ be the set of states you consider possible after observing your information. In our example, in states s_2 and s_4, Homer does not smile at you and so you consider $\{s_2, s_4\}$ to contain the possible states. Hence:

$$I(s_1) = \{s_1, s_3\},$$
$$I(s_2) = \{s_2, s_4\},$$
$$I(s_3) = \{s_1, s_3\},$$
$$I(s_4) = \{s_2, s_4\}.$$

If we take all the events on the right, and eliminate duplicates, then we end up with this partition of the set of states:

$$\{\{s_1, s_3\}, \{s_2, s_4\}\}.$$

In each state s, the set of states that you consider possible is the event in the partition that contains state s. That is, seeing whether Homer smiles tells you which event in the partition contains the true state.

I now claim that for an unboundedly rational agent who is aware of her information, information should be partitional as long as we have specified the model properly. Let me illustrate this using a related example. Suppose that Homer smiles if and only if he likes you and you are aware of this fact (at least if you think about it). Then we can return to the simpler set of states:

$$S = \left\{ \begin{array}{l} s_1 = \text{likes me}, \\ s_2 = \text{likes me not}. \end{array} \right\}$$

Seeing whether Homer smiles at you tells you exactly which is the true state ($I(s_1) = \{s_1\}$ and $I(s_2) = \{s_2\}$), and hence can be represented by the partition $\{\{s_1\}, \{s_2\}\}$ of S. Consider instead the following non-partitional information structure:

$$\begin{aligned} I(s_1) &= \{s_1\}, \\ I(s_2) &= \{s_1, s_2\}. \end{aligned}$$

Try to think of a story that almost fits the model and this information structure, and then try to see how the story has an inconsistency, at least for an unboundedly rational decision maker who is aware of her information.

For example, this information structure fits the following story. In state s_1, Homer smiles at me, and I think to myself: "Ahah! He smiled and this means he likes me for sure." In state s_2, Homer does not smile at me, but I do not think anything of it. Even though I know that Homer would smile at me if he liked me, I fail to draw the proper conclusion from the lack of a smile, and so still I consider states $\{s_1, s_2\}$ to be possible, just as I did before walking by Homer. You can probably imagine this kind of thing happening to you, but the story relies on incomplete reasoning, and hence bounded rationality, which we are not incorporating into our normative model.[2]

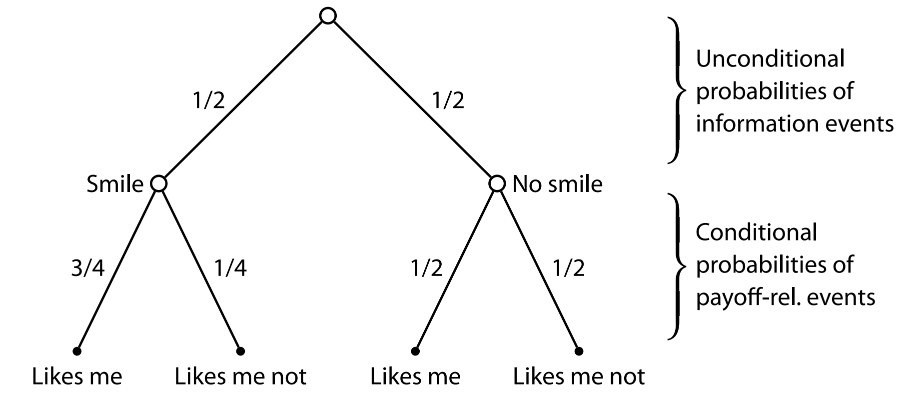
2.2 Bayes' Theorem

2.2.1 The observation of information and updating of beliefs

We can give a nice tree representation of the observation of information and the updating of beliefs.

2. As stated in Section 1.1, we use normative models as approximations for descriptive and prescriptive models because of their simplicity. As we moved from lotteries, to states with objective uncertainty, to states with subjective uncertainty, and to information and Bayesian updating, the decision problems have become more and more complex and the healthy level of scepticism of the realism of the normative model becomes higher and higher.

Figure 2.2



Conditional probabilities for the example.

We are given a set S of states with beliefs $\pi\colon S \to \mathbb{R}$. We partition the set of states in two ways:

1. Let $\{A_1, \ldots, A_m\}$ be a partition of the states into events that you *care about*. If there was no information, this would be a suitable collection of elementary events. In the example shown in Figure 2.1, this partition is $\{A_1, A_2\}$, where $A_1 = \{s_1, s_2\}$ and $A_2 = \{s_3, s_4\}$.
2. Let $\{E_1, \ldots, E_n\}$ be a partition of the states into events that you can *distinguish* with your information. In the example, this partition is $\{E_1, E_2\}$, where $E_1 = \{s_1, s_3\}$ and $E_2 = \{s_2, s_4\}$.

We can calculate the probability of each event in $\{A_1, \ldots, A_m\}$ conditional on each event in $\{E_1, \ldots, E_n\}$. For the example, we have

$$\begin{aligned} \pi(A_1 \mid E_1) &= 3/4 & \pi(A_1 \mid E_2) &= 1/2 \\ \pi(A_2 \mid E_1) &= 1/4 & \pi(A_2 \mid E_2) &= 1/2. \end{aligned}$$

We can also calculate the probability of observing each information event. For the example, we have

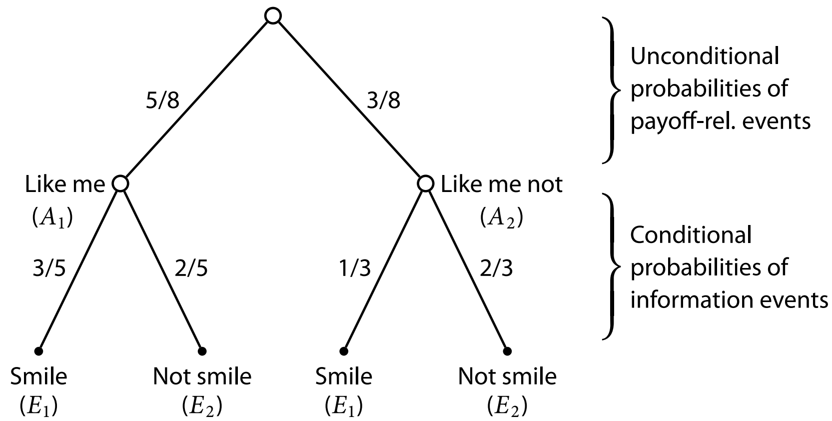
$$\pi(E_1) = 1/2 \quad \pi(E_2) = 1/2.$$

($\pi(E_1)$ is the unconditional probability that Homer smiles at you.) We can then draw the tree shown in Figure 2.2.

The first tier below the root has a node for each information event, and the probability of each of these events is drawn on the corresponding branch. The terminal nodes of each subtree are the payoff-relevant events. The branch to each node is labeled with the probability of the payoff-relevant event conditional on the subtree's information event. In general, the tree has the form shown in the bottom of Figure 2.4.

We worked with trees like this before, as compound lotteries. However, our calculations went in the opposite direction. When finding the reduced lottery of a compound lottery, we were given the probabilities of intermediate events and probabilities of outcomes conditional on these events, and we calculated first the probability of each terminal node and then the overall probability of outcomes that appeared at several terminal nodes. Here, we instead start with a probability measure on the set of states

Figure 2.3



Conditional probabilities of information events for example.

and calculated probabilities of intermediate events and probabilities of outcomes conditional on these events.

It is common to have to perform both of these calculations. This happens when you are given information about probabilities in a tree form, which shows the probability of each information event conditional on the payoff-relevant events. Calculating the probabilities of each payoff-relevant event conditional on the information events involves "inverting" the tree. In our example, suppose that instead of being given a table of probabilities about each state as in Figure 2.1, you are told the probability that Homer does or does not like you, and then the probability that Homer smiles at you given that he likes you or does not like you. These probabilities are shown in Figure 2.3.

To apply Bayes' Rule for calculating the probability of A_1 conditional on E_1,

$$\pi(A_1 \mid E_1) \;=\; \frac{\pi(A_1 \cap E_1)}{\pi(E_1)},$$

we first have to find the unconditional probability of E_1:

$$\begin{aligned}\pi(E_1) &= \pi(A_1)\pi(E_1 \mid A_1) + \pi(A_2)\pi(E_2 \mid A_1)\\ &= (5/8)(3/5) + (3/8)(1/3)\\ &= 1/2.\end{aligned}$$

In general, the information takes the form shown in the top of Figure 2.4. To calculate the conditional probabilities shown in the bottom of Figure 2.4 from the conditional probabilities shown in the top of Figure 2.4, we first calculate the probability of each terminal node $A_i \cap E_j$:

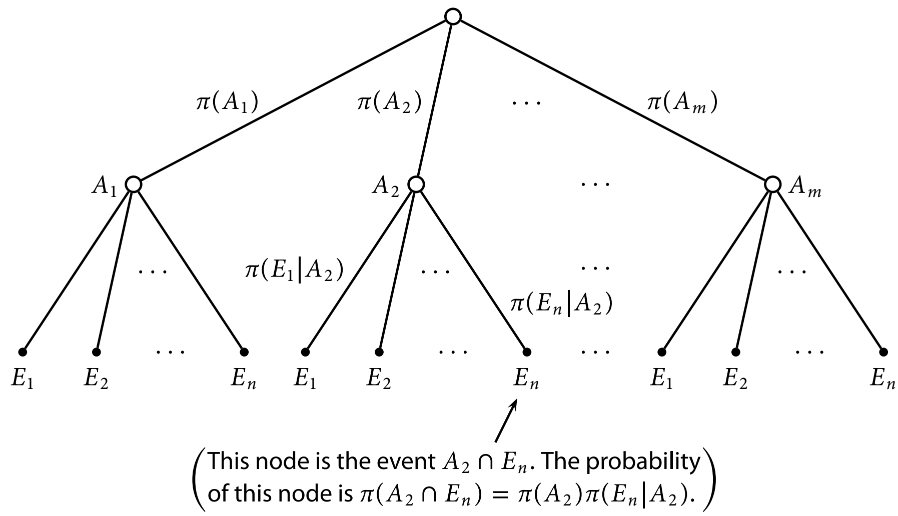
$$\pi(A_i \cap E_j) \;=\; \pi(A_i)\pi(E_j \mid A_i).$$

Then we sum those terminal nodes for E_j to find the overall probability of E_j:

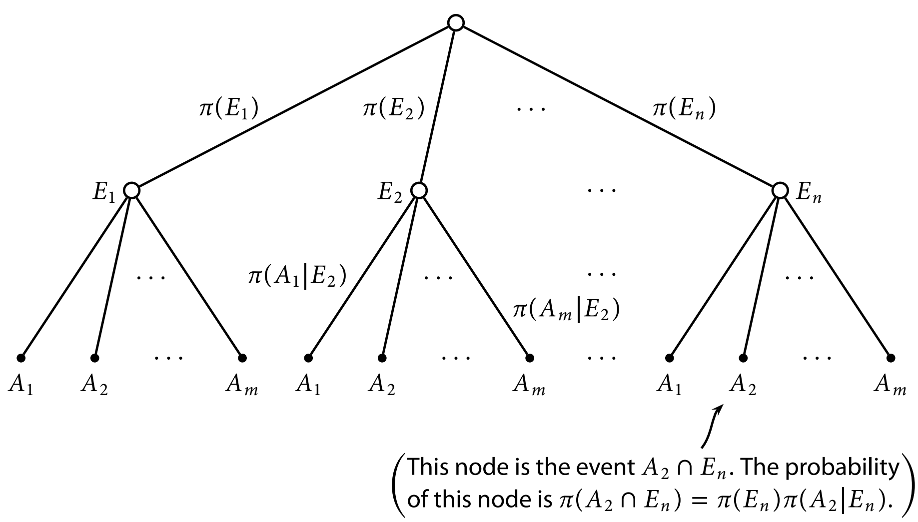
$$\pi(E_j) \;=\; \sum_{i=1}^{m} \pi(A_i \cap E_j) \;=\; \sum_{i=1}^{m} \pi(A_i)\pi(E_j \mid A_i).$$

Figure 2.4

A tree showing *unconditional* probabilities of *payoff-relevant* events $\{A_1, \ldots, A_m\}$, and *conditional* probabilities of *information* events $\{E_1, \ldots, E_n\}$:



A tree showing *unconditional* probabilities of *information* events $\{E_1, \ldots, E_n\}$, and *conditional* probabilities of *payoff-relevant* events $\{A_1, \ldots, A_m\}$:



Bayes' Theorem involves calculating the conditional probabilities shown in the bottom tree from the conditional probabilities shown in the top tree.

Now we apply Bayes Rule to find the probability of A_i conditional on E_j;

$$\pi(A_i \mid E_j) = \frac{\pi(A_i \cap E_j)}{\pi(E_j)} = \frac{\pi(A_i)\pi(E_j \mid A_i)}{\sum_{k=1}^{m} \pi(A_k)\pi(E_j \mid A_k)}. \tag{2.1}$$

This formula is called Bayes' Theorem. Note that all the steps only involve Bayes' Rule. Bayes' Theorem is just an application of the basic formula for conditional probabilities to a particular way in which the probability information is presented.

2.2.2 An application of Bayes' Theorem

Although Bayes' Theorem is just an application of conditional probabilities, the conclusions one obtains when using it are sometimes surprising.[3]

Here is a standard application of Bayes' Theorem. Suppose that you know that 5% of the population has HIV, and you have a test for detecting HIV that produces (type 1 and type 2) errors 5% of the time. If we draw someone randomly from the population and the person tests positive for HIV, what is the probability that the person has the virus?

The payoff relevant partition is

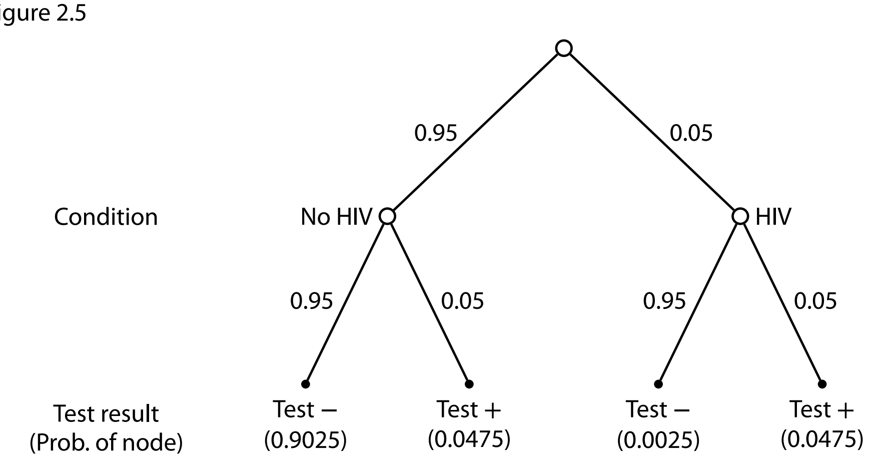
$$\left\{ \begin{array}{lcl} A_1 & = & \text{“No HIV”} \\ A_2 & = & \text{“HIV”} \end{array} \right\}.$$

The information partition is

$$\left\{ \begin{array}{lcl} E_1 & = & \text{“Test −”} \\ E_2 & = & \text{“Test +”} \end{array} \right\}.$$

We are given, for example, $\pi(A_1) = 0.95$, and $\pi(E_2 \mid A_1) = 0.05$. We can show these probabilities in tree form, as in Figure 2.5.

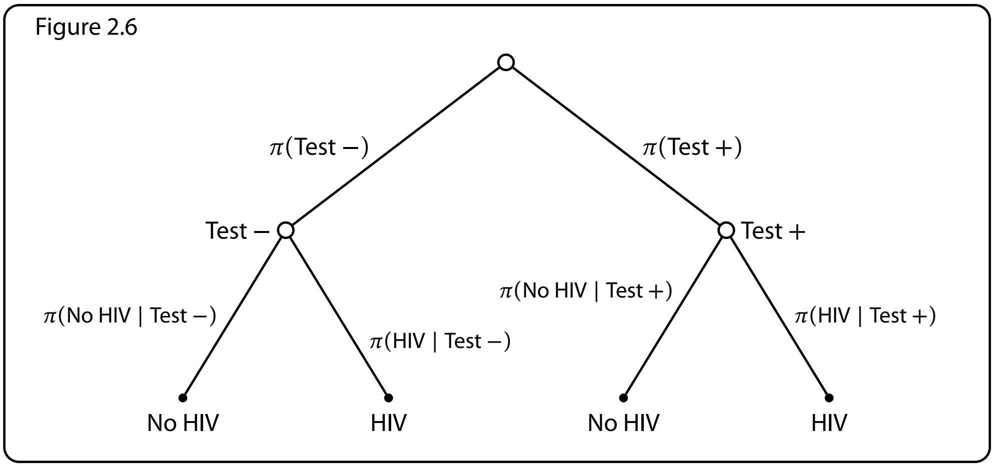
Figure 2.5



The terminal nodes are the events $A_i \cap Ej$, which can be our elementary events or states.

We are asked to find the probability that a person has the virus, conditional on testing positive. This information is part of the tree in Figure 2.6.

3. The fact that the answers can surprise people is yet more evidence that people are not perfect Bayesians!



Recall the formula for conditional probability:

$$P[\text{Has HIV} \,|\, \text{test} +] \;=\; \frac{P[\text{Has HIV and tests} +]}{P[\text{tests} +]}.$$

We were not told directly the probability that a person would both have HIV and test positive, but this is easy to calculate from the tree:

$$\begin{aligned} P[\text{Has HIV and tests} +] \;&=\; P[\text{Has HIV}]P[\text{Tests} + \,|\, \text{Has HIV}] \\ &=\; (0.05)(0.95) \;=\; 0.0475. \end{aligned}$$

That is, to find the probability of a state, we multiply the probabilities along the branches that lead to that state. All such probabilities are shown below the terminal nodes in the tree.

We were also not told directly the probability that a person would test positive. We have to sum the probabilities of all the states in which a person tests positive, calculating the probability of each state in the same way.

$$\begin{aligned} P[\text{tests} +] \;&=\; P[\text{Has HIV and tests} +] + P[\text{No HIV and tests} +] \\ &=\; (0.05)(0.95) + (0.95)(0.05) \;=\; 0.095. \end{aligned}$$

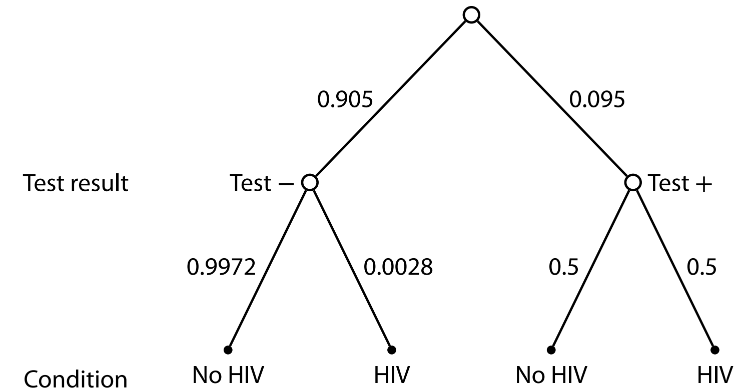
Hence,

$$P[\text{Has HIV} \,|\, \text{test} +] \;=\; \frac{0.0475}{0.095} \;=\; 0.5.$$

In this example, the surprise is that the small error from the test grew into much uncertainty about whether the person has HIV, due to the fact that few people in the population actually have the virus. Most people if they had to answer the original question quickly, without performing the calculations of Bayes' Theorem, would give a higher answer than 1/2. The test still provided significant information, because before the test the probability that the person had HIV was only 0.05.

We can calculate other probabilities and conditional probabilities in the same way. This gives us the inverse of the tree in which we were given initially given the probabilities in Figure 2.7.

Figure 2.7



2.2.3 Setting up the right state space: Example 1

Here is an application of Bayes' Theorem involving subjective uncertainty. There are three prisoners—Larry, Mo, and Curly—in different cells of a prison (in different cells). One of them will be executed, but Larry does not know which one. His subjective beliefs put equal probability on each prisoner. He asks the prison guard to tell him who will be executed. The prison guard replies that he will only tell Larry the name of one of the other prisoners (Mo or Curly) who will not be executed. Larry says OK, and the prison guard tells him that Mo will not be executed. Larry reasons that he has just learned that either he or Curly will be executed. Since it is equally likely that he or Curly will be executed, he figures that the probability of his execution has just increased from 1/3 to 1/2!

Larry's reasoning seems faulty, because the guard was going to name one of the prisoners no matter what. How could the guard's reply provide any information about whether Larry would be executed? The correct conditional probability must be 1/3. But this means that the conditional probability that Curly will be executed has gone from 1/3 to 2/3. How do we convince Larry of these facts? How do we explain the source of his error? How do we explain why the guard's reply provides no information about Larry's choice of execution, but does provide information about Curly's chance of execution?

Let's first come up with a (faulty) model that might lie behind Larry's calculations. He appears to be using (implicitly or explicitly) a model in which there are three states, identified by the initial of the person who will be executed:

$$S = \{\mathrm{L}, \mathrm{M}, \mathrm{C}\}.$$

His prior beliefs are

$$\pi(\mathrm{L}) = \pi(\mathrm{M}) = \pi(\mathrm{C}) = 1/3.$$

He reasons that the guard's reply tells him that the true state is in the event $\{\mathrm{L}, \mathrm{C}\}$. Applying Bayes' Rule,

$$\pi(\mathrm{L} \mid \{\mathrm{L}, \mathrm{C}\}) = \frac{\pi(\mathrm{L})}{\pi(\{\mathrm{L}, \mathrm{C}\})} = \frac{1}{2},$$

he calculated that the conditional probability of his execution is 1/2.

Let's check the information structure. By Larry's reasoning, when the state is "Mo," the guard says "Curly," and Larry's information is {L, C}. That is:

$$I(\mathrm{M}) = \{\mathrm{L}, \mathrm{C}\}.$$

By the same reasoning,

$$I(\mathrm{C}) = \{\mathrm{L}, \mathrm{M}\}.$$

Already we are in trouble. The information sets {L, C} and {L, M} are neither disjoint nor equal, and hence cannot be part of a partition. There is further trouble when we try to define $I(\mathrm{L})$. The information set in each state should be uniquely defined in a proper model. However, when the state is Larry, the guard might say Mo or Curly, and so Larry might think his information set is {L, C} or {L, M}.

This suggests that Larry's set of states is not rich enough. The original set of states (elementary events) given above can be the partition of the states into payoff relevant events:

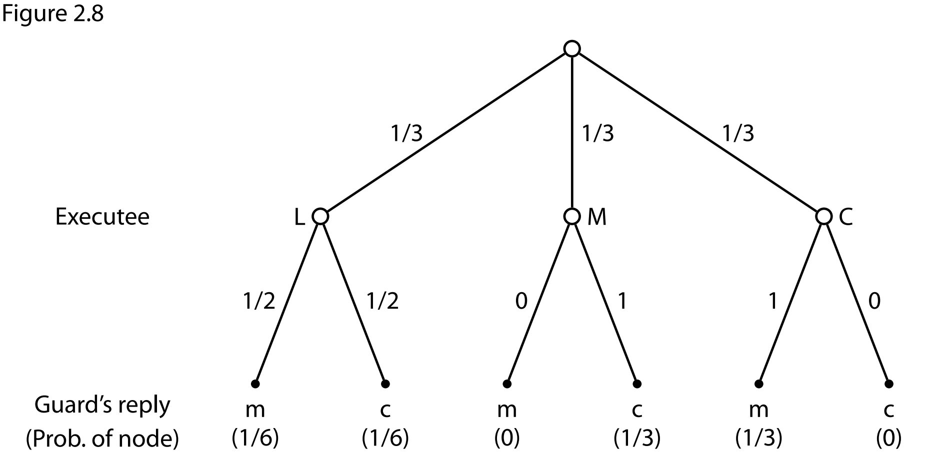
$$\left\{ \begin{array}{rcl} \mathrm{L} & = & \text{Larry} \\ \mathrm{M} & = & \text{Mo} \\ \mathrm{C} & = & \text{Curly} \end{array} \right\}.$$

The informational events are identified by the guard's reply, and are identified with the lower case initial of the name:

$$\left\{ \begin{array}{rcl} \mathrm{m} & = & \text{Mo} \\ \mathrm{c} & = & \text{Curly} \end{array} \right\}.$$

We know the prior probabilities of who will be executed. The probabilities of the guard's reply conditional on who will be executed are straightforward (or are they?). For example, the probability that the guard says "Curly" when Mo is to be executed is 1; if Larry is to be executed, then the guards says "Mo" or "Curly" with equal probability. These probabilities are shown in Figure 2.8.

Figure 2.8



Then the probability that Larry will be executed conditional on the guard's replying "Mo" is

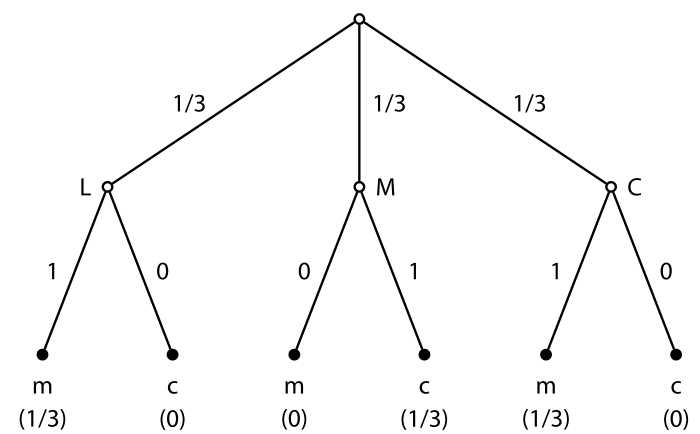
$$\begin{aligned} \pi(\mathrm{L} \mid \mathrm{m}) &= \frac{\pi(\mathrm{L} \cap \mathrm{m})}{\pi(\mathrm{m})} = \frac{\pi(\mathrm{L} \cap \mathrm{m})}{\pi(\mathrm{L} \cap \mathrm{m}) + \pi(\mathrm{M} \cap \mathrm{m}) + \pi(\mathrm{C} \cap \mathrm{m})} \\ &= \frac{1/6}{(1/6) + 0 + (1/3)} = \frac{1}{3}. \end{aligned}$$

On the other hand, the guard's reply does depend on whether Curly will be executed, and so the conditional probability of this event is different from the prior probability:

$$\pi(\mathrm{C} \mid \mathrm{m}) = \frac{\pi(\mathrm{C} \cap \mathrm{m})}{\pi(\mathrm{m})} = \frac{1/3}{1/2} = \frac{2}{3}.$$

While all this may sound like the right answer, it is possible that Larry would get some useful information from the guard's reply. Larry might have other beliefs about what the guard will say when Larry is to be executed. As an extreme case, suppose that the guard is monosyllabic and Larry knows that, given a choice of saying "Mo" or "Curly", the guard will always say "Mo". Then the probability that the guard says "Mo" conditional on Larry being executed is 1, rather than 1/2:

Figure 2.9



In this case, you can see that the probability that Larry is going to be executed conditional on the guard saying "Mo" is indeed 1/2. The reply "Mo" is bad news. But the reply "Curly" would be good news; the probability that Larry is going to be executed conditional on that reply is 0.

You may have seen this problem before. In another variant (the "Monte Hall" puzzle), a contestant on "Let's Make a Deal" chooses a door, in the hope of finding the door with a car behind it (the other two doors are empty). Then, as usual, Monte Hall opens one of the unchosen doors and shows that it is empty. The contestant is now given a chance to switch, but she reasons that there is no point in doing so because the car is still behind her door or the remaining door with equal probabilities. As an exercise, relate this puzzle to the previous problem, and determine whether there is any advantage to switching.

2.2.4 Setting up the right state space: Example 2

Our second example—of an application of Bayes' Theorem that requires first carefully constructing the state space—is another famous puzzle. It can only be understood by introducing subjective beliefs and it provides a nice critique of the principle of insufficient reason.

I show you two envelopes, and tell you that I have put twice as much money in one as in the other, but I don't tell you how much I put in or which envelope contains more. I let you pick an envelope. Then I give you the opportunity to switch envelopes. Not knowing what to do, you consult your local statistician, who tells you to switch, giving the following explanation:

> You know that there is a fifty-fifty chance that you have picked the envelope with the larger amount of money. Now, let x be the amount that is in the envelope you picked. If you switch, you get $x/2$ with probability 1/2, and $2x$ with probability 1/2. The expected value of the amount of money in the other envelope is thus $1.25x$, which is greater than the amount in your envelope.

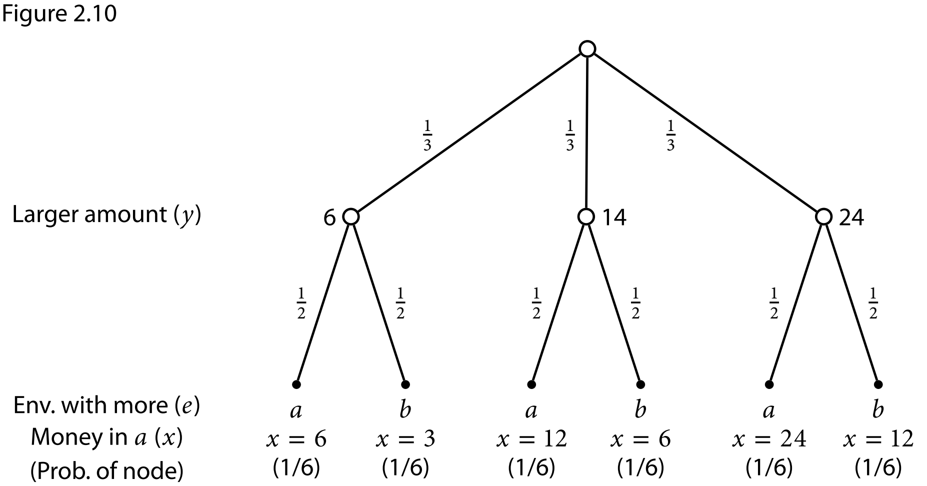
There is no question that the answer is wrong (if it were right, then after switching you would be better off switching again, and again, . . .). What we want to do is to see the flaw in the argument explicitly.

First we need a states-of-nature model. Let's give the envelopes names: a is the envelope in your hand, and b is the other envelope. A state of nature is given by $\langle e, y\rangle$, where e is the name of the envelope with the larger amount of money, and y is the amount of money in that envelope. For example, states $s_1 = \langle a, 10\rangle$ and $s_2 = \langle b, 30\rangle$ correspond to the following:

	s_1	s_2
\$ in your envelope	10	15
\$ in other envelope	5	30

Suppose that you believe that y is 6, 12, or 24 with equal probability. The probability that envelope a has the larger amount of money is independent of y. Hence, we can depict the uncertainty as in Figure 2.10.

Figure 2.10



Let's first see whether you should switch envelopes if you first get to observe how much is in the envelope you chose. Think for a moment about whether there might be some reason to switch.

In the erroneous argument given by the "statistician", it was assumed that the probability of having the envelope with the larger amount of money was independent of how much money was in your envelope. This is where he made his mistake. a and b are not independent of x, even though they are independent of y! In the example,

$$\begin{aligned} P[a\,|\,x=3] &= 0 \\ P[a\,|\,x=6] &= 1/2 \\ P[a\,|\,x=12] &= 1/2 \\ P[a\,|\,x=24] &= 1 \end{aligned}$$

Hence, the expected value of switching ($E[x_b]$, where x_B is the amount of money in envelope B) can be greater than or less than x:

$$\begin{array}{rcllcl} E[x_b\,|\,x=3] &=& P[a\,|\,x=3](1.5)+P[b\,|\,x=3](6) &=& 6 & > & 3 \\ E[x_b\,|\,x=6] &=& P[a\,|\,x=6](3)+P[b\,|\,x=6](12) &=& 7.5 & > & 6 \\ E[x_b\,|\,x=12] &=& P[a\,|\,x=12](6)+P[b\,|\,x=12](24) &=& 15 & > & 12 \\ E[x_b\,|\,x=24] &=& P[a\,|\,x=24](12)+P[b\,|\,x=24](48) &=& 12 & < & 24 \end{array}$$

In this example, you should switch if your envelope has \$3 or \$6, or \$12, and keep your envelope if it has \$24.

When we do not observe the amount of money in the envelope, we can calculate the expected value of switching by averaging the expected value of switching conditional on each value of x, given the probability of x. Then we get

$$\begin{aligned} E[x_b] &= P[x=3]E[x_b\,|\,x=3]+P[x=6]E[x_b\,|\,x=6] \\ &\qquad +P[x=12]E[x_b\,|\,x=12]+P[x=24]E[x_b\,|\,x=24], \\ &= (1/6)(6)+(1/3)(7.5)+(1/3)(15)+(1/6)(12), \\ &= 10.5. \end{aligned}$$

This is the same kind of calculation the statistician was using, but he substituted in the wrong conditional probabilities. We get the right answer, which is that the expected value of switching is equal to the expected value of keeping the envelope:

$$E[x] = (1/6)(3)+(1/3)(6)+(1/3)(12)+(1/6)(24) = 10.5.$$

Note that it is impossible that $E[a\,|\,y]$ be 1/2 for all y. This would imply that

$$\pi[y=2] = \pi[y=4] = \pi[y=8] = \pi[y=16] = \ldots;$$

if any of these probabilities is positive, then the sum of all the probabilities must be greater than one. That is, it is impossible to have a uniform probability measure on an unbounded range.

The fact that the resolution of the problem requires subjective beliefs that are not uniform over the money values might bother you. The problem, as it is presented to us, provides no information about how the money values are chosen. "How can I have beliefs? How can I give one value of money a higher probability than another, when I am totally ignorant?" It is the hypothetical nature of the puzzle that leads you to this objection. In any actual occurrence of the situation, you would have beliefs based on all sorts of information. For example, if I came to your class and did an experiment based on this puzzle with real money, would you switch after observing that your envelope has \$1000? You would probably be very surprised to see that much money, but also be quite sure that you have the envelope with the larger amount of money in it. That is, although you put low probability on $y = 1000$, you put even lower probability on $y = 2000$. If, on the other hand, you knew that the envelopes only contained bills and your envelope had \$1 or \$3, then you would know that your envelope had the smaller amount of money and so you would want to switch.

Exercise 2.1. Read Puzzle 2, from the Winter 1990 issue of the *Journal of Economic Perspectives*. Both the question and the solution are attached. You should derive and explain the solution clearly and explicitly.

a. Start by drawing a tree representation of the uncertainty, with the first level resolving whether or not the patient takes zomepirac, and the second level resolving whether the patient dies from zomepirac, dies from other causes, or does not die at all. Use symbols to label the probabilities of all the branches.

b. In terms of these symbols, give the formula for the probability that the woman died from zomepirac, conditional on her having died. (This is the probability that the puzzle asks you to calculate.)

c. State what is known for sure about the probabilities in the tree, and why.

d. State what is approximately known about the probability that a patient dies conditional on not taking zomepirac.

e. Using this approximation and the other information, you can now use the formula from (b) to find the solution to the puzzle.

2.3 Independence and Exchangeability

2.3.1 Introduction

Chapter 11 in David Kreps' *Notes on the Theory of Choice* is about independence, exchangeable random variables, and de Finetti's theorem. These are fundamental topics in statistics. They are also topics which highlight the differences between objective and subjective uncertainty, and the various philosophies of probability and uncertainty. Kreps gives an animated and interesting discussion of these differences.

The purpose of these notes is to review and practice the applications of exchangeable random variables. As is my approach in most of this book, I will consider only simple probability measures (finitely many possible outcomes), as this is sufficient for conveying all the intuition.

2.3.2 Independence

Let $\{\tilde{x}_1, \tilde{x}_2, \ldots\}$ be an infinite sequence of random objects, all taking values in the same finite set X. Think of these as the outcomes of potentially infinitely many repetitions of an experiment or trial. For each example, $\tilde{x}_i$ could be the outcome of the i'th coin toss, or the height of the i'th person drawn randomly from a population, or the outcome of the i'th game of squash played between the same competitors.

The joint distribution of the first n objects $\{\tilde{x}_1, \dots, \tilde{x}_n\}$ tells me, for any list of values $\{x_1, \dots, x_n\}$, the probability that simultaneously $\tilde{x}_1 = x_1$, $\tilde{x}_2 = x_2$, ..., and $\tilde{x}_n = x_n$. (Here, and in other places, I will take the first n objects even though the same would apply to any n objects, even if non-consecutive.) If I were considering a bet on the outcomes of the first two games of squash such that I win the bet if one play wins both games but I lose if the players split games, then I would be interested in the joint distribution of $\tilde{x}_1$ and $\tilde{x}_2$. The distribution of just one $\tilde{x}_i$ is called its marginal distribution. If I were contemplating a bet on the outcome of the second game of squash only, I would only be interesting in the marginal distribution of $\tilde{x}_2$.

The random objects $\{\tilde{x}_1, \tilde{x}_2, \dots\}$ are independent if the joint distributions are the products of the marginal distributions. That is,

$$P[\tilde{x}_1 = x_1, \tilde{x}_2 = x_2, \dots, \tilde{x}_n = x_n] = P[\tilde{x}_1 = x_1]P[\tilde{x}_2 = x_2] \cdots P[\tilde{x}_n = x_n].$$

A nice feature of independence is that we can calculate the joint distributions from the marginal distributions. These means that there is less information to keep track of.

The random objects $\{\tilde{x}_1, \tilde{x}_2, \dots\}$ are identically distributed if each $\tilde{x}_i$ has the same marginal distribution. By itself, this does not help us much. However, when the objects are both identically *and* independently distributed (IID), the common marginal distribution is all the information we need to know the joint distribution of any of the random objects.

One implication of the IID property is that the probability of a given string of outcomes does not depend on the order of the outcomes. For example,

$$\begin{aligned} &P[\tilde{x}_1 = \mathrm{W}, \tilde{x}_2 = \mathrm{W}, \tilde{x}_3 = \mathrm{T}] \\ &\qquad = P[\tilde{x}_1 = \mathrm{T}, \tilde{x}_2 = \mathrm{W}, \tilde{x}_3 = \mathrm{W}] = P[\tilde{x}_i = \mathrm{W}]^2 P[\tilde{x}_i = \mathrm{T}] \end{aligned} \tag{2.2}$$

That is, the probability of a given string of outcomes depends only on the number of times each outcome appears in the string.

Furthermore, it becomes easy to calculate the probability of a particular frequency of outcomes for the first n trials. An analogue of equation (2.2) gives us the probability of any outcome of the first 5 games of soccer that has 2 wins, 2 ties and 1 loss. To get the overall probability that the first 5 games have 2 wins, 2 ties and 1 loss, we multiply this probability times the number of outcomes with this frequency of wins, ties and losses. The number of such outcomes is the number of permutations of $\{\mathrm{W}, \mathrm{W}, \mathrm{T}, \mathrm{T}, \mathrm{L}\}$ (5 objects, 2 of which are of one kind, 2 of which are of another kind, and 1 of which is of another kind), which is equal to

$$\frac{5!}{2! \cdot 2! \cdot 1!} = 30.$$

If the probability of a win is 1/2, the probability of a tie is 1/6, and the probability of a loss is 1/3, then the probability of 2 wins, 2 ties and 1 loss in the first 5 games (or any given 5 games) is thus

$$(30)\left((1/2)^2(1/6)^2(1/3)^1\right) = 5/72.$$

When there are only two possible outcomes, then these method for calculating the probability of a given frequency of outcomes reduces to the binomial distribution. Suppose that each game must end in a win or a loss, and let θ be the probability of a win. Then the probability of any outcome of n games that has m wins is $\theta^m(1-\theta)^{n-m}$. (The probability of a loss is $1-\theta$ and there are $n-m$ losses.) The number of ways in which the n games can have m wins is $\binom{n}{m} = \frac{n!}{m!(n-m)!}$. Hence the probability of m wins out of n games is

$$b(m; n, \theta) = \binom{n}{m}\theta^m(1-\theta)^{n-m}.$$

Because the case of just two possible outcomes is arithmetically the simplest, I will focus on this case in subsequent sections However, having more outcomes is conceptually the same.

2.3.3 Exchangeability

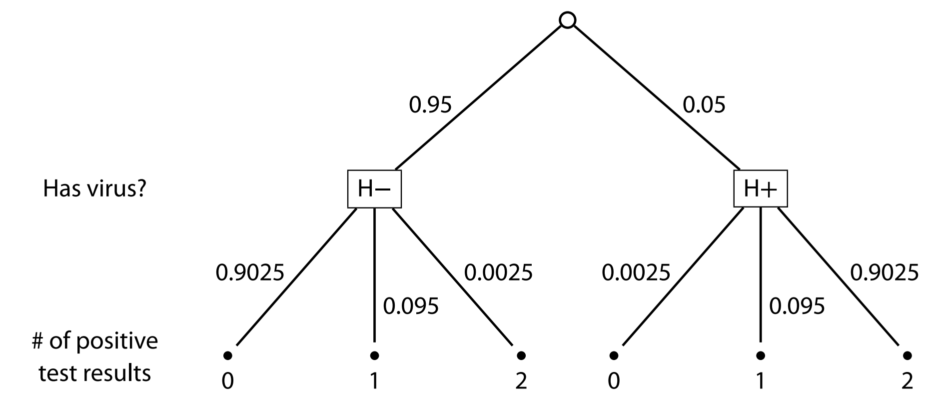
There are many situations of repeated experiments or trials where you think that the outcomes are identically and independently distributed, but you do not know what the common marginal distribution is! The phrase "you think that the outcomes are IID" is a little confusing here. Conditional on something you do not know, your subjective probabilities are IID. However, given that you do not know the common marginal distribution, your unconditional (prior) subjective probabilities do not satisfy independence. Instead, we just say that they are *exchangeable*.

Here is an extreme example that illustrates this distinction. Suppose that someone has a two-headed coin and two-tailed coin. The person puts both coins in a box, shakes the box, and then pulls out one of the coins but does not show it to you. Let's look at your beliefs about the outcomes of a sequence of tosses of the coin. With probability 1/2 the coin has two heads, and with probability 1/2 the coin has two tails. Thus, the probability that the first toss is heads is 1/2. Similarly, if you have to bet only on whether the second toss is heads or tails, you would place probability 1/2 on heads. These marginal probabilities are the same as when a normal coin is tossed. However, because the outcomes of the tosses of a normal coin are independent, the probability of getting heads on the first toss and tails on the second toss is 1/4 (1/2 times 1/2). In the case of the two-headed or two-tailed coin, on the other hand, this probability is 0. Hence, your beliefs (probabilities) about the outcomes of a sequence of tosses of the coin are not independent. However, conditional on the coin having two heads, your beliefs are trivially IID, because you know that each toss will come up heads.

Now a more interesting example. Recall the example of testing for the HIV virus in the handout on "Bayes Theorem" and in the column of *Ask Marilyn*. Modify the example by testing the same person n times. Let's suppose for now that false test results are only due to random laboratory errors, and do not have to do with the person who was being tested. Then, conditional on knowing that the person has the HIV virus, the test results are IID with probability 0.95 of a positive result. Conditional on knowing that the person does not have the HIV virus, the test results are IID with probability 0.05 of a positive result. However, we do not know whether the person has the virus.

The overall uncertainty we face can be represented in a two-level tree, in which the first level determines whether or not a person has the HIV virus, and the second level gives the number of positive results out of n trials. Here is the tree for $n = 2$:

Figure 2.11



The conditional probabilities in the second level are computed using the binomial distribution, because, conditional on whether or not the person has HIV, the test results are IID. Hence, we get some of the arithmetic advantages of working with IID test results, but overall we have to consider several IID distributions (one for each branch at the top level); if we already knew whether or not the person had HIV, we would only have to calculate one binomial distribution.

To find out the overall probability that both test results are positive, we have to add up the probabilities of the terminal nodes in the tree for which both test results are negative. Let Y_n be the random variable that is equal to the total number of positive test results after n trials. Then the question we have asked is the probability that $Y_n = 2$, and the answer is[4]

$$P[Y_n = 2] = P[\mathrm{H-}]P[Y_n = 2|\mathrm{H-}] + P[\mathrm{H+}]P[Y_n = 2|\mathrm{H+}] \tag{2.3}$$

$$= (0.95)(0.0025) + (0.05)(0.9025) = 0.0475 \tag{2.4}$$

4. This is an example of "undoing" conditional probabilities to get unconditional probabilities. We have done this kind of exercise several times, particular when applying Bayes Theorem. The general rule is that if $A_1, \ldots, A_k$ are events that are disjoint and whose union is the entire set of states (i.e., it always the case that one and only one of the events is true), and if B is any other event, then

$$P[B] = \sum_{i=1}^{k} P[A_i]P[B|A_i].$$

2.3.4 Exchangeability and Bayesian Inference

We can also apply Bayes Theorem to calculate our revised belief that the person has HIV, after observing two positive test results:

$$\begin{aligned} P[\mathrm{H+}|Y_n = 2] &= \frac{P[\mathrm{H+} \text{ and } Y_n = 2]}{P[Y_n = 2]} \\ &= \frac{P[\mathrm{H+}]P[Y_n = 2|\mathrm{H+}]}{P[Y_n = 2]} \\ &= \frac{(0.05)(0.9025)}{0.0475} \\ &= 0.95 \end{aligned}$$

The two positive tests are much more conclusive than a single positive test.

What we have just been doing looks identical to the Bayes inference problem that was on the handout "Bayes Theorem", except that we did two tests for the virus instead of one. Why is exchangeability important now that we have more than one test result? First, we could claim that only the total number of positive results, rather than the order of results, was important. Second, we could use the binomial distribution (or, with more than two possible outcomes, some other distribution based on IID random variables) to calculate the probabilities of the test results. If we repeated the tests 10 times, these simplifications would be very important. Instead of keeping track of order of the results (which gives $2^{10} = 1,024$ possible outcomes) we only have to keep track of the total number of positive results (which gives only 11 possible outcomes). We can mechanically calculate the conditional probabilities using the binomial theorem:

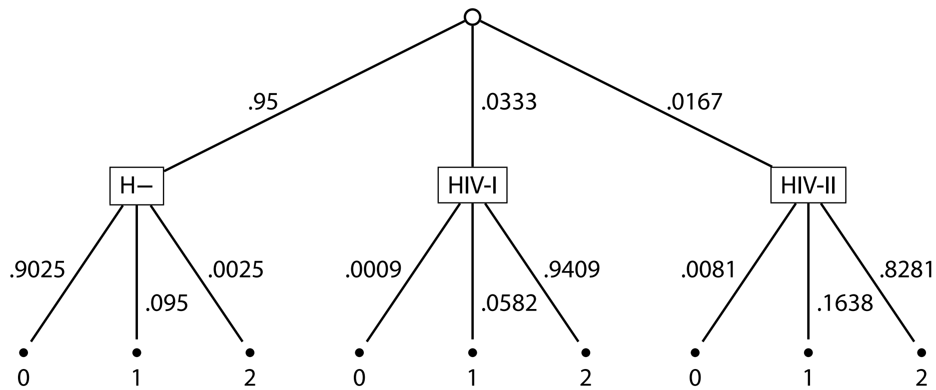
$$P[Y_n = 7|\mathrm{H+}] = \frac{10!}{7! \cdot 3!}(0.95)^7(0.05)^3 = 0.010475.$$

One has to be careful about jumping to the conclusion that the trial results are exchangeable. For example, if you look at the results of tennis games, you will find that there is correlation due to transient factors such as a player's fatigue at one moment. Also, even if trial results are exchangeable, you have to be careful when identifying *what* you must condition on for tests results to be IID.

For example, go back and read the *Ask Marilyn* article. The author notes that we cannot conclude that the test results are independent, because factors that cause an error for a person are likely to cause an error for the person in repetitions of the test. If these factors are transient, then the results are not exchangeable. If these factors are the same in every repetition, then the results may be exchangeable, but conditioning the probabilities just on whether or not the person has HIV will not be enough to get IID test results. We have to also condition the probabilities on these other factors.

Suppose that there are actually two kinds of HIV viruses, I and II. 2/3 of the those with HIV have virus I, and 1/3 have virus II. For people with HIV-I, the test produces a false negative 3% of the time. For people with HIV-II, the test produces a false negative 9% of the time. For people without HIV, the test produces a false positive 5% of the time, as before. The probability of a negative test result conditional on having HIV (either kind) is 5% (why?). Therefore, the tree I drew in the "Bayes' Theorem" handout, for one trial, is still correct. However, the tree I drew above for two trials is not longer correct. To restore independence of the trials, we have to condition our beliefs more finely; even though our goal is to draw inferences about whether the person has HIV, without distinguishing among the viruses, we cannot reap the benefits of conditional independence without conditioning also on whether the person has HIV-I or HIV-II. The correct tree is shown in Figure 2.12.

Figure 2.12



Now after observing two positive results, the probability that the person has either HIV virus is[5]

$$P[\mathrm{H}+\,|\,Y_2 = 2] \;=\; \frac{(0.0333)(0.9409) + (0.0167)(0.8281)}{(0.95)(0.0025) + (0.0333)(0.9409) + (0.0167)(0.8281)} \;=\; 0.950038.$$

2.3.5 What about Kreps' Formula on Page 155?

Look at the formula for $P[Y_{10} = 7]$ that is on the top of page 155 in Kreps' *Notes on the Theory of Choice*. Look at my equation (2.3) for $P[Y_2 = 2]$. What is the difference? For one thing, Kreps' formula has $n = 10$ and $m = 7$ whereas mine has $n = 2$ and $m = 2$; to make the formulas more comparable, let's rewrite mine for the case of $n = 10$ and $m = 7$:

$$P[Y_{10} = 7] \;=\; P[\mathrm{H}-]P[Y_{10} = 7\,|\,\mathrm{H}-] + P[\mathrm{H}+]P[Y_{10} = 7\,|\,\mathrm{H}+] \qquad (2.5)$$

Let α be the probability of a positive test result in each trial. Then $P[Y_{10} = 7\,|\,\mathrm{H}-]$ depends only on the value of α given H− ($\alpha = 0.05$), and $P[\mathrm{H}+]P[Y_{10} = 7\,|\,\mathrm{H}+]$ depends only on the value of α given "H+" ($\alpha = 0.95$). Therefore, we can rewrite equation (2.5) as

$$P[Y_{10} = 7] \;=\; P[\alpha = 0.05]P[Y_{10} = 7\,|\,\alpha = 0.05] + P[\alpha = 0.95]P[Y_{10} = 7\,|\,\alpha = 0.95]. \qquad (2.6)$$

5. After reading *Ask Marilyn*, you may have conjectured that this number would fall rather than increase. However, her claim is based on the assumption that there are factors that make some non-HIV people more likely to have false positive reports than others. I only modeled the case where some HIV people are more likely to have false negative reports than others.

Let's generalize this a bit and assume that α can take on k values $\alpha_1, \ldots, \alpha_k$ (e.g., at the end of the previous section, we considered $k = 3$). Then equation (2.6) becomes

$$P[Y_{10} = 7] = \sum_{i=1}^{k} P[\alpha_i] P[Y_{10} = 7 \mid \alpha = \alpha_i]. \tag{2.7}$$

$P[Y_{10} = 7 \mid \alpha = \alpha_i]$ is given by the binomial theorem, and so equation (2.7) becomes

$$P[Y_{10} = 7] = \sum_{i=1}^{k} P[\alpha_i] \frac{10!}{7!3!} \alpha_i^7 (1 - \alpha_i)^3. \tag{2.8}$$

This, in turn, is just

$$P[Y_{10} = 7] = E\left[\frac{10!}{7!3!} \alpha_i^7 (1 - \alpha_i)^3\right] \tag{2.9}$$

given our beliefs about α. This is a particular example of the formula that is near the bottom of page 154.

Now a technical detail you do not need to know for this course, except if you want to take the last step to understanding Kreps' formula on page 155. To cover the cases where α can have finitely many or infinitely many values, we can summarize the beliefs about α by a cumulative distribution function F (e.g., $F(\gamma)$ is the probability that α is less than or equal to γ). If F is differentiable, then the derivative f is the density function, and the expected value in equation (2.9) is computed by integrating:

$$P[Y_{10} = 7] = \int_0^1 \frac{10!}{7!3!} \alpha^7 (1 - \alpha)^3 f(\alpha)\, d\alpha. \tag{2.10}$$

In more general versions of integration that can handle the case where F is not differentiable, one common notation is to write $dF(\alpha)$ instead of $f(\alpha)\, d\alpha$. With this substitution, plus a change in Greek letters from α to γ, we get Kreps' formula:

$$P[Y_{10} = 7] = \int_0^1 \frac{10!}{7!3!} \gamma^7 (1 - \gamma)^3 f(\gamma)\, d\gamma. \tag{2.11}$$

2.3.6 What about de Finetti's Theorem?

I gave a definition of exchangeability in Section 2.3.3. Actually, the original definition of exchangeability is that the joint distribution function does not change when the order of the trials is reversed. See page 154 for details. De Finetti's proved that this condition is equivalent to the one I stated in Section 2.3.3.

Even when each trial has only finitely many outcomes, de Finetti's Theorem gets complicated. The reason as that there are still infinitely possible values for the entire sequence of trials. For example, if the outcome of each trial is 0 or 1, then an outcome for the entire sequence of trials is an infinite sequence of 0's and 1's. This looks like binary expansion of a real number, and the binary expansion of any real number could be the outcome for the sequence of trials. Thus, "size" of the set of outcomes has the cardinality of the continuum.

But you do not need to know the theorem in order to apply exchangeability, as we have done above. This is why I asked you to skip over de Finetti's Theorem when reading Kreps' Chapter 11. If you want to learn more about it, you will have to take an advanced probability or statistics class.

2.3.7 What about subjective uncertainty?

The ideas of exchangeability and applications to Bayesian inference do not require that uncertainty be subjective. This topic is taught in statistics courses that make no reference to subjective uncertainty. So why did we study this right after studying subjective uncertainty?

The long and interesting answer to this question is found in Kreps' Chapter 11. I will just give a very brief and simple answer. The reason is that conditional independence property of the exchangeable processes typically has some "objectivist" foundations, but the beliefs about the underlying parameters are typically subjective. For example, in Kreps' thumbtack example, we can repeatedly toss the thumbtack to measure the empirical distribution of heads and tails. However, we cannot repeat this whole scenario over and over again to get an objective measure of what our prior beliefs about the bias of the talk should be.

Chapter 3

Risk Preferences

3.1 Money outcomes and risk aversion

3.1.1 Monetary outcomes and random variables

Recall the states-of-nature model, in which S is a finite set of states, $\pi\colon S \to \mathbb{R}$ is a probability measure on S representing beliefs, and X is a finite set of outcomes. An act is a function $\tilde{x}\colon S \to X$. We now allow the set X to be infinite, so as not to be too restrictive a priori about the potential outcomes. However, to keep the statistics simple, we will almost always continue to assume that, for each act or lottery, the set of *possible* outcomes is finite. This set is called the *support* and denoted $\operatorname{supp}(\tilde{x})$ for a random object or act $\tilde{x}$, or $\operatorname{supp}(P)$ for a distribution P.

It is common in applications of the model for outcomes to be amounts of money or of a good. The set X is then a subset of the real line and an act is a *random variable*, which is the name in statistics for a real-valued random object. Because we can take the weighted average of real numbers, we can define the expected value of a random variable $\tilde{x}\colon S \to X$ to be

$$E[\tilde{x}] = \sum_{s \in S} \pi(s)\tilde{x}(s).$$

The expected value of $\tilde{x}$ depends only on its distribution P; we can define the expected value of such a distribution by $E[P] = \sum_{x \in \operatorname{supp}(P)} P(x)x$.

If we list the states as $S = \{s_1, \ldots, s_n\}$, then we can represent a random variable $\tilde{x}$ by the vector $\langle x_1, \ldots, x_n \rangle$, where $x_s = \tilde{x}(s)$ for each state s. Just as we can add vectors and multiply them by scalars, so we can perform such operations on random variables. If $\tilde{y}$ and $\tilde{z}$ are random variables and α and β are numbers, then $\alpha\tilde{y} + \beta\tilde{z}$ is the random variable $\tilde{x}$ defined by[1]

$$\tilde{x}(s) = \alpha\tilde{y}(s) + \beta\tilde{z}(s).$$

Using the vector representation, we have

$$\langle x_1, \ldots, x_n \rangle = \langle \alpha y_1 + \beta z_1, \ldots, \alpha y_n + \beta z_n \rangle.$$

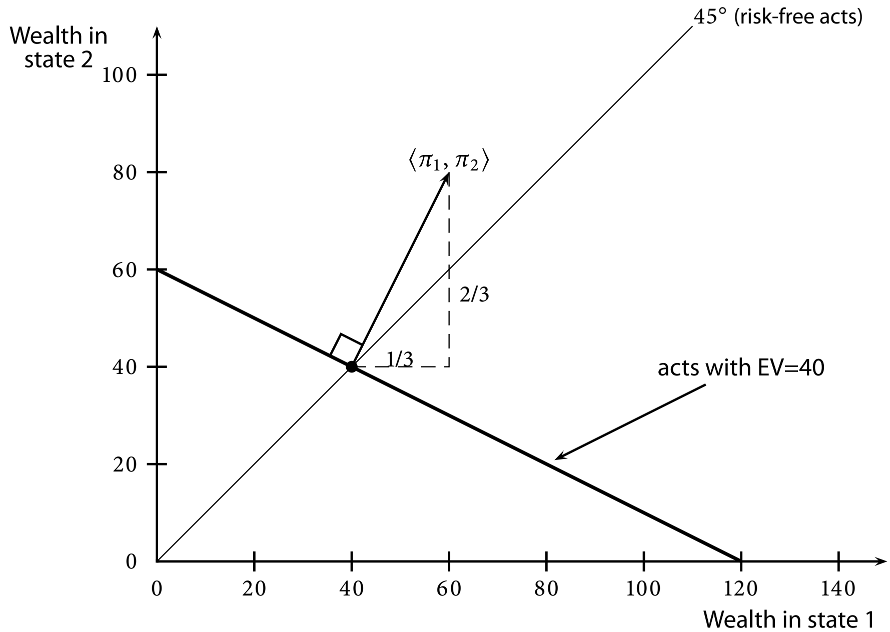
When there are only two states, 1 and 2, each act is given by a pair $\langle x_1, x_2 \rangle$ of numbers, where x_s is the monetary outcomes in state s. It is then possible to draw the set of acts on the plane. For example, suppose you are considering how much to invest in Zenith and the states are

1 = "Zenith wins HDTV contract",

2 = "Zenith does not win HDTV contract",

1. See the Appendix for the distinction between this operation and the mixing of probability measures.

Figure 3.1



Points on the plane are monetary outcomes of acts with two states of the world. Probability of state 1 is 1/3 and of state 2 is 2/3. 45° line is the set of risk-free acts. The line through $\langle 40, 40\rangle$, perpendicular to $\langle 1/3, 2/3\rangle$ is the set of acts that have expected value of 40. The scale of the vector $\langle \pi_1, \pi_2\rangle$ has been exaggerated to make the picture clearer; only the orientation of this vector matters.

which occur with probabilities π_1 and π_2, respectively. Let $X = [0, \infty)$, let x_1 be your net wealth in state 1, and let x_2 be your net wealth in state 2. An act is then given by a point $\langle x_1, x_2\rangle$ in the positive quadrant. This is like the consumption set in standard consumer theory with two goods, except that instead of having goods like quiche and beer, we have money in state 1 and money in state 2. The risk-free acts—those that have the same wealth in states 1 and 2—are the diagonal (the 45° line).

Fix a $\bar{x} \in [0, \infty)$ and consider the set of acts $\langle x_1, x_2\rangle$ whose expected value equals $\bar{x}$, i.e., for which

$$\pi_1 x_1 + \pi_2 x_2 \;=\; \bar{x}. \tag{3.1}$$

Equation (3.1) resembles the budget constraint in standard consumer theory when the price of good 1 is π_1, the price of good 2 is π_2, and income is $\bar{x}$. In vector notation, Equation (3.1) can be written

$$\langle \pi_1, \pi_2\rangle\langle x_1, x_2\rangle \;=\; \bar{x}.$$

Thus, the set of acts whose expected value is $\bar{x}$ is the line through $\langle \bar{x}, \bar{x}\rangle$ that is perpendicular to the vector $\langle \pi_1, \pi_2\rangle$. Figure 3.1 shows the set of acts whose expected value is \$40 when the probability of state 1 is 1/3 and the probability of state 2 is 2/3.

3.1.2 Attitudes toward risk

Assume, until stated otherwise, that X is an interval of the real line representing monetary outcomes and that preferences are state independent.

People use the term "risk" informally to mean the *possibility* of a *bad* outcome. Suppose I am climbing a mountain and the outcomes are "live" or "die." Sure, I prefer to live for sure than to face a chance of death, and in that sense I am "averse" to the "risk" of death. By the same token, I prefer to win a \$1M lottery for sure than to face a chance of not winning the lottery, and so I am "averse" to the "risk" of losing the lottery.

However, if X is an interval of numbers, so that the random value of an act is well-defined, we can ask a more subtle question. When faced with a random prospect P, would the decision maker prefer to avoid uncertainty by getting instead the expected value of the lottery for sure (i.e., getting instead the "lottery" that puts probability 1 on $\mathrm{EV}(P)$)? We classify the answers to this question as follows. The decision maker is said to be

$$\left\{\begin{array}{rll} \textit{risk averse} & \text{if} & \mathrm{EV}(P) \succ P \\ \textit{risk neutral} & \text{if} & \mathrm{EV}(P) \sim P \\ \textit{risk loving} & \text{if} & \mathrm{EV}(P) \prec P \end{array}\right\}$$

for every risky (non-deterministic) lottery P.

These definitions presume that preferences over acts reduce to preferences over lotteries, which would not be true if preferences were state-dependent. Two acts $\tilde{x}_1$ and $\tilde{x}_2$ that have the same expected value $\bar{x}$ might have the ranking

$$\tilde{x}_1 \succ \bar{x} \succ \tilde{x}_2$$

if $\tilde{x}_1$ has higher values in the states where money is most desired (like the net reimbursements from an insurance policy) and $\tilde{x}_2$ has higher values in the states where money is least desired.

We will usually assume that decision makers are risk averse because this is what we commonly observe. Hence, this is a *descriptive*, rather than normative, assumption. There is nothing dumb about not being risk averse; we just do not see it very-loving behavior very often, except in gambling situations.

3.1.3 Risk aversion and concavity of the utility function

Let $u\colon X \to \mathbb{R}$ be the decision maker's VNM utility function. By definition, the decision maker is risk averse if, for every lottery with outcomes $x_1, \ldots, x_n$ and probabilities $\alpha_1, \ldots, \alpha_n$, respectively,

$$\underbrace{u\left(\sum_{i=1}^{n} \alpha_i x_i\right)}_{\text{Utility of expected value}} > \underbrace{\sum_{i=1}^{n} \alpha_i u(x_i)}_{\text{Expected value of utility}} \tag{3.2}$$

If there are just two outcomes, with probabilities α and $1-\alpha$, respectively, this condition is

$$u(\alpha x_1 + (1-\alpha)x_2) > \alpha u(x_1) + (1-\alpha)u(x_2). \tag{3.3}$$

When equation (3.3) holds for all distinct $x_1, x_2 \in X$ and all $\alpha \in (0,1)$, a function $u\colon X \to \mathbb{R}$ is said to be *strictly concave*. One can show that equation (3.2) holds for all lotteries if and only if equation (3.3) holds for all lotteries with two outcomes. We therefore have Proposition 1.

Proposition 1. *A VNM expected utility maximizer is risk averse (resp., risk neutral, risk loving) if and only if her VNM utility function is strictly concave (resp., linear, strictly convex).*

A generalization of risk neutrality and risk aversion is weak risk aversion. A decision maker is *weakly risk averse* if $\mathrm{EV}(P) \succsim P$ for all lotteries P. For weak risk aversion, the strict inequalities in equations (3.2) and (3.3) become weak inequalities, and we have the definition of a (*weakly*) *concave* function. Hence, a decision maker is weakly risk averse if and only if her VNM utility function is concave.

A characterization of a concave function is that the area *under* the graph of the function is a convex set.[2] A function that is concave but not strictly concave has "flat" regions in the graph, or could even be linear. If $u\colon X \to \mathbb{R}$ is differentiable, then u is concave if and only if $u''(x) \le 0$ for all $x \in X$. If $u''(x) < 0$ for all $x \in X$, then u is strictly concave. For a differentiable utility function u, we will take $u'' < 0$ to be the definition of strict concavity, even though it is possible for a differentiable function to satisfy equation (3.3) and have $u''(x) = 0$ for some isolated $x \in X$.

Figure 3.2 shows the strictly concave utility function

$$u(x) = (49 + x)^{1/2},$$

for which $X = [-49, \infty)$. Observe that the shaded area below the graph is convex and that

$$u''(x) = -\frac{1}{4}(49 + x)^{-3/2} < 0.$$

Let $\tilde{x}$ be an act whose possible outcomes $x_1 = -40$ and $x_2 = 51$ occur with probabilities $\alpha_1 = 4/7$ and $\alpha_2 = 3/7$, respectively. The expected value of this lottery is

$$E[\tilde{x}] = \alpha_1 x_1 + \alpha_2 x_2 = \frac{4}{7}(-40) + \frac{3}{7}(51) = -1.$$

The expected utility of this lottery for $u(x) = (49 + x)^{1/2}$ is

$$E[u(\tilde{x})] = \alpha_1 u(x_1) + \alpha_2 u(x_2) = \frac{4}{7}(3) + \frac{3}{7}(10) = 6.$$

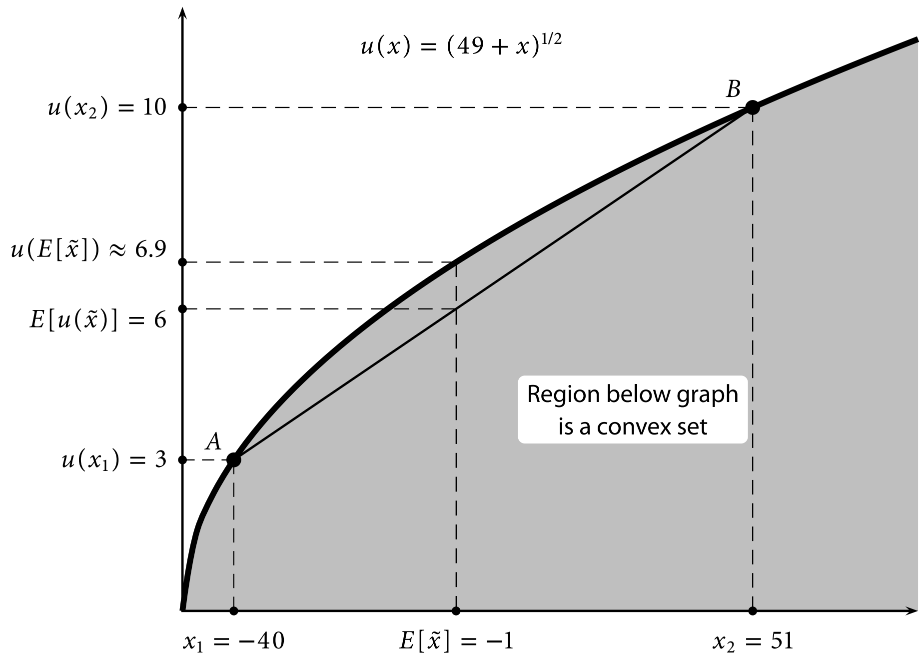
Figure 3.2 shows the points

$$\begin{aligned} A &= \langle x_1, u(x_1)\rangle = \langle -40, 3\rangle \\ B &= \langle x_2, u(x_2)\rangle = \langle 51, 10\rangle \end{aligned}$$

on the graph of u, and a line connecting these two points. If we move along this line $\alpha_1 = 4/7$ of the way from A to B, we reach the point C whose x coordinate is the

2. For points $\{x_1, \ldots, x_n\} \subset \mathbb{R}^n$ and positive numbers $\lambda_1, \ldots, \lambda_n$ that sum to 1, $\sum_{i=1}^n \lambda_i x_i$ is said to be a *convex combination* of $\{x_1, \ldots, x_n\}$. A set $X \subset \mathbb{R}^n$ is said to be convex if it contains all convex combinations of points in X. It is enough to check convex combinations for any pair of points. That is, a set is convex if and only if it contains all convex combinations of every *two* points in X (i.e., the line connecting every two points in X). For example, a sphere and a disk are convex, but a torus and a circle are not.

Figure 3.2



A concave utility function $u(x) = (49 + x)^{1/2}$, and an act that pays $-\$40$ and $\$51$ with probabilities 4/7 and 3/7, respectively.

expected value $\alpha_1 x_1 + \alpha_2 x_2 = -1$ of the act, and whose y coordinate is the expected utility $\alpha_1 u(x_1) + \alpha_2 u(x_2) = 6$ of the act. Since the line lies *below* the graph of the function, you can see that $u(E[\tilde{x}]) > E[u(\tilde{x})]$.

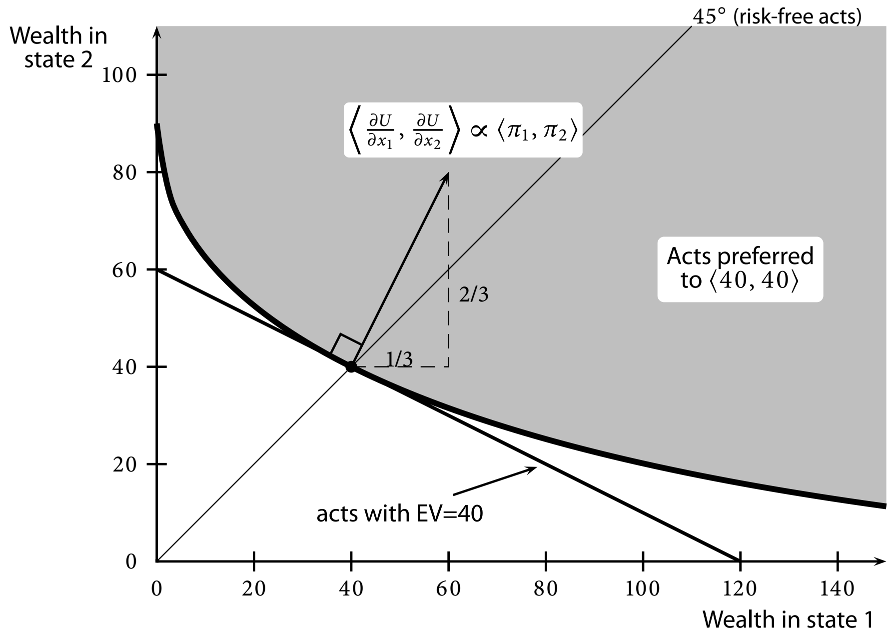
Figure 3.3 gives a different graphical view of risk aversion, drawn on the set of acts when there are two states. The graph shows the set of acts whose expected value is 40, as in Figure 3.1. It also shows the set of acts that are weakly preferred to the risk-free act $\langle 40, 40 \rangle$. Here are some properties of this set:

- If the decision maker is risk averse, then $\langle 40, 40 \rangle$ is strictly preferred to all the risky acts whose expected value is 40. Hence, the weakly-preferred set only intersects the $\text{EV} = 40$ line at the 45° line.
- Assuming that more money is preferred to less, the weakly-preferred set must lie above the $\text{EV} = 40$ line.
- One can show that the weakly-preferred set is convex.

Hence, the weakly-preferred set must look like the shaded region in Figure 3.3. The boundary of this set is the indifference curve through $\langle 40, 40 \rangle$, and it is seen to be tangent to the $\text{EV} = 40$ line at the point $\langle 40, 40 \rangle$.

We can also derive this tangency property using calculus, assuming that the VNM utility function u is differentiable. Recall that the budget line in standard consumer theory is perpendicular to the price vector. Therefore, the condition that the indifference curve through an optimal consumption bundle x be tangent to the budget line means that the indifference curve and the budget line are perpendicular to the same vector at x; hence, if $U \colon \mathbb{R}^n_+ \to \mathbb{R}$ is the consumer's utility function and $\langle p_1, \ldots, p_n \rangle$ is

Figure 3.3



Probability of state 1 is $\pi_1 = 1/3$ and of state 2 is $\pi_2 = 2/3$. Shaded region is set of acts that are preferred to the risk-free act $\langle 40, 40 \rangle$.

the price vector, then

$$\left\langle \frac{\partial U}{\partial x_1}(x), \ldots, \frac{\partial U}{\partial x_n}(x) \right\rangle \propto \langle p_1, \ldots, p_n \rangle.$$

Let's apply this property to the utility function

$$U(x_1, x_2) = \pi_1 u(x_1) + \pi_2 u(x_2) \tag{3.4}$$

on the set of acts. The gradient at a point $\langle x_1, x_2 \rangle$ is

$$\left\langle \frac{\partial U}{\partial x_1}(x_1, x_2), \frac{\partial U}{\partial x_2}(x_1, x_2) \right\rangle = \langle \pi_1 u'(x_1), \pi_2 u'(x_2) \rangle.$$

Along the 45° line (or the diagonal, in higher-dimensional cases), $x_1 = x_2 = \bar{x}$ for some $\bar{x} \in \mathbb{R}$. Hence $u'(x_1) = u'(x_2) = u'(\bar{x})$. This implies that

$$\left\langle \frac{\partial U}{\partial x_1}(\bar{x}, \bar{x}), \frac{\partial U}{\partial x_2}(\bar{x}, \bar{x}) \right\rangle = u(\bar{x})\langle \pi_1, \pi_2 \rangle \propto \langle \pi_1, \pi_2 \rangle.$$

That is, the gradient points in the same direction as the vector of probabilities. As discussed in Section 3.1.1, a set of points with the same expected value is a line perpendicular to $\langle \pi_1, \pi_2 \rangle$. Hence, at $\langle 40, 40 \rangle$, both the $EV = 40$ line and the indifference curve through $\langle 40, 40 \rangle$ are perpendicular to $\langle \pi_1, \pi_2 \rangle$, which implies that the $EV = 40$ line and the indifference curve are tangent.

Exercise 3.1. A risk-averse decision maker has initial wealth of \$10,000 and needs to leave \$5,000 in the hotel safe. Not being the nicest hotel, there is still a 1/4 chance of theft. The hotel will only insure against theft if the decision maker buys insurance. The hotel will sell the decision maker insurance at the rate of \$.25 per dollar of coverage.

a. On a graph showing wealth in each state, mark the act (state-dependent wealth) the DM faces if he buys no insurance and mark the act he faces if he buys full insurance (\$5,000 of coverage).

b. Draw all the acts the DM can choose from, by varying the amount of coverage (including "negative" coverage and excess coverage, but without letting wealth in either state be negative).

c. How much coverage should the DM buy? Give the most direct explanation you can.

3.1.4 Certainty equivalents and risk premia

Suppose that a decision maker prefers more money to less (preferences over pure outcomes are strictly monotone). Then, for each lottery P, there is a unique amount of money $\mathrm{CE}(P)$ such that the decision maker is indifferent between P and getting $\mathrm{CE}(P)$ for sure. $\mathrm{CE}(P)$ is called the *certainty equivalent* of P. If the decision maker is risk neutral, the certainty equivalent is equal to the expected value of the lottery. However, if the decision maker is risk averse and P is risky, then these are not equal and the expected value is strictly preferred to the certainty equivalent. Since we are assuming that preferences are strictly monotone, the expected value must be greater than the certainty equivalent. The difference,

$$\mathrm{RP}(P) \;=\; \mathrm{EV}(P) - \mathrm{CE}(P),$$

is called the *risk premium* of lottery P. Since $\mathrm{CE}(P)$ is the smallest amount of money the decision maker would accept in exchange for the risky lottery P, the risk premium is the maximum amount of money the decision maker would be willing to give up in *expected value* in order to avoid the riskiness of P.

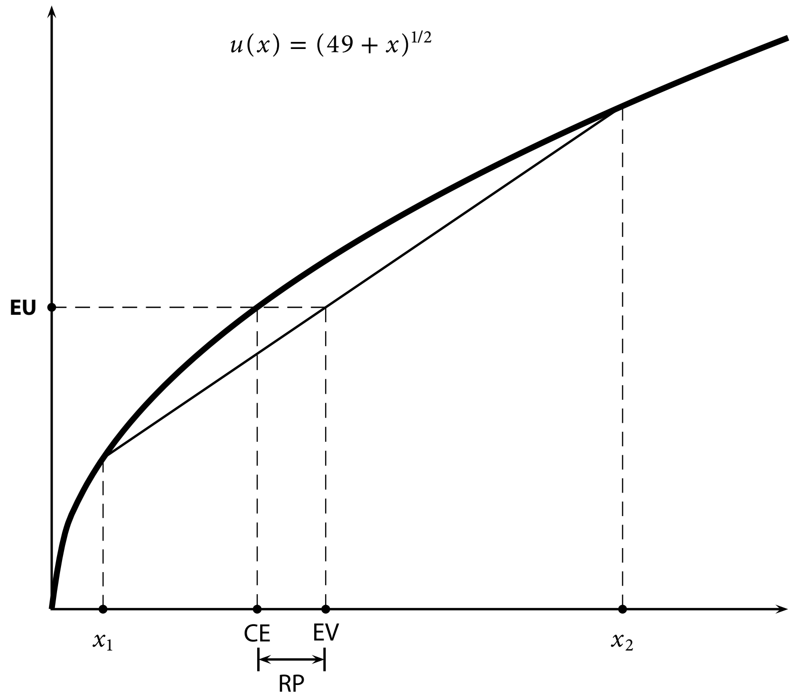
To calculate the certainty equivalent of a lottery, we just find the expected utility, and then take the inverse of the expected utility to find the sure amount that gives the same utility. Recall the example, from Section 3.1.3, with $u(x) = (49 + x)^{1/2}$ and the act $\tilde{x}$ that yields -40 and 51 with probabilities 4/7 and 3/7, respectively. We calculated that the expected utility of the act is 6. Hence, the certainty equivalent is

$$\mathrm{CE}(\tilde{x}) \;=\; u^{-1}(E[u(\tilde{x})] \;=\; u^{-1}(6) \;=\; -13.$$

We also calculated that the expected value of $\tilde{x}$ is -1. Hence, the risk premium is 12. The certainty equivalent and risk premium are shown in Figure 3.4.

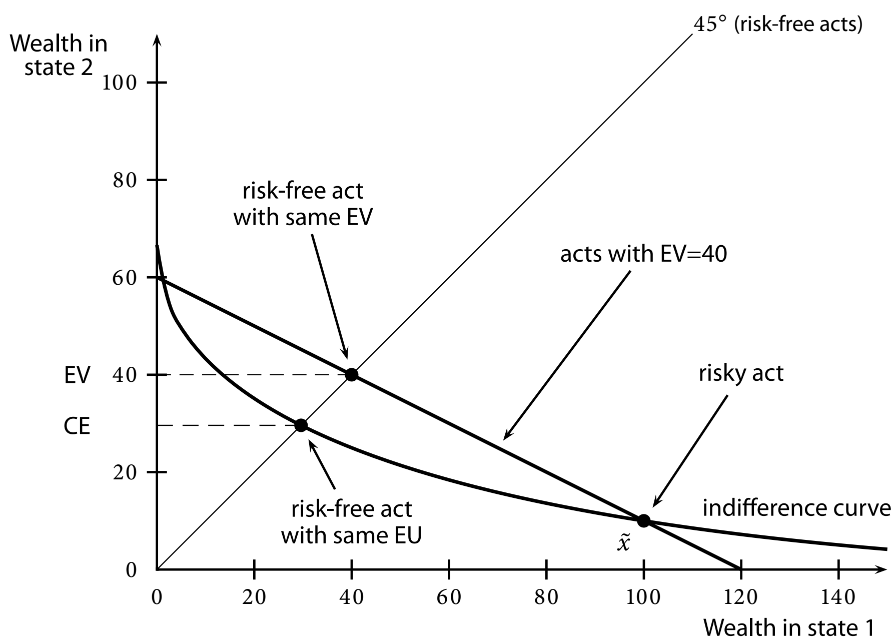
Figure 3.5 illustrates the certainty equivalent and risk premium of an act when there are two states. It shows an indifference curve through a risky act. The certainty equivalent is the intersection of the indifference curve with the 45° line.

Figure 3.4



The expected utility (EU), expected value (EV), certainty equivalent (CE), and risk premium (RP) of the lottery shown in Figure 3.4.

Figure 3.5



Probability of state 1 is 1/3 and of state 2 is 2/3. Graphs shows a risky act, the indifference curve through the act, the risk-free act with the same utility (intersection of indifference curve and $45°$ line), and the risk-free act that has same expected value.

Exercise 3.2. Suppose that a person maximizes his expected utility, with the utility function given by $u(z) = z^{1/2}$. Suppose that the person engages in a risky venture which leaves him with either \$81 or \$25, with equal probability. What is the certainty equivalent of this business venture? What is the risk premium?

Exercise 3.3. Suppose that Mao is an expected utility maximizer, with the VNM utility function $u(x) = \log(x)$, for $x > 0$ (use natural log).

a. What is Mao's certainty equivalent of the following lottery:

Probability	Money
.4	30
.5	100
.1	500

b. What is the risk premium Mao is willing to pay to insure against this uncertain prospect?

c. Suppose Mao has \$1,200,000 in wealth, and decides to become a backsliding oil prospector. He finds a tract of land for sale for \$1,000,000 dollars, which will produce no return at all if no oil is found, or will yield \$10,000,000 of income (net of operating expenses but not of the cost of the land) if oil is found. Let p be the probability that oil is found. Specify the two lotteries that result from the actions, "buy that land" and "not buy the land". What probability p of finding oil would make Mao exactly indifferent between buying the land and not buying the land?

Exercise 3.4. A person has VNM utility function $u(z) = \log_{10} z$. She has initial wealth of \$10,000 and has become a finalist in a lottery such that her ticket will pay off \$990,000 with probability 1/2, and \$0 with probability 1/2. What is the minimum amount of money she would be willing to receive in exchange for the ticket? Show your calculations.

3.2 Comparing the risk aversion of people

3.2.1 Interpersonal comparisons of risk aversion

Let Ingrid and Hobbes be two weakly risk-averse expected utility maximizers with strictly increasing utility. We say that Ingrid is *as risk averse as* Hobbes if, for every act, Ingrid's risk premium is as large as Hobbes' risk premium. If Ingrid's risk premium is larger than Hobbes' whenever the act is risky, then Ingrid is said to be *more risk averse than* Hobbes.

You can see that Ingrid is as risk averse as Hobbes if and only if, when comparing a risky act $\tilde{x}$ to a risk-free act $\tilde{y}$, if Ingrid (weakly) prefers $\tilde{x}$ to $\tilde{y}$, then so does Hobbes.

Note that "as risk averse as" is an incomplete ordering of decision makers. For some lotteries, Ingrid's risk premium may be higher than Hobbes', while the opposite may be true for other lotteries; then we can say neither that Ingrid is as risk averse as Hobbes nor that Hobbes is as risk averse as Ingrid.

When the VNM utility functions u_i and u_h of Ingrid and Hobbes are differentiable, there is a simple way to check whether Ingrid is as risk averse as Hobbes. Let

$$R_A(x) = -\frac{u''(x)}{u'(x)}.$$

This is called the *Arrow-Pratt measure of absolute risk aversion*. Ingrid is as risk averse as (more risk averse than) Hobbes if and only if Ingrid's R_A is as large as (greater than) Hobbes' for all x.[3]

For example, suppose $u_i = x^\alpha$ and $u_h = x^\beta$, where $0 < \alpha < \beta < 1$. Then

$$-\frac{u_i''(x)}{u_i'(x)} = \frac{1-\alpha}{x} \qquad \text{and} \qquad -\frac{u_h''(x)}{u_h'(x)} = \frac{1-\beta}{x}.$$

Since $\alpha < \beta$,

$$-\frac{u_i''(x)}{u_i'(x)} > -\frac{u_h''(x)}{u_h'(x)}$$

for all x. Hence, Ingrid is more risk averse than Hobbes.

If we are interested in who has a greater risk premium for a lottery with just two outcomes x_1 and x_2, then it is not enough to compare R_A for just x_1 and x_2. We have to check the condition for all x between x_1 and x_2 as well.

When comparing the risk aversion of two people, we might define outcomes to be final wealth. However, it is more relevant to define outcomes to be net transactions. Then we can compare the risk premia of an investment for two people who have different levels of baseline wealth.

3.2.2 Intrapersonal comparisons of absolute risk aversion

When outcomes are net transactions, we can also ask how a single person's risk aversion changes with her baseline wealth. That is, how does the risk premium of an act $w + \tilde{x}$, with outcomes defined as final wealth, vary with w? If the risk premium *decreases* (resp., *remains constant, increases*) as w increases for every risky $\tilde{x}$, then the DM is said to exhibit *decreasing* (resp., *constant, increasing*) *absolute risk aversion*.

A DM with VNM utility u has decreasing (resp., constant, increasing) absolute risk aversion if and only if $R_A(x)$ is decreasing (resp., constant, increasing) as a function of x.

3. You might have guessed that comparing the second derivatives of the utility functions would work; $u'' < 0$ is a characterization of risk aversion, and so perhaps Ingrid is as risk averse as Hobbes if $u_i''(x) \leq u_h''(x)$ for all x. However, for any constant $\alpha > 0$, the VNM utility function αu_i also represents Ingrid's risk preferences, but $\alpha u_i'' \neq u_i''$ if $\alpha \neq 1$.

Not all risk preferences can be classified this way. There may be two gambles $\tilde{x}_1$ and $\tilde{x}_2$ and four baseline levels of wealth $w_{11} < w_{12}$ and $w_{21} < w_{22}$ such that

$$\begin{aligned} RP(W_{11} + \tilde{x}_1) &< RP(W_{12} + \tilde{x}_1), \text{ and} \\ RP(W_{21} + \tilde{x}_2) &> RP(W_{22} + \tilde{x}_1). \end{aligned}$$

As an exercise, show that this is the case for the following piecewise-linear concave utility function:

$$u(x) = \begin{cases} 2x & x \leq 1000, \\ x + 1000 & x \geq 1000. \end{cases}$$

It is widely observed that people exhibit decreasing absolute risk aversion. If we take this as given and also assume that everyone has the same VNM utility function over final wealth, then wealthier people are less risk averse than poorer people with respect to their net transactions. This link between wealth and risk aversion is also observed empirically—not as a rule, but as a general pattern.

In spite of this empirical evidence, economists often make the simplifying assumption that people have constant absolute risk aversion (CARA). All utility functions with this property are of the form

$$u(x) = -e^{-\lambda x}$$

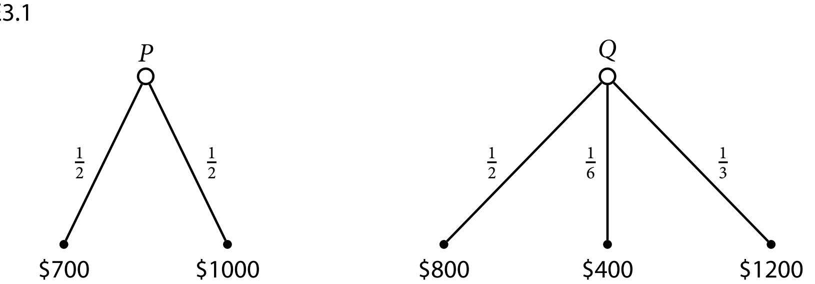
(negative exponential), where λ is a positive constant that is equal to $R_A(x)$ for all x and is called the *coefficient of absolute risk aversion*. Such a utility function is called a CARA utility function. For empirical work, if one restricts attention to CARA utility functions then there is only one parameter to be estimated. It may be a reasonable approximation when modeling investment decisions, even if a rich investor would have much different risk preferences if he were poor, as long as the outcomes of the investments are not likely to make the investor poor.

Exercise 3.5. A risk-averse VNM decision maker has *decreasing* absolute risk aversion. Her certainty equivalent for a lottery that pays \$0 and \$800 with probabilities 1/3 and 2/3, respectively, is \$500.

a. Which does she prefer, to get \$400 and \$1200 with probabilities 1/3 and 2/3, respectively, or to get \$900 for sure? (Explain.)

b. How does she rank the lotteries in Figure E3.1? (Explain.)

Figure E3.1



Exercise 3.6. Kreps, *A Course ...*, p. 131, Problem 15.

Exercise 3.7. Consider a person who has the following piecewise-linear utility function:

$$u(z) = \begin{cases} 2z & z < \$1,000 \\ 1000 + z & z \geq \$1,000. \end{cases}$$

Graph this utility function. Does the person have increasing or decreasing absolute risk aversion over the domain of her utility function (e.g., 0 to $2,000)? Explain. Do not try to apply a mechanical criterion, such as the measures of risk aversion that use differentiation. Instead, directly apply the definition of increasing and decreasing absolute risk aversion.

3.2.3 Intrapersonal comparisons of relative risk aversion

[Incomplete]

Exercise 3.8. Consider a person who has *decreasing absolute* risk aversion and *constant relative* risk aversion. Let $\tilde{x}$ be the risky net profit of a particular business venture, and let w be the person's initial wealth, so that the person's total wealth given the venture is $w + \tilde{x}$. Suppose that if $w = 200$, then the person's risk premium for the venture $\tilde{x}$ is $10.

a. Can you say whether the risk premium if the initial wealth were $300 would be greater than, less than or equal to $10?

b. What is the risk premium for a risky venture $3\tilde{z}$ if the initial wealth is 600? (Explain)

3.3 Comparing the riskiness of acts

3.3.1 Dominance for risk-averse decision makers

If you know a VNM utility function and the distributions of two acts $\tilde{x}$ and $\tilde{y}$, then it is a simple matter to calculate the expected utility of each act and choose the one with the higher expected utility. However, it is often useful to be able to determine the ranking of two acts even when you have only partial information about the utility function. This is the purpose of the various *dominance* criteria we have studied and will study. For example, suppose you have been able to check a mathematical condition that tells you that every risk-averse decision maker prefers act $\tilde{x}$ to act $\tilde{y}$. This helps you as follows.

1. If you are an economist or are otherwise interested in predicting the choices made

by decision makers, you can predict a risk-averse individual's choice between $\tilde{x}$ and $\tilde{y}$, without knowing the individual's exact preferences over risky prospects.
2. If you have been delegated to make a decision for your risk-averse boss, then you can eliminate $\tilde{y}$ from consideration even if you do not know exactly what your boss's risk preferences are.
3. If you are making a decision for yourself, but are not sure even of your own risk preferences, you can at least eliminate from consideration $\tilde{y}$.[4]

Similarly, when choices are ranked by statewise dominance (Section 1.2.3), you can predict choices or make delegated choices, knowing only that that the interested party's utility is strictly increasing. When choices are ranked by first-order stochastic dominance (Section 1.3.6), you can predict choices or make delegated choices knowing only the interested party's beliefs and that utility is strictly increasing.

Here are the dominance criteria that presume that a decision maker is risk averse:

Definition 1. Let $\tilde{x}$ and $\tilde{y}$ be two acts with known distributions.

- $\tilde{x}$ is *less risky* than $\tilde{y}$ if every risk-averse decision maker prefers $\tilde{x}$ over $\tilde{y}$.
- $\tilde{x}$ *second-order stochastically dominates* $\tilde{y}$ if every risk-averse decision maker with strictly increasing utility prefers $\tilde{x}$ over $\tilde{y}$.

The first criteria, called *decreasing risk*, is weaker than second-order stochastic dominance (s.o.s.d.) because it is not assumed that the decision maker has increasing utility. The reason I focus on decreasing risk in this part, rather than simply s.o.s.d., is not because it bothers me to assume that utility is increasing. Instead, it is because I can then focus on what risk aversion tells us, independently of monotonicity.

3.3.2 Variance as a measure of risk

The criteria of decreasing risk and second-order stochastic dominance are not very useful unless we have some mathematical characterizations that allow us to identify when one act dominates another.

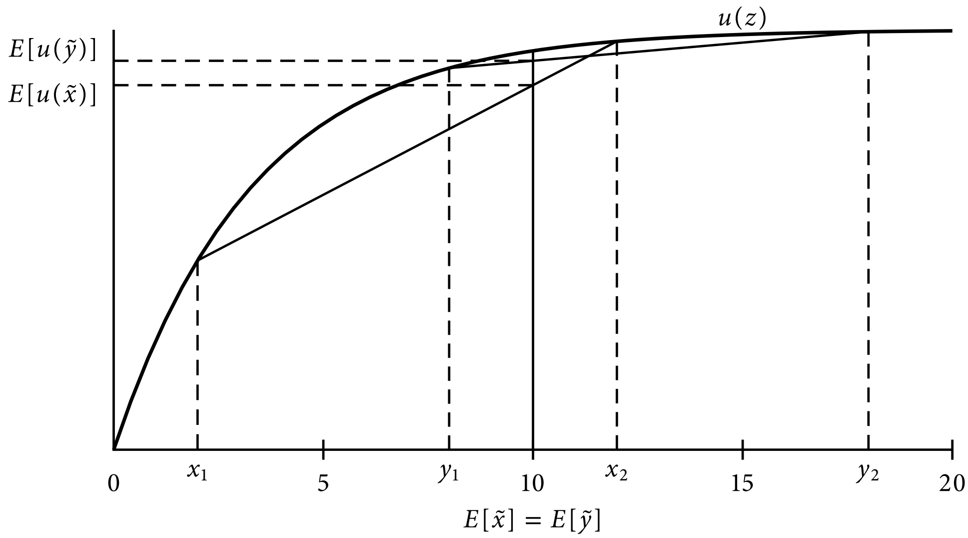
Consider decreasing risk. Since this criteria does not presume that utility is increasing, a necessary condition for $\tilde{x}$ to be less risky than $\tilde{y}$ is that $E[\tilde{x}] = E[\tilde{y}]$. If $E[\tilde{x}] > E[\tilde{y}]$, for example, there will always be a decision maker with decreasing utility and sufficiently low risk aversion that she prefers $\tilde{y}$ to $\tilde{x}$.

Let $\mathrm{Var}(\tilde{x})$ be the variance of an act $\tilde{x}$. Given two acts $\tilde{x}$ and $\tilde{y}$ with the same mean, you might conjecture that $\tilde{x}$ is less risky than $\tilde{y}$ if and only if $\mathrm{Var}(\tilde{x}) < \mathrm{Var}(\tilde{y})$. After all, informally speaking, variance is a measure of the variability of a random variable, which in turn is closely related to risk, Furthermore, if $\tilde{x}$ is risk free and $\tilde{y}$ is not (so that $\tilde{x}$ is surely less risky than $\tilde{y}$), then $\mathrm{Var}(\tilde{x}) = 0 < \mathrm{Var}(\tilde{y})$, .

It would be very nice if this conjecture were correct. Variance is easy to calculate and we could unequivocally rank every pair of acts that have the same mean. Unfortunately, variance does not work as a measure of decreasing risk.

4. Note how I have inverted expected utility theory. Expected utility was introduced as a *representation* of rational preferences over lotteries or acts (normative theory). Now it is a *method* for determining preferences (prescriptive theory).

Figure 3.6



An example of a utility function and two acts $\tilde{x}$ and $\tilde{y}$ that have the same mean and the same variance such that $\tilde{y}$ gives higher expected utility than $\tilde{x}$.

To demonstrate this, it suffices to find two acts $\tilde{x}$ and $\tilde{y}$ with the same mean and variance and one risk-averse decision maker who strictly prefers $\tilde{x}$ to $\tilde{y}$. Figure 3.6 shows such an example. The acts $\tilde{x}$ and $\tilde{y}$ have the following distributions:

x	Prob(x)
2	1/5
12	4/5

y	Prob(y)
8	4/5
18	1/5

The mean of both acts is 10 and the variance of both acts is 16:

$$\mathrm{Var}(\tilde{x}) = \tfrac{1}{5}(10-2)^2 + \tfrac{4}{5}(10-12)^2 = 16,$$
$$\mathrm{Var}(\tilde{y}) = \tfrac{4}{5}(10-8)^2 + \tfrac{1}{5}(10-18)^2 = 16.$$

Yet we can see in the figure that $E[u(\tilde{x})] < E[u(\tilde{y})]$. (I could also find a utility function such that $E[u(\tilde{x})] > E[u(\tilde{y})]$.)

The acts $\tilde{x}$ and $\tilde{y}$ violate an easy-to-check *necessary* condition for $\tilde{x}$ to be less risky than $\tilde{y}$. The support of $\tilde{x}$ must be *nested* in the support of $\tilde{y}$. This means that (a) the lowest possible value of $\tilde{x}$ is as high as the lowest possible value of $\tilde{y}$ and (b) the highest possible value of $\tilde{x}$ is as low as the highest possible value of $\tilde{y}$. In this example, the support of $\tilde{x}$ is not nested in the support of $\tilde{x}$ because $x_1 < y_1$; the support of $\tilde{y}$ is not nested in the support of $\tilde{x}$ because $y_2 > x_2$.

Variance is not completely useless for checking decreasing risk because another *necessary* condition for $\tilde{x}$ to be less risky than $\tilde{y}$ is that $\mathrm{Var}(\tilde{x}) < \mathrm{Var}(\tilde{y})$. Hence, if we find that $\mathrm{Var}(\tilde{y}) < \mathrm{Var}(\tilde{x})$, we cannot be sure that $\tilde{y}$ is less risky than $\tilde{x}$, but we do know that $\tilde{x}$ is *not* less risky than $\tilde{y}$.

To see that this is true, we just have to find one utility function such that, for acts with the same mean, the one with the lowest variance has the highest expected utility.

The quadratic utility function

$$u(x) = -x^2 + 2ax$$

has this property. Observe that

$$E[u(\tilde{x})] = -E[\tilde{x}^2] + 2aE[\tilde{x}].$$

Since $E[\tilde{x}^2] = \mathrm{Var}(\tilde{x}) + E[\tilde{x}]^2$,

$$E[u(\tilde{x})] = 2aE[\tilde{x}] - E[\tilde{x}]^2 - \mathrm{Var}(\tilde{x}).$$

Hence, $E[u(\tilde{x})] > E[u(\tilde{y})]$ if $E[\tilde{x}] = E[\tilde{y}]$ and $\mathrm{Var}(\tilde{x}) < \mathrm{Var}(\tilde{y})$.

In summary, although variance is not a sufficient measure of decreasing risk, we have found three necessary conditions for $\tilde{x}$ to be less risky than $\tilde{y}$:

1. $E[\tilde{x}] = E[\tilde{y}]$;
2. support of $\tilde{x}$ is nested in the support of $\tilde{y}$;
3. $\mathrm{Var}(\tilde{x}) < \mathrm{Var}(\tilde{y})$.

The analogous necessary conditions for $\tilde{x}$ to second-order stochastically dominate $\tilde{y}$ are the following:

1. $E[\tilde{x}] \geq E[\tilde{y}]$;
2. the lowest possible value of $\tilde{x}$ is as large as the lowest possible value of $\tilde{y}$;
3. $\mathrm{Var}(\tilde{x}) \leq \mathrm{Var}(\tilde{y})$.

Exercise 3.9. Consider the example in Figure 3.6, which illustrates that variance is not a sufficient measure of risk. In the example, $\tilde{y}$ is preferred to $\tilde{x}$. Draw a similar example with the same acts but a different concave utility function such that $\tilde{x}$ is preferred to $\tilde{y}$.

3.3.3 Characterization of decreasing risk

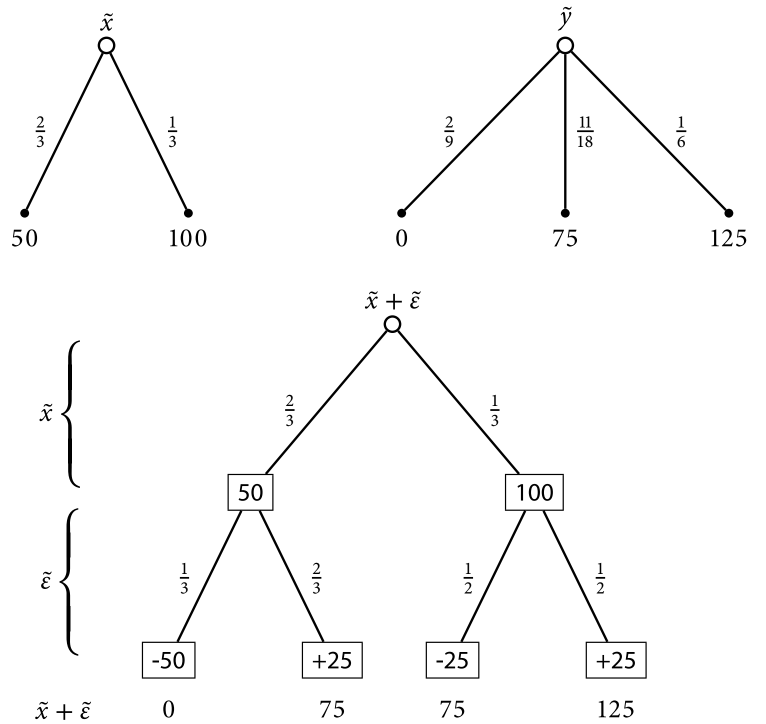
Here is a criterion that does determine whether a random variable is less risky than another:

PROPOSITION 2. *An act $\tilde{x}$ is less risky than an act $\tilde{y}$ if and only if there is a random variable $\tilde{\varepsilon}$ such that:*

1. $E[\tilde{\varepsilon} \mid \tilde{x} = x] = 0$ *for all x in the support of $\tilde{x}$ (i.e., $E[\tilde{\varepsilon} \mid \tilde{x}] = 0$); and*
2. *$\tilde{y}$ and $\tilde{x} + \tilde{\varepsilon}$ have the same distribution (written $\tilde{x} \stackrel{d}{=} \tilde{y} + \tilde{\varepsilon}$).*

Once you understand what these conditions mean, the criterion is intuitive. For example, consider the acts $\tilde{x}$ and $\tilde{y}$ shown at the top of Figure 3.7. It may not be obvious that every risk-averse decision maker prefers $\tilde{x}$ to $\tilde{y}$. But now consider the tree at the bottom of the figure. It shows a two-stage gamble. In the first stage, you receive winnings (or losses) $\tilde{x}$. Instead of stopping, you gamble again and win $\tilde{\varepsilon}$, which is added to $\tilde{x}$. The distribution of $\tilde{\varepsilon}$ depends on the realization of $\tilde{x}$, but for each possible realization x, the second-stage gamble has a conditional expected value of 0. That is,

Figure 3.7



A random-variable characterization of decreasing risk. $\tilde{y}$ has the same distribution as the two-stage gamble $\tilde{x} + \tilde{\varepsilon}$, and $E[\tilde{x}|\tilde{\varepsilon}] = 0$. Hence, $\tilde{x}$ is less risky than $\tilde{y}$.

$E[\tilde{\varepsilon}\,|\,\tilde{x} = x] = 0$. If you are risk averse, you would always prefer to stop after learning $\tilde{x}$. Participating in the second stage subjects you to additional risk, but never increases the expected value of your winnings. Hence, overall you prefer the one-stage gamble $\tilde{x}$ to the two-stage gamble $\tilde{x} + \tilde{\varepsilon}$. You are indifferent between $\tilde{x} + \tilde{\varepsilon}$ and $\tilde{y}$ if these two gambles have the same distribution, which you can verify for this example. Therefore, you also prefer $\tilde{x}$ to $\tilde{y}$.

We will use this "random-variable" characterization of decreasing risk several times in this book because it is sometimes very straightforward to apply in "abstract" settings. As an example, we have the following proposition.

PROPOSITION 3. *If two random variables are normally distributed and have the same mean, then the one with the lower variance is less risky.*

(That is, if we restrict attention to normally distributed acts, variance *is* a measure of risk.)

Proof. We just need the following property of the normal distribution. If $\tilde{x} \sim N(\mu_x, \sigma_x^2)$ (read this "$\tilde{x}$ is normally distributed with mean μ_x and variance σ_x^2") and $\tilde{y} \sim N(\mu_y, \sigma_y^2)$ and if $\tilde{x}$ and $\tilde{y}$ are independent, then

$$\tilde{x} + \tilde{y} \;\sim\; N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2).$$

(Read this: "The sum of two normally distributed random variables is normally distributed; if the random variables are independent, the mean and variance of the sum are equal to the sums of the means and of the variances, respectively.") Let $\tilde{x}$ and $\tilde{y}$ be normally distributed with the same mean μ and with variances σ_x^2 and σ_y^2, respectively. Assume $\sigma_x^2 < \sigma_y^2$. Let $\tilde{\varepsilon}$ be a random variable that is distributed $N(0, \sigma_y^2 - \sigma_x^2)$ and is independent of $\tilde{x}$. Then

$$\tilde{x} + \tilde{\varepsilon} \sim N(\mu_x + 0, \sigma_x^2 + (\sigma_y^2 - \sigma_x^2)) = N(\mu_y, \sigma_y^2).$$

That is, $\tilde{x} + \tilde{\varepsilon}$ has the same distribution as $\tilde{y}$. Furthermore, since $\tilde{\varepsilon}$ is independent of $\tilde{x}$, $E[\tilde{\varepsilon} \mid \tilde{x}] = E[\tilde{\varepsilon}] = 0$. Hence, $\tilde{x}$ is less risky than $\tilde{y}$. □

If you are just given two simple random variables, applying the random-variable condition is a tedious linear programming problem. There is another characterization of decreasing risk that is easier to check, especially if you can program your computer to do the calculations. Let F and G be the *cumulative* distribution functions of acts $\tilde{x}$ and $\tilde{y}$, respectively. Then $\tilde{x}$ is less risky than $\tilde{y}$ if and only if $\tilde{x}$ and $\tilde{y}$ have the same mean and

$$\int^x (F(t) - G(t))\, dt \leq 0$$

for all x, with strict inequality for some x. However, we do not use this characterization in this book; hence, I will not give an explanation or an example.

The characterizations of s.o.s.d. are similar. In the random-variable condition, we replace $E[\tilde{\varepsilon} \mid \tilde{x}] = 0$ by $E[\tilde{\varepsilon} \mid \tilde{x}] \leq 0$. In the integral condition, we drop the requirement that $\tilde{x}$ and $\tilde{y}$ have the same mean.

Exercise 3.10. You are trying to decide how to invest \$5,000. Only money outcomes matter and your preferences over money are state independent. Here are four investment opportunities, together with the possible outcomes:

	"Investment"	*Possible outcomes*
A	Buy \$5,000 of bonds from the Hungarian State Bank	lose \$5000 or win \$1000
B	Bet \$2000 that a presidential candidate will pledge to raise taxes	lose \$2000 or win \$50000
C	Buy \$5000 of euros	lose \$500 or win \$800
D	Send your delinquent son to college	lose \$5000

You know that each outcome can occur with positive probability and you know that the expected payoffs for investments A, B and C are the same (but you don't know what this mean is).

State whether each of the following second-order stochastic dominance relations is true, false or uncertain (given the information you have) and give a brief explanation:

1. A s.o.s.d. B;
2. C s.o.s.d. A;
3. A s.o.s.d. D.

Exercise 3.11. What is wrong with this reasoning? Suppose $\tilde{x}$ and $\tilde{y}$ have the same mean and $\tilde{y}$ is less risky than $\tilde{x}$, so that we can write

$$\tilde{x} \stackrel{d}{=} \tilde{y} + \tilde{\varepsilon}$$

for some $\tilde{\varepsilon}$ such that $E[\tilde{\varepsilon}|\tilde{y}] = 0$. Then we can also write

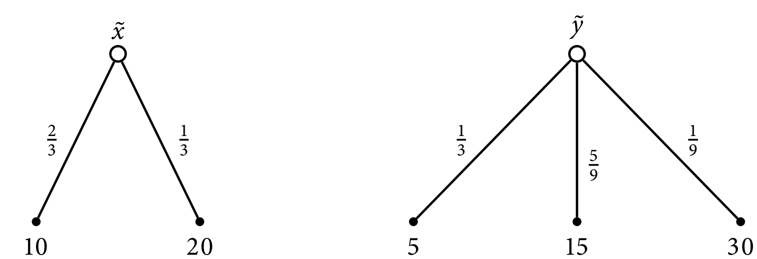
$$\tilde{y} \stackrel{d}{=} \tilde{x} - \tilde{\varepsilon}.$$

Since the expected value of $\tilde{\varepsilon}$ is 0, so is the expected value of $-\tilde{\varepsilon}$. Therefore, y is also risker than x.

Exercise 3.12. A consumer must choose between (A) a sure payment of \$400 and (B) a gamble with prizes \$0, \$100, \$600 and \$1000 with probabilities 0.25, 0.1, 0.4 and 0.25, respectively. All you know is that (i) the consumer satisfies the VNM axioms for this kind of lottery, (ii) she is risk averse, (iii) she prefers more money over less, and (iv) her risk premium for a gamble (C) with prizes \$0 and \$1,000, equally likely, is \$100. Show that the consumer therefore must prefer B over A. (If your answer is getting complicated, you are on the wrong track. The idea of increasing risk is useful.)

Exercise 3.13. Show that the random prospect $\tilde{x}$ in Figure E3.2 is less risky than $\tilde{y}$ by showing that $\tilde{y} \stackrel{d}{=} \tilde{x} + \tilde{\varepsilon}$ with $E[\tilde{\varepsilon}|\tilde{x}] = 0$:

Figure E3.2



Exercise 3.14. Let P be a lottery that pays \$20 with probability 1/3 and \$40 with probability 2/3. Let Q be a lottery that pays \$10 with probability 1/6, \$30 with probability 11/18, and \$60 with probability 2/9. Show that P is less risky than Q by showing that there are random variables $\tilde{x}$, $\tilde{y}$ and $\tilde{\varepsilon}$ such that (i) P is the distribution of $\tilde{x}$, (ii) Q is the distribution of $\tilde{y}$, (iii) $E[\tilde{\varepsilon}|\tilde{x}] = 0$, and (iv) $\tilde{y}$ and $\tilde{x} + \tilde{\varepsilon}$ have the same distribution.

Exercise 3.15. Let $\tilde{w}$ be independent of both $\tilde{x}$ and $\tilde{y}$. $\tilde{x}$ and $\tilde{y}$ are two random prospects that a decision maker with utility u is choosing between. $\tilde{w}$ is the decision maker's other wealth. Assume that u is increasing and concave and assume that $\tilde{x}$ second-order stochastically dominates $\tilde{y}$.

a. Applying the law of iterated expectations, show that

$$E[u(\tilde{w}+\tilde{x})] > E[u(\tilde{w}+\tilde{y})].$$

b. Conclude that $\tilde{w}+\tilde{x}$ second-order stochastically dominates $\tilde{w}+\tilde{y}$.

c. Give an example that illustrates that this is not true if $\tilde{w}$ is not independent of $\tilde{x}$ and $\tilde{y}$.

d. Explain how this result might help you if you are managing a portfolio for someone whose risk preferences you do not fully know (e.g., for yourself!) or if you are trying to predict the behavior of agents whose preferences you do not fully know.

3.3.4 Mean-variance analysis

In some of the theory and practice of finance and portfolio choice, it is assumed that the investor's preferences over acts are represented by a utility function $U(\mu, \sigma^2)$ of the mean and variance of the acts, which is increasing in the mean and decreasing in the variance. This approach to portfolio selection is called *mean-variance analysis*. It is the basis, for example, of the capital-asset pricing model (CAPM).

In Section 3.3.2, we saw that if we start with expected utility maximization, the preferences over acts or lotteries generally do not reduce to a function of just the mean and variance of the acts or lotteries. That is, mean-variance analysis is generally not consistent with expected utility maximization. In this section, I will give some justification for using mean-variance analysis anyway. First, let's consider its nice properties.

1. The set of distribution functions over the real numbers is huge (infinite dimensional). With mean-variance analysis, distributions are parameterized just by their mean and variance. Hence, the set of distributions reduces to a 2-dimensional set. This simplifies theoretical modeling.
2. The reduction of the set of distributions to a 2-dimensional set makes practical and empirical finance much easier because it reduces the information that is needed about asset returns. The distribution of asset returns are typically predicted based on past returns. The empirical mean and variance of the past returns can be used as an estimate of the mean and variance of future returns. With mean-variance analysis, these are the only aspects of the asset returns that must be estimated, whereas with general expected utility all moments of the asset returns are relevant. Although it is possible to estimate higher-order moments, the precision of these estimates decreases quickly with the order of the moment.
3. The problem of quantifying the distribution of assets is closely related to the econometrician's problem of deducing investor's utility functions. At first, it may not seem that it is less difficult to estimate a utility function over the mean and variance of the lotteries (two variables) than it is to estimate a utility function over money. However, we do not directly observe the strength of investors' preferences

for money; instead, we observe choices of risky portfolios. Unless we have estimates of the investors' beliefs about the distributions of the asset returns, we cannot use these choices to make estimates of the utility functions. Obviously, estimating investors' beliefs about distributions is even harder than estimating the distributions (since we do not know what information the investors have). It is much easier if we only have to estimate what the investors' estimates of the mean and variance are.

4. Mean-variance analysis provides a dominance criterion ($\tilde{x}$ mean-variance dominates $\tilde{y}$ if $E[\tilde{x}] \geq E[\tilde{y}]$ and $\mathrm{Var}(\tilde{x}) \leq \mathrm{Var}(\tilde{y})$) that is much more powerful than second-order stochastic dominance. Although second-order stochastic dominance implies mean-variance dominance, the converse is not true. Furthermore, whereas to show second-order stochastic dominance we need to know the entire distribution of the lotteries (e.g., we need to integrate their cumulative distribution functions), we can rank lotteries by their mean and variance knowing only these two moments.

Given these advantages, we may be willing to use mean-variance analysis even if it is just an approximation (after all, as a positive theory, expected utility is also just an approximation). Under what conditions is it a good approximation?

We can get two answers to this question from results of earlier sections. The first case is when all available acts are (approximately) normally distributed. Because normal distributions are completely parameterized by the mean and variance, it is possible to write preferences over a set of normally distributed acts in terms of preferences over the means and variances of the acts. Normal distributions have three additional properties that make this fact useful. First, we showed in Section 3.3.3 that risk-averse decision makers prefer normal distributions with lower variance. Second, linear combinations of normally distributed acts are normally distributed.[5] This means, for example, that if asset returns are normally distributed, then the return on a portfolio of assets is normally distributed. Third, the central limit theorem tells us that many empirical distributions are, in fact, approximately normal.

Mean-variance preferences are particularly simple (linear) when the acts are normally distributed and utility exhibits constant absolute risk aversion ($u(x) = -e^{-\lambda x}$). Preferences over acts with mean μ and variance σ^2 are then represented by the function

$$U(\mu, \sigma^2) = \mu - \tfrac{\lambda}{2}\sigma^2.$$

The second case in which mean-variance analysis is a good approximation is when the decision maker's utility function is approximately quadratic:

$$u(x) = 2ax - x^2.$$

We saw is Section 3.3.2 that, for any act $\tilde{x}$,

$$E[u(\tilde{x})] = 2aE[\tilde{x}] - E[\tilde{x}]^2 - \mathrm{Var}(\tilde{x}).$$

In fact, the quadratic utility function is the only one for which preferences over all distributions are mean-variance preferences. However, the quadratic utility function

5. These two properties hold for elliptical distributions, which include the normal distributions.

has some peculiar properties. It reaches a maximum at $x = a$ and decreases for $x > a$. Even for $x < a$, this utility function exhibits *increasing* absolute risk aversion, which is inconsistent with empirical observations.

The use of quadratic utility functions is usually accompanied by an assumption that the outcomes of the acts fall in the region where the utility function is increasing. This is just a dissimulated claim that quadratic utility is good as a local approximation. The idea is that we can use the Taylor expansion of the utility function and drop terms higher than the second power, which leaves a quadratic function that approximates the utility function. Assuming that u is three times continuously differentiable, the first two terms

$$v(x) \;=\; u(\bar{x}) + u'(\bar{x})(x - \bar{x}) + \tfrac{1}{2}u''(\bar{x})(x - \bar{x})^2$$

of the Taylor expansion of u around $\bar{x}$ is an approximation of u for x close to $\bar{x}$. Calculating $E[v(\tilde{x})]$ and simplifying by adding and multiplying by constants, we obtain the following mean-variance preferences:

$$2\left(\bar{x} - \frac{u'(\bar{x})}{u''(\bar{x})}\right) E[\tilde{x}] - E[\tilde{x}]^2 - \mathrm{Var}(\tilde{x}). \tag{3.5}$$

Chapter 4

Market Decisions in the Presence of Risk

In this chapter, we look at a variety of decision problems for participants in markets in which uncertainty and risk (but not information and inference) are important. The chapter has two broad goals:

- To learn and practice techniques for solving or characterizing the solutions to such decision problems. In pursuit of this goal, we will even study decision problems that will not reappear in later chapters.
- To discover some properties of demand for state-contingent contracts—such as financial assets or insurance contracts—in preparation for Chapter 5, where we study financial and insurance markets.

4.1 Demand for a state-contingent contract

One market decision that is obviously related to uncertainty is demand for a state-contingent contract. Such a contract might be a gamble, a share of common stock or some other financial instrument, or insurance. In this section, we are interested in cases in which it is possible to increase the wager in a gamble without changing the odds, or to buy multiple units of a stock or financial asset without affecting its price, or to buy arbitrary amounts of insurance coverage at a fixed rate.

4.1.1 Gambles

A simple state-contingent contract is a gamble. A gamble with variable wagers can be represented by a random variable $\tilde{x}$ that is equal to the net winnings per dollar wagered. For example, suppose that a gambler at a horse race is considering betting that a horse Lucky Star will win a race and the odds on Lucky Star are 10:1. Then $\tilde{x} = 10$ if Lucky Star wins and $\tilde{x} = -1$ otherwise. The gamble is said to be fair if $E[\tilde{x}] = 0$, favorable if $E[\tilde{x}] > 0$, and unfavorable if $E[\tilde{x}] < 0$.

Suppose the gambler bets α dollars. This results in state-contingent net winnings $\alpha\tilde{x}$, which is the gambler's net transaction or net trade with the racetrack. In the absence of any gamble, the bettor has baseline wealth $\tilde{w}$, which is the bettor's initial (pre-contracting) state-dependent allocation of wealth. Hence, after betting $\alpha\tilde{x}$ dollars, the gambler's final (post-contracting) state-contingent allocation of wealth is $\tilde{w} + \alpha\tilde{x}$.

Suppose that baseline wealth is non-random. Denote its constant value by w. If u is the gambler's VNM utility function, then he chooses the wager α^* that solves the following decision problem:

$$\max_{\alpha} \; E[u(w + \alpha\tilde{x})] \tag{4.1}$$

If it is possible to take either side of the gamble at the same odds, then α can be positive or negative. Otherwise, there is the constraint that $\alpha \geq 0$.

This gambling story is of pedagogical rather than practical interest to us. Real gambling, such as in casinos and at horse races, usually involves making bets that incur risk but have zero or negative expected net payoff, and this is not consistent with our maintained assumptions that the decision maker cares only about monetary payoffs (as opposed to the fun of playing the game) and is risk averse. Nevertheless, we will see that this decision problem (equation (4.1)) is a reduced form of several decision problems of economic interest.

4.1.2 Portfolio selection

Portfolio selection involves allocating monetary resources to current consumption and to the many available instruments for borrowing and saving or investing money. In Chapter 6, we will see a broad overview of these instruments and the portfolio selection problem. Here, we consider a simple version in which there is one investment period, the money W to be invested is fixed, and there are two financial assets, one riskless (e.g., bank accounts or government bonds) and one risky (e.g., corporate or municipal bonds, stocks or commodity features).

Let q_0 be the current cost of one unit of the riskless asset and let Y_0 be the money received at the end of the investment period for each unit of the asset. For example, if the asset is a bank account and one unit means one dollar put in the account today, then $q_0 = 1$ and Y_0 is one plus the interest rate. If the asset is a zero-coupon bond, then Y_0 is the face value of the bond and q_0 is its current price in the bond market.

Let q_1 be the price of the risky asset and let $\tilde{Y}_1$ be the state-contingent money received at the end of the investment period for each unit of this asset. For example, if the asset is a share of stock, then q_1 is the current price per share in the stock market and $\tilde{Y}_1$ is the dividend plus the price per share at the end of the investment period. Y_0 and $\tilde{Y}_1$ are called the payoffs of the riskless and risky assets, respectively.

A portfolio is $\langle\theta_0, \theta_1\rangle$, where θ_0 and θ_1 are the units of the riskless and risky assets, respectively, that the investor purchases. The investor's budget constraint in the asset market is that the cost of the portfolio equal the wealth to be invested:

$$q_0\theta_0 + q_1\theta_1 = W.$$

The payoff or value of the portfolio $\langle\theta_0, \theta_1\rangle$ at the end of the investment period is

$$\theta_0 Y_0 + \theta_1 \tilde{Y}_1.$$

Since the amount to be invested, and hence current consumption, are fixed, the investor cares only abut the payoff of the portfolio. If the investor is an expected utility maximizer with state-independent utility u, then the decision problem is

$$\begin{aligned} \max_{\theta_0, \theta_1} \quad & E\left[u(\theta_0 Y_0 + \theta_1 \tilde{Y}_1)\right] \\ \text{subject to:} \quad & q_0\theta_0 + q_1\theta_1 = W. \end{aligned} \tag{4.2}$$

We can reformulate this as a single-variable unconstrained maximization problem by solving the constraint for θ_0,

$$\theta_0 = \frac{W - q_1\theta_1}{q_0},$$

and substituting this into the portfolio payoff:

$$\theta_0 Y_0 + \theta_1 \tilde{Y}_1 = \frac{W - q_1\theta_1}{q_0} Y_0 + \theta_1 \tilde{Y}_1 = W\frac{Y_0}{q_0} + (q_1\theta_1)\left(\frac{\tilde{Y}_1}{q_1} - \frac{Y_0}{q_0}\right) \tag{4.3}$$

We can simplify equation (4.3) by replacing asset payoffs and prices by returns. An asset's (*total*) *return* is the payoff per dollar invested, or the asset payoff divided by the asset price. Then $R_0 = Y_0/q_0$ is the *riskless return* and $\tilde{R}_1 = \tilde{Y}_1/q_1$ is the return on the risky asset. Let $\alpha = q_1\theta_1$ be the amount of money invested in the risky asset. Then we can write the payoff equation (4.3) of the portfolio as

$$WR_0 + \alpha(\tilde{R}_1 - R_0). \tag{4.4}$$

Let $w = WR_0$. This is the future value of the wealth W to be invested; it is also the investor's baseline wealth if the investor does not buy any of the risky asset. Let $\tilde{x} = \tilde{R}_1 - R_0$. This is called the *excess return* of the risky asset. Then the value of the portfolio is

$$w + \alpha\tilde{x}. \tag{4.5}$$

We can see that the gambling problem in equation (4.1) is a reduced form of this portfolio selection problem in equation (4.2). The "gamble" $\tilde{x}$ is favorable ($E[\tilde{x}] > 0$) if the expected value of the risky return is greater than the riskless return.

If it is impossible to sell the risky asset ("go short"), then there is a constraint that $\alpha \geq 0$. If it is impossible to sell the riskless asset (which we can interpret as taking out a bank loan for the purpose of investing in the risky asset), then there is a constraint that $\alpha \leq W$. Selling an asset short may seem like an extreme action, but everyone who issues a bond or stock or takes out a loan is doing exactly that. Hence, it is best to think carefully before imposing restrictions on α.

4.1.3 Insurance

Another common class of state-contingent policies is insurance contracts. With some types of insurance, such as life insurance, it is possible to choose from a range of levels of coverage at roughly the same premium rate.

As a market transaction, buying such insurance is like buying a financial asset (i) whose future payoff $\tilde{Y}_1$ is the state-contingent reimbursement by the insurance company to the policy holder (per unit of coverage) and (ii) whose current price q_1 is the premium (per unit of coverage). For example, a life insurance policy might cost \$2 per \$1000 of coverage. Then $q_1 = 2$ and $\tilde{Y}_1 = 1000$ in the event of death and $\tilde{Y}_1 = 0$ otherwise.

If the only other instrument for borrowing and saving money is a riskless asset with future value Y_0 and current price q_0 and if wealth W to be saved or used for premiums is fixed, then purchasing insurance is like purchasing the risky asset in the portfolio selection problem. The post-contracting allocation can be written $w + \alpha\tilde{x}$, where w is the baseline wealth in the absence of coverage, α is the number of dollars spent on premiums, and the "excess return" $\tilde{x}$ is the future value of the net transactions with the insurance company, per dollar spent on premiums. The insurance is said to be *actuarially fair, favorable,* or *unfavorable* if $E[\tilde{x}] = 0$, $E[\tilde{x}] > 0$, or $E[\tilde{x}] < 0$, respectively.

When studying insurance, it is simplest to ignore the intertemporal aspect (premiums are paid up front and reimbursements are received later) by assuming that $R_0 = 1$ (the real interest rate is 0). Then

$$\alpha\tilde{x} = \alpha(\tilde{R} - 1) = \theta_1(\tilde{Y}_1 - q_1).$$

(θ_1 is the number of units of coverage and $\tilde{Y} - q_1$ is the net transaction per dollar spent on insurance.) We simplify by dropping the subscript from θ_1, $\tilde{R}_1$, and $\tilde{Y}_1$. Then

$$\alpha\tilde{x} = \alpha(\tilde{R} - 1) = \theta(\tilde{Y} - q).$$

In this case, the insurance is actuarially fair if the expected value of the reimbursements, $E[\tilde{Y}]$, equals the premium q.

Our model has a problem; it is too similar to the portfolio selection model. Insurance is usually actuarially unfair. However, a risk-averse decision maker with state-independent utility and state-independent wealth would not take on a gamble with $E[\tilde{x}] \leq 0$ hence would not buy insurance that is at best actuarially fair. The demand for insurance is due to either state-dependent preferences over money or state-dependent wealth. You are to explore insurance with state-dependent preferences in an exercise. I will consider the case of state-dependent wealth. This means that the baseline wealth is a random variable $\tilde{w}$, and the decision maker chooses coverage to solve.

$$\max_{\alpha} \quad E[u(\tilde{w} + \alpha\tilde{x})]. \tag{4.6}$$

I also assume that it is possible to completely insure against the fluctuations in wealth. This means that there is some level $\bar{\alpha}$ of coverage such that final wealth $\tilde{w} + \bar{\alpha}\tilde{x}$ is non-random (and hence equal to $E[\tilde{w}] + \bar{\alpha}E[\tilde{x}]$ for sure).

For example, suppose that a person faces the risk of a monetary loss of L dollars, which might be due to theft, fire or a lawsuit. Let W be wealth without the loss (state s_1), so that $W - L$ is wealth with the loss (state s_2). Suppose he can buy insurance that pays \$1 in the event of the loss and costs q per unit of coverage. Then the following table shows how we determine the allocation (act) that the decision maker faces when he purchases α units of coverage:

State s	Baseline wealth $\tilde{w}(s)$	Reimbursements per unit $\tilde{Y}(s)$	Net transaction per unit $\tilde{x}(s) = \tilde{Y}(s) - q$	Final wealth given α units $\tilde{z}(s) = \tilde{w}(s) - \alpha\tilde{x}(s)$
s_1	W	0	$-q$	$W - \alpha q$
s_2	$W - L$	1	$1 - q$	$W - L + \alpha - \alpha q$

The insurance is actuarially fair if q is equal to the probability of state s_2. The decision maker gets full coverage by buying L units of insurance. His final wealth is then $W - qL$ in both states.

If, as a mental exercise, we treat full coverage as the status quo, this insurance problem with state-dependent wealth is equivalent to the gambling and portfolio selection problems. With full coverage, the decision maker's "baseline wealth" is non-random. *Reducing coverage is then like taking a gamble or making a risky investment.* To see this formally, we just have to redefine the variables. Let the choice variable be the number β of units below full coverage: $\beta = \bar{\alpha} - \alpha$. Let the baseline wealth $\bar{w}$ be the risk-free wealth given full coverage: $\bar{w} = \tilde{w} + \bar{\alpha}\tilde{x}$. Let $\tilde{y}$ be the change in net transactions by reducing coverage by one unit: $\tilde{y} = -\tilde{x}$. Then

$$\tilde{w} + \alpha\tilde{x} \;=\; \tilde{w} + (\bar{\alpha} - \beta)\tilde{x} \;=\; \tilde{w} + \bar{\alpha}\tilde{x} - \beta\tilde{x} \;=\; \bar{w} + \beta\tilde{y}.$$

If the insurance is actuarially unfavorable ($E[\tilde{x}] < 0$), then reducing coverage away from full coverage is like taking on a favorable gamble ($E[\tilde{y}] > 0$).

Exercise 4.1. You have just moved to the United States from Mexico. Your retirement savings consist of peso-denominated bonds issued by the Mexican government. You will then spend your retirement money in the United States. Suppose that there is no chance of default by the Mexican government, and your bonds will be worth 1 million pesos at maturity. However, the value of your nest egg in US$ is uncertain because of exchange-rate uncertainty. Suppose also that the future value of the peso is either 5 pesos/dollar (state 1) or 2 pesos/dollar (state 2), and you believe these states will occur with probabilities $\pi_1 = 1/3$ and $\pi_2 = 2/3$, respectively.

You can acquire a forward contract that requires you to exchange, for each unit of the contract you acquire, 1000 pesos for US$350. The current price of this contract is zero.

In the questions below, let an allocation or act be the state-dependent dollar value of your retirement funds, after liquidating any holdings of the forward contract. Assume that you are a VNM expected utility maximizer with risk-averse, state-independent preferences and differentiable utility.

You are to draw some things on the graph in Figure 4.1.

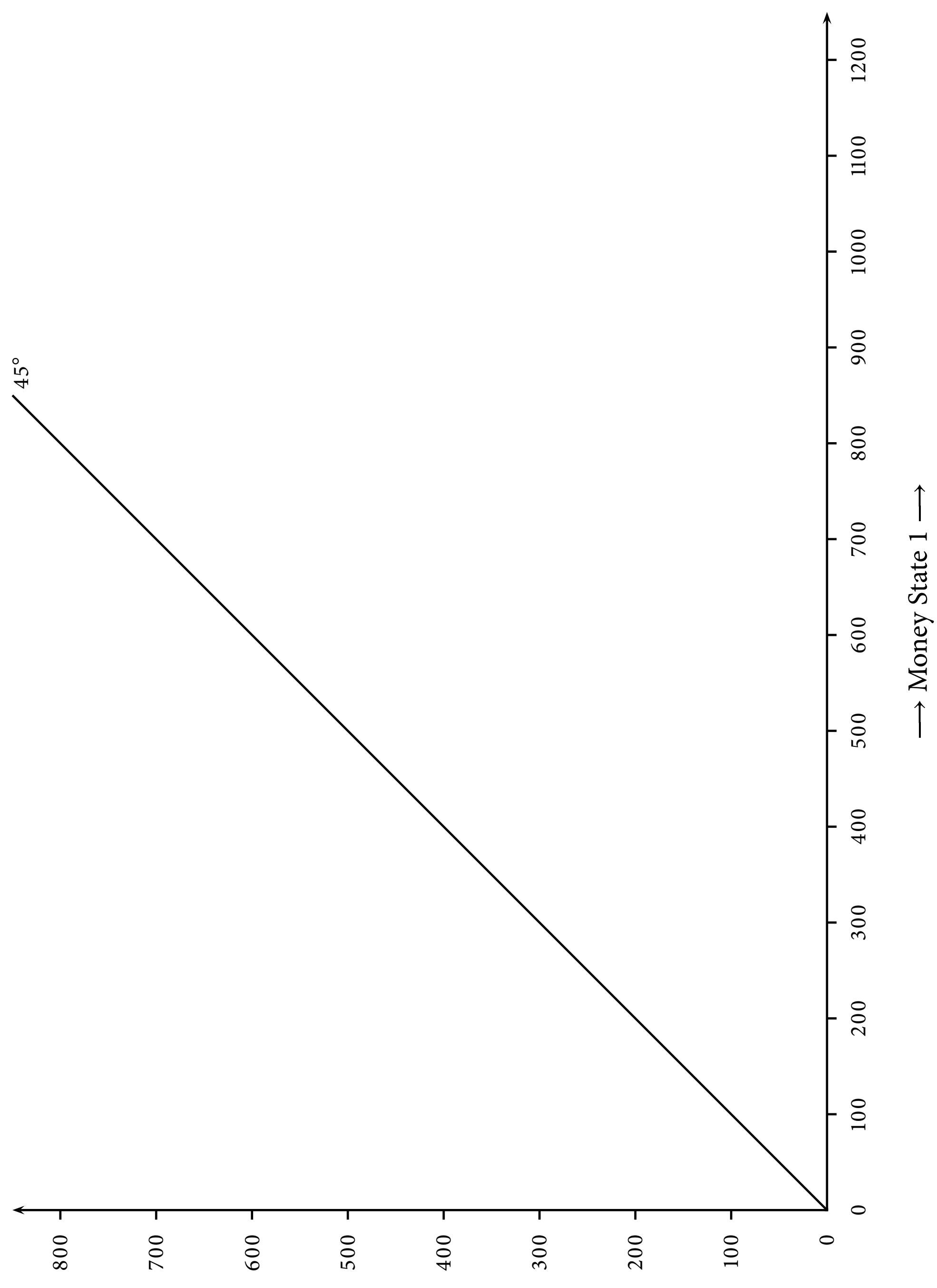
a. Mark your baseline allocation on the graph.

b. Derive the budget constraint your allocations must satisfy, and identify the state prices.

c. Draw the budget line (assuming you can both buy and sell the forward contract) and a vector from the budget line pointing in the direction of the state prices.

d. Using just the information you have been given, what can you say about the position of the optimal allocation in the budget line?

e. Select a possible optimal allocation. Draw a plausible indifference curve through the allocation. Indicate the slope of the indifference curve where it crosses the 45° line by drawing a vector perpendicular to the indifference curve at that point.



Graph for Problem 4.1.

4.2 Techniques for characterizing market decisions

4.2.1 Elimination of dominated choices

We can often say a lot about the solution to a decision problem—even when we do not have full information about the objectives and constraints—by ruling out dominated choices.

For example, recall the gambling problem:

$$\max_{\alpha} \; E[u(w + \alpha\tilde{x})]. \tag{4.7}$$

Let α^* be the solution. Assume that the bettor is risk averse and has increasing utility.

Suppose that it is possible to wager both positive and negative amounts at the same odds. Knowing only that the decision maker is risk averse, we can conclude that $\alpha^* = 0$ if $E[\tilde{x}] = 0$. The act when $\alpha = 0$ is the risk-free act w. For $\alpha \neq 0$, the decision maker faces the risky act $w + \alpha\tilde{x}$. Since $E[w + \alpha\tilde{x}] = w + \alpha E[\tilde{x}] = w$, the risk averse decision maker prefers to get w for sure than $w+\alpha\tilde{x}$. This is an example of a dominance argument.

The implication for the portfolio selection problem is that, if the expected return of the risky asset equals the riskless return, then the investor should only purchase the riskless asset. The implication for the insurance problem is that, if the insurance is actuarially fair, then the consumer should buy full insurance ($\beta^* = 0$).

Similarly, suppose that $E[\tilde{x}] < 0$. Then the act for $\alpha = 0$ second-order stochastically dominates the act for $\alpha > 0$, since the former has no risk and has a higher expected value than the latter. Therefore, $\alpha^* \leq 0$. If $E[\tilde{x}] > 0$, then the same argument tells us that $\alpha^* \geq 0$.

If $\tilde{x}(s) > 0$ for all s, then $w + \alpha'\tilde{x}$ statewise dominates $w + \alpha''\tilde{x}$ whenever $\alpha' > \alpha''$. Because the decision maker always prefers more of the gamble, there is no solution to the problem in equation (4.1) unless there is an exogenous limit on the size of α. This is an example of an *arbitrage opportunity*. If α can be negative, then arbitrage is also possible if $\tilde{x}(s) < 0$ for all s. In the portfolio problem, arbitrage is possible if the return of the risky asset is higher than or lower than the riskless return in every state.

Here is a more sophisticated dominance argument. Suppose that an investor wants to invest W dollars in two risky assets that have returns $\tilde{R}_1$ and $\tilde{R}_2$, respectively. Let α_1 and α_2 be the dollars invested in the two assets. Then the return on the portfolio is $\alpha_1\tilde{R}_1 + \alpha_2\tilde{R}_2$. Given that the amount to be invested is fixed, the investor is maximizing the expected utility of the portfolio return:

$$\begin{array}{ll} \max\limits_{\alpha_1,\alpha_2} & E\left[u(\alpha_1\tilde{R}_1 + \alpha_2\tilde{R}_2)\right] \\ \text{subj. to:} & \alpha_1 + \alpha_2 = W \end{array}$$

Proposition 1. *Perfect diversification ($\alpha_1 = \alpha_2 = W/2$) is optimal if $\tilde{R}_1$ and $\tilde{R}_2$ are symmetrically distributed.*

Two random variables $\tilde{x}$ and $\tilde{y}$ are symmetrically distributed if their joint distribution is the same when the roles for $\tilde{x}$ and $\tilde{y}$ are reversed.[1] For example, identically and independently distributed random variables are symmetrically distributed. The proof involves showing that the portfolio is less risky when $\alpha_1 = \alpha_2 = W/2$.

Proof. We can make the problem a single-variable unconstrained maximization problem by letting $\alpha = \alpha_2$ and substituting the constraint $\alpha_1 = W - \alpha$ into the portfolio return. Then the return when α is invested in asset 2 is

$$\tilde{R}_{(\alpha)} = (W-\alpha)\tilde{R}_1 + \alpha\tilde{R}_2.$$

I will show that $\tilde{R}_{(W/2)}$ is *less risky* than $\tilde{R}_{(\alpha)}$ for $\alpha \neq W/2$. I will use the random-variable characterization of increasing risk.

Let

$$\begin{aligned}\tilde{\varepsilon} &= \tilde{R}_{(\alpha)} - \tilde{R}_{(W/2)}\\ &= (W-\alpha)\tilde{R}_1 + \alpha\tilde{R}_2 - (W/2)\tilde{R}_1 - (W/2)\tilde{R}_2\\ &= (\tfrac{W}{2} - \alpha)(\tilde{R}_1 - \tilde{R}_2).\end{aligned}$$

Since $\tilde{R}_{(\alpha)} = \tilde{R}_{(W/2)} + \tilde{\varepsilon}$, it is also true that $\tilde{R}_{(\alpha)} \stackrel{d}{=} \tilde{R}_{(W/2)} + \tilde{\varepsilon}$. The proof is completed by showing that $E\big[\tilde{\varepsilon} \,\big|\, \tilde{R}_{(W/2)}\big] = 0$.

Observe that

$$\begin{aligned}E\big[\tilde{\varepsilon} \,\big|\, \tilde{R}_{(W/2)}\big] &= E\big[\tilde{\varepsilon} \,\big|\, (W/2)(\tilde{R}_1 + \tilde{R}_2)\big]\\ &= E\big[\tilde{\varepsilon} \,\big|\, \tilde{R}_1 + \tilde{R}_2\big]\\ &= (\tfrac{W}{2} - \alpha)E\big[\tilde{R}_1 - \tilde{R}_2 \,\big|\, \tilde{R}_1 + \tilde{R}_2\big].\end{aligned}$$

Since $\tilde{R}_1$ and $\tilde{R}_2$ are symmetrically distributed,[2] $E\big[\tilde{R}_1 - \tilde{R}_2 \,\big|\, \tilde{R}_1 + \tilde{R}_2\big] = 0$. Therefore, $E\big[\tilde{\varepsilon} \,\big|\, \tilde{R}_{(W/2)}\big] = 0$. □

Exercise 4.2. Suppose that in a market for state-contingent contracts with two states, there are three assets, with payoffs and prices as follows:

		Payoffs	
Asset	*Price*	*State 1*	*State 2*
x	\$200	\$100	\$300
y	\$150	\$225	\$75
z	\$150	\$120	\$160

1. For simple random variables, this means that

$$\text{Prob}\,[\tilde{x} = z_1 \,\&\, \tilde{y} = z_2] = \text{Prob}\,[\tilde{y} = z_1 \,\&\, \tilde{x} = z_2]$$

for any numbers z_1 and z_2.

2. For any random variables $\tilde{x}$ and $\tilde{y}$ that are symmetrically distributed,

$$E\big[\tilde{x} - \tilde{y} \,|\, \tilde{x} + \tilde{y}\big] = E\big[\tilde{y} - \tilde{x} \,|\, \tilde{y} + \tilde{x}\big].$$

The left-hand side is minus the right-hand side, and so both must be zero. That is, $E\big[\tilde{x} - \tilde{y} \,|\, \tilde{x} + \tilde{y}\big] = 0$.

The problem is to choose the optimal portfolio given a fixed total investment. Assume that the states have equal probability. Assume first that short sales are not possible. What can you say about the optimal portfolio? Now assume that short sales are possible (it is possible to buy a negative amount of an asset), and again describe the optimal portfolio. Note: This is a trick question.

4.2.2 First-order conditions

For any single-variable maximization problem

$$\max_{x \in X} V(x)$$

with a continuously differentiable objective function V, a necessary condition for an interior point x^* of X to be a solution is that $V'(x^*) = 0$. This is called the first-order condition. Any solution x^* to the first-order condition $V'(x^*) = 0$ is a global solution to the maximization problem if V is concave (e.g., if $V''(x) \leq 0$ for all x). This is called the second-order condition.

Solving the first-order condition can be a simple and mechanical method for finding the solution to such a maximization problem, even if the method does not provide much intuition. The only trick to applying it to maximization problems in which there is uncertainty is knowing how to differentiate a function that contains the expectations operator. For example, the objective function in a maximization problem might have the form

$$V(\alpha) = E\big[f(\alpha, \tilde{x})\big],$$

where $\tilde{x}$ is a random variable. In the problem in equation (4.1) of choosing a wager α in a gamble for which the net winnings are $\tilde{x}$, $f(\alpha, \tilde{x}) = u(w + \alpha\tilde{x})$ and the objective function is $V(\alpha) = E[u(w + \alpha\tilde{x})]$. Note that α, the choice variable, cannot be a random variable since supposedly the decision maker controls α. Note also that $V(\alpha)$ is an ordinary function; thanks to the expectations operator, there is nothing uncertain about the value of the objective function once the DM picks α. That is, the realization of the DM's VNM utility may be uncertain, but the DM's *expected* utility is *not*.

4.2.3 Differentiation of expected values

Fortunately, finding the derivative of $V(\alpha)$ is not tricky at all. Just remember the following mantra: *The derivative of the expected value is equal to the expected value of the derivative*. That is,

$$\underbrace{\frac{d}{d\alpha} E\big[f(\alpha, \tilde{x})\big]}_{\text{derivative of EV}} = \underbrace{E\left[\frac{d}{d\alpha} f(\alpha, \tilde{x})\right]}_{\text{EV of derivative}}. \tag{4.8}$$

When calculating $\frac{d}{d\alpha} f(\alpha, \tilde{x})$, $\tilde{x}$ is teated as a constant (whose value is random). For example, suppose that $\tilde{x}$ equals 2 or 3 with equal probability. Then

$$E\big[f(\alpha, \tilde{x})\big] = \frac{1}{2} f(\alpha, 2) + \frac{1}{2} f(\alpha, 3),$$

and so

$$\begin{aligned}\frac{d}{d\alpha}E[f(\alpha,\tilde{x})] &= \frac{d}{d\alpha}\left(\frac{1}{2}f(\alpha,2)+\frac{1}{2}f(\alpha,3)\right)\\ &= \frac{1}{2}\frac{d}{d\alpha}f(\alpha,2)+\frac{1}{2}\frac{d}{d\alpha}f(\alpha,3)\\ &= E\left[\frac{d}{d\alpha}f(\alpha,\tilde{x})\right].\end{aligned}$$

In the gambling problem, where $V(\alpha) = E[u(w+\alpha\tilde{x})]$,

$$V'(\alpha) = \frac{d}{d\alpha}E\left[u(w+\alpha\tilde{x})\right] = E\left[\frac{d}{d\alpha}u(w+\alpha\tilde{x})\right] = E\left[\tilde{x}u'(w+\alpha\tilde{x})\right]. \qquad (4.9)$$

Therefore, the first-order condition is

$$E[\tilde{x}u'(w+\alpha\tilde{x})] = 0. \qquad (4.10)$$

The second-order condition is easy to check. If $\frac{d^2}{d\alpha^2}f(\alpha,\tilde{x}) < 0$ for all possible values of $\tilde{x}$, then

$$V''(\alpha) = E\left[\frac{d^2}{d\alpha^2}f(\alpha,\tilde{x})\right] < 0.$$

(In general, if f is concave in α for all possible $\tilde{x}$, then $E[f(\alpha,\tilde{x})]$ is concave in α.) In this gambling problem,

$$\frac{d^2}{d\alpha^2}u(w+\alpha\tilde{x}) = \tilde{x}^2u''(w+\alpha\tilde{x}), \qquad (4.11)$$

which is negative for $\tilde{x} \neq 0$ since $\tilde{x}^2 > 0$ and $u'' < 0$.

Suppose, for example, that $\tilde{x} = 2$ with probability π and $\tilde{x} = -1$ with probability $1-\pi$, and that $u(z) = \log z$. Then

$$V(\alpha) = (1-\pi)\log(w-\alpha)+\pi\log(w+2\alpha).$$

Therefore,[3]

$$V'(\alpha) = (1-\pi)\left(\frac{-1}{w-\alpha}\right)+\pi\left(\frac{2}{w+2\alpha}\right).$$

The solution to the first-order condition $V'(\alpha) = 0$ is

$$\alpha^* = \frac{(3\pi-1)w}{2}.$$

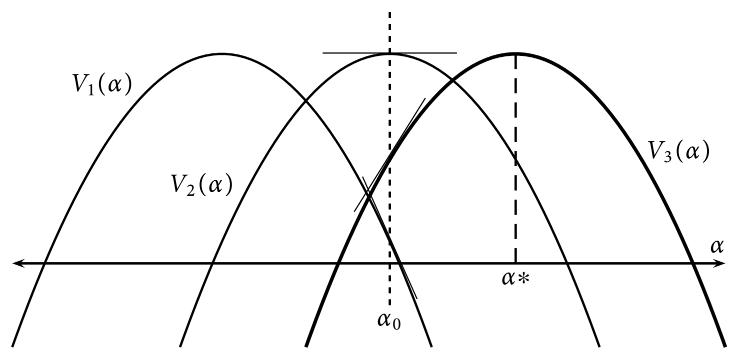
When the gamble is fair, $\pi = 1/3$ and $\alpha^* = 0$. When the gamble is favorable, $\pi > 1/3$ and $\alpha^* > 0$.

3. You can see that this is the expected value of the derivative of $u(w+\alpha\tilde{x})$, as given by equation (4.9), since $-1/(w-\alpha) = \frac{d}{d\alpha}\log(w-\alpha)$ is the derivative of $u(w+\alpha\tilde{x})$ when $\tilde{x} = -1$ and $2/(w+2\alpha) = \frac{d}{d\alpha}\log(w+2\alpha)$ is the derivative when $\tilde{x} = 2$.

4.2.4 Local risk neutrality

Solving the first-order condition is obviously useful for calculating examples like the one above. We can also prove some general theorems using derivatives. Suppose that we want to show that the solution to $\max_{\alpha} V(\alpha)$ is greater than some number α_0. It suffices to show that V is concave and $V'(\alpha_0) > 0$, because then the graph of V must look something like V_3 below, instead of V_1 or V_2:

Figure 4.1



You can see from this picture that

$$\begin{aligned} V_1'(\alpha_0) < 0 &\rightarrow \alpha^* < \alpha_0 \\ V_2'(\alpha_0) = 0 &\rightarrow \alpha^* = \alpha_0 \\ V_3'(\alpha_0) > 0 &\rightarrow \alpha^* > \alpha_0. \end{aligned}$$

We can use this to prove that if u is differentiable (and concave and increasing), then it is always optimal to take some stake in a favorable gamble, even if the gamble is risky. In other words, when the gamble is favorable, there is always some stake small enough that the risk premium of the gamble is lower than the positive expected return. This property is known as *local risk neutrality*.

Differentiability of u is not a mere technical assumption that simplifies the proof. The mathematical intuition behind the result is that, because u is differentiable, it can be locally approximated by a linear—and hence risk-neutral—utility function. We will see a graphical example in Section 4.2.5 in which u is not differentiable and local risk neutrality does not hold.

Proposition 2. *Assume that $u\colon \mathbb{R} \to \mathbb{R}$ is differentiable and concave and that $u' > 0$. Assume that $\tilde{x}$ is a random variable with $E[\tilde{x}] > 0$. Let α^* be the solution to*

$$\max_{\alpha} E[u(w + \alpha\tilde{x})].$$

Then $\alpha^* > 0$.

Proof. Let $V(\alpha) = E[u(w + \alpha\tilde{x})]$, which is the objective function in the maximization problem. $V(\alpha)$ is concave, as shown in equation (4.11). We need to show that $V'(0) > 0$.

From equation (4.9),

$$V'(\alpha) = E\left[\tilde{x}u'(w + \alpha\tilde{x})\right].$$

Warning: For arbitrary values of α, we cannot put a sign on $E[\tilde{x}u'(w+\alpha\tilde{x})]$. Writing

$$E[\tilde{x}u'(w+\alpha\tilde{x})] \;=\; E[\tilde{x}]E[u'(w+\alpha\tilde{x})] \;>\; 0$$

is a mistake because $E[\tilde{y}\tilde{z}] = E\left[\tilde{y}\right]E[\tilde{z}]$ for two random variable $\tilde{y}$ and $\tilde{z}$ if and only if $\tilde{y}$ and $\tilde{z}$ are *uncorrelated* (by definition). If $\alpha > 0$ then $\tilde{x}$ and $u'(w+\alpha\tilde{x})$ are negatively correlated: when $\tilde{x}$ is higher, $u'(w+\alpha\tilde{x})$ is lower because u is concave.

However, we are only asked to find $V'(\alpha)$ at $\alpha = 0$:

$$V'(\alpha)\mid_{\alpha=0} \;=\; E\left[\tilde{x}u'(w)\right] \;=\; \underset{+}{E[\tilde{x}]}\,\underset{+}{u'(w)} \;>\; 0.$$

□

Here are the implications of this result for portfolio selection and demand for insurance:

- In the portfolio selection problem, if the expected return of the risky asset is greater than the expected demand of the riskless asset, then the decision maker should buy some amount of the risky asset.
- In the insurance problem, if the insurance is actuarially unfavorable, underinsuring is like taking on a favorable gamble. Therefore, the decision maker should not fully insure ($\beta^* > 0$).

Exercise 4.3. Consider a portfolio selection problem in which a risk-averse investor has \$1 of wealth to invest, and there are two risky assets available whose gross returns, per dollar invested, are $\tilde{x}$ and $\tilde{y}$. Assume that $\tilde{x}$ and $\tilde{y}$ are independent, and have the same mean, although they may not be identically distributed. Show that the investor will not put all his money in the same asset (e.g., not all in $\tilde{x}$). This will involve differentiation, and you have to use the fact that for a random variable $\tilde{z}$ (that is not constant) and a decreasing function f, $E[\tilde{z}f(\tilde{z})] < E[\tilde{z}]E[f(\tilde{z})]$.

Exercise 4.4. We showed that if utility is differentiable, then a person is willing to accept some share of a favorable gamble. E.g., we concluded that if insurance is not actuarially fair, then a person will not buy full insurance, and if an investor divides his portfolio among a riskless asset and a risky asset with higher expected return, then he will invest at least some amount in the risky asset.

a. Suppose that one can buy insurance that is actuarially fair, except for a fixed fee that does not depend on the extent of coverage. Will a risk-averse person buy full insurance if he buys any at all? Explain.

b. Suppose that there is a fixed broker's fee on stock market transactions, that does not depend on the size of the transaction. Is it still true that an investor will put at least some of his wealth into a risky stock whose expected return is higher than the return on the riskless asset?

c. Exercise 4.3 shows that, if an investor divides his portfolio among a riskless assets and several risky assets with independent returns that are higher than the return on the riskless asset, then if the investor holds any of the riskless asset, he also holds some amount of each of the risky assets.

We can roughly say that putting money into a bank account is a riskless investment. There are also zillions of risky investments out there in the world with roughly independent returns and with expected returns that are higher than the return on a bank account.

Do you have any money in a bank account? Do you also hold a little bit of each of the zillions of risky investments mentioned above, as our theory would predict? Why not? (At most a short paragraph is sufficient.)

4.2.5 Quiche and beer

The third technique for characterizing solutions to such decision problems is a graphical tool. When the decision problem reduces to the choice of monetary allocations (acts) with two states, we can represent the allocations as points on the plane, as in Section 3.1.1. We can then illustrate the optimal choice by drawing the choice set or "budget" set and indifference curves. The picture can end up looking like pictures from consumer theory, but with a different interpretation. Instead of choosing consumption bundles consisting of quiche and beer, the decision maker chooses consumption bundles consisting of money in state 1 and money in state 2.

For example, consider the gambling problem with net winnings $\tilde{x}$. Suppose that there are two states, s_1 and s_2. Assume that arbitrage is not possible, which means that $\tilde{x}(s_1)$ and $\tilde{x}(s_2)$ are not both positive or both negative. Without loss of generality, we can assume $\tilde{x}(s_1) > 0$ and $\tilde{x}(s_2) < 0$.

Let z_1 and z_2 be final wealth in states s_1 and s_2, respectively, so that an allocation is a point $\langle z_1, z_2 \rangle$ on the plane. Given a stake α in the gamble, the investor's allocation is

$$z_1 = w + \alpha\tilde{x}(s_1) \tag{4.12}$$

$$z_2 = w + \alpha\tilde{x}(s_2) \tag{4.13}$$

As we vary α, we trace out the set of possible acts, which I will call the DM's budget set. By looking at this budget set, we can see that betting involves exchanging money in state 2 for money in state 1.

There are two ways to describe the budget set using vectors. One way is to write an act as

$$\langle z_1, z_2 \rangle = \langle w, w \rangle + \alpha\langle \tilde{x}(s_1), \tilde{x}(s_2) \rangle.$$

This means that the budget set is a line through $\langle w, w \rangle$ that extends in the direction $\langle \tilde{x}(s_1), \tilde{x}(s_2) \rangle$. This is drawn in Figure 4.2 for $\tilde{x}(s_1) = 1$ and $\tilde{x}(s_2) = -1/2$, assuming that $w = 8$.

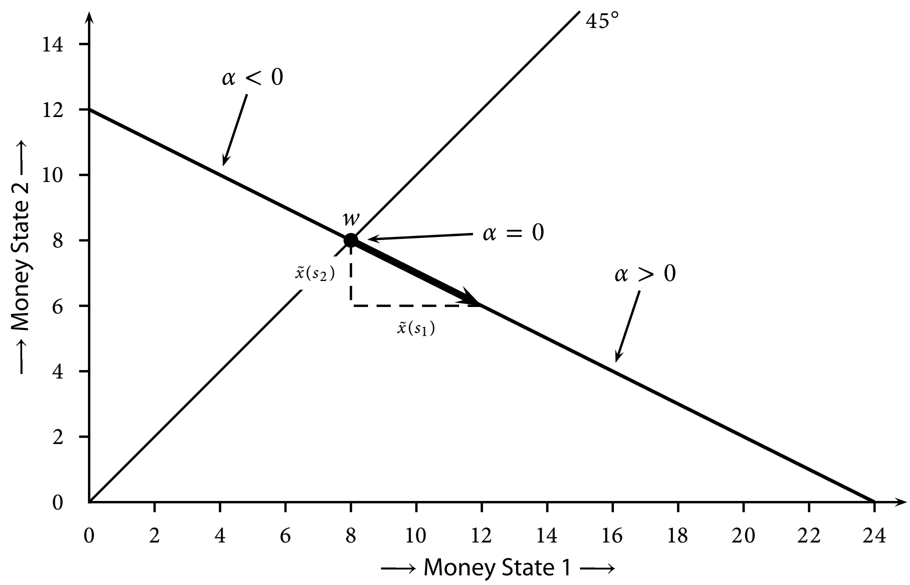
Another way is to write the "budget constraint" for an act is by solving equation (4.12) for α,

$$\alpha = \frac{z_1 - w}{\tilde{x}(s_1)}$$

and plugging the answer into equation (4.13).

$$z_2 = w + \frac{z_1 - w}{\tilde{x}(s_1)}\tilde{x}(s_2)$$

Figure 4.2



The budget set for the bet that pays \$2 in state s_1 and −\$1 in state s_2 when baseline wealth is \$10 in both states.

Rewriting yields

$$-\tilde{x}(s_2)z_1 + \tilde{x}(s_1)z_2 = -\tilde{x}(s_2)w + \tilde{x}(s_1)w.$$

If we let $p_1 = -\tilde{x}(s_2)$ and $p_2 = \tilde{x}(s_1)$, then we obtain

$$p_1 z_1 + p_2 z_2 = p_1 w + p_2 w.$$

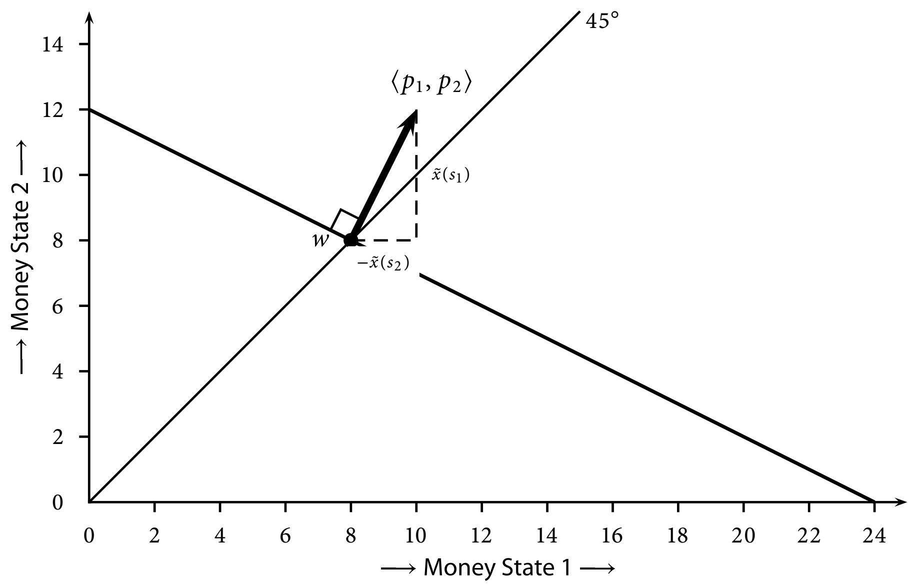
This looks like the budget constraint of a trader in a market for quiche and beer if consumption of quiche and beer are z_1 and z_2, respectively, if the prices of quiche and beer are p_1 and p_2, respectively, and if the trader arrives to the market with w units of quiche and w units of beer (this is called the trader's endowment). Hence, we can call p_1 and p_2 *state prices*; they are the implicit prices of money in states s_1 and s_2, respectively.[4] These state prices depend on the asset prices and payoffs in the portfolio selection problem and on the insurance premium rates in the insurance problem. In the example illustrated in Figure 4.2, $p_1 = 1$ and $p_2 = 2$. That is, money in state 2 is twice as "expensive" as money in state 1, meaning that a trader has to give up \$2 in state 1 to get \$1 in state 2.

As when the goods are quiche and beer, the budget line passes through the endowment (it is always possible to not trade or gamble and just get the baseline level of wealth) and is perpendicular to the price vector. This is illustrated in Figure 4.3.

Consider what happens when the gamble changes (because asset prices change or the premium rate changes). If $\tilde{x}(s_1)$ remains fixed at 1 and $\tilde{x}(s_2)$ changes to −1, then

4. This reduction of the gambling problem with a single gamble to a standard consumer choice problem does not generalize to a larger number of states unless the number of gambles increases with the number of states. In consumer theory, if there are three states, the budget set is a plane that is perpendicular to the price vector. With a single gamble but three states, the budget set in the 3-dimensional set of acts is still a line.

Figure 4.3



The budget set is perpendicular to the state prices.

the ratio p_1/p_2 of state prices increases from 1/2 to 1. The budget line rotates around $\langle w, w\rangle$ as shown in Figure 4.4.

Note that I have not yet mentioned the probabilities of the states. The probabilities affect the preferences over allocations, not the set of affordable allocations. In Section 3.1.3, we learned the following properties of each indifference curve for state-independent preferences:

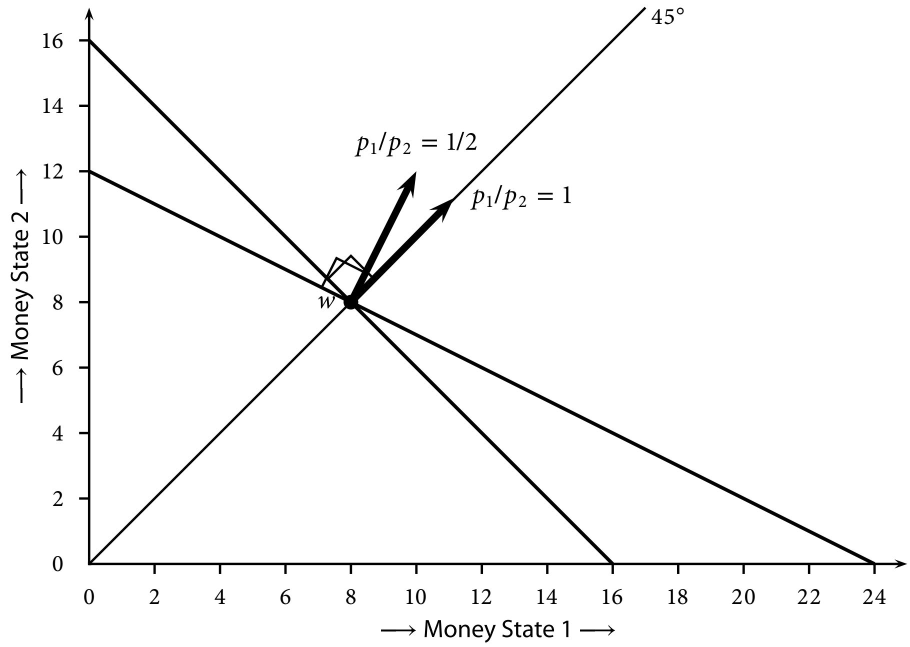
- Where it crosses the 45° line, it is perpendicular to the vector of probabilities, and hence it is tangent to the line containing those allocations with the same expected value.
- If the decision maker is risk neutral, the indifference curve is equal to this line.
- If the decision maker is risk averse, the indifference curve is strictly convex and lies above this line.

When the gamble is fair ($E[\tilde{x}] = 0$) is zero, the state-prices vector is proportional to (points in the same direction as) the vector of probabilities. The fact that the investor should not buy any of the risky asset in this case is illustrated in Figure 4.5. Every affordable act other than the baseline-wealth act on the 45° line lies below the indifference curve through the baseline wealth.

Suppose that expected value of the excess returns is greater than zero. Then the budget line is no longer tangent to the indifference curve through the baseline wealth, and hence it cuts above this indifference curve. Therefore, there is some demand for the risky asset that gives higher utility than no demand. This is illustrated in Figure 4.6.

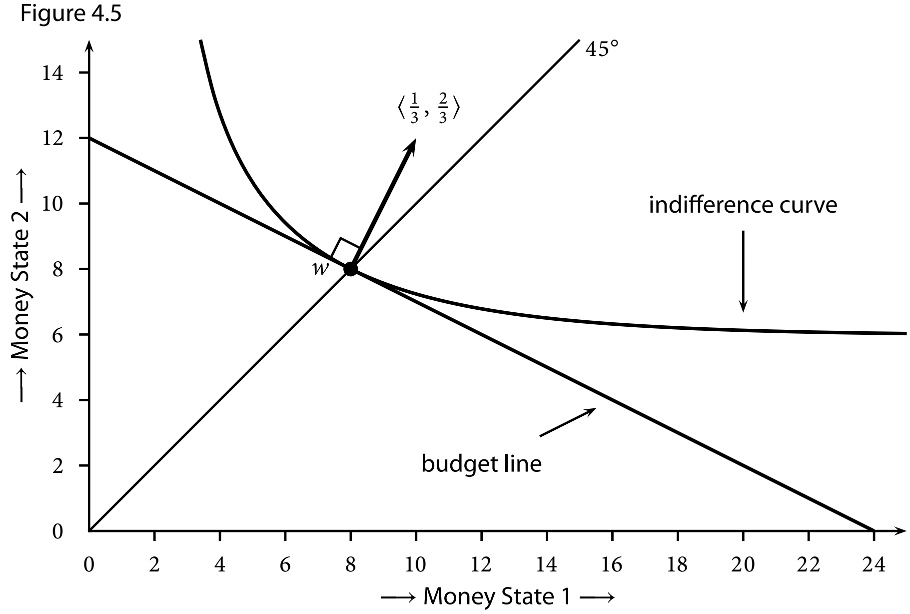
Try to visualize how, no matter how little state prices differ from the probabilities, the budget line must cut above the indifference curve through the baseline wealth. This visualization requires that you zoom in on the area around the baseline wealth, letting the smooth indifference curve look flatter and flatter as you do so. This is not possible if

Figure 4.4



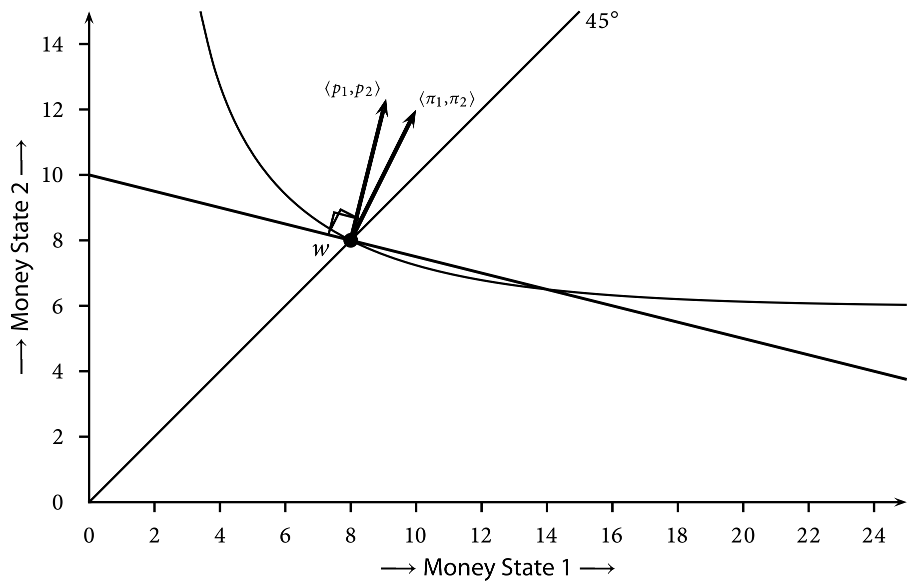
Change in the budget set when the price of the risky asset changes from 2 to 5/2, and thus the relative state prices change from $p_1/p_2 = 1/2$ to $p_1/p_2 = 1$.

Figure 4.5



Optimal choice in the portfolio selection problem is the baseline allocation if the expected excess return is zero. In this example, probabilities and state prices are both (proportional to) $\langle 1/3, 2/3 \rangle$.

Figure 4.6



Local risk neutrality. Even a small change positive expected return of the risky asset causes the budget line to rotate and cut above the indifference curve through the baseline wealth. State prices are no longer equal to the probabilities.

the indifference curve has a kink at the baseline wealth, which can happen if the utility function is not differentiable. Figure 4.7 illustrates that local risk neutrality does not hold in this case.

In Section 4.1.3, we showed how the insurance problem can be made equivalent to the portfolio problem. Hence, state prices and the diagrams work pretty much the same way. There is one difference, however. A change in the price of insurance causes the budget line to rotate around the true random baseline wealth $\tilde{w}$, whereas the risk-free wealth w_0 shifts. If $\tilde{x}$ is the net transaction with the insurance company per dollar spent on premiums,, then the budget constraint in the two-state case is

$$-\tilde{x}(s_2)z_1 + \tilde{x}(s_1)z_2 \;=\; -\tilde{x}(s_1)\tilde{w}(s_1) + \tilde{x}(s_1)\tilde{w}(s_2).$$

If state s_1 is the state where there is a loss, then $\tilde{x}(s_1) > 0$ and $\tilde{x}(s_2) < 0$. Let $p_1 = -\tilde{x}(s_2)$ and $p_2 = \tilde{x}(s_2)$. The budget constraint is then

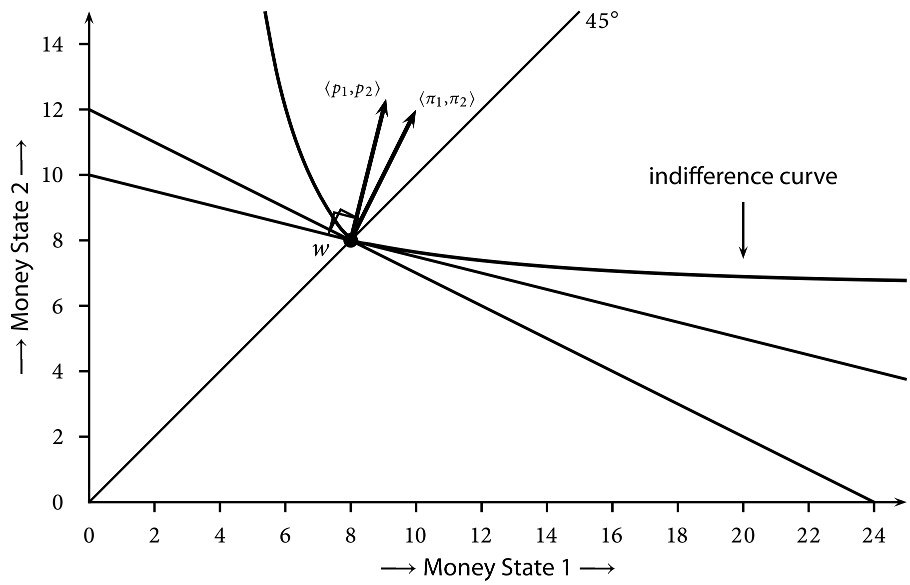
$$p_1 z_1 + p_2 z_2 \;=\; p_1\tilde{w}(s_1) + p_2\tilde{w}(s_2).$$

The endowment point $\langle \tilde{w}(s_1), \tilde{w}(s_2)\rangle$ thus has different amounts of money in the two states.

Suppose that you face a risk of fire, which you treat as a purely monetary loss of $\$60K$. Without the fire (state s_2), your wealth is $\tilde{w}(s_2) = \$80K$, but with the fire (state s_1), your wealth is $\tilde{w}(s_1) = \$20K$. For the sake of drawing a picture, let's assume that the probability of the fire is quite high: $\pi(s_1) = 1/3$.

If q, the premium per dollar of coverage, is equal to 1/3, then the insurance is actuarially fair. The net transactions per dollar of premium are $\tilde{x}(s_1) = 2$ and $\tilde{x}(s_2) = -1$. Hence, the state prices are $p_1 = 1$ and $p_2 = 2$, which are proportional to the state

Figure 4.7



A violation of local risk neutrality because the utility function is not differentiable.

probabilities $\pi(s_1) = 1/3$ and $\pi(s_2) = 2/3$. Full insurance gives wealth \$60 for sure, and so the budget line crosses the 45° line at $\langle 60, 60 \rangle$.

If instead $q = 1/2$, the insurance is actuarially unfavorable. The net transactions are $\tilde{x}(s_1) = 1$ and $\tilde{x}(s_2) = -1$, and the state prices are $p_1 = p_2 = 1$. Full insurance gives wealth \$50K for sure. Compared to $q = 1/3$, the budget line rotates around the endowment $\langle 20, 80 \rangle$ and intersects the 45° line at $\langle 30, 30 \rangle$. Money in the state s_1 becomes relatively more "expensive," and the decision maker responds by reducing coverage. The budget lines for $q = 1/3$ and $q = 1/2$, and the indifference curves through the corresponding optima, are shown in Figure 4.8.

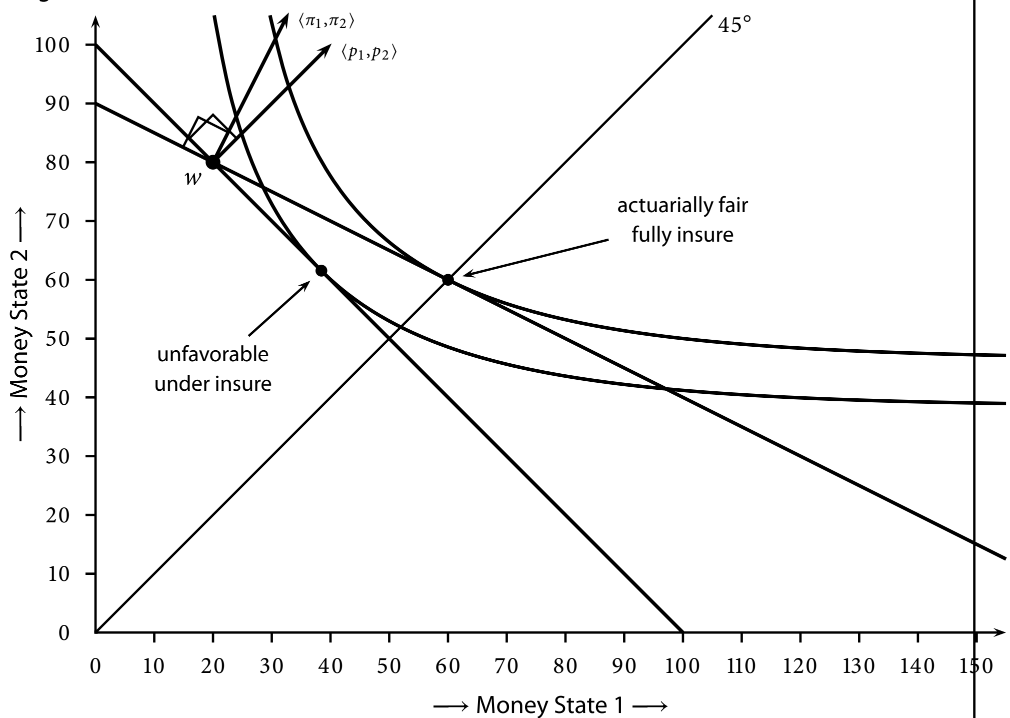
4.3 Comparative statics

Models have *endogenous* variables, which are determined in equilibrium by the model, and *exogenous* parameters, which are fixed outside the model. For example, in a duopoly model, the firm's production technologies are fixed and the market prices are determined by the model. In a model of consumer optimization, prices are fixed and the consumer's demand is determined by the model.

Comparative statics involves comparing the equilibrium values of the endogenous variables for different values of the exogenous parameters.[5] For example, in consumer theory we might compare a consumer's demand at one price with her demand at another price.[6] In this section, I will present some basic methods for comparing the

5. This is called sensitivity analysis in some other disciplines.
6. The answer might be phrased, "If the price changes from p to p', then demand shifts from x to x'." However, this sentence should not be interpreted as describing a dynamic process by which demand responds to

Figure 4.8



A shift in the premium in the insurance problem. The zero-insurance, baseline wealth is state-dependent. The budget line rotates around this baseline wealth.

solutions to single-variable maximization problems, with application to problems involving risk.

4.3.1 Changes to the decision maker's welfare

z Consider a class of single-variable maximization problems in which the choice variable is α and within which the objective function and constraint set may depend on exogenous parameters. Consider two instances of the problem:

$$\begin{array}{cc} \max\limits_{\alpha \in A_1} V_1(\alpha) & \max\limits_{\alpha \in A_2} V_2(\alpha) \\ \text{(Problem 1)} & \text{(Problem 1)} \end{array} \tag{4.14}$$

Let α_1^* and α_2^* be the solutions to Problems 1 and 2, respectively, and let $v_1^* = V_1(\alpha_1^*)$ and $v_2^* = V_2(\alpha_2^*)$ be the corresponding maximum values.

If V_1 and V_2 both measure the decision maker's utility in a comparable way, then we can ask whether the decision maker is better off in Problem 1 or Problem 2. For example, Problems 1 and 2 might both be portfolio selection problems, but with different excess returns $\tilde{x}_1$ and $\tilde{x}_2$:

$$V_1(\alpha) = E[u(w + \alpha\tilde{x}_1)] \qquad V_2(\alpha) = E[u(w + \alpha\tilde{x}_2)]\ .$$

a shift in prices.

If $\tilde{x}_2$ second-order stochastically dominates $\tilde{x}_1$, does the investor attain higher expected utility on Problem 2 than in Problem 1? That is, is $v_2^* > v_1^*$?

This is true if we can show that, whatever the investor does in Problem 1, he can do better in Problem 2, perhaps even if he does the same thing as in Problem 1. Formally, to show that $v_2^* \geq v_1^*$ or $v_2^* > v_1^*$, it suffices to find some $\alpha \in A_2$ such that $V_2(\alpha) \geq V_1(\alpha_1^*)$ or $V_2(\alpha) > V_1(\alpha_1^*)$. In particular, we can start by checking whether $\alpha_1^* \in A_2$ and $V_2(\alpha_1^*) \geq V_1(\alpha_1^*)$ or $V_2(\alpha_1^*) > V_1(\alpha_1^*)$.

For example, in the comparison of the portfolio selection problems, for any investment $\alpha \neq 0$, $w + \alpha\tilde{x}_2$ second-order stochastically dominates $w + \alpha\tilde{x}_1$, and hence $E[u(w+\alpha\tilde{x}_2)] > E[u(w + \alpha\tilde{x}_1)]$. In particular, $V_2(\alpha_1^*) > V_1(\alpha_1^*)$, and hence $v_2^* > v_1^*$, if $\alpha_1^* \neq 0$. In words, the decision maker is better off in Problem 1 than in Problem 2 even if he just chooses in Problem 2 the investment that was optimal in Problem 1 (and he may be able to do even better by adjusting his investment).

4.3.2 Changes to the decision maker's action

Comparing the solutions α_1^* and α_2^* is typically more difficult. We might ask, for example, whether the optimal investment is higher in the portfolio problem with s.o.s.d. excess returns. I will present a method for answering such a question and two examples where the method works. However, I will not attempt a thorough treatment of this topic because typically it is either very difficult or impossible to obtain an answer without specific assumptions on the utility functions or distributions.

Suppose we want to show that $\alpha_2^* > \alpha_1^*$. Suppose that V_1 and V_2 are both concave and differentiable. Then, from Section 4.2.2, it suffices to show that $V_2'(\alpha_1^*) > 0$. If α_1^* is an interior solution, then $V_1'(\alpha_1^*) = 0$, and hence it suffices to show that $V_2'(\alpha_1^*) > V_1'(\alpha_1^*)$.

Most textbooks that discuss comparative statics for decision problems under uncertainty present the following example about precautionary savings, because it is one of the few that works out with only a minor extra assumption—that the third derivative of the utility function be negative. Here is the scenario

> Consider a household that lives two periods, with income y_1 in period 1 and uncertain income $\tilde{y}_2$ in period 2. The household chooses consumption in period 1, saves or borrows the difference between period 1 income and period 1 consumption at a known return R, and consumes whatever income/wealth that remains in period 2. Denote consumption in period 1 by c_1 and the uncertain consumption in period 2 by $\tilde{c}_2$. Utility is $u_1(c_1) + u_2(c_2)$, and the household maximizes expected utility $u_1(c_1) + E[u_2(\tilde{c}_2)]$. Assume that u_1 and u_2 are differentiable, strictly increasing, and strictly concave, and defined for all real numbers (so we don't have to worry about whether consumption is negative in period 2 when period 2 income is low).

First we formulate the problem as an unconstrained maximization problem with choice variable c_1. The household's budget constraint is

$$\tilde{c}_2 = R(y_1 - c_1) + \tilde{y}_2 \ .$$

By substituting the budget constraint into the objective function, we get an unconstrained maximization problem:

$$\max_{c_1} \quad u_1(c_1) + E\left[u_2(R(y_1 - c_1) + \tilde{y}_2)\right] . \tag{4.15}$$

Denote the objective function by $V(c_1)$. Now we take the first and second derivatives of the objective function:

$$V'(c_1) = u_1'(c_1) - RE[u_2'(R(y_1 - c_1) + \tilde{y}_2)] \tag{4.16}$$

$$V''(c_1) = u_1''(c_1) + R^2E[u_2''(R(y_1 - c_1) + \tilde{y}_2)] \tag{4.17}$$

We can see that $V''(c_1) < 0$ for all c_1, and hence V is concave, since $u_1'' < 0$ and $u_2'' < 0$. Now we will show that when $\tilde{y}_2$ is less risky, $V'(c_1)$ is higher for all c_1, and hence the optimal c_1 is higher as long as $u_2''' > 0$. Therefore, when future income is less risky, savings is lower. This response of savings to risk is called the precautionary motive for savings.

Let $\tilde{y}_2'$ be less risky than $\tilde{y}$. Write $V'(c_1; \tilde{y}_2)$ and $V'(c_1; \tilde{y}_2')$ to indicate that the derivative depends on the distribution of period 2 income. The derivatives $V'(c_1; \tilde{y}_2)$ and $V'(c_1; \tilde{y}_2')$ have a common term, $u_1'(c_1)$, and hence we can ignore it when checking which derivative is higher. I.e., we want to show that

$$-E[u_2'(R(y_1 - c_1) + \tilde{y}_2)] < -E[u_2'(R(y_1 - c_1) + \tilde{y}_2')] \tag{4.18}$$

or equivalently, that

$$E[u_2'(R(y_1 - c_1) + \tilde{y}_2)] > E\left[u_2'(R(y_1 - c_1) + \tilde{y}_2')\right] \tag{4.19}$$

Since $\tilde{y}_2'$ is less risky than $\tilde{y}_2$, (by definition) $E[f(\tilde{y}_2')] > E[f(\tilde{y}_2)]$ for any concave f, which is the same as saying that $E[f(\tilde{y}_2')] < E[f(\tilde{y}_2)]$ for any convex f. Thus, inequality (4.19) holds if $u_2'(r(y_1 - c_1) + y_2)$ is convex in y_2. $f(y)$ is convex in y if and only if $f(a + y)$ is convex in y for any constant a. Thus, (4.19) holds if u_2' is a convex function. The second derivative of u_2' is u_2''', and so u_2' is convex if $u_2''' > 0$.

At this point, you probably do not want to see what a hard problem looks like. You may also be wondering why I called $u''' > 0$ a "minor assumption." While $u'' < 0$ has the interpretation of risk aversion, who would want to make claims about the third derivative of their utility function? Actually $u''' > 0$ also has some empirical content. First, you can check that constant or decreasing absolute risk aversion implies $u''' > 0$ (see, for example, the graph of u' for a constant relative risk aversion utilty function in Figure 4.9. Second, you can also check that u'' cannot always be negative, because (given that $u'' < 0$) this would imply that u' is eventually negative.

(That doesn't mean, of course, that u_2''' is negative for all values, as is seen in Figure 4.10. However, it does mean that the assumption that u''' is convex is not totally outrageous, and under this assumption, we can conclude that consumption is lower and thus savings is higher with uncertain income compared to with certain income. This corresponds to the intuitive notion that one motive for saving is precautionary savings to guard against fluctuations in future income.

Suppose we try the same trick to show that in the portfolio problem, the optimal investment is higher when the excess return is less risky? We have to check that

$$V'(\alpha) = E\left[\tilde{x}u'(w + \alpha\tilde{x})\right]$$

Figure 4.9

The graph of $u'(x)$ when $u(x) = x^{1/2}$. $u'(x)$ is convex.

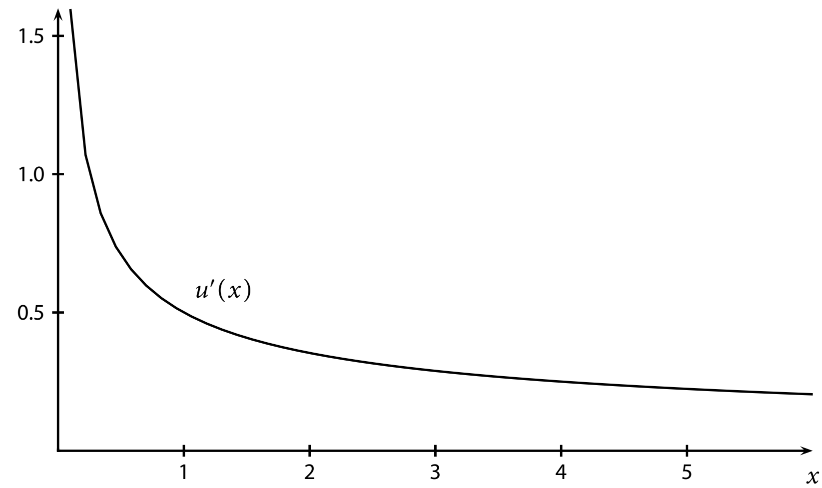
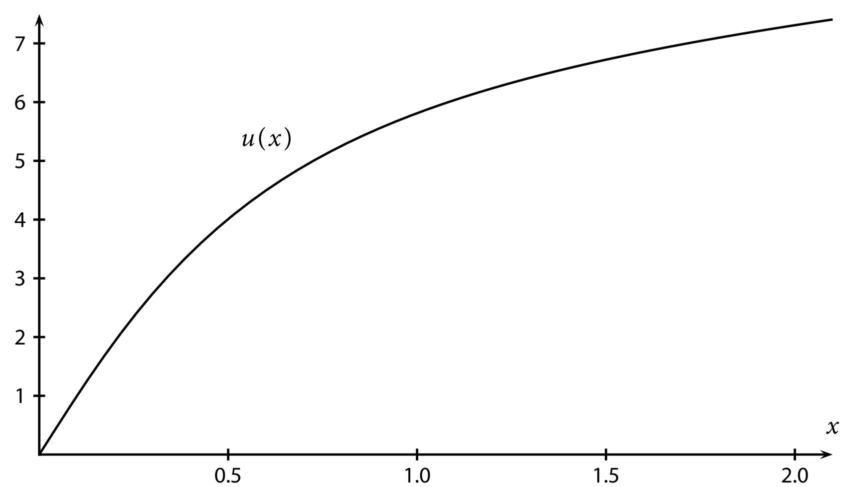
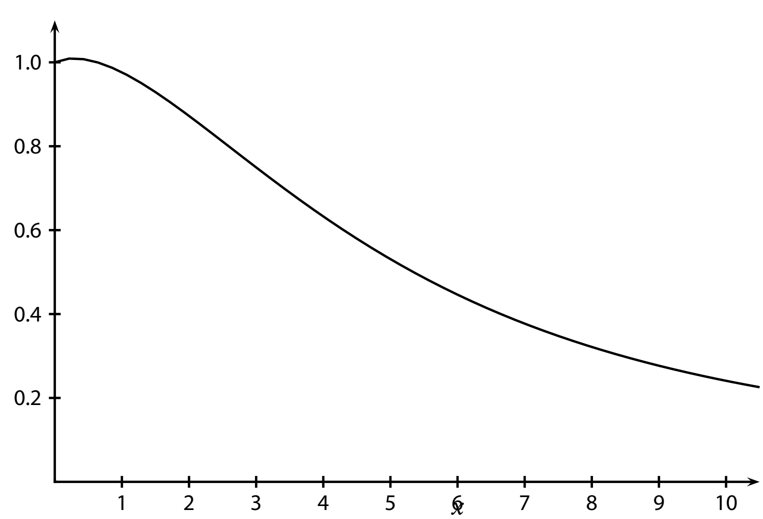


Figure 4.10

The graph of $u(x) = (20/(x^{3/2} + 20))x + (x^{3/2}/(x^{3/2} + 20))x^{1/2}$:



The graph of $u'(x)$ for the same utility function. Note that $u'(x)$ is not convex:



becomes larger when $\tilde{x}$ is less risky. This means that $xu'(w+\alpha x)$ is concave. Unfortunately, there is no simple condition on u that guarantees that this condition is satisfied. It is not true for the CARA utility function, for example. However, you can verify that $xu'(w+\alpha x)$ is concave when $u(z) = z^b$ for some $b \in (0,1)$.

Exercise 4.5. Show that if a utility function u is differentible and strictly increasing and exhibits constant or decreasing absolute risk aversion, then $u''' > 0$.

Exercise 4.6. Consider a firm that produces a quantity y of a single good at known cost $c(y)$. Profit when the price is p is $py - c(y)$. We compare two cases: (a) when the price is uncertain, equal to the random variable $\tilde{p}$ (the "uncertainty case"), and (b) when the price is certain, equal to $\bar{p} = E\left[\tilde{p}\right]$ for sure (the "certainty case").

a. Scenario: The firm chooses the level of output before observing the price of output. The owner of the firm is risk neutral with respect to profits.

1. State the maximization problems for the two cases and determine whether or not the two problems are equivalent.
2. How do the solutions and values of the two maximization problems compare?

b. Scenario: The firm chooses the level of output before observing the price of output. The owner of the firm is a risk-averse expected utility maximizer with respect to profits, for a strictly increasing and strictly concave utility function u. That is, utility when output is y and the price is p is $u(py - c(y))$.

1. Let $U(y)$ be the objective function for the uncertainty case and let $V(y)$ be the objective function for the uncertainty case. State the formulae for U and V. For which case is the maximization problem equivalent to when the owner is risk neutral?
2. Assuming only that each problem has a non-zero solution, compare the *values* of the maximization problems.
3. Prove: If y_1 and y_2 are solutions with and without uncertainty, respectively, then $y_1 \leq y_2$.
   (You should not use calculus nor add any auxilary assumptions on u. Instead, you should show that, if y_2 is a solution to the certainty problem and $y_1 > y_2$, then the profit given y_2 second-order stochastically dominates the profit given y_1. For this purpose, you can use the fact that, if $\tilde{x}$ is a non-degenerate random variable and if $\alpha_1, \alpha_2, \beta_1, \beta_2$ are such that $\alpha_1 > \alpha_2$ and $E\left[\alpha_1\tilde{x} - \beta_1\right] \leq E\left[\alpha_2\tilde{x} - \beta_2\right]$, then $\alpha_2\tilde{x} - \beta_2$ second-order stochastically dominates $\alpha_1\tilde{x} - \beta_1$.)
4. Assume for the rest of this part that c and u are differentiable. Provide the formulae for $U'(y)$ and $V'(y)$ and simplify the first-order condition in the certainty case.
5. Can you determine whether $U'(y) < V'(y)$ or $U'(y) > V'(y)$ for all y?
6. Show that, if y_2 is an interior solution in the certainty case, then $U'(y_2) < 0$.
7. Conclude that, if y_1 and y_2 are solutions with and without uncertainty, respectively, then $y_1 < y_2$.

c. Scenario: The firm chooses output *after* observing the price. The firm is risk neutral with respect to profit. Viewed from before observing the price, the decision problem of the firm is to choose a *plan* $y\colon \mathbb{R}_+ \to \mathbb{R}_+$ which states the level of output $y(p)$ as a function of the observed price p.

Assuming only that each problem has a solution, show that the (ex-ante) *value* of the problem in the uncertainty case is higher than that of the problem in the certainty case. Show that the values are the same only if, in the uncertainty case, there is a solution in which the output level does not depend on the price.

Exercise 4.7. Suppose that a person can work at a random hourly wage $\tilde{w}$ (always strictly positive), with mean $\bar{w}$ (the person might work on a commission basis, or be an entrepreneur). The money earned is used to buy a single consumption good with price 1. Let the utility from working x hours and consuming c units be

$$U(c, x) = u(c) - x,$$

where u is a strictly increasing, strictly concave function. The person maximizes expected utility.

a. Denote by $V(x)$ the expected utility as a function of the number of hours worked. Write down $V(x)$ in terms of x, $\tilde{w}$ and u.

b. Write down the first-order conditions for expected utility maximization.

c. Verify that the second-order condition for a stationary point to be a unique global maximum is satisfied.

d. Determine whether the person will work less or more the riskier the wage is, assuming that $u(c) = c^{1/2}$.

Exercise 4.8. A risk-averse expected-utility maximizer has initial wealth w and utility function u. She faces a risk of a financial loss of L dollars, which occurs with probability π. An insurance company offers to sell a policy that costs P dollars per dollar of coverage (per dollar paid back in the event of a loss). Denote by x the number of dollars of coverage.

a. Give the formula for her expected utility $V(x)$ as a function of x.

b. Suppose that $u(z) = -e^{-\lambda z}$, $\pi = 1/4$, $L = 100$, and $P = 1/3$. Write $V(x)$ using these values. There should be three variables, x, λ and w. Find the optimal value of x, as a function of λ and w, by solving the first-order condition (set the derivative of the expected utility with respect to x equal to zero). (The second-order condition for this problem holds but you do not need to check it.) Does the optimal amount of coverage increase or decrease in λ?

c. Repeat b, but with $P = 1/6$.

d. You should find that for either b or c, the optimal coverage is increasing in λ, and that in the other case it is decreasing in λ. Reconcile these two results.

e. The optimal x in your answers to b and c should not have depended on w. Why not?

f. Return to the general scenario. We have shown that a decision maker with differentiable utility should accept some stake in a favorable gamble. Using this fact, find the conditions on π and L under which the optimal level of coverage is (i) greater than L, (ii) equal to L, and (iii) less than L. Be clear, concise and *explicit*. You do not need to reprove the fact, and your answer should not involve any differentiation or even an expression for the decision maker's expected utility.

g. What does this problem tell you about whether, in practice, it is typically optimal to get full coverage for a financial loss?

Exercise 4.9. In Exercise 4.8, the monetary loss has two possible values, 0 and L. More generally, the monetary loss can be a random variable $\tilde{z} \geq 0$ with many possible positive values. There are two ways to extend the idea of partial coverage. One is to have coverage that pays a fraction α of each loss. The other is to have a deductible δ, such that if $\tilde{z} \leq \delta$, the insurance pays nothing, and if $\tilde{z} > \delta$, the insurance pays $\tilde{z} - \delta$.

Suppose that the loss can be 0, \$300 or \$900, each occurring with equal probability. Compare an actuarially fair policy that covers a fraction 1/2 of each loss with an actuarially fair policy that has a deductible of \$300. What is the premium charged in each case? What are the three possible outcomes for each policy? Show that the policy with fractional coverage is less risky than the policy with a deductible.

4.4 State-dependent utility

We can always define the set X of outcomes so that preferences are state-independent (we make sure the outcome captures everything the decision maker cares about). However, if we wish to restrict the set of outcomes to monetary values, then we will come across situations in which preferences over money are naturally state-dependent. In this section, you are asked to explore decision with state-dependent utility on your own, through several guided exercises.

Exercise 4.10. Think a moment about the following question: Should a risk-averse mother buy an insurance policy on her son's life, if that policy is actuarially fair?

Well, you might reason that the death of her son is a "risk", and being risk-averse, she should buy fair insurance against this risk.

The problem with this reasoning is that risk aversion is defined with respect to utility over money (or some one-dimensional outcome), and so we cannot decide a priori how a risk-averse person will treat other "risks" in her life that give her state-dependent preferences over money.

Let's suppose that in addition to her son's life/death, all the mother cares about is money. Then an outcomes can be written $\langle z, s \rangle$, where z is an amount of money, and

s is either

$$\begin{aligned} s_1 &= \text{son dies, or} \\ s_2 &= \text{son does not die.} \end{aligned}$$

Let $u(z, s)$ be the mother's VNM utility function.

s is also the state of the world. Preferences over outcomes as defined above are state independent because the preference-relevant aspects of the state are included in the outcomes. However, for this exercise, it is simpler to specify an outcome as just money, and let preferences be state dependent. It will also simplify the exercise if we let the outcomes be net transactions. I will now adopt these interpretations. This means that an act is a pair $\langle z_1, z_2 \rangle$, where z_1 is the net transaction in state 1 and z_2 is the net transaction in state 2.

Let π be the probability that her son dies. Suppose the mother can buy a life insurance policy on her son, which costs π dollars per \$1 of coverage. I.e., buying α units of insurance costs $\alpha\pi$ dollars, and pays out α dollars if the son dies. The policy is thus actuarially fair. We can allow α to be positive or negative; negative α means that the mother receives money when her son lives and she pays the company when her son dies (let's hope the insurance company will not go out of its way to collect on the policy).

In the questions below, $V(\alpha)$ is the mother's expected utility as a function of the level α of coverage.

a. If the mother buys α units of insurance, what act does she face? For $\pi = 1/3$, draw a picture of the acts she can face as α varies from -1,000 to 1,000. (This is part of the budget set in the space of acts.)

b. Suppose

$$u(z, s) = v(z) + w(s) ,$$

where v is a concave concave function. In words, the marginal utility of money is independent of the death of the son, and preferences over money exhibit risk aversion.

Write the formula for $V(\alpha)$. Group terms that depend on α and terms that do not. How much insurance should the mother buy?

(Be explicit in your answer. *Do not differentiate.* This involves basic ideas of risk and risk aversion and does not require solving first-order conditions.)

Illustrate graphically your answer by drawing a possible indifference curve in the space of acts through the optimal act in the budget set.

c. Suppose

$$u(z, s) = v(z + w(s)) ,$$

where v is concave, and $w(s_2) > w(s_1)$. In words, money and the son's life are perfect *substitutes*, with the imputed monetary value of the son's life equal to $w(s_2) - w(s_1)$. Because v is concave and $w(s_2) > w(s_1)$, the marginal utility of money is *higher* when her son dies.

Write the formula for $V(\alpha)$. Group terms that depend on α and terms that do not. How much insurance should the mother buy?

(Be explicit in your answer. *Do not differentiate.* This involves basic ideas of risk and risk aversion and does not require solving first-order conditions.)

Draw a possible indifference curve through the optimal act in the budget set for the case where $w(s_2) = 1000$ and $w(s_1) = 200$.

d. Suppose

$$u(z, s) = v(z)w(s) \,,$$

where $w(s_2) > w(s_1) > 0$ and v is strictly increasing and concave. In words, money and the son's life are *complements*. The marginal utility of money is *lower* when her son dies.

Write the formula for V. Assume that v is differentiable. Write down the first and second derivatives of V. Show that $V''(\alpha) < 0$, so that V is concave, and show that $V'(\alpha)|_{\alpha=0} < 0$. Does this imply that the optimal α is positive or negative? Explain. Draw a plausible graph of V that is consistent with what you have found.

Draw a possible preferred act and a possible indifference curve through the act.

e. Consider the following three cases:

1. Marginal utility of money is higher in state s_1 than in state s_2. In particular, $u'(0; s_1) > u'(0; s_2)$.
2. Marginal utility of money is the same in state 1 as in state 2. In particular, $u'(0; s_1) = u'(0; s_2)$.
3. Marginal utility of money is lower in state 1 than in state 2. In particular, $u'(0; s_1) < u'(0; s_2)$.

For each case, (i) find the sign of $V'(\alpha)|_{\alpha=0}$, (ii) infer from this whether the optimal amount of insurance is positive, and (iii) state whether the case applies to part b, c, or d.

f. Which of these scenarios seems more likely? What would you do?

g. Which case do you think best fits a man who doesn't love or even live with his wife but relies on her for the salary she earns?

Exercise 4.11. We have considered insuring against a monetary loss (e.g., life insurance when family members depend on the insured's income, getting partners to share risks in a business venture, disability insurance that protects against lost income, liability insurance that protects against lawsuits, sharing the riskiness of income with family members). Exercise 4.10, we look at insurance when the risk is simply something that may affect your preferences over money (e.g., a child's death, getting a date). In this problem, we discuss insurance when the risk affects your utility from acquiring a specific good. It is the specificity of the good that makes us want to model this differently from the previous question.

Examples:

Health insurance The value of knee surgery depends on whether you have knee problems or not.

Auto collision insurance The value of a car repair or a second car depends on whether your first car gets damaged in an accident.

Fire insurance The value of rebuilding your house or buying a new one depends on whether your house burns down.

In the simplest model, there are two goods, x and y, such that your utility from purchasing x is affected by some risk, and y represents everything else. Assume that the VNM utility for x and y is separable. I.e.,

$$u(x, y; s) = u_x(x, s) + u_y(y)$$

Let there be two states, s_1 and s_2. State s_2 is when something bad happens that makes consumption of good x more important. Specifically, suppose there is $v(x)$ such that

$$\begin{aligned} u_x(x, s_1) &= v(x) \\ u_x(x, s_2) &= 2v(x). \end{aligned}$$

Assume further that both v and u_y are strictly concave in x and y, respectively, and that they are differentiable.

Finally, assume your baseline income I is state independent and that you can buy actuarially fair insurance that reimburses you in state s_2.

You are to show the following:

- Consumption of y is the same in both states.
- Consumption of x is higher in state s_2 than in state s_1.
- Demand for the insurance is positive.

I will not walk you through the solution, but I will give a few suggestions on how to answer this question. Think of this as a consumer choice problem with four goods: x in state 1, x in state 2, y in state 1, y in state 2. Write down the consumer's utility function over these four goods, and the consumer's budget constraint. The prices of these four goods that appear in the budget constraint are a function of the prices of x and y (which are not state dependent) and of the state prices.

You can then answer the first two parts of the question using the fact that at an optimum, the marginal rates of substitution are equal to the relative prices. That is, for goods k and j, $\frac{\partial u/\partial x_j}{\partial u/\partial x_k} = \frac{p_j}{p_k}$. Define any additional notation that you use.

Exercise 4.12. Consider a life insurance policy that costs p dollars per dollar of coverage. The relevant set of states of the world is {die, live}. We can think of the insurance as an asset that pays \$1 in state "die" and \$0 in state "live". Suppose that there is also a riskless asset that costs \$1 and pays R dollars in either state. Assume that the riskless asset can be sold short (i.e., it is possible to take out bank loans).

a. Describe the portfolio that has 1,000 units of insurance and has zero net cost. I.e., say how many units of each asset are in this portfolio, and give the payoff on the portfolio in each state.

b. Explain, by way of an example, why the VNM model does not apply to preferences over portfolio payoffs in this model.

Exercise 4.13. Suppose Biff has asked Bonnie out for a date on Friday, and is awaiting her reply. He figures that Bonnie will accept with probability 1/3. In the meantime, he negotiates with his father for some cash for the date. His father, who can never give a straight answer, offers him the following options:

A If Bonnie accepts the date, Biff gets \$40. If Bonnie does not accept the date, Biff gets \$10.

B If Bonnie accepts the date, Biff gets \$15. If Bonnie does not accept the date, Biff gets \$24.

Show that B second-order stochastically dominates A. Assuming that Biff is risk averse in his preferences over money lotteries, can we conclude that Biff would/should take offer B? Why or why not? What is the relationship between this question and the theory of insurance?

Chapter 5

Markets for state-contingent contracts

We say there are gains from trade for a group of people if they can make trades that leave one or all of the people better off, but no one worse off. Gains from trade are due to heterogeneity. For example, there are gains from trade between countries because countries have different relative quantities of the factors of production and different preferences over final goods.

State-contingent contracting, which we call risk sharing, is a form of trade; people are trading wealth in the various states. Section 5.1 studies the gains from state-contingent contracting which exist because people have heterogeneous state-dependent wealth and state-dependent preferences. In the previous chapter (e.g., Section 4.1.2), we considered demand for assets by risk-averse traders with state-independent wealth and preferences. Such traders will only buy risky assets if the expected excess return is positive. But none of these traders will sell such an asset, which must therefore be issued by traders with state-dependent wealth or preferences, who are willing to be paying a premium in expected wealth in return for reducing their risk. Examples of such traders are the purchasers of insurance who we modeled in Section 4.1.3.

There are many ways in which people share risks through state-contingent contracts. The first way that comes to mind are insurance contracts. The risks for the insured are shared among the many shareholders or policy holders (in the case of a mutual insurance company). Competitive insurance markets are studied in Section 5.2. Financial markets are another very important way. For example, banks sell bundles of mortgages in asset markets in order to diversify risks: firms issue common stock to share the risks among shareholders; farms sell commodity futures to insure against uncertainty after harvests; and multinational firms buy forward contracts for foreign currency in order to reduce their exposure to the risk of fluctuations in foreign exchange rates. We study financial markets in Chapter 6.

Social institutions, social norms, and other informal arrangements are also extremely important means for sharing risks. Governments provide disaster relief, subsidized care of the mentally handicapped, and welfare payments, without which we would bear much greater risks in our lives. Family members help those with financial misfortune and share in financial success. Such co-insuring is also common between families in traditional communities such as villages in developing countries. We will not study these informal mechanisms, but their importance should not be underestimated.

5.1 Gains from sharing risks

5.1.1 Pareto efficiency and individual rationality

Imagine an abstract setting in which there are n agents, with names $i = 1, 2, \ldots, n$, and a set X of social outcomes. X might be the result of political elections or international agreements about environmental protection, or might be a decision about what a family will have for dinner or what TV show a family will watch. Each of the agents has preferences over X, which we represent by a utility function. Suppose that the agents meet and collectively choose one of the outcomes from X. Assume also that each agent can guarantee himself a utility level u_i just by refusing to be part of an agreement. Agent i's no agreement outcome is called his *outside option*, and u_i is his *reservation utility*. For example, in markets, u_i is the utility from not trading at all.

If the agents can meet and collectively choose an outcome without any transaction costs, then here are two properties we may expect of any agreement:

1. Each agent should at least get her reservation utility, since otherwise she should refuse to participate in the agreement. The outcome is then said to be *individually rational*.
2. If all parties know each other's preferences and reservation utilities, then it should be impossible to find some other outcome that at least one party prefers and no party likes less. The outcome is then said to be *Pareto efficient*.

Consider two outcomes A and B. If at least one person strictly prefers B over A and no one strictly prefers A to B, then B is said to *Pareto dominate* A (or to be *Pareto superior* to A or to be a *Pareto improvement* over A). Hence, a feasible outcome A is Pareto efficient if there is no feasible Pareto dominating outcome.

Let's return to the specific setting of risk sharing. Each agent's outside option is the state-dependent allocation $\tilde{w}_i$ of wealth she gets if she does not trade or sign any state-contingent contracts. This is called her endowment or pre-contracting allocation. Contracting involves state-contingent transfers of wealth between the agents. If agent i gets net transfers $\tilde{x}_i$, then his final or post-contracting allocation is $\tilde{w}_i + \tilde{x}_i = \tilde{z}_i$.

An allocation for the n agents is a list specifying the allocation for each of the agents. Because the net transfers have to balance ($\sum_{i=1}^n \tilde{x}_i = 0$), the total wealth in each state for a feasible allocation $\langle \tilde{z}_1, \ldots, \tilde{z}_n \rangle$ must equal the total wealth for the endowment $\langle \tilde{w}_1, \ldots, \tilde{w}_n \rangle$. That is, $\sum_{i=1}^n \tilde{z}_i = \sum_{i=1}^n \tilde{w}_i$. An allocation is individually rational if each agent weakly prefers her allocation to her endowment. An allocation is Pareto efficient if there is no further feasible trade that leaves at least one agent better off and no one worse off.

5.1.2 Depicting trade with an Edgeworth box

Consider the following story, which is an allegory for more important risk-sharing situations:

> Soze and Keyser are at summer camp together. On Friday afternoon, the campers will receive a little money from their parents which they use to buy

candy and such from the camp store. It is now Friday morning. Soze knows she will get \$10, because her parents always send \$10. However, Keyser's parents are unpredictable. Both campers believe that Keyser will get \$16 (state 1) or \$6 (state 2), with equal probability.

There is a nice graphical method for illustrating trade and gains from trade, called an Edgeworth box. We draw the set of trades, and then simultaneously show each parties preferences over the trades. For the set of trades to be 2-dimensional, we need there to be two agents and two states of the world. A trade is then a transfer in state 1 and a transfer in state 2.

Here is how it works for the example of Keyser and Soze. We start by drawing the set of allocations for Keyser, which is shown at the top of Figure 5.1. I have also marked his endowment $\tilde{w}_k = \langle 16, 6\rangle$, and the allocation $\tilde{z}_k = \langle 12, 9\rangle$ he gets if he signs a contract with Soze whereby Keyser gives Soze 4 dollars in state 1 and Soze gives Keyser 3 dollar in state 2.

Then we draw the set of allocations for Soze, as shown in the middle of Figure 5.1. I have marked her endowment $\tilde{w}_s = \langle 10, 10\rangle$ and the allocation $\tilde{z}_s = \langle 14, 7\rangle$ she gets when she signs the same contract described above.

What we would like to do is combine these two pictures. The trick is that, when we know Keyser's allocation, then we know Soze's allocation from the resource constraint. Suppose Keyser gets wealth 14 in state 1 and 8 in state 2. Since total wealth is 26 in state 1 and 16 in state 2, we know that Soze's wealth is 12 in state 1 and 0 in state 2.

We can thus superimpose the two sets of allocations, as shown at the bottom of Figure 5.1. We rotate Soze's axes 180° degrees, because more for Keyser means less for Soze. Soze's origin is the point where Keyser gets all the wealth: $\langle 26, 16\rangle$. The endowment points and the post-trade allocations line up. The dimensions of the box are the total endowment: 26×16. The two 45° lines do not line up because when one party has a risk-free allocation, the other party is bearing all the risk.

Now we can depict both parties preferences over feasible allocations in the Edgeworth Box. The top of Figure 5.2 shows the allocations weakly preferred to the endowment by Keyser as a grey region in Keyser's set of acts. The middle of Figure 5.2 shows the same set for Soze as a crosshatched region. At the bottom, we see how these two appear in the Edgeworth Box. The intersection of these two regions is the set of allocations that are individually rational.

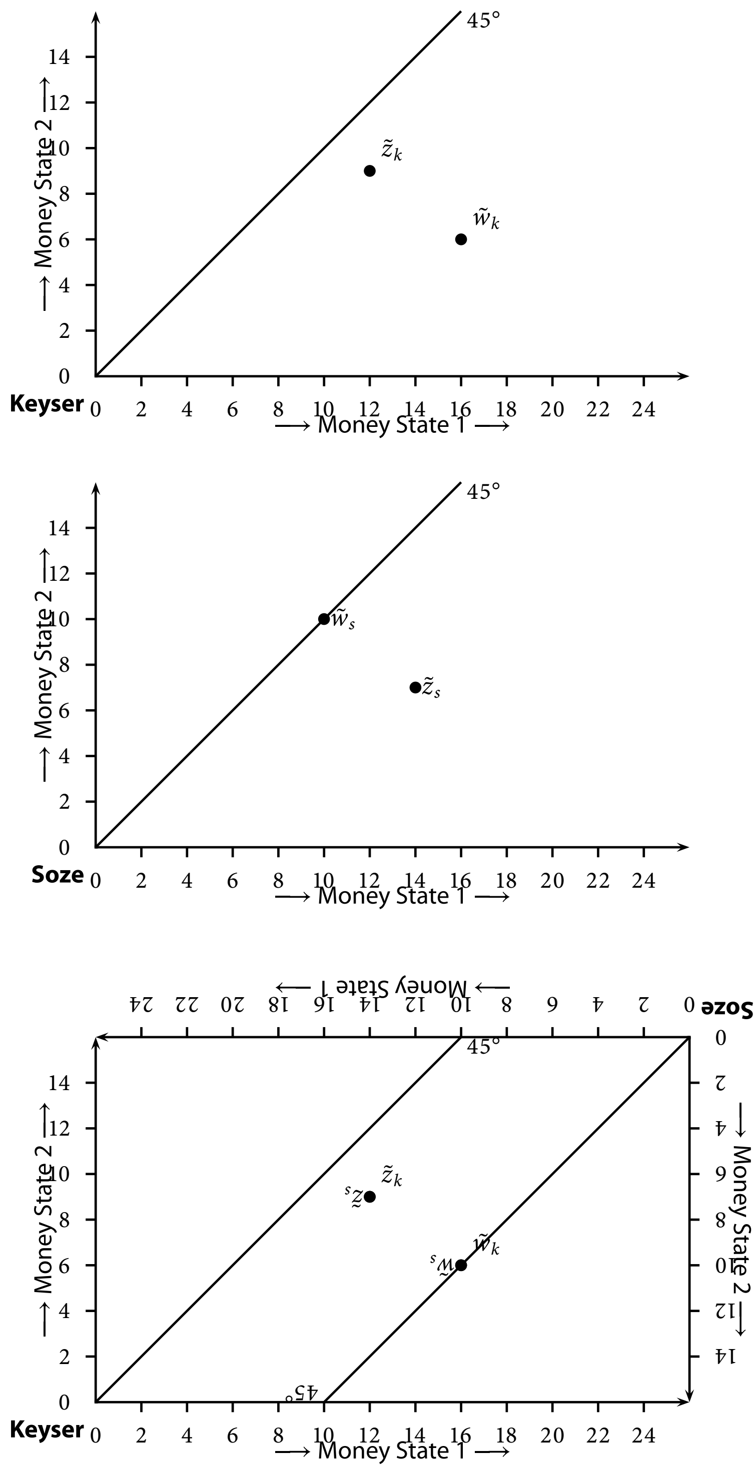
If an allocation is Pareto efficient, then the set of acts strictly preferred by Soze must not intersect the set of acts strictly preferred by Keyser. This means that the indifference curves through the allocation are tangent, in the sense that they touch at the allocation but do not cross each other. This is shown in Figure 5.3

In an Edgeworth box, the set of allocations that is Pareto efficient is called the *contract curve*. This is because it typically is a curve, and because we expect people to sign contracts that are Pareto efficient. The contract curve is shown in Figure 5.4. The portion that lies between the two indifference curves through the endowment is the set of individually rational and Pareto efficient allocations.

5.1.3 Some properties of efficient risk sharing

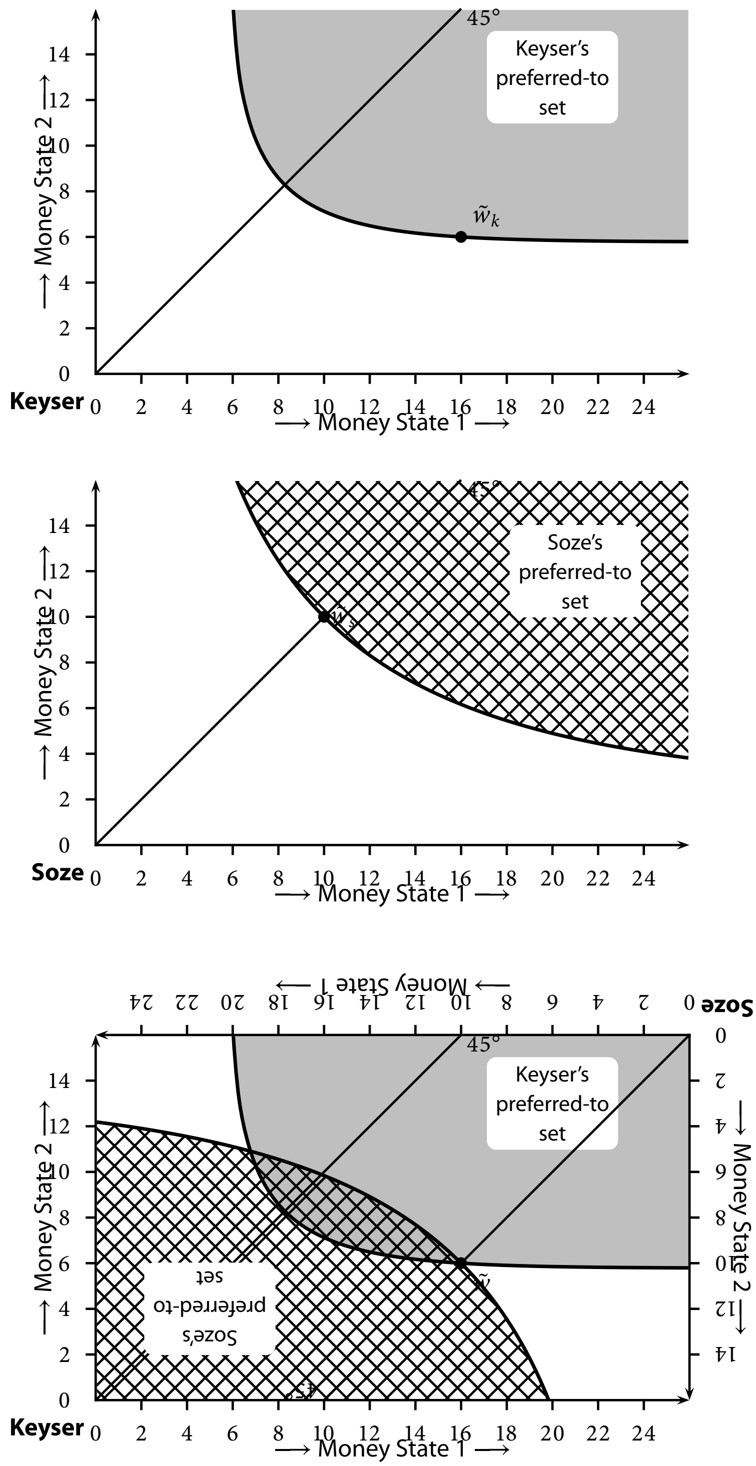
The following is assumed through this section:

Figure 5.1



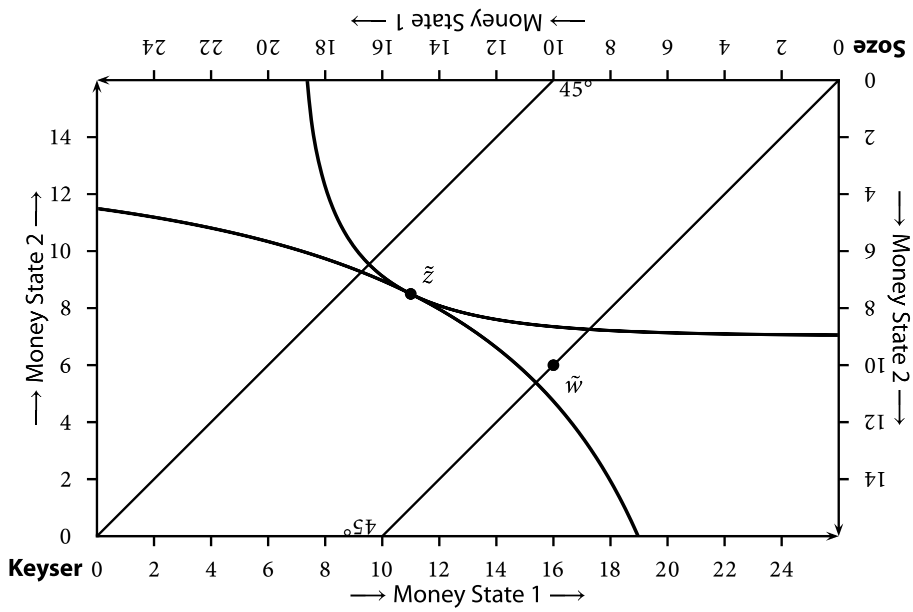
The sets of allocations for Keyser and Soze, and their superimposition. $\tilde{w}_i$ is the endowment of each agent, and $\tilde{z}_i$ is the allocation each agent ends up with if Keyser gives Soze 4 dollars in state 1 and Soze gives Keyser 3 dollar in state 2.

Figure 5.2



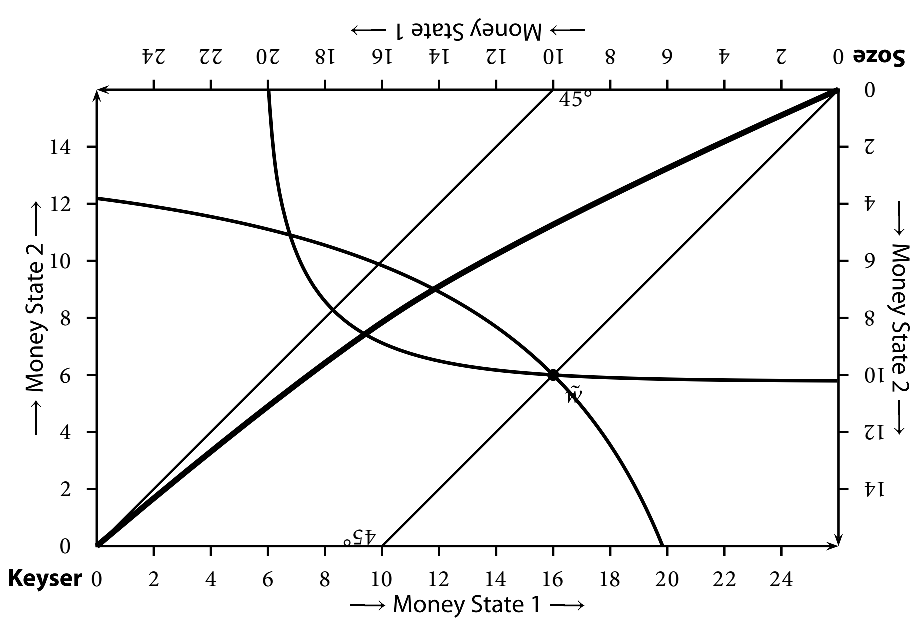
The indifference curve and preferred-to set the Keyser's endowment and for Soze's endowment, and their superimposition.

Figure 5.3



The indifference curves through a Pareto optimal allocation must be tangent, since the preferred to sets cannot intersect.

Figure 5.4



The line that runs diagonally is the set of Pareto efficient allocations for the example (which depends on the utility functions I am using). In an Edgeworth box, this set is called the contract curve.

Assumption 1. All agents have state-independent preferences and have the same beliefs.

Even when the endowment of each agent is random, the total endowment might not be. For example, consider two firms competing for a defense contract. One and only one firm will get the contract. For each firm, sales are risky. However, the total sales of the firms are not. We say that, although there is *individual risk*, there is no *aggregate risk*.

Proposition 1. *Suppose the agents are risk averse and there is no aggregate risk. For any efficient allocation, each agent's allocation is riskless.*

This is the first of three propositions we will prove in this section. Each proposition and proof has the same logic. The conclusion of each proposition has the form: "If an allocation is efficient, then it has Property Q." In this first proposition, Property Q is that no agent bears any risk. Rather than proving directly "If an allocation is efficient, then it has Property Q," we prove the contrapositive: "If an allocation does not have Property Q, then it is not efficient." You can see that the original statement and its contrapositive are logically equivalent, which is why it is OK to prove the contrapositive. In the proof, we then have to show that an allocation is not efficient. To do so, we construct a Pareto dominating allocation, using the assumption that the original allocation does not have Property Q.

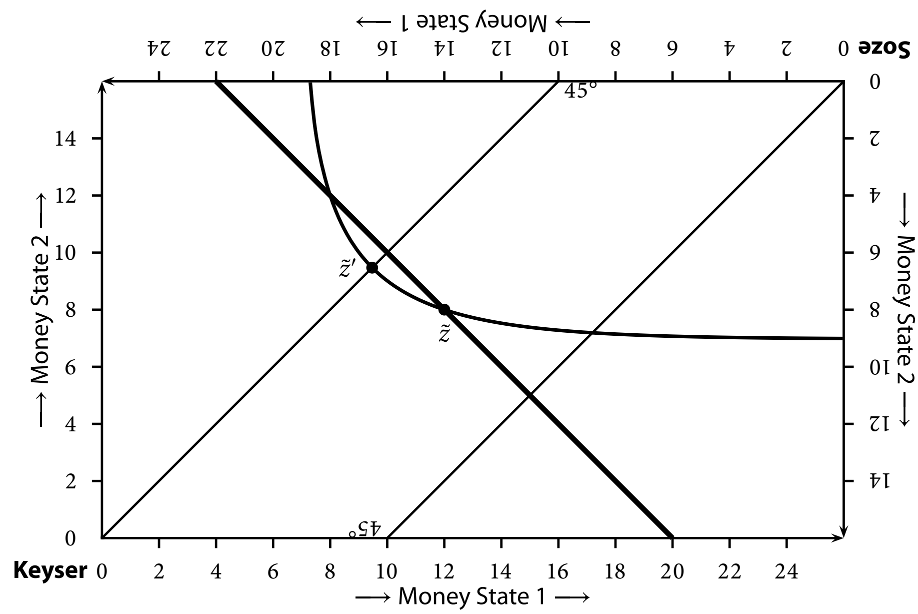
Proof. Suppose all agents are risk averse and there is no aggregate risk. Let $\langle \tilde{z}_1, \ldots, \tilde{z}_i \rangle$ be a feasible allocation such that the allocation for at least one person is random. $\sum_{i=1}^{n} \tilde{z}_i$ is not random, since there is no aggregate uncertainty. If each agent gets $E[\tilde{z}_1]$ for sure, then total wealth is the same $w \sum_{i=1}^{n} E[\tilde{z}_i] = E[\sum_{i=1}^{n} \tilde{z}_i] = \sum_{i=1}^{n} \tilde{z}_i$, and hence the allocation is also feasible. Everyone weakly prefers $E[\tilde{z}_i]$ to $\tilde{z}_i$, and anyone for whom $\tilde{z}_i$ is risky strictly prefers $E[\tilde{z}_i]$ to $\tilde{z}_i$. Hence, the original allocation is not Pareto efficient. □

In an Edgeworth box, when there is no aggregate risk the box is square and the 45° lines for the two agents coincide. The only place the two agents' indifference curves can be tangent is along the 45° line, because the slopes of the indifference curves along this line are determined by the probabilities of the two states.

Proposition 2. *If one or more of the agents are risk neutral, then no risk-averse agent bears any risk in an efficient allocation.*

Proof. Suppose that the allocation of one of the risk-averse agents is random. Suppose that this agent and a risk neutral agent sign a contract such that the risk-averse agent's resulting allocation is non-random and equal to his certainty equivalent. Then the expected value of his allocation has fallen by the amount of his risk premium for the original allocation, but (by definition of the certainty equivalent) his is neither better nor worse off. In contrast, the expected value of the risk-neutral agent's allocation has increased by the amount of the risk premium, and so she is better off. Thus, this trade is a Pareto improvement, which means that the original allocation was not Pareto

Figure 5.5



Keyser is risk averse and Soze is risk neutral. The straight line through $\tilde{z}$ is Soze's indifference curve through this allocation; the curved line is Keyser's indifference curve. The allocation $\tilde{z}$, in which Keyser bears risk, is Pareto dominated by $\tilde{z}'$, because Keyser is no worse off but Soze is better off. Any of the allocations in the region bounded by the two indifference curves also Pareto dominates $\tilde{z}$.

efficient. □

This fact is illustrated in the Edgeworth box in Figure 5.5. Soze is the risk-neutral agent. Recall that an indifference curve for agent Soze is just a line perpendicular to the vector of probabilities. This vector of probabilities for Soze points in the opposite direction as the vector of probabilities for Keyser, but they are colinear. Hence, an indifference curve for risk averse Keyser can only be tangent to Soze's indifference curve on Keyser's 45° line, but not elsewhere. Figure 5.5 shows an allocation $\tilde{z}$ in which Keyser bears risk, and a Pareto dominating allocation $\tilde{z}'$, like the one constructed in the proof of the proposition.

The next proposition shows local risk neutrality in action. The assumption that the total endowment is random means that it is not the same in each state and hence there is aggregate risk.

Proposition 3. *If the total endowment is random and every agent is risk averse, then every risk-averse agent with differentiable utility bears some risk in an efficient allocation.*

Proof. Suppose one agent, who I will call Soze, bears no risk. Let her allocation be z_s. Because the total endowment is random, there is some other agent, which I will call Keyser, whose allocation $\tilde{z}_k$ is random. Let CE_k be the allocation that is equal to Keyser's certainty equivalent for $\tilde{z}_k$. Let $\tilde{x} = \tilde{z}_k - CE_k$, and suppose Keyser trades

$\alpha\tilde{x}$ to Soze, so that Keyser's allocation is $\tilde{z}_k - \alpha\tilde{x}_k$ and Soze's allocation is $z_s + \alpha\tilde{x}$. Since Keyser is risk averse, $CE_k < E[\tilde{z}_k]$, and hence $E[\tilde{x}] > 0$. Since Soze is locally risk neutral, there is α such that $0 < \alpha < 1$ and Soze prefers $z_s + \alpha\tilde{x}$ to z_s. Keyser is indifferent between $\tilde{z}_k$ and CE_k. $\tilde{z}_k + \alpha x_s$ is on the line connecting $\tilde{z}_k$ and CE_k, and so Keyser weakly prefers $\tilde{z}_k + \alpha x_s$ to both $\tilde{z}_k$ and CE_k. Hence, the trade $\alpha\tilde{x}$ is a Pareto improvement. □

This is illustrated at the bottom of Figure 5.2 in a Edgeworth box for the case where Soze bears no risk and Keyser bears risk. Because there is aggregate uncertainty, Keyser's and Soze's 45° lines do not coincide. The initial allocation $\tilde{w}$ is on Soze's 45° line. Her indifference curve through $\tilde{w}$ is thus perpendicular to the vector of probabilities. Keyser's indifference curve through $\tilde{w}$ cannot be perpendicular at $\tilde{w}$ to the probabilities since $\tilde{w}$ is not on Keyser's 45° line. Hence, the two indifference curves cross, and there are further gains from trade.

Exercise 5.1. What is wrong (and what is right) with the following: "Options markets are like gambling halls—what one person wins, another loses. Therefore, they are socially wasteful, especially given that they are costly to operate."

Exercise 5.2. Consider the Edgeworth box in Figure 5.2 for state-contingent trading between two traders with two states. The dimensions of the Edgeworth box tells you that the total wealth in states 1 and 2 is 26 and 16, respectively, but otherwise you are not told the initial allocation of wealth. The other information you have is that both traders have state-independent preferences, are risk averse, and assign the same probabilities to the states.

The following two questions are related, and it will help you to think about them together. Your answer should include an explanation, and you should draw on the graph to illustrate the answer.

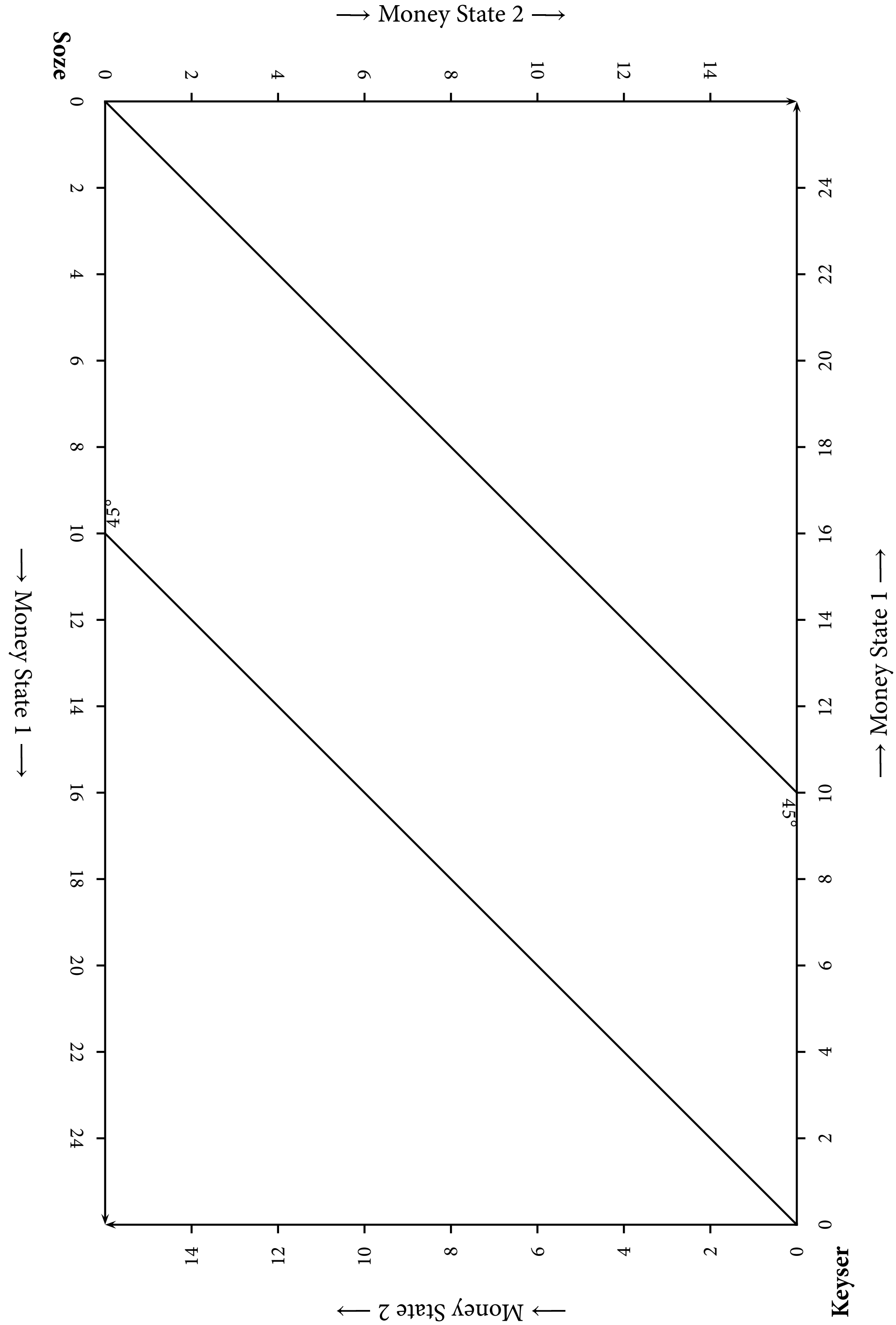
1. With just this information, what can you tell me about the equilibrium allocation? (That is, you should be able to identify a region in the box where the equilibrium allocation must lie).
2. Suppose that I tell you that the probabilities of the two states are equal. What can you tell me about the relative state prices?

5.2 Insurance market

The following is a very rough sketch of insurance markets.

The important characteristic of insurance markets is that they are means to share many relatively small and similar but uncorrelated or *idiosyncratic* risks.

Each person who shares a small amount in a large pool of such risks actually bears little risk. Therefore, as an approximation we can assume that insurance companies are



Graph for Exercise 5.2.

risk neutral. It follows that, in the absence of administrative costs, insurance policies in competitive insurance markets are actuarially fair. The rest of this section elaborates on the sharing of many small risks.

5.2.1 Sharing risks in a small pool

Suppose that there are n people who face identically and independently distributed risks $\tilde{x}^1, \ldots, \tilde{x}^n$, with mean $\bar{x}$ (final wealth is baseline wealth plus $\tilde{x}^i$).

If $n = 2$, and these two people share the risks equally, then they end up with the gamble

$$(1/2)\tilde{x}^1 + (1/2)\tilde{x}^2$$

This is less risky than either $\tilde{x}^1$ or $\tilde{x}^2$, as we showed in a portfolio problem where one invests in two assets with IID returns. (Although this may not be the optimal risk-sharing rule, because the risk preferences of the two people may differ.)

5.2.2 Sharing risks in a large pool

Returning to the general case of n people, if these risks are shared exactly, then each person faces the risk:

$$\tilde{X}_n = (1/n) \sum_{i=1}^{n} \tilde{x}^i$$

We can also show that the larger is n, the less risky is $\tilde{X}_n$. Hence, there are gains to increasing the number of people in the pool.

Proposition 4. (Weak Law of Large Numbers) *For all $\varepsilon > 0$,*

$$P\left\{|\tilde{X}_n - \bar{x}| > \varepsilon\right\} \to 0 \qquad \textit{as} \qquad n \to \infty.$$

This means that for any given amount, however small, the probability that a person share of the pooled risk deviates from the mean of the risk by this amount goes to zero as the number of participants goes to infinity.

This also means that the expected utility from participating in the pool converges to the utility of the expected value of the risk as $n \to \infty$. That is, by participating in the pool, you can effectively dissipate all risk if the pool gets large enough.

The weak law of large numbers requires independence of risks and uniformly bounded variances, but not that risks be identically distributed.

5.2.3 A caveat about the Law of Large Numbers

The riskiness of the average wealth $(1/n)\sum_{i=1}^{n} \tilde{x}^i$ decreases with n, *but the riskiness of the total wealth $\sum_{i=1}^{n} \tilde{x}^i$ does not.*

To illustrate, if the $\tilde{x}^i$'s all have variance σ^2, then the variance of the average is $(1/n)\sigma^2$, but the variance of the total wealth is $n\sigma^2$.

Therefore, we cannot say that insurance companies are approximately risk neutral simply because they have many customers with independently distributed risks.

Rather, they are approximately risk neutral because they have many shareholders who together bear the independently distributed risks of their many customers.

By sharing many such risks among many people, each person sharing the risk bears very little risk. This is why, as an approximation, we will treat insurance as if they were risk neutral. It then follows that, in the absence of administrative costs and informational asymmetries, insurance policies in competitive insurance markets will be actuarially fair.

Exercise 5.3. You can take units in this problem to be in millions of dollars.

Suppose that an investment opportunity matures in one year and pays \$1 with probability 9/10 and \$0 with probability 1/10. The profits of the investment are initially owned by a single individual, but that individual has decided to sell the opportunity to investors, for whatever reason. For simplicity, suppose that the owner issues a single infinitely-divisible share, whose endogenous market price is p (i.e., p dollars gets you ownership of the entire return on the investment). Suppose that there is one other riskless asset whose return per dollar invested is exactly \$1 and whose price is fixed exogenously at \$1. The buyers of the firm will "sell" this riskless asset (borrow money) in order to pay for the shares of stock and the seller of the firm will buy this riskless asset (loan money) with the proceeds of the sale of the stock. (That is, there is no consumption in the period in which the trading occurs.)

Suppose that there are N risk-averse investors with the same CARA utility function $u(z) = -e^{-\lambda z}$. We can ignore the investors initial wealth because with CARA utility wealth does not affect risk preferences.

a. Let $V(\theta, p)$ be an investor's expected utility when purchasing θ units of the asset at price p.

b. Let $\theta(p)$ be each investor's demand for the asset, as a function of the price. Derive $\theta(p)$ by differentiating $V(\theta, p)$ with respect to θ and solving the first-order condition. (You do not need to check the second-order condition.) Your answer should give θ as a function of p and λ, and should have a logarithm.

c. The equilibrium condition is that $N\theta(p) = 1$. From this condition, solve for p as a function of N and λ.

d. Show that (for fixed N) as $\lambda \downarrow 0$, i.e., as the investors become more risk neutral, the price increases to 9/10, which is the expected return on the asset.

e. Show that (for fixed $\lambda > 0$) as $N \uparrow \infty$, i.e., as there are more and more investors, the price increases to 9/10.

f. What does this say about whether large, publically traded corporations are less risk averse than individually owned companies?

Chapter 6

Asset Markets

6.1 The nature of asset markets

6.1.1 Asset markets versus bilateral contracting

We have seen that preferences over market transactions can be affected by many random factors such as random endowments of labor and wealth and random productivity. We have also seen how this creates gains from sharing risks via state-contingent contracts.

Such state-contingent contracting occurs in a variety of ways. In many cases, such as contracting between insurance companies and customers, between landowners and sharecroppers, and between banks and borrowers, the contracts are negotiated directly between two parties and the contracts are non-transferable. The disadvantage of such bilateral contracting is that it is cumbersome if risks are to be shared among many parties or if many parties face similar risks. For example, it would be impractical for the government to finance debt by writing a separate contract with every potential investor, or for each corn grower to individually negotiate corn forward contracts with speculators.

Trading of financial assets is a means of state-contingent contracting that has lower transaction costs when many individuals are involved. A financial asset is *a transferable contract with fixed terms* whose price is determined endogenously in asset markets. Because the terms are fixed, there is less to negotiate in the markets. Financial assets are better than bilateral contracts when the contingencies of the contract (e.g., whether or not a firm is bankrupt) are easily observable by the many potential investors and when there is enough volume to justify the costs of issuing an asset. For these reasons, home mortgages are not financed by selling bonds, but rather are financed bilaterally between a homeowner and a bank; the bank takes care of verifying the payment or default of the mortgage.

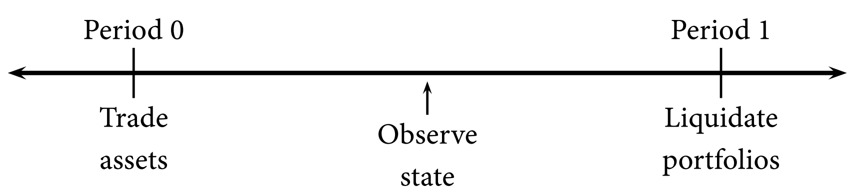
In this Section, we will study the trading of financial assets in asset markets. We will see that in the absence of transaction costs or problems with verifying payments, all state-contingent contracting could be done through trading of financial assets. It is also true that in the absence of transaction costs, all contracting could be done through bilateral, non-transferable contracts. The mix of financial assets and bilateral contracts is due to the actual transaction costs and information asymmetries, but we will have no more to say about this here.

6.1.2 Intertemporal versus "interstate" trading

Asset markets have two roles. The one we are mainly interested in here is risk-sharing. The state-contingent contracts allow traders to reduce risk by shifting their wealth between states (i.e., lose money in some states in return for gaining money in other states). However, another purpose of asset markets is to allow traders to shift wealth intertemporally (between one period and another). Some assets, such as a contract that guarantees future sale of oil at a fixed price, mainly shift wealth between states rather than between periods. Markets that are mainly tools for risk-sharing are often called *forward markets*. Other assets, such as government bonds issued to finance public spending or corporate bonds issued to build a new factory, are mainly devices for raising capital (for the borrowers) or saving money (for the lenders). Markets that are mainly tools for such intertemporal trading are called *capital markets*.

In the previous paragraph, we attempted to separate the roles of risk-sharing and intertemporal trading. However, in practice, these two roles of asset markets are inseparable. On the one hand, capital contracts always involve some kind of state-contingent payments, even if simply because the ability of the borrower or issuer to repay is state-dependent. On the other hand, state-contingent contracts must always be entered into *before* the state is learned by the parties, and typically some payment is made when the contract is signed.

A consequence is that a model of asset trading must be intertemporal and there must be uncertainty. Our goal is to develop a *simple* model, and so we assume that there are only two periods, $t = 0$ and $t = 1$. Uncertainty is represented by S states, $s = 1, \ldots, S$. Assets are traded in period 0, then everyone learns the state, and then the assets mature and payoffs are paid in period 1. The sequence of events is illustrated here:



By assuming that there are only two periods, we cannot study the affect of gradual revelation of information on asset prices over time and the relationship between short-term and long-term assets. Also, by assuming that no one knows the state when asset trading occurs and everyone knows the state when the assets mature, we are abstracting from asymmetries of information (heterogeneous information of traders) that are common in asset markets. Hence, the model does not capture insider trading, adverse selection, or moral hazard. We will touch on some of these asymmetries later in the course, but not in the context of a general model of asset markets.

6.1.3 Asset markets versus spot markets

In addition to asset markets, there are *spot markets* in which goods are traded and exchanged immediately. You have probably heard of the spot market for oil, which refers to oil that is traded at prices that adjust daily and are not fixed in advance by any contract. Our use of the term "spot market" includes the spot market for oil, but it is much more general. Your daily purchases of food and clothing are also spot-market transactions because the prices are not agreed to in advance and you pay for the goods immediately.

In our model, traders use whatever wealth they have left over from asset trading in period 0 to buy goods in the period 0 spot market, for their immediate assumption. When period 1 arrives, the traders observe the state, liquidate their portfolios, and go off to the spot commodity market with the payoffs on their portfolios, and whatever other wealth they have. Even if there were no asset trading, the equilibrium prices and trades in the spot market in period 1 would be state-dependent, because uncertainty affects the traders' endowments of goods (e.g., quality of labor) and preferences over goods (e.g., demand for building construction is affected by earthquakes). However, the asset trades have an additional state-dependent effect on spot market prices because the portfolio payoffs, which represent transfers of wealth between traders, are state-dependent.

Conversely, the expected spot prices in each state affect the asset trades, because the spot prices determine what commodities can be bought with the portfolio payoffs. This has brought us *almost* full circle: spot prices depend on asset prices, and asset prices depend on *anticipated* spot prices. A common modeling assumption is that traders anticipate spot prices correctly (i.e., they anticipate correctly how spot prices depend on the state). This is a type of *rational expectations* assumption. With this assumption, we have indeed come full circle, and the asset prices and state-dependent spot prices are determined simultaneously in equilibrium. This is called an *equilibrium of plans, prices and price expectations*, because, in equilibrium, traders have the same expectations about future spot prices, they observe the asset prices in the asset markets, the asset trades are balanced, and anticipated (planned) commodity trades in each spot market are balanced.

This should give you a flavor of the link between asset markets and spot commodity markets, but to simplify this discussion we will not attempt to model the simultaneous determination of asset and spot prices. Instead, we treat the commodity prices as being exogenous. It is then possible to derive preferences over money in each state from preferences over consumption. Suppose z_0 is a trader's money in the asset market (in period 0); z_0 is the trader's baseline wealth in period 0 minus the cost of the portfolio. Let z_s be the trader's wealth in state s in period 1; this is the trader's baseline wealth in state s plus the payoff on the trader's portfolio in state s. Then let

$$U_i(z_0, z_1, \dots, z_S)$$

be the trader's maximum expected utility from consumption given that he spends his wealth optimally in each spot market at the exogenous prices. U might be consistent with expected utility maximization (but with state-dependent preferences) and might be additive over time.

6.1.4 Nominal versus real payoffs

We now need a description of the assets. The short story is that an asset is described by its payoff in each state. Let's go over the long story too.

We distinguish between nominal assets (also called financial securities) and real assets. Nominal assets are those whose payoffs are given in dollars. Examples are bonds, bank loans, and insurance contracts. Real assets are those whose payoffs are given in goods. Examples are commodity futures, stocks and gold. The payoffs of some assets are both nominal and real. For example, if I sign a contract to buy one million barrels of oil on June 1 at \$30 a barrel, then I have an asset whose payoff is 1 million barrels of oil and −30 million dollars. The discussion below on both nominal and real payoffs applies to such mixed-payoff assets.

Once we know the prices in the spot markets, we can translate real payoffs into nominal payoffs, and vice-versa. The nominal payoff of a real asset in state s is the value, at spot prices in state s, of the goods the asset pays out in state s. The real payoff of a nominal asset in state s is the set of commodities that can be purchased at the spot prices in state s with the dollars the asset pays out in state s.

Because wealth and preferences are state-dependent, we cannot evaluate the riskiness of portfolio payoffs without knowing the state-dependent wealth and preferences. Insurance policies are examples of contracts whose volatile payoffs actually reduce risks for one of the parties because of state-dependent preferences. The variability of spot prices is another fact that makes the riskiness of asset payoffs difficult to evaluate. For example, suppose there is a nominal asset whose payoff is one dollar in every state, and a real asset whose payoff is a fixed bundle of commodities in each state. These might both look like riskless assets. However, suppose the payoff of the real asset is a sack of coffee, whose relative spot price is volatile. Then the consumption possibilities of a trader who holds just this asset will also be highly volatile. On the other hand, if the general price level (i.e., the rate of inflation) is highly volatile, then the consumption possibilities of a trader who holds just the nominal asset will be highly volatile. When prices are volatile, the least risky asset is a real asset whose payoff is a broad consumption basket, such as an asset whose payoff is tied to the Consumer Price Index. Such indexed assets are common in countries with hyperinflation.

If we were going to determine the equilibrium spot commodity prices and asset prices simultaneously, then this distinction between real and nominal assets would be very important in our model, because the nominal payoff of real assets and the real payoff of nominal assets would be endogenous. However, we have opted to treat spot market prices as fixed. We can specify the payoffs of all assets in terms of their dollar values at the spot-market prices. Variability of the spot-market price of coffee is reflected in variability of the nominal payoff of coffee futures. Variability of inflation rates is reflected in the state-dependent utility of money.

6.1.5 Forward contracts versus tangible assets

Assets can also be divided into forward contracts and tangible assets. A forward contract is one where one party promises to deliver money or goods to another party in the future. A tangible asset is some storable commodity such as gold, land or machinery to which one has title.[1] With the possible exception of money, all nominal assets are

1. The term "paper asset" has nothing to do with this distinction. Anything in which ownership is recorded on paper, be it a commodity futures contract or ownership in a firm, looks to journalist like a paper asset. Thus, they would call shares of stock a paper asset, but we will call it a tangible asset. An asset does not have to be in your living room or back yard to be tangible.

forward contracts.[2]

The purchases and sales of forward contracts must always balance, and so their net supply is zero. Furthermore, until the asset is traded for the first time, each trader's initial holding is zero. For example, a life insurance contract is like an asset that pays \$1 if you die and \$0 otherwise. When you meet with the insurance agent, neither you nor your insurance company hold any of this asset. If you decide to purchase \$100,000 of coverage, your holding is 100,000 and your insurance company's holding is $-100,000$. The total of everyone's holding is still zero. Suppose that you pay the insurance company by borrowing money from the bank, at a zero rate of interest. Bank loans are assets that pay \$1 in states in which you do not default, and \$0 in the other states. You have sold 100,000 units of the asset and the bank has bought 100,000 units. Again, the total supply is zero before and after trade.

A tangible asset, on the other hand, is more like bananas at a spot market. The total supply is positive, and it is impossible to hold a negative amount. However, for each tangible asset there can be a parallel forward contract that promises delivery of the asset in the future, and hence has the exact same payoff as the tangible asset. For example, there are stocks (tangible assets) and stock options (forward contracts). There is no difference between buying gold at $t = 0$ and holding it until $t = 1$, and buying a forward contract that promises delivery of gold at $t = 1$. There is a small difference between buying stock and buying a forward contract for stock; buying stock gives you voting rights and the control that accompanies ownership. However, this distinction is not part of our model, and it is fairly accurate to unify the market for a tangible asset with the market for its parallel forward contract. This means that short sales (negative holdings) of tangible assets (and hence all assets) are allowed, and the only difference between tangible assets and forward contracts is that the net supply of tangible assets is positive.

6.2 Market equilibrium

6.2.1 The general model

Tangible assets can be created through production, while financial assets are created in the marketplace. For example, Motorola decides whether to float a bond, investment banks decide whether to issue derivative securities, and the Chicago Mercantile Board decides whether to have trading in corn futures. Financial theory is roughly divided into corporate finance, which studies in part how corporations choose which financial assets to issue, and asset pricing, which studies asset markets assuming that the available assets are given exogenously. It is quite complex to combine these two topics. We take the asset pricing approach, and study the trading of exogenously given assets.

Let there be J assets. (Generally, I will name the asset $1, 2, \dots, J$, but when there

2. Are dollars a tangible nominal asset, or are they a forward contract in which the government agrees to redeem the dollars for real commodities? This may seem like an easy question, but books are written about this and we leave you to think about this question on your own. For our model, the distinction is unimportant, as long as we remember that money is one of the available assets.

are two or three assets I will name them a, b, c.) As described above, each asset j is characterized by its nominal payoff $\tilde{Y}_j(s)$ in each state $s = 1, \ldots, S$. $\tilde{Y}_j(s)$ is the amount of money you get in state s per unit of the asset you hold; if you hold a negative amount of the asset (e.g., you issued the asset or sold it short) then you pay this amount. These payoffs are typically not neatly listed on the asset titles, but instead are implicit. For example, a bond issued by the city of Trenton might say that it pays $1,000 upon maturity. However, this is only the payoff in those states in which the city does not default. It might pay $800 if federal funding is eliminated and $600 if the New Jersey capital is moved to Rocky Hill. An ounce of gold and a share of ownership of a firm do not have any payoffs listed on them, but their payoffs are determined by the state-dependent spot market prices and the firm's state-dependent productivity.

A trader's portfolio is a list $\theta = \langle \theta_1, \ldots, \theta_J \rangle$ of the amount of each asset the trader holds. Traders come to the market with an initial portfolio, which is the result of past asset trading that does not appear in our model. For simplicity, let's assume everyone initially holds no assets. This is like assuming that there are no tangible assets and that all forward contracts are being issued for the first time. At the expense of extra notation, we could remove this assumption and our conclusions would not change.

Let q_j be the price of asset j. Because each trader has no initial endowment of assets, his surplus in the asset market when he acquires portfolio $\langle \theta_1, \ldots, \theta_J \rangle$ is minus the cost $q_1\theta_1 + \cdots + q_J\theta_J$ of his portfolio. This cost can be either positive or negative. The payoff on this portfolio in state s is $\theta_1\tilde{Y}_1(s) + \cdots + \theta_J\tilde{Y}_J(s)$. This payoff can also be positive or negative.

Assume that the agents trade competitively, which means that they choose their portfolios taking asset prices as given.

Consider the portfolio selection problem of a typical trader. This trader cares only about the dollar value z_0 of consumption in period 0 and the dollar value z_s of consumption in each state s in period 1, according to a utility function

$$U(z_0, z_1, \ldots, z_S).$$

He starts with a baseline wealth $w_0, w_1, \ldots, w_S$ (this is his initial allocation or endowment), which he can modify by trading assets. If he purchases a portfolio $\langle \theta_1, \ldots, \theta_J \rangle$, then he ends up with

$$z_0 = w_0 - (q_1\theta_1 + \ldots + q_J\theta_J)$$

available for consumption in period 0 after trading, and he ends up with

$$z_s = w_s + \theta_1\tilde{Y}_1(s) + \ldots + \theta_J\tilde{Y}_J(s)$$

in state s in period 1, after liquidating his portfolio. Therefore, he selects a portfolio that solves the following maximization problem:

$$\begin{aligned} \max_{\theta_1,\ldots,\theta_J} \quad & U(z_0, z_1, \ldots, z_S) \qquad (6.1) \\ \text{subject to:} \quad z_0 &= w_0 - (q_1\theta_1 + \ldots + q_J\theta_J) \\ z_s &= w_s + \theta_1\tilde{Y}_1(s) + \ldots + \theta_J\tilde{Y}_J(s) \quad \forall s = 1, \ldots, S. \end{aligned}$$

Our definition of an equilibrium will look like the definition of an equilibrium in competitive commodity markets. The two ingredients are individual optimality and market clearing. That is, the asset market is in equilibrium when the competitive (price-taking) asset demands at the prevailing asset prices balance. In commodity markets, "balance" means that the total consumption equals total resources. In these asset markets, where the net initial supply of each asset is zero, "balance" means that the total demand of each asset is zero: for each unit bought, there must be a unit sold.

DEFINITION 1. Asset prices $\langle q_1, \ldots, q_J \rangle$ and a portfolio $\langle \theta_1^i, \ldots, \theta_J^i \rangle$ for each trader $i = 1, \ldots, n$ are a *financial equilibrium* if

Individual optimality	$\langle \theta_1^i, \ldots, \theta_J^i \rangle$ solves equation (6.1) for $i = 1, \ldots, n$.
Market clearing	$\sum_{i=1}^n \theta_j^i = 0$ for $j = 1, \ldots, J$.

6.2.2 A two-dimensional model

Before studying this general model, let's consider a simple version that is designed so that we can draw some pictures. I will call it the $2 \times 2 \times 2 \times 2$ model because it has:

- Two periods: 0 and 1.
- Two states: 1 and 2.
- Two assets: a and b.
- Two traders: k (Keyser) and s (Soze).

Unlike in the general model, there is no consumption or baseline wealth in period 0 when asset trading occurs. This can be unintuitive at times, but it is necessary so that an allocation is just a point on the plane, giving money in state 1 and money in state 2. One story that fits this $2 \times 2 \times 2 \times 2$ model is that of the campers Keyser and Soze in the beginning of Section 5.1.2. It is also unintuitive that there are competitive asset markets when there are just two traders. It may help to imagine that there are many identical Keyser brothers and Soze sisters at the camp.

The period 0 budget constraint, which is

$$z_0 = w_0 - (q_a \theta_a + q_b \theta_b)$$

when there is consumption in period 0, becomes instead

$$0 = q_a \theta_a + q_b \theta_b.$$

That is, each trader must finance any purchases of one of the assets by selling the other asset. Then the portfolio selection problem is

$$\begin{aligned} \max_{\theta_a, \theta_b} \quad & U(z_1, z_2) && \text{(PSP)} \\ \text{subject to:} \quad 0 &= q_a \theta_a + q_b \theta_b \\ z_1 &= w_1 + \theta_a \tilde{Y}_a(1) + \theta_b \tilde{Y}_b(1) \\ z_2 &= w_2 + \theta_a \tilde{Y}_a(2) + \theta_b \tilde{Y}_b(2). \end{aligned}$$

This maximization problem has an interesting property. The trader cares about his real allocation z_1, z_2, but he chooses a portfolio θ_a, θ_b. The real allocation is linked to the portfolio through the constraints. In solving this problem, the trader can (i) look at how the constraints determine his set of feasible allocations, then (ii) pick the best of these allocations, and then (iii) figure out what portfolio gets him this allocation.

To determine the set of feasible allocations, we need to combine the budget constraints by solving out θ_a and θ_b. Let's start by solving the asset market constraint for θ_a as a function of θ_b:

$$\theta_a = -q_b\theta_b/q_a.$$

Now substitute this into the two spot-market constraints:

$$\begin{aligned} z_s &= w_s - \frac{q_b\theta_b}{q_a}\tilde{Y}_a(s) + \theta_b\tilde{Y}_b(s) \\ &= w_s + q_b\theta_b\left(\frac{\tilde{Y}_b(s)}{q_b} - \frac{\tilde{Y}_a(s)}{q_a}\right). \end{aligned}$$

$\tilde{R}_b(s) = \tilde{Y}_b(s)/q_b$ is the *return* (payoff per dollar invested in the asset) of asset b in state s , and

$$\tilde{x}(s) = \tilde{R}_b(s) - \tilde{R}_a(s) = \frac{\tilde{Y}_b(s)}{q_b} - \frac{\tilde{Y}_a(s)}{q_a}$$

is called the excess return of asset b over asset a. Then we can write the two spot-market budget constraints as follows:

$$z_1 = w_1 + (q_b\theta_b)\tilde{x}(1) \tag{6.2}$$

$$z_2 = w_2 + (q_b\theta_b)\tilde{x}(2). \tag{6.3}$$

Digression on arbitrage The excess return $\tilde{x}(s)$ is the payoff on my portfolio if I buy \$1 of asset b and sell \$1 of asset a. If $\tilde{x}(1)$ and $\tilde{x}(2)$ are both positive, it is possible to have unlimited wealth in both states by increasing the dollars $q_b\theta_b$ invested in asset b and sold of asset a. This is an example of *arbitrage*. For example, suppose the asset prices and payoffs are

	Asset a	Asset b
Price	4	2
Payoff state 1	3	2
Payoff state 2	1	3

Then the returns are

	Returns		
	Asset a	Asset b	Excess
State 1	3/4	1	1/4
State 2	1/4	3/2	5/4

By selling one unit of asset a and buying two units of asset b, I get a payoff of 1 in state 1 and 5 in state 2. I can get unlimited consumption in both states by buying arbitrarily large quantities of this portfolio.

If $\tilde{x}(1)$ and $\tilde{x}(2)$ are both negative, unlimited wealth is attained by increasing the dollars $q_b\theta_b$ sold of asset b and purchased of asset a. Arbitrage is also possible if the excess return is zero in one state and positive or negative in the other. Looking ahead to equilibrium, we can see that there can be no equilibrium when arbitrage is possible because traders will demand unlimited amounts of arbitrage portfolios. Therefore, we assume that arbitrage is not possible, which means that either the excess return in both states is zero or the excess return is positive in one state and negative in the other.

Digression on incomplete markets When the excess return is zero in both states, it is impossible to make state-contingent trades by trading assets. Any portfolio with zero cost also has a zero payoff in both states. This can happen if and only if the payoffs of asset b are proportional to the payoffs of asset a. For example, suppose the payoffs are as follows:

	Payoffs	
	Asset a	Asset b
State 1	1	2
State 2	3	6

The payoffs of asset b are twice the payoffs of asset a. The excess returns are:

$$\tilde{x}(1) = \frac{2}{q_b} - \frac{1}{q_a}$$

$$\tilde{x}(2) = \frac{6}{q_b} - \frac{3}{q_a} = 3\tilde{x}(1).$$

Thus, if $q_b \neq 2q_a$, the excess returns are both positive or negative, and arbitrage is possible. If $q_b = 2q_a$, then the excess returns are both zero. Buying one unit of asset b is equivalent to buying one unit of asset a. It is as if there were only one asset, which is not enough for state-contingent trade. This is an example of incomplete markets, which is discussed in Section 6.3.3. For now, we simply assume that the payoffs of asset b are not proportional to the payoffs of asset a, which means that we have complete asset markets.

In summary, with no-arbitrage asset prices and complete markets, $\tilde{x}(1)$ and $\tilde{x}(2)$ are unequal to zero and have opposite signs. We can then combine the two constraints in equation (6.2) and (6.3) by solving the state-1 constraint for $q_b\theta_b$,

$$q_b\theta_b = \frac{z_1 - w_1}{\tilde{x}(1)},$$

and substituting the answer into the state-2 constraint:

$$z_2 = w_2 + \frac{z_1 - w_1}{\tilde{x}(1)}\tilde{x}(2).$$

Rearranging:

$$-\frac{1}{\tilde{x}(1)}z_1 + \frac{1}{\tilde{x}(2)}z_2 = -\frac{1}{\tilde{x}(1)}w_1 + \frac{1}{\tilde{x}(2)}w_2. \tag{6.4}$$

If $\tilde{x}(1) > 0$ and $\tilde{x}(2) < 0$, then let

$$p_1 = \frac{1}{\tilde{x}(1)} \quad \text{and} \quad p_2 = -\frac{1}{\tilde{x}(2)}.$$

If instead $\tilde{x}(1) < 0$ and $\tilde{x}(2) > 0$, let

$$p_1 = -\frac{1}{\tilde{x}(1)} \quad \text{and} \quad p_2 = \frac{1}{\tilde{x}(2)}.$$

Either way, p_1 and p_2 are positive and the budget constraint in equation (6.4) is

$$p_1 z_1 + p_2 z_2 = p_1 w_1 + p_2 w_2.$$

As in the portfolio selection and insurance demand problems of Section 4.2.5, we have reformulated the decision problem so that it looks like a standard consumer demand problem, for a household that consumes two goods (e.g., corn and leisure/labor) that have prices p_1 and p_2, and whose income is the value of its initial endowment of these goods. Rather than choosing consumption of corn and leisure/labor; or consumption of oranges and apples, the trader chooses money in state 1 and consumption in state 2, when the "prices" of consumption in states 1 and 2 are p_1 and p_2, respectively. p_1 and p_2 are called the *state prices*.

In summary, the reduced form of the traders portfolio selection problem is the following consumption choice problem:

$$\begin{aligned} \max_{z_1, z_2} \quad & u(z_1, z_2) && \text{(CCP)} \\ \text{subject to:} \quad & p_1 z_1 + p_2 z_2 = p_1 w_1 + p_2 w_2. \end{aligned}$$

I call (CCP) a reduced form of (PSP) because the solutions to both problems give the same allocations and portfolios.

Proposition 1. *Let q_a and q_b be asset prices, and let p_1 and p_2 be the corresponding state prices. Let $\langle \theta_a, \theta_b \rangle$ be a portfolio, and let $\langle z_1, z_2 \rangle$ be the corresponding allocation:*

$$\begin{aligned} z_1 &= w_1 + \theta_a \tilde{Y}_a(1) + \theta_b \tilde{Y}_b(1) \\ z_2 &= w_2 + \theta_a \tilde{Y}_a(2) + \theta_b \tilde{Y}_b(2). \end{aligned}$$

Then $\langle \theta_a, \theta_b \rangle$ solves (PSP) given $\langle q_a, q_b \rangle$ if and only if $\langle z_1, z_2 \rangle$ solves (CCP) given $\langle p_1, p_2 \rangle$.

Because there is no consumption in period 0 of this artificial model, assets are bartered rather than bought or sold using dollars. A consequence is that only relative asset prices q_b/q_a, which determine the exchange ratio, matter. Similarly, in (CCP), only relative state prices

$$\frac{p_2}{p_1} = \frac{\tilde{x}(1)}{\tilde{x}(2)}$$

matter.

Suppose that asset payoffs are as follows:

Payoffs

	Asset a	Asset b
State 1	1	1
State 2	1	3

If $q_a = 1$ and $q_b = 5/2$, then the returns are

Returns

	Asset a	Asset b	Excess
State 1	1	2/5	-3/5
State 2	1	6/5	1/5

Therefore, the relative state prices are

$$\frac{p_2}{p_1} = -\frac{-3/5}{1/5} = 3.$$

Having translated the asset prices into state prices, suppose that a trader's baseline wealth is $w_1 = w_2 = 4$ and that her solution to (CCP) when $p_2/p_1 = 3$ is $z_1 = 10$ and $z_2 = 2$. Now let's translate this consumption choice into a portfolio selection. Her portfolio payoff is $z_s - w_s$ in state s, and hence is 6 in state 1 and -2 in state 2. The portfolio $\langle\theta_a, \theta_b\rangle$ that yields this payoff is the solution to

$$\begin{aligned} 6 &= \theta_a + \theta_b \\ -2 &= \theta_a + 3\theta_b. \end{aligned}$$

The answer is $\theta_a = 10$ and $\theta_b = -4$.

6.2.3 Real and financial equilibrium in the two-dimensional model

Definition 2. A *financial equilibrium* is a pair $\langle q_a, q_b\rangle$ of asset prices and portfolios $\langle\theta_a^k, \theta_b^k\rangle$ and $\langle\theta_a^s, \theta_b^s\rangle$ for Keyser and Soze such that

Individual optimality $\langle\theta_a^i, \theta_b^i\rangle$ solves (PSP) for trader i, given $\langle q_a, q_b\rangle$.
Market clearing $\theta_a^k + \theta_a^s = 0$ and $\theta_b^k + \theta_b^s = 0$.

The equivalence between (PSP) and (CCP) suggest a possible equivalence between equilibrium in the asset market and an equilibrium in a hypothetical market in which traders directly exchange consumption in state 1 for consumption in state 2. Here is the definition of equilibrium for this hypothetical market:

Definition 3. A *real equilibrium* is a pair $\langle p_1, p_2\rangle$ of state prices and allocations $\langle z_1^k, z_2^k\rangle$ and $\langle z_1^s, z_2^s\rangle$ for Keyser and Soze such that

Individual optimality $\langle z_1^i, z_2^i\rangle$ solves (CCP) for trader i, given $\langle p_1, p_2\rangle$.
Market clearing $z_1^k + z_1^s = w_1^k + w_1^s$ and $z_2^k + z_2^s = w_2^k + w_2^s$

If you find the idea of direct trading of consumption in the two states too hypothetical, you can also interpret a real equilibrium as a financial equilibrium when the asset payoffs are *canonical*:

Payoffs	Asset a	Asset b
State 1	1	0
State 2	0	1

Asset a is the same as consumption in state 1 and asset b is the same as consumption in state 2.

For any prices q_a and q_b, the excess returns are

$$\tilde{x}(1) = -1/q_a \qquad \tilde{x}(2) = 1/q_b.$$

Therefore, relative state prices are equal to the relative asset prices.

$$\frac{p_2}{p_1} = -\frac{\tilde{x}(1)}{\tilde{x}(2)} = \frac{q_b}{q_a}.$$

This makes sense, since buying a unit of asset a is equivalent to buying a unit of consumption in state 1. When the assets have these payoffs, the market is as similar to a market for apples and bananas as is possible. You cannot actually have units of consumption in states 1 and 2 sitting on tables for people to buy, but with these assets you can have a table with coupons for consumption in state 1 and a table with coupons for consumption in state 2.

PROPOSITION 2. *Let $\langle q_a, q_b\rangle$ be asset prices with corresponding state prices $\langle p_1, p_2\rangle$. Let $\langle \theta_a^k, \theta_b^k\rangle$ and $\langle \theta_a^s, \theta_b^s\rangle$ be portfolios that yield allocations $\langle z_1^k, z_2^k\rangle$ and $\langle z_1^s, z_2^s\rangle$, respectively. Then $\langle q_a, q_b\rangle$ and $\langle \theta_a^k, \theta_b^k\rangle$ and $\langle \theta_a^s, \theta_b^s\rangle$ are a financial equilibrium if and only if $\langle p_1, p_2\rangle$ and $\langle z_1^k, z_2^k\rangle$ and $\langle z_1^s, z_2^s\rangle$ are a real equilibrium.*

Proof. We claimed in the previous section that $\langle \theta_a^i, \theta_b^i\rangle$ solves (PSP) given $\langle q_a, q_b\rangle$ if and only if $\langle z_1^s, z_2^s\rangle$ solves (CCP) given $\langle p_1, p_2\rangle$. We need to also show that portfolios are balanced if and only if the allocations are balanced. That is

$$\begin{array}{rcl} \theta_a^k + \theta_a^s & = & 0 \\ \theta_b^k + \theta_b^s & = & 0 \end{array} \iff \begin{array}{rcl} z_1^k + z_1^s & = & w_1^k + w_1^s \\ z_2^k + z_2^s & = & w_2^k + w_2^s. \end{array}$$

For each agent's budget constraint,

$$z_s^i - w_s^i = \theta_a^i \tilde{Y}_a(s) + \theta_b^i \tilde{Y}_b(s).$$

Therefore:

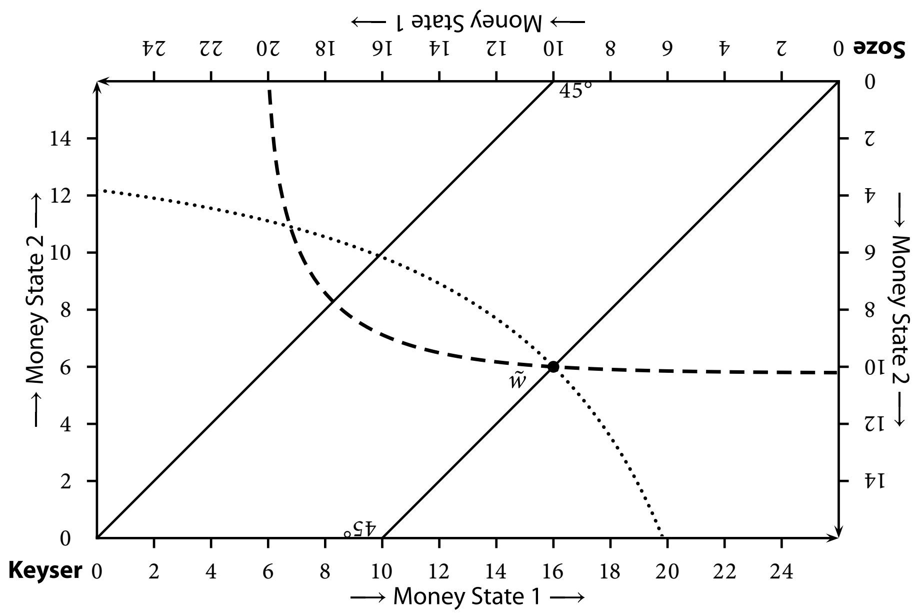
$$\begin{aligned} (z_1^k - w_1^k) + (z_1^s - w_1^s) &= (\theta_a^k + \theta_a^s)\tilde{Y}_a(1) + (\theta_b^k + \theta_b^s)\tilde{Y}_b(1) \\ (z_2^k - w_2^k) + (z_2^s - w_2^s) &= (\theta_a^k + \theta_a^s)\tilde{Y}_a(2) + (\theta_b^k + \theta_b^s)\tilde{Y}_b(2). \end{aligned}$$

You can see that the left-hand side of each equation is zero (real markets clear) if $\theta_a^k + \theta_a^s = 0$ and $\theta_b^k + \theta_b^s = 0$ (financial markets clear). I will not prove the converse, as it requires some facts from linear algebra. □

This means that, to find a financial equilibrium, we can look for a real equilibrium and then "unravel" the state prices and allocations to find the asset prices and portfolios. There are two advantages to doing this:

- Finding a real equilibrium is easier, because we already know preferences over allocation and we can use the standard tools of equilibrium analysis, such as Edgeworth's boxes.
- The real equilibrium does not depend on the asset payoffs. Hence, if we want to compare the financial equilibrium for two different asset structures, we can first find the real equilibrium and then translate it into the financial equilibrium for each of the asset structures.

Figure 6.1



An Edgeworth box, showing Keyser's endowment $\tilde{w}^k = \langle 16, 6\rangle$ and Soze's endowment $\tilde{w}^s = \langle 10, 10\rangle$, and indifference curves through the endowments. (Keyser's indifference curve is dashed and Soze's is dotted.)

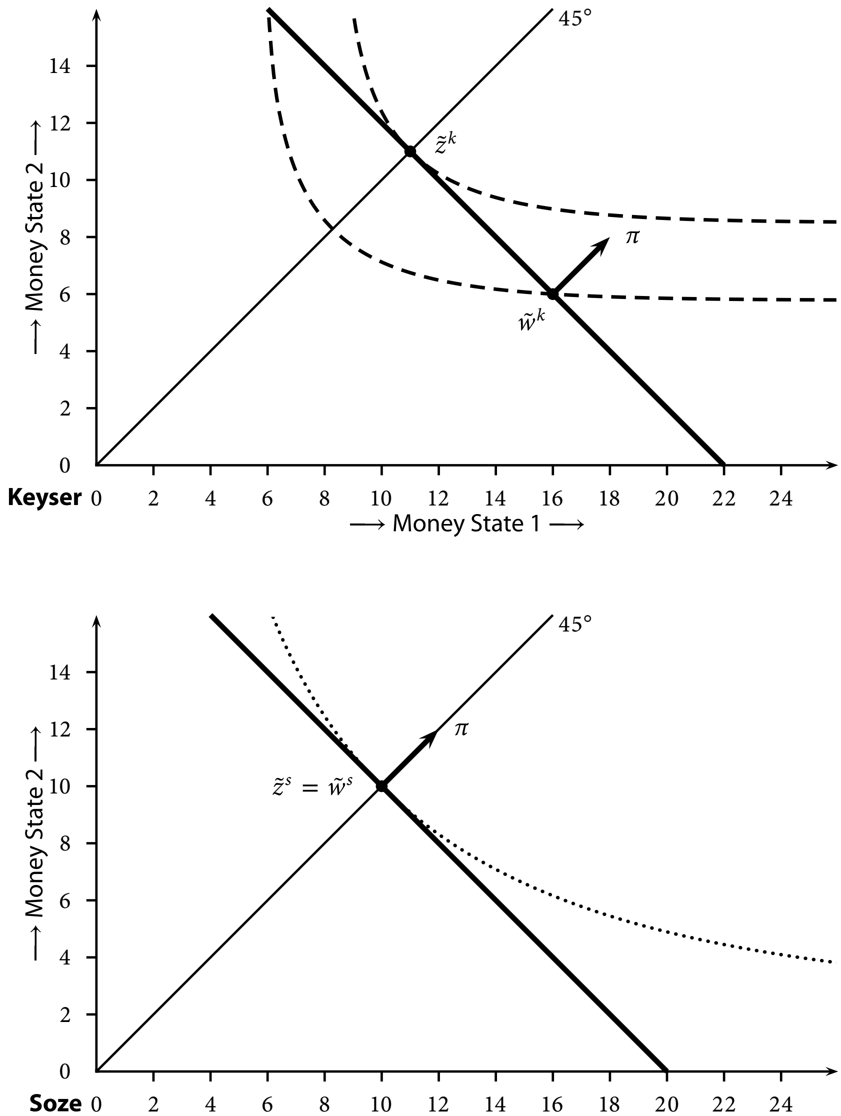
I will do one example. The first step is to find the real equilibrium, which I will do using Edgeworth boxes. This is just a pure exchange equilibrium, which you have studied in a microeconomics course, but I will not rely upon perfect recall. If you need additional refreshing on the topic, consult any microeconomics text.

Consider first the Edgeworth box economy in Figure 6.3, which is the same as the one in Figure 5.2. Keyser's endowment is $\langle 16, 6\rangle$ and Soze's endowment is $\langle 10, 10\rangle$. The figure shows the endowment and the two indifference curves through the endowment. The traders believe the states are equally probable, and they have state-independent, risk averse preferences.

First, we will use the Edgeworth box to see that the relative state prices $p_2/p_1 = 1$ are not equilibrium prices. The beauty of the Edgeworth box is that we can simultaneously show the budget lines and consumption decisions for both traders, but let's start with each individual trader. The top of Figure 6.2 shows Keyser's endowment and indifference curve through the endowment, plus his budget line and indifference curve through his preferred allocation on the budget line. The budget line passes through his endowment and is perpendicular to the vector $\langle 1, 1\rangle$ of state prices. Since the prices are the same in both states, just like the probabilities, the vector of prices points in the same direction as the vector of probabilities, and all allocations on the budget line have the same expected value. (The budget line is the fair-odds line.) Since Keyser is risk averse, he chooses the risk-free allocation, where his budget line crosses his 45° line.

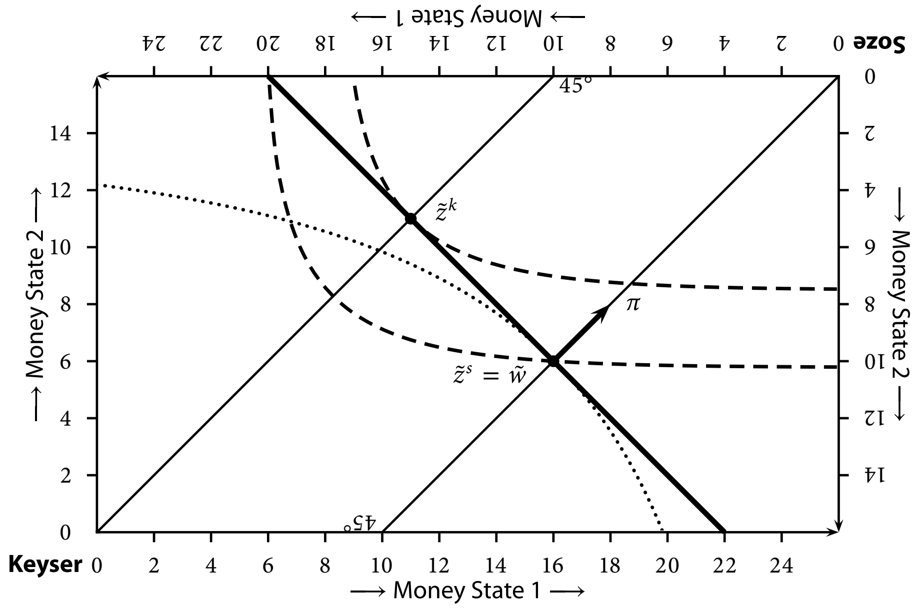
The bottom of Figure 6.2 shows Soze's endowment, her indifference curve through her endowment, and her budget line. Like Keyser, Soze is risk averse and prefers the riskfree allocation; for Soze, this is her endowment and so she chooses not to trade at

Figure 6.2



Individual consumption decisions when state prices are equal to the probabilities, and hence all points on the budget line have the same expected value. The top graph shows Keyser's choice and the bottom figure shows Soze's choice. Both decision makers have state-independent and risk-neutral preferences, and hence choose the risk-free allocation.

Figure 6.3



An Edgeworth box showing demand for consumption in each state by both traders when state prices are equal to the probabilities ($p_2/p_1 = \pi_1/\pi_2 = 1$). The dashed lines are Keyser's indifference curves and the dotted line is Soze's indifference curve. Both traders have state-independent, risk-averse preferences. Hence, both traders demand a risk-free allocation. Since the aggregate endowment is random, the demands do not equal the allocations, and so these prices are not equilibrium prices.

all.

When we superimpose these two pictures to make an Edgeworth box, we obtain Figure 6.3. Observe that the budget lines for the two traders coincide. This is a useful feature of Edgeworth boxes. For market clearing, the total demand must equal the total endowment. The Edgeworth box is set up so that the total endowment equals the dimensions of the box. This implies, by example, that there is market clearing if and only if the two traders' demands coincide in the box. Then, for example, the demand for good 1 by Soze—which is the distance from Keyser's vertical axis to Keyser's demand—plus the demand for good 1 by Soze—which is the distance from Soze's vertical axis to Soze's demand—equals the total endowment of good 1—which is the distance between the two vertical axes. For the state prices shown in Figure 6.3, the demands do not coincide and the relative state prices equal to the relative probabilities are not equilibrium prices.

We can guess where the equilibrium allocation lies by looking at the Edgeworth box in Figure 6.1. Since trade is voluntary, the equilibrium allocations must be individually rational, which means they must lie in the region bounded by the two indifference curves through the endowment. Observe that the budget line passes through this region only if $p_2/p_1 > \pi_2/\pi_1$. The intuition is that, in equilibrium, Keyser is going to share risk with Soze. For Soze to take on this risk, she must get a positive expected return. The equilibrium and relative state prices are $p_2/p_1 = 2$, and the equilibrium allocations are $\tilde{z}^k = \langle 11, 8.5\rangle$ and $\tilde{z}^s = \langle 15, 7.5\rangle$. The top of Figure 6.4 shows that the equilibrium allocations are individually rational. The bottom of the figure shows the indifference curves through the equilibrium allocation.

Note that I have not yet told you what assets are traded in the asset market; we do not need to know this in order to determine the real equilibrium (as long as there are two assets whose payoffs are not proportional to each other). Nevertheless, having found the real equilibrium allocations, we can determine what the portfolio payoffs must be in the financial equilibrium, because each trader's portfolio payoff must equal his or her net trade in each state:

Trader	State	Allocation z_s^i	Endowment w_s^i	Portfolio payoff $z_s^i = w_s^i$
Keyser	1	11	16	−5
	2	8.5	6	2.5
Soze	1	15	10	5
	2	7.5	10	−2.5

Note that Soze's portfolio payoffs are the opposite (sign) of Keyser's payoffs, which must be true for the net trades to balance.

Let's find the financial equilibrium when the asset has the following payoffs:

	Payoffs	
	Asset a	Asset b
State 1	1	1
State 2	1	3

To solve for the relative asset prices, we can use the equation

$$\frac{p_2}{p_1} = -\frac{\tilde{x}(1)}{\tilde{x}(2)}.$$

For the equilibrium relative state prices $p_2/p_1 = 2$ and the asset payoffs, we get the following equation that we solve for q_b/q_a:

$$\begin{aligned} 2 &= -\frac{\frac{1}{q_b} - \frac{1}{q_a}}{\frac{3}{q_b} - \frac{1}{q_a}} \\ 6 - 2q_b/q_a &= q_b/q_a - 1 \\ \frac{q_b}{q_a} &= \frac{7}{3}. \end{aligned}$$

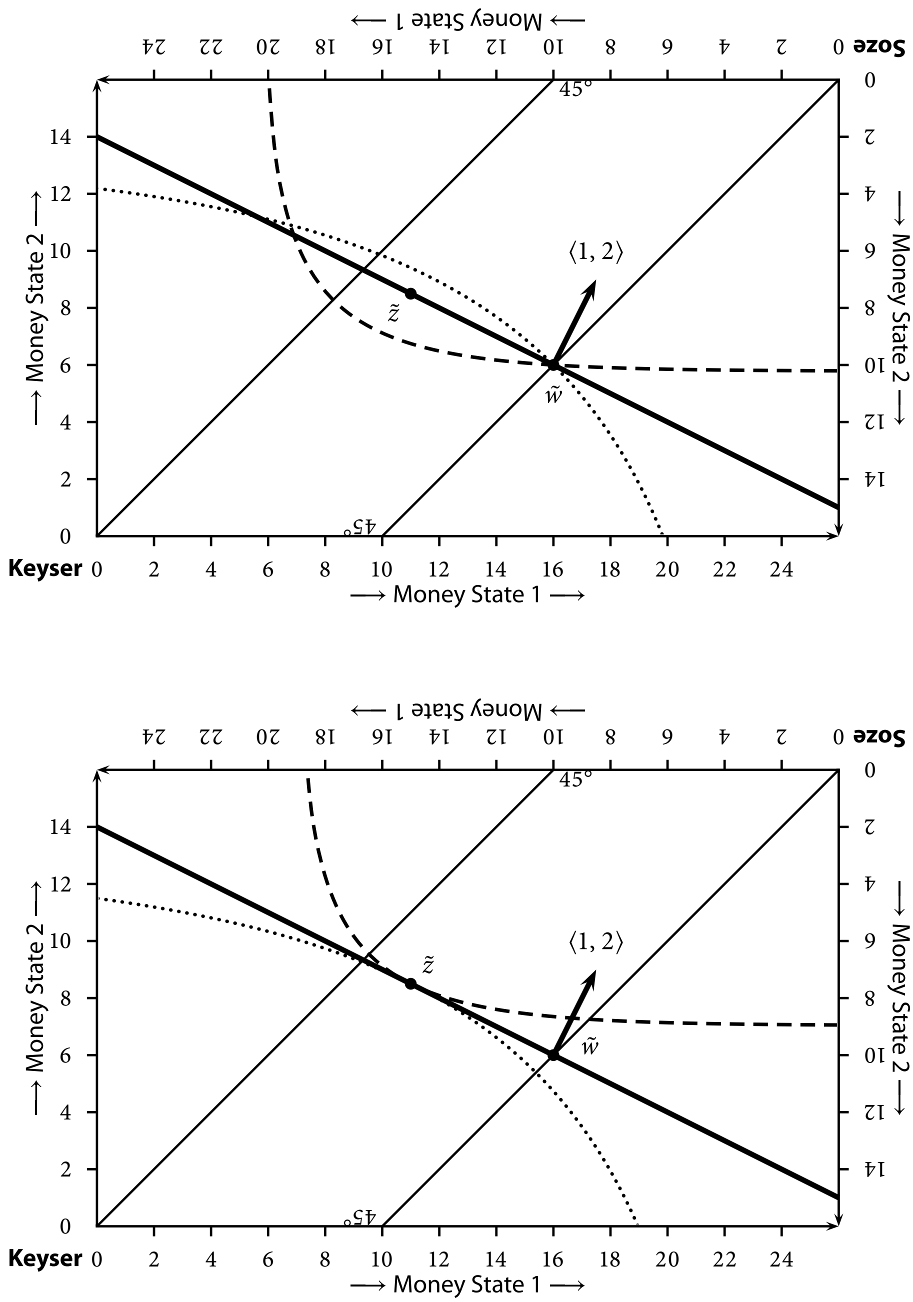
Here is an easier method for calculating relative prices of the assets. If state prices are p_1 and p_2, and an asset pays off $\tilde{Y}(1)$ and $\tilde{Y}(2)$, then the state-price value of this asset's payoff is $p_1\tilde{Y}(1) + p_2\tilde{Y}(2)$. Therefore, relative asset prices should be

$$\frac{q_b}{q_a} = \frac{p_1\tilde{Y}_b(1) + p_2\tilde{Y}_b(2)}{p_1\tilde{Y}_a(1) + p_2\tilde{Y}_a(2)}.$$

For this example, we have

$$\frac{q_b}{q_a} = \frac{(1\times 1) + (2\times 3)}{(1\times 1) + (2\times 1)} = \frac{7}{3}.$$

Figure 6.4



The equilibrium in the Edgeworth box. The state prices are $p_2/p_1 = 2$. The equilibrium allocation is $\tilde{z}^k = \langle 11, 8.5\rangle$ and $\tilde{z}^s = \langle 15, 7.5\rangle$. The dashed and dotted curves are Keyser's and Soze's indifference curves, respectively. The top graph shows the indifference curves through the endowment; the equilibrium allocation is individually rational. The bottom graph shows the indifference curves through the equilibrium allocation; the equilibrium-allocation is Pareto efficient.

Amazingly enough, it worked.

Now let's solve for the equilibrium portfolios. Note that this calculation is independent from the calculation of asset prices. That is, to calculate the equilibrium asset prices, we only need to know the equilibrium state prices. To calculate the equilibrium portfolio, we only need to know the equilibrium allocations.

We need to find portfolios that generate the equilibrium portfolio payoffs. That is, for Keyser we find θ_a^k and θ_b^k such that

$$\begin{aligned} z_1^k - w_1^k &= \theta_a^k \tilde{Y}_a(1) + \theta_b^k \tilde{Y}_b(1) \\ z_2^k - w_2^k &= \theta_a^k \tilde{Y}_a(2) + \theta_b^k \tilde{Y}_b(2). \end{aligned}$$

For this example, these equations are

$$\begin{aligned} -5 &= \theta_a^k + \theta_b^k \\ 2.5 &= \theta_a^k + 3\theta_b^k. \end{aligned}$$

The solution is $\theta_a^k = -35/4$ and $\theta_b^k = 15/4$. We could do similar calculations for Soze, but *why* bother: Since these are equilibrium portfolios, $\theta_a^s = -\theta_a^k = 35/4$ and $\theta_b^s = -\theta_b^k = -15/4$.

Here are the same calculations for canonical payoffs:

	Payoffs	
	Asset a	Asset b
State 1	1	0
State 2	0	1

As you might suspect, the calculations in this case are trivial. The asset prices are equal to the corresponding state prices: $q_b/q_a = p_2/p_1 = 2$. The portfolios match the portfolio payoffs

$$\begin{aligned} \langle \theta_a^k, \theta_b^k \rangle &= \langle z_1^k - w_1^k, z_2^k - w_2^k \rangle &= \langle -5, 2.5 \rangle \\ \langle \theta_a^s, \theta_b^s \rangle &= \langle z_1^s - w_1^s, z_2^s - w_2^s \rangle &= \langle 5, -2.5 \rangle. \end{aligned}$$

When one of the traders is risk neutral, it is easy to determine equilibrium asset prices and state prices. The state prices must equal the probabilities, and the asset prices must be such that the expected return of each asset is the same. The proof of this is left as an exercise.

Exercise 6.1. The attached graphs show an Edgeworth box with agents Keyser and Soze, for an asset market with two states s_1 and s_2 and no consumption in period 0. The endowment and the indifference curves through the endowment are drawn. Both traders are have state-independent preferences and are risk averse.

a. What is the endowment of each trader? Which trader(s) have a risky endowment?

b. On the first graph, label which indifference curve belongs to Soze and which belongs to Keyser.

c. On the first graph, draw the budget line when state prices are equal. Mark the optimal allocation from the budget line for each of the traders, and illustrate the optimality by drawing the indifference curve of each trader through his/her optimal allocation.

d. Referring to the graph to support your argument, explain why, for the given endowments and for any risk-averse preferences, equal state prices cannot be equilibrium state prices, and that equilibrium state prices must satisfy $p_2/p_1 > 1$.

e. The equilibrium state prices for the preferences I am using are approximately $p_2/p_1 = 2$, and the equilibrium allocations are approximately $\tilde{z}^k = \langle 11, 8.5\rangle$ and $\tilde{z}^s = \langle 15, 7.5\rangle$. Draw the budget line for these prices, mark the equilibrium allocations, and draw plausible indifference curves showing that these allocations are optimal for each trader. What is each trader's portfolio return?

f. Suppose there are two assets, a and b. Asset a's payoff is 1 in each state, and asset b's payoff is 1 is state s_1 and 3 in state s_2. Given the equilibrium state prices and allocations stated above, derive the equilibrium asset prices and portfolios.

g. Suppose that instead, asset a's payoff is 1 in state s_1 and 0 in state s_2, and asset b's payoff is 0 in state s_1 and 1 in state s_2. Derive the equilibrium asset prices and portfolios.

6.3 Complete versus incomplete markets

6.3.1 Arbitrage and linearly dependent assets

We return to the more general model of Section 6.2. Just by looking at the budget constraint, we can make some important observations about equilibrium asset prices. We assume that traders like more money over less in every period and state. Then, in equilibrium, there cannot be a portfolio that gives a non-negative surplus or payoff in every period and state, and a positive surplus or payoff in some period or state. Such a portfolio is called an *arbitrage opportunity*. If such a portfolio existed, each trader could do better by buying more of the portfolio. If everyone is buying this arbitrage portfolio, asset markets cannot clear; furthermore, there is not even a solution to the portfolio selection problem because it is always possible to get more wealth in at least one period or state, without giving up money in any other period or state. Asset prices for which there are no arbitrage opportunities are called *no-arbitrage prices*. Only no-arbitrage prices can be equilibrium prices.[3]

For example, suppose that there are two states and three assets, with the following prices and payoffs:

	Asset a	Asset b	Asset c
Price	200	150	150
Payoff state 1	100	225	140
Payoff state 2	300	75	140

3. At least this is the case in our idealized, fractionless, competitive market. In real markets, there may be small, transitory arbitrage opportunities that are quickly dissipated by traders who, by trying to profit from these opportunities, push prices towards no-arbitrage prices.

Consider the portfolio $\langle 3/8, 1/2, -1\rangle$. The cost of this portfolio is

$$(3/8)200 + (1/2)150 + (-1)150 \;=\; 0.$$

The payoff on this portfolio is strictly positive in each state:

$$\begin{aligned} z_1 &= (3/8)100 + (1/2)225 + (-1)140 \;=\; 10 \\ z_2 &= (3/8)300 + (1/2)75 + (-1)140 \;=\; 10. \end{aligned}$$

Hence, it is an arbitrage portfolio.

In this example, the asset payoffs are *linearly dependent*. This means that the payoff of one of the assets is equal to the payoff of a portfolio containing only the other assets. That is, there are θ_a and θ_b such that

$$\begin{aligned} \tilde{Y}_c(1) &= \theta_a \tilde{Y}_a(1) + \theta_b \tilde{Y}_b(1) \\ \tilde{Y}_c(2) &= \theta_a \tilde{Y}_a(2) + \theta_b \tilde{Y}_b(2). \end{aligned}$$

These equalities hold when $\theta_a = 7/20$ and $\theta_b = 7/15$. The no-arbitrage condition implies that the prices of portfolios that have the same payoffs must be the same. In this example, this means that

$$q_c \;=\; q_a\theta_a + q_b\theta_b \;=\; (7/20)q_a + (7/15)q_b.$$

Therefore, the no-arbitrage condition imposes restrictions on the prices of linearly dependent assets. The use of these restrictions is called arbitrage pricing.

Note that if $q_c = q_a\theta_a + q_b\theta_b$, then eliminating asset 3 does not limit the possible portfolio payoffs. Instead of buying or selling a unit of asset 3, the traders can buy or sell the portfolio $\langle\theta_a, \theta_b\rangle$ of the remaining two assets. Therefore, whenever some assets are linearly dependent, there are *redundant assets*. (Note that any one of the three assets can be written as a portfolio of the other two, and hence any one of them can be eliminated. There are redundancies, but we cannot single out one of the assets as *the* redundant asset.)

Here is an example of what is called binomial option pricing. Suppose that assets include a riskless bond (asset a), a risky stock (asset b), and a call option on the stock with exercise price K (asset c). The option gives the holder the right to pay K for the stock; it is exercised only when the payoff (dividend plus value) of the stock exceeds K, in which case the payoff of the option is $\tilde{Y}_b(s) - K$. Hence, for any state s the payoff is

$$\tilde{Y}_c(s) \;=\; \max\{\tilde{Y}_b(s) - K,\, 0\}.$$

Suppose there are two states, 1 and 2, in which the stock's payoffs are 1 and 3, respectively, and that the call option has an exercise price of 5/2. Then the payoffs $\tilde{Y}_c$ of the call option are $\tilde{Y}_c(1) = 0$ and $\tilde{Y}_c(2) = 1/2$. Assume that the bond pays \$1 in both states. Then the matrix fo payoffs is

	Payoffs		
	Bond	Stock	Option
State 1	1	1	0
State 2	1	3	1/2

The option has the same payoff as the portfolio with $\theta_1 = -1/4$ and $\theta_2 = 1/4$. Therefore, the prices of the bond, stock and option satisfy

$$q_c = q_b/4 - q_a/4.$$

Exercise 6.2. In the general case of binomial option pricing:

1. The price of the bond and the stock are q_1 and q_2, respectively;
2. The payoff of the bond is Rq_1 in both states, where R is the riskless return;
3. The payoff of the stock is $R^L q_b$ in state 1 and $R^H q_b$ in state 2, where $R^L < R^H$, R^L is the return in the "bad" state, and R^H is the return in the good state.
4. The price of the call option on the stock is q_3 and the strike price is K, where $R^L q_2 < K < R^H q_2$.

Derive the no-arbitrage price of the option as a function of q_1, q_2, R, R^H, R^L and K.

6.3.2 No-arbitrage asset prices and state prices

Given asset prices $\langle q_1, \ldots, q_J \rangle$, a vector $\langle p_1, \ldots, p_S \rangle$ of positive numbers is a *state-price vector* if, for each asset j,

$$q_j = p_1 \tilde{Y}_j(1) + p_2 \tilde{Y}_j(2) + \cdots + p_S \tilde{Y}_j(S). \tag{6.5}$$

Imagine that you are in a fruit market with S varieties of fruit, in which the price of fruit s is p_s. If someone is selling coupons that entitle you to $\tilde{Y}_j(1)$ units of fruit 1, $\tilde{Y}_j(2)$ units of fruit 2, and so on, then the value if the coupon is q_j, as defined in equation (6.5). Thus, we can think of component p_s of a state-price vector as the price of consumption in state 2.

We already saw the use of state prices for characterizing the equilibrium of the $2 \times 2 \times 2 \times 2$ model; we will see similar application later, and will use state prices to derive the CAPM asset-pricing model. Another use of state prices is to price linearly dependent assets. It is easy to show that if $\langle p_1, \ldots, p_s \rangle$ is a state-price vector for a set of assets, and if we add additional, linearly dependent assets, then its no-arbitrage price must be also be given by equation (6.5).

The purpose of this section is to show a neat trick for finding the set of no-arbitrage prices, and then to show that asset prices satisfy no arbitrage if and only if there is a state-price vector.

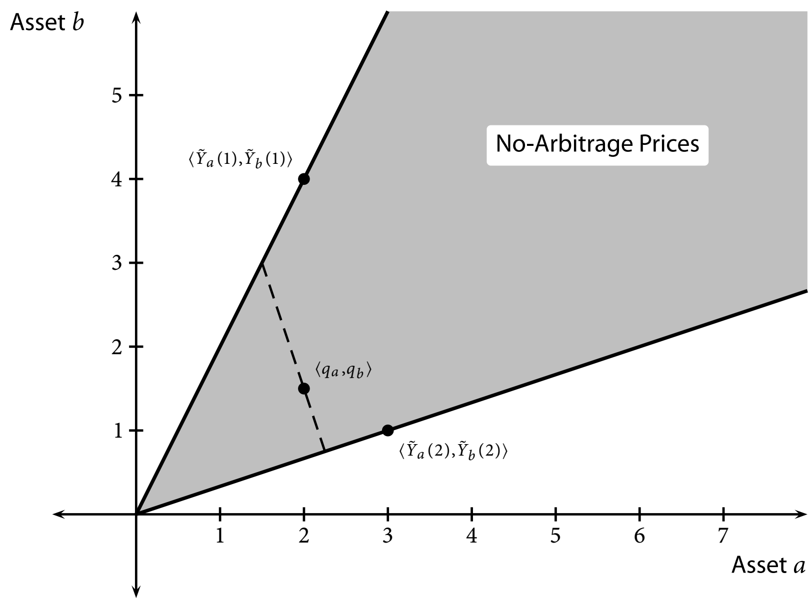
Let's start with the punchline. Suppose I ask you to draw on the plane the set of asset prices $\langle q_a, q_b \rangle$ that do not permit arbitrage when the payoffs are as follows:

Payoffs

	Asset a	Asset b
State 1	2	4
State 2	3	1

Take a moment to try to work out the answer.

Figure 6.5



The set of no-arbitrage prices.

Here is an easy way to answer this question. Take the two *state* payoff vectors:

$$\begin{aligned} \tilde{Y}(1) &= \langle \tilde{Y}_a(1),\ \tilde{Y}_b(1) \rangle &= \langle 2, 4 \rangle \\ \tilde{Y}(2) &= \langle \tilde{Y}_a(2),\ \tilde{Y}_b(2) \rangle &= \langle 3, 1 \rangle. \end{aligned}$$

Mark them on the plane, and draw the two rays form the origin through the points. Now shade in the region between the rays, as shown in Figure 6.5. This region (not including the rays) is called the *open convex cone* generated by $\langle \tilde{Y}_a(1),\ \tilde{Y}_b(1) \rangle$ and $\langle \tilde{Y}_a(2),\ \tilde{Y}_b(2) \rangle$. *It is the set of no-arbitrage prices.*

A point is in this convex cone if and only if it is a linear combination of the points $\langle \tilde{Y}_a(1), \tilde{Y}_b(1) \rangle$ and $\langle \tilde{Y}_a(2), \tilde{Y}_b(2) \rangle$, with positive weights. That is, if and only if there are $p_1 > 0$ and $p_2 > 0$ such that

$$\langle q_a, q_b \rangle = p_1 \langle \tilde{Y}_a(1), \tilde{Y}_b(1) \rangle + p_2 \langle \tilde{Y}_a(2), \tilde{Y}_b(2) \rangle. \tag{6.6}$$

These positive weights are state prices. Equation (6.6) is really two equations, one for each of the components of the vectors:

$$q_a = p_1 \tilde{Y}_a(1) + p_2 \tilde{Y}_a(2) \tag{6.7}$$

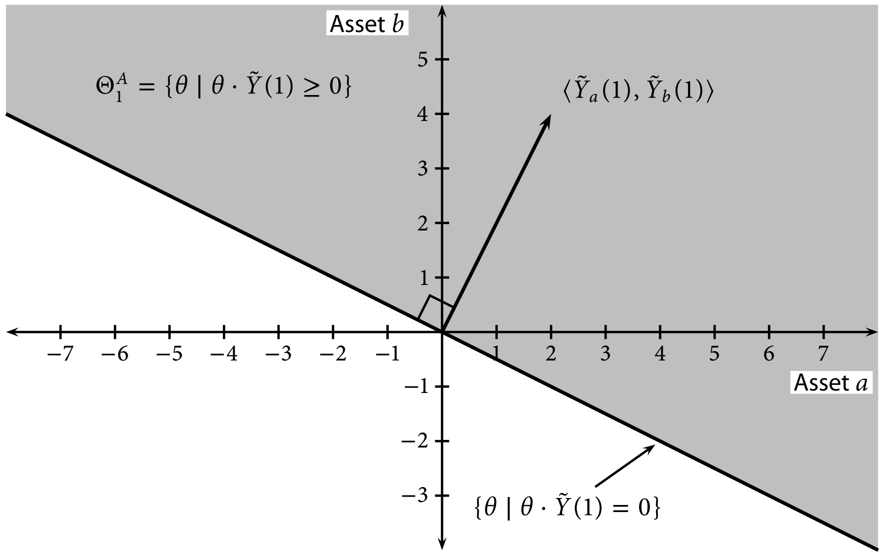
$$q_b = p_1 \tilde{Y}_b(1) + p_2 \tilde{Y}_b(2). \tag{6.8}$$

Observe that equations (6.7) and (6.8) are examples of equation (6.5).

In summary:

Proposition 3. *Asset prices $\langle q_a, q_b \rangle$ satisfy no-arbitrage if and only if there is a state-price vector.*

Figure 6.6



The set Θ_1^A of portfolios with non-negative payoff in state 1.

Now let's take a step back and illustrate graphically that the set of no-arbitrage prices is the open, convex cone generated by the state-payoff vectors. First, we draw the set Θ^A of "potential" arbitrage portfolios, which is the set of portfolios with non-negative payoff in both states. These are the portfolios that could be arbitrage portfolios for the wrong prices. This set is the intersection of the sets Θ_1^A and Θ_2^A of portfolios that have a non-negative payoff in states 1 and 2, respectively. Recall that for a portfolio $\theta = \langle \theta_a, \theta_b \rangle$, the payoff in state s is the dot product $\theta \cdot \tilde{Y}(s)$ of θ and the state payoff vector $\tilde{Y}(s) = \langle \tilde{Y}_a(s), \tilde{Y}_b(s) \rangle$. Then

$$\begin{aligned}
\Theta_1^A &= \{\theta \in \mathbb{R}^2 \mid \theta \cdot \tilde{Y}(1) \geq 0\} \\
\Theta_2^A &= \{\theta \in \mathbb{R}^2 \mid \theta \cdot \tilde{Y}(2) \geq 0\} \\
\Theta^A &= \Theta_1^A \cap \Theta_2^A.
\end{aligned}$$

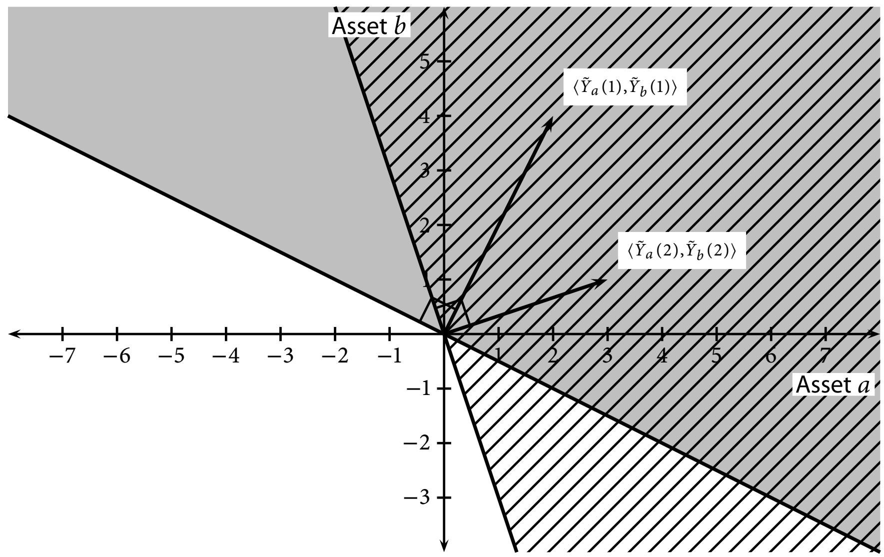
The condition $\theta \cdot \tilde{Y}(1) = 0$ means that θ is orthogonal or perpendicular to $\tilde{Y}(1)$. Figure 6.6 shows $\tilde{Y}(1)$ and the line of points perpendicular to $\tilde{Y}(1)$.[4] $\theta \cdot \tilde{Y}(1) > 0$ means that θ lies "above" the line perpendicular to $\tilde{Y}(1)$, that is, on the side of $\tilde{Y}(1)$. This region, which (together with the line) is Θ_1^A, is shaded in Figure 6.6. Figure 6.7 shows Θ_1^A and Θ_2^A, and their intersection Θ^A.

You can check that each portfolio in Θ^A not only has a non-negative payoff in each state, but also a positive payoff in at least one state. Therefore, asset prices $q = \langle q_a, q_b \rangle$ satisfy no-arbitrage if and only if the cost $q \cdot \theta$ of each portfolio in Θ^A is positive. That is, the set Q^{NA} of no-arbitrage prices is given by

$$Q^{NA} = \{q \in \mathbb{R}^2 \mid q \cdot \theta > 0 \text{ for all } \theta \in \Theta^A\}.$$

4. Note that on these axes, which are labeled "Asset a" and "Asset b," we are drawing both portfolios $\langle \theta_a, \theta_b \rangle$ and state-payoff vectors $\langle \tilde{Y}_a(1), \tilde{Y}_b(1) \rangle$.

Figure 6.7



The sets Θ_1^A and Θ_2^A of portfolios with non-negative payoff in states 1 and 2, respectively. Θ_1^A is shaded, and Θ_2^A has diagonal lines. Their intersection is the set Θ^A of potential arbitrage portfolios.

The boundaries of this set are perpendicular to the boundaries of Θ^A, as seen in Figure 6.8. Take portfolios θ' and θ'' on the two boundaries. The asset prices such that $q\theta' > 0$ are the region that lies above the line perpendicular to θ'; the asset prices such that $q\theta'' > 0$ are the region above the line perpendicular to θ''. The intersection is Q^{NA}.

Observe, from Figure 6.7, that the ray perpendicular to θ' passes through $\tilde{Y}(1)$, and the ray perpendicular to θ'' passes through $\tilde{Y}(2)$. Therefore, Q^{NA} is the region between the rays that pass through $\tilde{Y}(1)$ and $\tilde{Y}(2)$, as we set out to demonstrate.

Exercise 6.3. Consider a model of an asset market in which there are 2 states (1 and 2) and 2 assets (a and b) with the following payoffs:

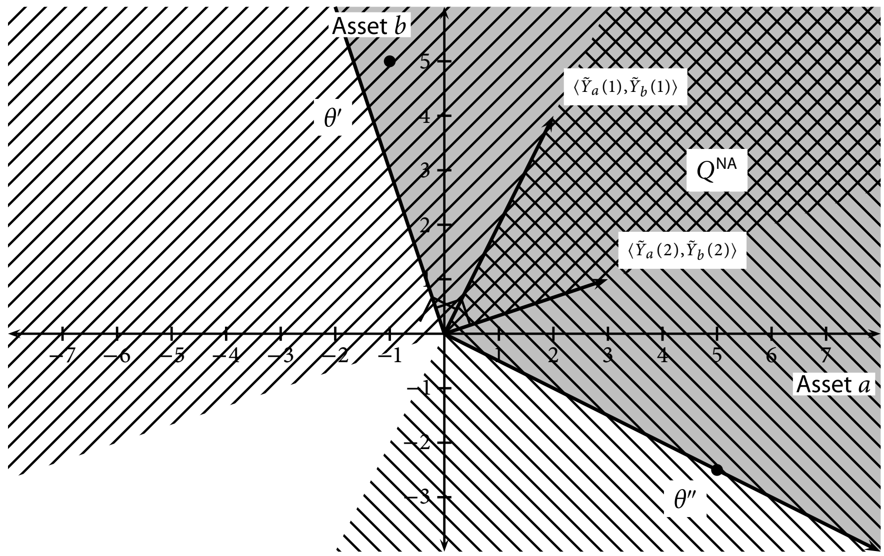
	Payoffs	
	Asset a	Asset b
State 1	1	3
State 2	2	1

a. Verify that the asset prices $q_a = q_b = 5$ are no-arbitrage prices by calculating the state prices.

b. Graph the entire set of no-arbitrage prices. Explain what you are doing, and specify whether the boundary of the region you draw is in the set of no-arbitrage prices.

c. Suppose that you learn that a trader's baseline wealth is \$12 in state 1 and \$10 in state 2, and, after liquidating her portfolio, she has an allocation of \$16 in state 1 and

Figure 6.8



The shaded regions is the set Θ^A of potential arbitrage portfolios. These portfolios must have positive cost. The set of asset prices for which portfolio θ' has positive cost has lines going up to the right. The set of asset prices for which portfolio θ'' has positive cost has lines going down to the left. The intersection of these two sets—the crosshatch region—is the set of no-arbitrage prices. Its boundaries are the rays through $\tilde{Y}(1)$ and $\tilde{Y}(2)$.

\$8 in state 2. Calculate her portfolio.

6.3.3 Complete and incomplete markets

An important fact is that there can be at most S linearly independent assets. Therefore, there are necessarily redundant assets if the number of assets is greater than the number of states (as in the example). If there are exactly S linearly independent assets, then we say that markets are *complete*. If there are fewer than S linearly independent assets, then markets are *incomplete*. Markets are necessarily incomplete if there are fewer assets than states ($J < S$), but they can also be incomplete even if there are as many assets as states ($J \geq S$).

In our $2 \times 2 \times 2 \times 2$ model, we assumed that the payoffs of the two assets were not proportional, which means that the assets are linearly independent. Since there were only two states, markets were complete. Much of our analysis of that model depended on the completeness of markets, and does not hold when markets are incomplete.

In particular, we showed that the multiple budget constraints in the portfolio selection problem could be reduced to a single budget constraint in the consumption choice problem, with prices at consumption in each state given by state prices. Now that consumption in period 0 is allowed, the analogue of that single budget constraint is

$$z_0 + p_1 z_1 + p_2 z_2 + \ldots + p_s z_s = w_0 + p_1 w_1 + p_2 w_2 + \ldots + p_s w_s. \tag{6.9}$$

All prices are in period-0 dollars, which is why the price of consumption in period 0 is 1. $p_1, \ldots, p_s$ are state prices.

The budget set that satisfies a standard budget constraint such as in equation (6.9) has a dimension that is one less than the number of goods. For example, when there are two goods, the budget set is a line. When there are three goods, it is a plane. In this model, it is an S-dimensional "hyperplane." If I am choosing quantities of J independent assets, I only have enough degrees of freedom to trace out a J-dimensional set of allocations. If $J < S$, then my budget set is of a lower dimension than a standard budget set, and so I do not face a standard consumption choice problem.

However, didn't we show that for any no-arbitrage asset prices, there are corresponding state prices? Yes, but these state prices do not entirely determine the constraints on consumption if markets are incomplete. The actual budget set will be a strict subset of the standard budget set generated by the state prices, and there will be more than one state-price vector.

The model with incomplete markets is complex and unfamiliar; this is the message I am trying to convey in the previous paragraphs. However, when markets are complete, we are back to a familiar world in which the portfolio selection problem reduces to a standard consumption choice problem and the financial equilibrium is determined by the real equilibrium in a hypothetical market in which consumption in each state is traded directly.

The following fact will help us to show this:

Proposition 4. *When markets are complete, the equilibrium allocations are the same whatever are the S independent assets.*

Here is brief explanation. Suppose we start out with complete asset markets, and add a new set of S independent assets. Since among the original assets there were already S independent assets, the new assets are redundant and hence do not affect the allocations. Since the S new assets are independent, each of the original assets is equivalent to a portfolio of the new assets. Hence, we can then eliminate all the old assets without affecting the allocations. We can thus replace the original assets with any S independent assets without affecting the equilibrium allocations. (When assets are incomplete, on the other hand, the equilibria depend not simply on how many independent assets there are, but on what the payoffs of these assets are).

Let's choose the simplest asset payoffs:

	Payoffs				
Asset	State 1	State 2	…	State $S-1$	State S
1	1	0	…	0	0
2	0	1	…	0	0
$\vdots$	…	…	$\ddots$	…	…
$S-1$	0	0	…	1	0
S	0	0	…	0	1

That is, there are S assets, and asset s pays 1 in state s and 0 in the other states. These are called *canonical asset returns*. With canonical asset returns, the constraint

$$z_s \ = \ w_s + \theta_1 \tilde{Y}_j(s) + \ldots + \theta_J \tilde{Y}_J(s)$$

in the portfolio selection problem becomes

$$z_s - w_s \;=\; \theta_s.$$

The portfolio selection problem in equation (6.1) then reduces to

$$\begin{aligned} \max_{z_0, z_1, \ldots, z_s} \quad & u_i(z_0, z_1, \ldots, z_S) \\ \text{subject to:} \quad & z_0 + q_1 z_1 + \ldots + q_S z_S = w_0 + q_1 w_1 + \ldots + q_s w_s. \end{aligned}$$

For these canonical payoffs, the state prices are equal to the asset prices, and it is easy to find the reduced-form budget constraint.

6.3.4 Pareto optimality and suboptimality

You surely know and love the First Welfare Theorem, which says that the equilibrium allocations in the market for bananas, apples and oranges are Pareto efficient. Since the asset market is mathematically equivalent to the fruit market when assets are complete, we can conclude that the allocations in the asset market are also Pareto efficient, *without having to reprove the First Welfare Theorem!*

It is always amazing that decentralized trade in markets can result in efficient allocations, even if only under very unrealistic assumptions such as perfect competition, perfect rationality of economic agents, and no externalities. One of the unrealistic assumptions we made for the asset markets is that they are complete. One can show that when markets are not complete, the equilibrium allocations are almost never efficient. Consider an extreme case—when there are no assets at all. In this case, state-contingent trade is not possible. This is only efficient if the original state-dependent allocations of wealth are Pareto Optimal and there are no gains from trade. Now consider intermediate levels of incompleteness. The explanation of inefficiency is tricky for this case; here is just a taste. Trade is possible, but the budget sets are too small to guarantee that the marginal rates of substitution of wealth in the various periods and states are equal for all consumers. For example, with two states, an allocation is a point in 3-dimensional space. With complete markets, the budget set is a plane, but with incomplete markets it is a line or a point. Tangency of the indifference curves (surfaces) for such budget sets does not imply that all the households' indifference curves have the same slope at the points of tangency.

Are incomplete markets an important source of market failure? It is an empirical fact that markets are incomplete. Furthermore, there are other kinds of market incompleteness not in our model that also reduce efficiency, such as restrictions on short sales. However, a reason for this incompleteness is that transaction costs limit the number of assets that are traded. Could it be that if we model transaction costs and the endogenous creation of assets by financial intermediaries, we would find that markets are "optimally incomplete," given the transaction costs? This is a theoretical question which economists cannot yet answer. Furthermore, economists do not yet have empirical measurements of the welfare loss due to market incompleteness.

Exercise 6.4. This is a two-state example of arbitrage prices. Suppose that there is a riskless bond with payoff of \$1 in each state, and a risky stock with a payoff of \$2 in state 1 and \$5 in state 2.

a. Draw a graph showing the set of no-arbitrage prices for these two assets.

b. Suppose there is also a call option on the stock, with a strike price of \$3, which gives the buyer the option of acquiring the stock (with dividend) for \$3. What are the payoffs of this asset? What portfolio of the bond and the stock give the same payoffs? What is the no-arbitrage price of the option, as a function of the prices of the bond and the stock?

6.4 Capital Asset Pricing Model

State prices are useful for thinking about the determination of a real equilibrium, and for deriving many properties of asset prices and financial equilibria. However, they are not of direct empirical use because they are difficult to measure.

With additional assumptions on preferences and/or the distributions of asset payoffs and wealth, it is possible to derive a variety of empirically useful formulae about asset prices. In this section, we will derive the best known of these, the Capital Asset Pricing Model (CAPM).

Assume that markets are complete. We need the following two properties of complete markets:

- Equilibrium real allocations are efficient, and the shadow prices of consumption in each state is its state price.
- For any payoff $\tilde{Y}$, there is a portfolio with this payoff.

Let $p_1, \ldots, p_S$ be the equilibrium state prices. Then the equilibrium price of asset j is

$$q_j = p_1 \tilde{Y}_j(1) + \cdots + p_S \tilde{Y}_j(S). \tag{6.10}$$

Consider the equilibrium situation of a typical trader i. For now, I will omit the superscript i, to simplify notation. Let $\tilde{z} = \langle z_0, z_1, \ldots, z_S \rangle$ be the trader's equilibrium allocation. Then $\tilde{z}$ satisfies the trader's first-order conditions for utility maximization. This means that the ratio of the price (p_s) of consumption in state s to the price $(p_0 = 1)$ of consumption in period 0 must be equal to the ratio of the marginal utility in state s to the marginal utility in period 0:

$$p_s = \frac{\partial U / \partial z_s}{\partial U / \partial z_0} \qquad \text{for } s = 1, \ldots, S. \tag{6.11}$$

Assume that the trader is an expected utility maximizer with (for simpler notation) time-separable and time-homogeneous utility:

$$U(z_0, z_1, \ldots, z_S) = u(z_0) + \delta(\pi_1 u(z_1) + \cdots + \pi_S u(z_S)).$$

u is the VNM utility function, δ is the discount factor, and $\pi_, \ldots, \pi_S$ are beliefs. The right-hand side is the utility in period 0 plus the discounted value of the expected utility in period 1. Then equation (6.11) becomes

$$p_s = \frac{\delta \pi_s u'(z_s)}{u'(z_0)}. \tag{6.12}$$

and the asset-pricing formula in equation (6.10) becomes

$$q_j = \frac{\delta}{u'(z_0)}\left(\pi_1 u'(z_1)\tilde{Y}_j(1) + \cdots + \pi_S u'(z_S)\tilde{Y}_j(S)\right) = \frac{\delta}{u'(z_0)} E[\tilde{Y}_j u'(\tilde{z})]. \quad (6.13)$$

This is mildly interesting, showing how the asset price is higher when payoffs are positively correlated with marginal utilities (the asset payoff is high when marginal utility is high, which is when consumption is relatively scarce), but is still lacks empirical usefulness because the state-dependent equilibrium marginal utilities are even harder to measure than state prices.

Now assume that the trader has quadratic utility, which means that[5]

$$u(z) = z - \frac{1}{2}bz^2.$$

Recall that a quadratic utility function is eventually decreasing. We assume that it is increasing over any feasible wealth levels for the trader. Recall also that the quadratic utility function is the one function for which variance is a measure of risk. We do not use this property directly. Instead, the property that we make use of is that the derivative is linear:

$$u'(z) = 1 - bz.$$

Then equation (6.12) becomes

$$p_s = \frac{\delta \pi_s (1 - bz_s)}{1 - bz_0}. \quad (6.14)$$

This formula is useful only if we know the trader's utility function and state-dependent consumption, but by aggregating over the traders we can obtain a much more empirically useful formula in terms of aggregate consumption. We start by inserting the superscript i for trader i (on δ^i, b^i and z_s^i) in equation (6.14), and rearranging:

$$\frac{1}{\delta^i}\left(\frac{1}{b^i} - z_0^i\right) p_s = \pi_s \left(\frac{1}{b^i} - z_s^i\right).$$

Next we sum both sides of the equation over traders $i = 1, \ldots, n$:

$$\sum_{i=1}^{n} \frac{1}{\delta^i}\left(\frac{1}{b^i} - z_0^i\right) p_s = \sum_{i=1}^{n} \pi_s \left(\frac{1}{b^i} - z_s^i\right)$$

$$p_s = \pi_s \frac{\sum_{i=1}^{n} \frac{1}{b^i} - \sum_{i=1}^{n} z_s^i}{\sum_{i=1}^{n} \frac{1}{\delta^i}\left(\frac{1}{b^i} - z_0^i\right)}.$$

Because $\tilde{z}^i$ is part of an *equilibrium* allocation, total consumption must equal the total endowment. That is, $\sum_{i=1}^{n} z_s^i = w_s$, where $w_s = \sum_{i=1}^{n} w_s^i$ is the total endowment in state s. Let

$$A = \frac{\sum_{i=1}^{n} \frac{1}{b^i}}{\sum_{i=1}^{n} \frac{1}{\delta^i}\left(\frac{1}{b^i} - z_0^i\right)}$$

$$B = \frac{1}{\sum_{i=1}^{n} \frac{1}{\delta^i}\left(\frac{1}{b^i} - z_0^i\right)}.$$

5. This particular quadratic form is perfectly general, because it is a linear affine transformation of any more general quadratic function $a_0x^2 + a_1x + a_2$.

Then we have

$$p_s = \pi_s(A - Bw_s).$$

Therefore, equation (6.13) can be written

$$\begin{aligned} q_j &= \pi_1(A - Bw_1)\tilde{Y}_j(1) + \cdots + \pi_S(A - Bw_S)\tilde{Y}_j(S) \\ &= A(\pi_1\tilde{Y}_j(1) + \cdots + \pi_S\tilde{Y}_j(S)) + B(\pi_1 w_1\tilde{Y}_j(1) + \cdots + \pi_S w_S\tilde{Y}_j(S)) \\ &= A\ E[\tilde{Y}_j] - B\ E[\tilde{Y}_j\tilde{w}], \end{aligned}$$

where $\tilde{w}$ is the random aggregate endowment $\langle w_1, \ldots, w_S\rangle$ in period 1.

We could use this formula directly. Aggregate wealth as a correlation of an asset's payoff is easier to measure than quantities such as state prices. A and B are parameters that depend on the utility functions. They could be estimated from some data where q_j and $\tilde{Y}_j$ are known, and then used to price other assets. However, it is common to write this as a "beta" model.

First, divide the pricing equation through by q_j to obtain an equation involving the return $\tilde{R}_j = \tilde{Y}_j/q_j$ of the asset:

$$1 = A\,E[\tilde{R}_j] - B\,E[\tilde{R}_j\tilde{w}]. \tag{6.15}$$

Since markets are complete, there is a "market" portfolio whose payoff is perfectly correlated with the aggregate endowment, and hence has payoff $\tilde{Y}_m = \tilde{w}$. Let its price and return be q_m and $\tilde{R}_m = \tilde{Y}_m/q_m$, respectively. Then $\tilde{w} = q_m\tilde{R}_m$, and equation (6.15) becomes

$$1 = A\,E[\tilde{R}_j] - Bq_m\,E[\tilde{R}_j\tilde{R}_m]. \tag{6.16}$$

Recall the definition of covariance:

$$\mathrm{Cov}(\tilde{R}_j, \tilde{R}_m) = E[\tilde{R}_j\tilde{R}_m] - E[\tilde{R}_j]E[\tilde{R}_m].$$

Then equation (6.15) becomes

$$\begin{aligned} 1 &= A\,E[\tilde{R}_j] - Bq_m(\mathrm{Cov}(\tilde{R}_j, \tilde{R}_m) + E[\tilde{R}_j]\,E[\tilde{R}_m]) \\ 1 &= (A - Bq_m\,E[\tilde{R}_m])E[\tilde{R}_j] - Bq_m\,\mathrm{Cov}(\tilde{R}_j, \tilde{R}_m). \end{aligned}$$

This equation holds for a riskless portfolio with return R_0 in each state. Since $\mathrm{Cov}(R_0, \tilde{R}_m) = 0$, we have

$$\begin{aligned} 1 &= (A - Bq_m\,E[\tilde{R}_m])R_0 \\ 1/R_0 &= (A - Bq_m\,E[\tilde{R}_m]). \end{aligned}$$

Therefore, our pricing equation is

$$1 = \frac{1}{R_0}\,E[\tilde{R}_j] - Bq_m\,\mathrm{Cov}(\tilde{R}_j, \tilde{R}_m) \tag{6.17}$$

$$E[\tilde{R}_j] - R_0 = Bq_m\,\mathrm{Cov}(\tilde{R}_j, \tilde{R}_m)R_0. \tag{6.18}$$

For the market portfolio with return $\tilde{R}_m$, $\mathrm{Cov}(\tilde{R}_m, \tilde{R}_m) = \mathrm{Var}(\tilde{R}_m)$, and we have

$$\begin{aligned} E[\tilde{R}_m] - R_0 &= Bq_m\,\mathrm{Var}(\tilde{R}_m)R_0 \\ Bq_m &= \frac{E[\tilde{R}_m] - R_0}{\mathrm{Var}(\tilde{R}_m)R_0}. \end{aligned}$$

Substituting the right-hand side for Bq_m in equation (6.18):

$$E[\tilde{R}_j] - R_0 = \frac{\text{Cov}(\tilde{R}_j, \tilde{R}_m)}{\text{Var}(\tilde{R}_m)}(E[\tilde{R}_m] - R_0)$$
$$E[\tilde{R}_j] - R_0 = \beta_j(E[\tilde{R}_m] - R_0),$$

where

$$\beta_j = \frac{\text{Cov}(\tilde{R}_j, \tilde{R}_m)}{\text{Var}(\tilde{R}_m)}.$$

In the typical presentation of the CAPM, this formula is derived using mean-variance analysis. $\tilde{R}_m$ is the return on the "aggregate" portfolio, and β_j is called the *beta* of asset j. Our derivation gives a slightly different interpretation. $\tilde{R}_m$ is the return on a portfolio that is perfectly correlated with the aggregate portfolio as long as all variations in wealth are due to variable payoffs of tangible assets that are traded. In a multiperiod setting, there is a further distinction. Wealth is equal to consumption in this simple model. With more periods, wealth and consumption differ. In the generalization of the formula, we would be interested in the correlation of an asset's payoffs with aggregate *consumption*, leading to what is called a consumption-based CAPM.

Chapter 7

Contracting with Hidden Actions

7.1 Efficient contracts with moral hazard

Exercise 7.1. Consider the following problem of moral hazard between a principal and an agent (e.g., an employer and an employee). The agent works on a project that may result in a gross profit to the principal of either 1600 or 2500. The agent can exert low or high effort, denoted e_L and e_H, respectively. If low, the probability that the gross profit is 2500 is equal to 1/2; if high, that probability is 8/9.

The principal is risk neutral. The agent is an expected-utility maximizer who is risk averse with respect to money. His utility when receiving wage w and exerting effort e is

$$u(w,e) = \begin{cases} w^{1/2} & \text{if } e = e_L \\ w^{1/2} - 3 & \text{if } e = e_H. \end{cases}$$

The purpose of this exercise is to derive the entire set of efficient contract. This requires the use of numerical software. The main concepts this exercise illustrates are the constraints that define the first-best and second-best contracts.

a. Calculate the expect gross profit with low effort and with high effort.

b. Plot the frontier of the set of low-effort contracts and of the set of high-effort contracts for the case where there is no moral hazard (in utility space, i.e., worker's expected utility on one axis, and the employer's expected net profits on the other axis). Show your calculation.

c. Repeat the last question for the case of moral hazard.

d. Suppose that, if no contracting takes place, then the principal's profit is 0 and the agent's utility is 30. On a separate graph, plot the set of individually rational and efficient contracts both with and without moral hazard.

e. Pick a point in the interior of the set of IR and efficient contracts without moral hazard (neither party gets all the gains from trade). Now suppose that we switch to a regime with moral hazard. Indicate which IR and efficient contracts with moral hazard make both parties worse off than under the original outcome you picked.

Exercise 7.2. What is wrong (and what is right) with the following: "Basing teachers' pay on surprise classroom inspections or the results of student exams is exploitive. Teachers should strongly resist this when negotiating their contracts."

7.2 Contracts that give one party the gains from trade

Exercise 7.3. Assume, in Exercise 7.1, that the worker (you) has a reservation utility of 30 units, and that the employer gets all the gains from trade.

a. Suppose there is no moral hazard. What contract would the principal offer if he could get all the gains from trade?

b. Suppose there is moral hazard. What contract would the principal offer if he could get all the gains from trade?

Exercise 7.4. Reinterpret Exercise 7.1 as an example moral hazard in insurance markets. You are buying insurance against the theft of money from your house. Suppose that there is some chance that someone will enter your house and steal \$900 that you have lying around. Your total wealth is \$2,500. If you stay home all the time (high level of care, e_H) the probability of a theft is 1/9. If you go out often (low level of care, e_L), the probability of theft is 1/2. Your utility from money w and the level of care is

$$u(w, e) = \begin{cases} w^{1/2} & \text{if } e = e_L \\ w^{1/2} - 3 & \text{if } e = e_H. \end{cases}$$

a. What is the expected loss for each of the two levels of care.

b. Assume that the insurance companies make zero profits, so that you get all the gains from trade. What is the best contract you can design (i.e., what is the level of care, the level of coverage, and the premium) if there is no moral hazard?

c. Is this policy incentive compatible if the insurance companies cannot observe whether you leave the house alone?

d. With moral hazard, what is the optimal contract you can design which has as a "clause" that you go out often?

e. With moral hazard, what is the optimal contract you can design which has as a "clause" that you stay home? (It suffices to give the equations that define the optimal contract.)

Exercise 7.5. A friend has asked you to sneak a six-pack of beer for him into a concert. Being the kind of friend that you are, you cannot be trusted not to drink the beer yourself just before going into the concert. Unfortunately, if you do not drink the beer, then there is some possibility (1/10) that the beer will be confiscated by security on your way in. Thus, if you show up with no beer, your friend cannot tell whether you drank the beer or it was confiscated. (Remember, this is fiction.) You are going to offer a deal to your friend where the fee your friend pays for your service depends on whether you deliver the beer. Here are the important things you need to know in order to design an optimal contract:

1. Your utility if you get x dollars out of this transaction is

$$u(x) = \begin{cases} -e^{-.2x} & \text{if you don't drink the beer} \\ -e^{-.2(x+5)} & \text{if you drink the beer.} \end{cases}$$

   (x is positive if you receive a payment from your friend and negative if you pay money to your friend.) Thus (i) you are risk averse, with constant absolute risk aversion, and (ii) beer and money are perfect substitutes, with the six-pack being equivalent to \$5.
2. Your friend's utility if she gets x dollars out of this transaction is

$$v(x) = \begin{cases} x & \text{if she doesn't drink the beer.} \\ x+3 & \text{if she drinks the beer before the concert} \\ x+8 & \text{if she drinks the beer in the concert} \end{cases}$$

   (x is positive if she receives a payment from you and is negative if she pays money to you.) Thus (i) your friend is risk neutral, and (ii) beer and money are perfect substitutes, with beer being equivalent to \$8 if drunk in the concert and \$3 if drunk before the concert.

a. Suppose your friend will accept any deal you offer her for which her expected utility is at least 3, which is what she would get if she rejected your deal and just drank the beer before the concert.

Write down the two equations whose solution gives the optimal contract. (OPTIONAL: Solve the two equations numerically.)

b. Suppose the bargaining power is shifted to your friend. She makes a take-it-or-leave-it offer, and you accept the deal as long as your expected utility is at least −1 (i.e., $-e^0$), which is what you would get if you didn't make any deal with your friend. Write down the two equations whose solution gives the optimal contract for your friend. (This time, if you write the equations properly, you can solve them easily.) What is the first-best contract, and how do the first-best and second-best contracts compare in terms of the expected utility that you and your friend get?

Chapter 8

Monopolistic screening with hidden valuations

8.1 Nonlinear pricing

8.2 Differentiated products

Exercise 8.1. New Jersey state law prohibits self-service gas. The following example of monopolistic screening may be relevant!

In the following, imagine that gas is not a divisible good; e.g., a consumer either gets a fill up or does not.

Suppose that a town has a single gas station. Consider the pricing decisions of this gas station, taking the prices of gas in the surrounding towns as fixed. Suppose that there are two types of consumers, frugal and lazy. The lazy consumers will pay up to $1.20 per gallon, whereas the frugal ones will pay up to $1.00 per gallon. (The difference lies in the willingness of the consumers to drive to a neighboring town in order to save money.) The cost of gas, including operating expenses, for the gas station is $.90 per gallon.

a. If the station can charge different prices to frugal and lazy customers (e.g., because all frugal customers drive Hyundai and and all lazy ones drive Mercedes), what prices will it charge to each type?

b. Suppose instead that the gas station cannot distinguish between a frugal and lazy consumer. (Suppose also that it is illegal for the gas station to offer self-service gasoline.) What prices might the gas station charge, and what information would you need to know to determine which price is optimal?

c. Suppose again that the station cannot tell who is frugal and who is lazy, but the station can offer self-service gasoline. Furthermore, suppose that self-serve gas is an inconvenience to the customer, but is not any cheaper for gas stations to provide. Suppose that frugal customers are willing to pay up to $.01 per gallon for full serve, whereas lazy customers are willing to pay up to $.10 per gallon for full serve.

(Note: The insurance example studied in class is conceptually closely related, but I have simplified this problem by allowing only two levels of inconvenience, full-serve and self-serve, which would be analogous to allowing only two levels of coverage in the insurance example.)

1. For each of the following pricing strategies, find a better price strategy and explain

why it is better:

(a) Full-serve: \$1.15. Self-serve: \$.99.

(b) Full-serve: \$1.06. Self-serve: \$.99.

2. Then find the optimal pricing strategy, among those with which the station sells both full and self-serve gasoline. Specify for which consumers the self-selection constraint is binding and for which consumers the individual rationality constraint is binding.
3. Who is better off and who is worse off (among frugal customers, lazy customers, and the gas station), compared to when self-serve gasoline is not allowed?

8.3 Bundling

Chapter 9

Screening with adverse selection

9.1 The nature of adverse selection

Moral hazard arises when parties to a contract take actions after signing the contract that affect each other, and these actions are not observable to all parties.

Now we consider another kind of information asymmetry, when parties have different information *at the time of contracting*. This information is about *exogenous* characteristics of the agents (i.e., characteristics not having to do with the agents' future actions) that are relevant to the contractual relationship. Here are some leading examples:

- In product markets, the sellers typically have more information about the goods than the buyers.
- In labor markets, the potential employees have more information about their skills and productivity than the employers.
- In insurance markets, different individuals have different risks, for exogenous reasons.
  - The probability of a fire in a firm might depend on the condition of the internal wiring, on how smart the manager of the firm is, or on the neighborhood (if there is a threat of arson).
  - The probability of an accident depends on how good the driver is.
  - The probability of a heart attack depends on a person's current health.

  Insured parties know more about these risks than the insurers.

9.1.1 Adverse selection

Consider one of the contracting situations described above, in which there is one fully informed party with private information about her characteristics—which we call her type—and one uninformed party. If you are the uninformed party, then you have to take into account that the willingness of the other person to accept a contract may depend on her type. Therefore, only a selection of the possible types will agree to a contract. It is often the case that the people with whom it is least desirable to contract also have the worst outside options, and hence are the most likely ones to accept a contract. The selection of types who accept a contract is then a bad selection for the uninformed party, and this phenomenon is called *adverse selection*. Here are some examples:

- In the used-car market, a bad car is not only worth less to a buyer, but is is also worth less to a seller. Hence, the seller who would accept a fixed offer for their cars are those with the most cars.

- In labor markets, the least productive workers have the most difficulty making a living on their own or finding a decent job, and hence are the most likely ones to accept any job offer.
- In insurance markets, the highest-risk consumers benefit the most from buying insurance, and hence are the most likely to accept a given contract.

If you, as the uninformed party, do not take this adverse selection into account, you may be in for an unpleasant surprise. For example, if an insurance company calculates the actuarially fair premium based on the statistical average of the risks over the whole population of potential insured, then the high-risk parties will tend to buy the insurance but the low-risk parties will not. The statistical average of the risk for the group that actually buys the insurance will be higher than for the whole population, and so the insurance company ends up losing money.

If you are aware of this adverse selection problem, then you should interpert the willingness of someone to contend with you as a bad signal about the benefit to you of trading with that person. As Groucho Marx said, "I would never want to join a club that would have me as a member."

Here is an example, framed as a story about the used-car market. Suppose that you are considering buying a car from a man named Frank. You know that Frank's car is worth \$1,000 more to you than to him (because you just got a job to which you need to commute, and Frank just quit a job from which he was commuting). Let θ be the value of the car to you, which is a measure of the quality of the car.

Suppose that you are in a position to make a take-it-or-leave it offer to Frank. If you knew θ, you would just offer Frank θ for the car (or θ plus a penny). This is the smallest amount that Frank is willing to accept. Frank is then indifferent between trading and not trading, and you get all the gains from trade (you pay θ and the car is worth $\theta+1000$ to you). Trade always takes place, which is efficient since the car is always worth more to you than to Frank.

The problem is that Frank knows θ but you do not. You cannot rely on Frank to reveal θ; he will always say whatever gets him the highest offer, independently of his true θ. Therefore, your offer cannot be contingent on θ. In choosing your offer, you have to balance the amount you pay with the possibility of trade. For example, suppose that you are deciding whether to offer \$3,000 or \$4,000. If you offer \$3,000, Frank accepts whenever $\theta \leq 3000$, and so the car is worth between \$1,000 and \$4,000 to you. If you increase your offer to \$4,000, then you also purchase the car when θ is between 3,000 and 4,000. This is good, since the car is worth \$4,000 to \$5,000 to you for these values of θ. On the other hand, you still purchase the car when θ is between 0 and 3,000, but you have to pay \$1,000 more, which is bad.

Suppose that you are risk neutral and choose your offer x to maximize the expected value of your surplus, $V(x)$. You trade when $\theta \leq x$, and hence Prob $(\theta \leq x)$ is the probability of trade. Conditional on trade, the expected value of the product to you is $E[\theta \mid \theta \leq x] + 1000$, and so the expected value of your surplus is $E[\theta \mid \theta \leq x] + 1000 - x$. Hence,

$$V(x) \;=\; \text{Prob }(\theta \leq x)(E[\theta \mid \theta \leq x] + 1000 - x).$$

For example, suppose that θ is distributed uniformly between 0 and 10,000. (This means that $F(\theta) = \theta/10,000$ and $f(x) = 1/10,000$.) Then $E[\theta \mid \theta \leq x] = x/2$, and

Prob $(\theta \leq x) = x/10,000$. $F(x) = x/10000$. Therefore, expected surplus $V(x)$ is

$$V(x) = \frac{x}{10,000}\left(\frac{x}{2} + 1000 - x\right) = -\frac{x^2}{20,000} + \frac{x}{10}.$$

The first-order condition for maximizing $V(x)$ is

$$V'(x) = -\frac{x}{10,000} - \frac{1}{10} = 0.$$

The solution is $x = 1,000$.

Look at the inefficiency that is caused by this informational asymmetry! It is efficient (first-best) for you to always trade, but instead you only trade when $\theta \leq 1000$, or 1/10 of the time. Without the informational asymmetry, your surplus is $1,000 for sure. With the informational asymmetry, your expected surplus is just

$$V(1000) = -\frac{1000^2}{20000} + \frac{1000}{10} = 50.$$

(Conditional on trading, the expected value of θ is 500 and hence your expected surplus is 500; however, the probability of trade is just 1/10.)

On the other hand, when $\theta < 1000$, Frank is better off with the private information. When someone gets additional surplus because of private information, she is said to earn *informational rents*.

This example can be reinterpreted with the labor market or insurance stories (with monopsonistic employers or monopolistic insurance companies). In the labor market story, θ is the worker's reservation wage (the wage the worker can obtain elsewhere), and $\theta + 1000$ is the value of the worker to the firm. In the insurance story, θ is the certainty equivalent of the consumer's risky wealth, and $\theta + 1000$ is the expected value of the consumer's wealth (x is the certain wealth guaranteed by the insurance company to the consumer).

If the seller makes the take-it-or-leave-it offer, and hence, gets the gains from trade in the absence of asymmetric information, the analysis is quite different but again there is a welfare loss because trade does not always take place.

9.2 Screening

In Section 9.1.1, the buyer makes a single offer, and trade either takes place at this offer or does not take place at all. There is no point in offering several different amounts of money and letting the seller choose, because the seller will always choose the largest amount of money.

However, if there is an extra dimension along which contracting is possible, it may be possible to offer several different contracts, each of which is accepted for some values of θ. For example, applicants might be able to select from jobs with commission-based pay and jobs with fixed salaries. The good workers, for whom commissioned based pay has the highest expected value, may prefer commissions over salary, whereas the low-productivity workers prefer the fixed salary over the commissions. This characteristic-dependent selection of contracts is called *self-selection*, and the procedure of using menus of multi-dimensional contracts to induce self-selection is called *screening*.

A mathematically equivalent example is screening in insurance markets. Instead of high and low-productivity workers, we have low and high-risk consumer. Instead of a commission versus a fixed wage, we have partial coverage versus full coverage. In this section, I will develop a model of such screening, using the interpretation of insurance markets.

9.2.1 An insurance story

Consider the following fire-insurance example. There are many firms with a risk of a fire, which results in a loss of $L = \$100,000$. Each firm's gross profit without the loss (good state) is $w_g = \$120,000$, and with the loss (bad state) is $w_b = \$20,000$. If a firm buys an insurance contract which has a premium of p and provides coverage x when there is a fire, then its net profits z_g and z_b in the good and bad states, respectively, after adjusting for insurance premiums and reimbursements, are

$$\begin{aligned} z_g &= w_g - p \\ z_b &= w_b - p + x. \end{aligned}$$

Each firm is owned by a risk-averse expected utility maximizer whose state-independent utility function over net profit is $u(z) = z^{1/2}$.

There is some intrinsic characteristic of each firm that affects the probability of a loss, as discussed in the beginning of Section 9.1. For fraction α of the firms, called low-risk firms, the probability of a loss is $\pi^L = 1/5$. For the other fraction $1 - \alpha$ of the firms, called high-risk firms, the probability is $\pi^H = 1/2$.

There is one or more risk neutral insurance companies, or a risk-neutral government that provides insurance.

9.2.2 Efficient contracts when there is no adverse selection

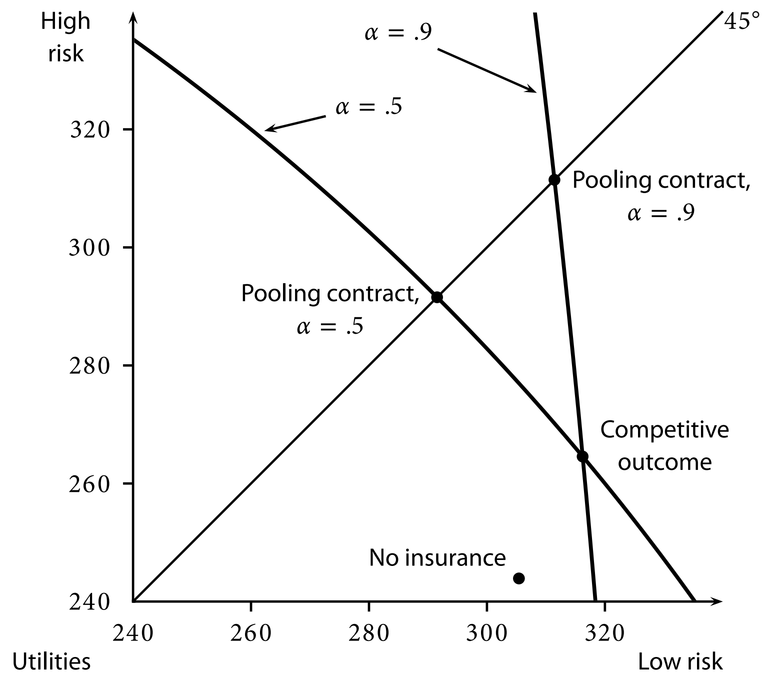
Consider what happens when each firm's type is observable to the insurance companies in a competitive insurance market. Then insurance companies can sell one policy C^H to high-risk firms, and another C^L to low-risk firms, and there really are two separate insurance markets.

With competitive insurance markets, the equilibrium policies provide full coverage and are actuarially fair:

Low risk Full coverage, at a premium of \$20,000. Utility is $100,000^{1/2} = 316$.
High risk Full coverage, at a premium of \$50,000. Utility is $70,000^{1/2} = 265$.

The competitive outcome is Pareto efficient with respect to the welfare of the three groups: low-risk firms, high-risk firms, and insurance companies. That is, it is impossible to make the members of one of these groups better off, without making the members of another group worse off. For example, suppose we want to make the low-risk firms better off. Since in the competitive equilibrium, these firms have full coverage and hence bear no risk, we would have to increase the expected value of these firm's income to make them better off. This means that either the expected value of the high-risk firms' income falls, which leaves them worse off, or the expected profits of the insurance companies fall, leaving them worse off.

Figure 9.1



Efficiency frontiers without adverse selection and with zero expected profits, for $\alpha = .5$ and $\alpha = .9$.

In order to consider the possibility of a wider range of Pareto efficient outcomes, imagine that instead of competitive insurance companies, there is only a risk-neutral government that provides insurance. This government can impose any insurance contracts it wants,[1] but it is interested in the welfare of the firms and also in its own expected profits or deficit. Hence it looks for menus $\{C^L, C^H\}$ of contracts that are efficient with respect to the expected utilities of low and high-risk firms and its own expected profits. Since the government is risk neutral, a menu of contracts is Pareto optimal if and only if all firms have full coverage. The premiums just determine the distribution of income between low-risk firms, high-risk firms, and the government.

The entire Pareto frontier, which is the set of utilities for the efficient outcomes, sits in three dimensions (each point is given by the expected utility for low-risk firms, the expected utility for high-risk firms, and the expected profits of the government), but we can look at a two-dimensional "slice" by fixing the expected profits of the government and just showing the trade off between the low-risk and high-risk firms' utilities. Such a slice is shown in Figure 9.1 for the case where the government earns zero expected profits. It is drawn for $\alpha = 1/2$ and $\alpha = 9/10$.

The menu where both types of firms get the same premium—and hence the same contract and utility—is called a *pooling outcome*. The other menus are called *separating outcomes*. In the pooling outcomes for which the government earns zero expected profits, the common premium p is the *average* expected loss over all the firms. For

1. However, we assume that it treats all low-risk firms alike and all high-risk firms alike.

$\alpha = 1/2$, this average expected loss is

$$(1/2)(1/5)(100,000) + (1/2)(1/2)(100,000) = 35,000.$$

Each firm's utility is $u(85,000) = 292$. When $\alpha = 9/10$, this average expected loss is lower because there are fewer high-risk firms:

$$(9/10)(1/5)(100,000) + (1/10)(1/2)(100,000) = 23,000.$$

Each firm's utility is thus higher: $u(97,000) = 311$.

Compare this with the competitive outcomes, which does not depend on α because it does not involve any subsidies from one type of firm to the other:

Utilities	*low risk*	*high risk*
no insurance	305	244
competitive outcome	316	265
pooling contract $\alpha = 1/2$	292	292
pooling contract $\alpha = 9/10$	311	311

In this example, the low-risk firms prefer no insurance to the pooling contract, when $\alpha = 1/2$, but the opposite is true when $\alpha = 9/10$.

Moving from the bottom-right to the top-left of either of the Pareto frontiers in Figure 9.1, the other points are found by lowering the premiums for high-risk firms and raising the premiums for the low-risk firms, but maintaining full coverage for all firms and keeping the average premium, $\alpha p^L + (1-\alpha)p^H$, equal the average expected loss, $\alpha\pi^L L + (1-\alpha)\pi^H L$.[2]

9.2.3 Self-selection constraints with adverse selection

Suppose now that the insurance companies or the government cannot observe which firms are low risk and which are high risk. The government can no longer dictate who gets what contract. Instead, it can only offer a menu of contracts, and let each firm pick the one it wants.[3] This is called *self-selection*.

Recall that when studying moral hazard, it was convenient for us to work with implicit contracts that contained unenforceable clauses, so that contracts had the same clauses with and without moral hazard. With moral hazard, the unenforceable clauses had to be self-enforcing, or incentive compatible, which was an additional feasibility constraint on contracts. We will take a similar approach here. As when there is no hidden information, the government designs a menu $\{C^L, C^H\}$ of contracts, where C^L

2. Equivalently, the average final wealth z^L and z^H of the low-risk and high-risk firms must equal the average expected value of their endowments:

$$\begin{aligned}
\alpha z^L + (1-\alpha)z^H &= \alpha\left(\tfrac{4}{5}120,000 + \tfrac{1}{5}20,000\right) + (1-\alpha)\left(\tfrac{1}{2}120,000 + \tfrac{1}{2}20,000\right).\\
\alpha z^L + (1-\alpha)z^H &= \alpha(100,000) + (1-\alpha)(70,000).\\
z^H &= 70,000 + \frac{\alpha}{1-\alpha}(100,000 - z^L).
\end{aligned}$$

Hence, we can trace out the Pareto frontier $\langle u(z^L), u(z^H)\rangle$ for the values of z^L and z^H that satisfy the governments zero-profit condition.

3. But as a technical simplification, we assume that when firms are indifferent between contracts, they choose the contract the government would like them to choose.

is for low-risk firms and C^H is for high-risk firms. This designation is feasible only if the firms voluntarily choose their designated contracts, which means that the following *self-selection constraints* must be satisfied:

$$C^L \succeq^L C^H \qquad \text{and} \qquad C^H \succeq^H C^L,$$

where $\succeq^L$ and $\succeq^H$ are the preferences over contracts of the low-risk and high-risk firms, respectively. We might say "constrained-efficient" or "second-best," to distinguish the efficient outcomes given these self-selection constraints from the "efficient" or "first-best" outcomes that are efficient when there is no adverse selection and hence no self-selection constraints. This is analogous to constrained efficiency with moral hazard, where we had to take into account the incentive compatibility constraints.

If $C_L = C_H$, i.e., if both types of firms end up with the same contract, then we say the menu or outcome is *pooling*. Otherwise, we say it is *separating*. When there is adverse selection, separating outcomes cannot have full insurance for both types of firms, because then contracts would differ only by the premium charged and all firms would choose whichever contract has the lowest premium. Reducing coverage hurts the high-risk firms more than the low-risk firms. Hence, *we induce separation by giving less coverage to the low-risk firms than to the high-risk firms*.

To show this formally (but graphically), we need to compare the preferences of the firms over contracts. Each contract yields a unique state-dependent allocation $\langle z_g, z_b \rangle$, which is the same whether the contract is accepted by a low-risk or high-risk firm, and each allocation is the result of a unique contract. Hence, we can just identify contracts with the allocation they yield, which is more intuitive when depicting preferences and indifference curves.

Figure 9.2 shows two indifference curves in the set of allocations. Stop and decide which is an indifference curve for low-risk firms, and which is an indifference curve for high-risk firms. If you are having trouble, consider the following question. Start at the allocation where the two curves cross, and imagine that you reduce consumption in the good state by \$20,000. Who needs to get more extra money in the bad state to end up indifferent to the initial allocation? Try also to compare the slopes of the indifference curves at the point where they cross.

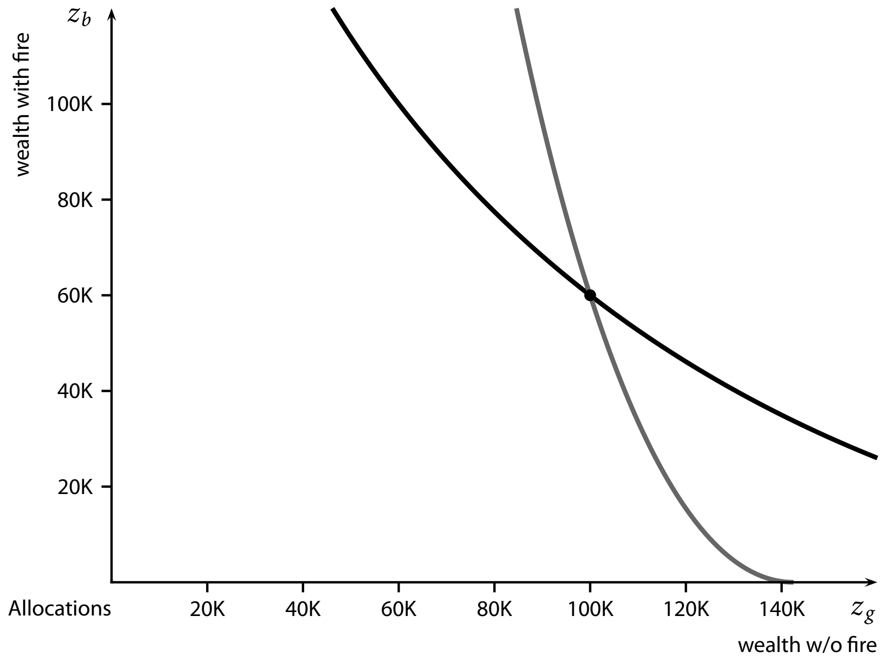
Here is the answer: The low-risk firms care less about money in the bad state than high-risk firms do, since the bad state is less likely for them than for the high-risk firms. Therefore, they need more extra money in the bad state to make up for a loss of money in the good state. This means that the gray curve is the low-risk indifference curve and the black curve is the high-risk indifference.

To calculate the slope of the indifference curves, we should first write down the firms' utility functions over allocations $\langle z_g, z_b \rangle$ of wealth in the good and bad states, respectively. Let $U^L(z_g, z_b)$ and $U^H(z_g, z_b)$ be the utility functions of the low-risk and high-risk firms, respectively:

$$\begin{aligned} U^L(z_g, z_b) &= (1-\pi^L)u(z_g) + \pi^L u(z_b) &= \tfrac{4}{5}z^{1/2} + \tfrac{1}{5}z^{1/2} \\ U^H(z_g, z_b) &= (1-\pi^H)u(z_g) + \pi^H u(z_b) &= \tfrac{1}{2}z^{1/2} + \tfrac{1}{2}z^{1/2}. \end{aligned}$$

Then the gradients at the point $\langle z_g, z_b \rangle$ where the two curves cross, which are perpen-

Figure 9.2



Which indifference curve is for a low-risk firm and which is for a high-risk firm?.

dicular to the indifference curves at that point, are

$$\left\langle \frac{\partial U^L}{\partial z_g}, \frac{\partial U^L}{\partial z_b} \right\rangle = \left\langle \frac{4}{5}u'(z_g), \frac{1}{5}u'(z_b) \right\rangle$$
$$\left\langle \frac{\partial U^H}{\partial z_g}, \frac{\partial U^H}{\partial z_b} \right\rangle = \left\langle \frac{1}{2}u'(z_g), \frac{1}{2}u'(z_b) \right\rangle.$$

The low-risk gradient points further down than the high-risk gradient, and hence the indifference curve is steeper. This is shown in Figure 9.3. Since the low-risk indifference curve is always steeper than the high-risk indifference curve where they cross, they can cross only once. This is called the *single-crossing property*.

In summary, we have shown:

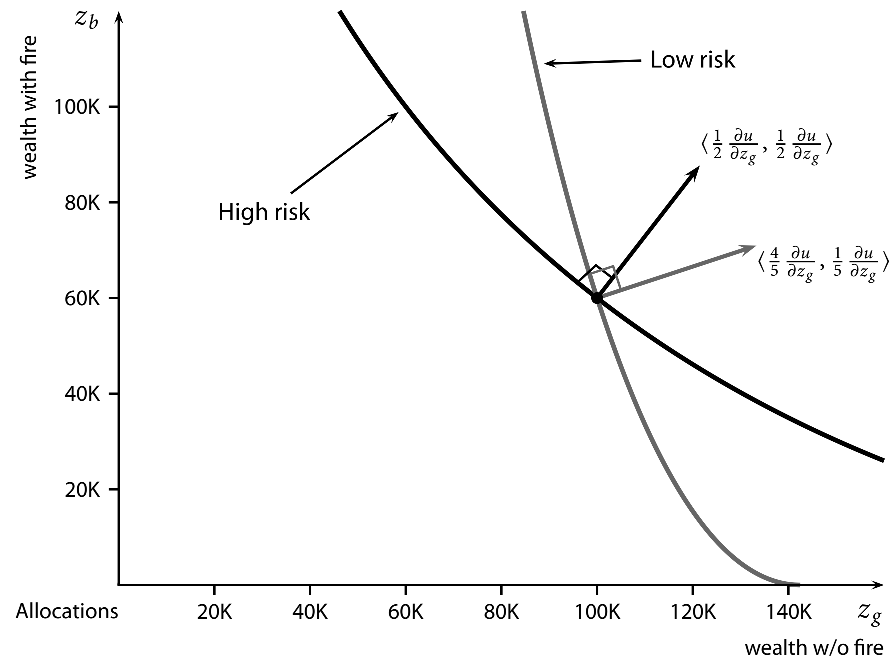
Proposition 1. *An indifference curve for a lower risk firm and an indifference curve for a high-risk firm can cross only once, and the low-risk indifference curve is steeper at the crossing point.*

Note the graphical significance of the self-selection constraints for a menu $\{C^L, C^H\}$ of contracts:

- $C^L \succsim^L C^H$ means that C^H lies below the low-risk indifference curve through C^L.
- $C^H \succsim^H C^L$ means that C^L lies below the high-risk indifference curve through C^H.

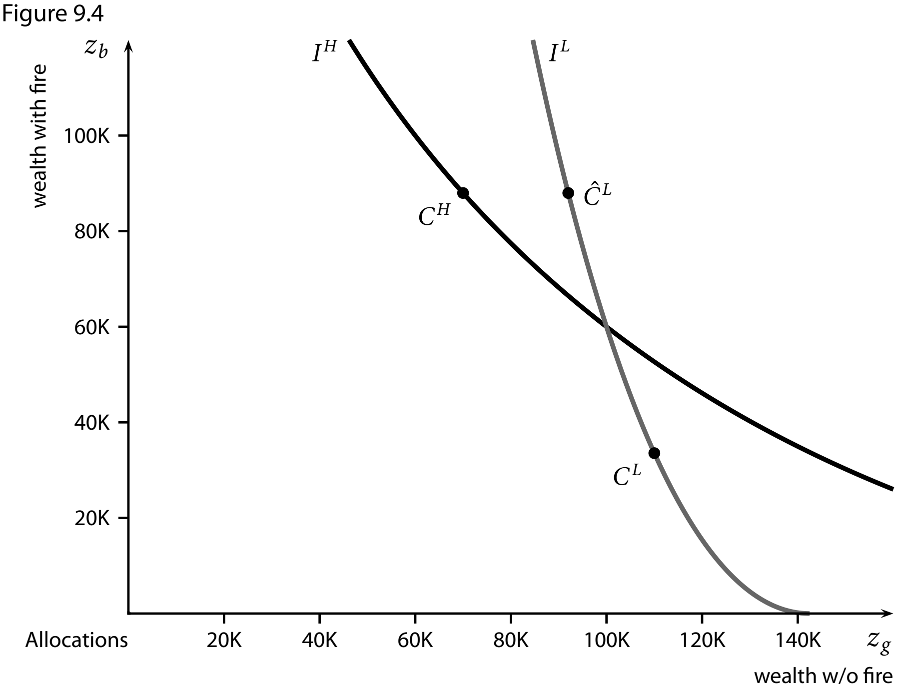
Figure 9.4 shows a menu $\{C^L, C^H\}$ that satisfies the self-selection constraints, and a menu $\{\hat{C}^L, C^H\}$ that violates the self-selection constraint of high-risk firms.

Figure 9.3



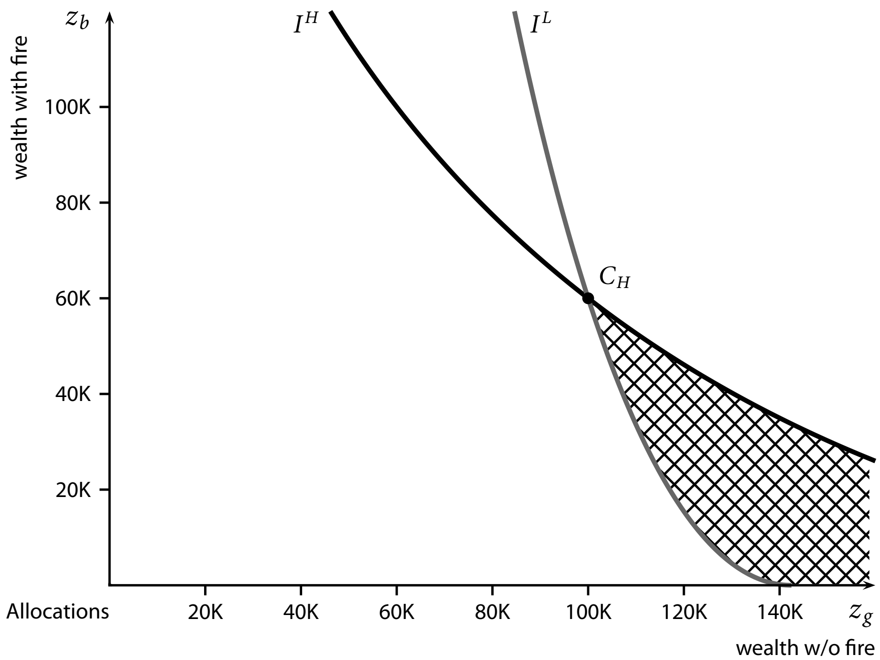
Money in the bad state is relatively more important to the high-risk firms than the low-risk firms because the bad state is more likely for the high-risk firms.

Figure 9.4



The menu $\{C^L, C^H\}$ satisfies both self-selection constraints. The menu $\{\hat{C}^L, C^H\}$ violates the self-selection constraints of high-risk firms, because high-risk firms prefer $\hat{C}^L$ to C^H.

Figure 9.5



The crosshatch region is the set of contracts C^L such that $\{C^L, C^H\}$ satisfies the self-selection constraints. For separating menus, the low-risk firms get less coverage than the high-risk firms.

Let's pick a contract (allocation) C^H for the high-risk firms, and ask which contracts C^L for the low-risk firms are such that $\{C^L, C^H\}$ satisfies the self-selection constraints.

- $\{C^L, C^H\}$ satisfies the low-risk self-selection constraint if C^L lies *above* the *low-risk* indifference curve through C^H.
- $\{C^L, C^H\}$ satisfies the high-risk self-selection constraint if C^L lies *below* the *high-risk* indifference curve through C^H.

Both constraints are satisfied in the intersection of these two regions, which is the crosshatched region in Figure 9.5.

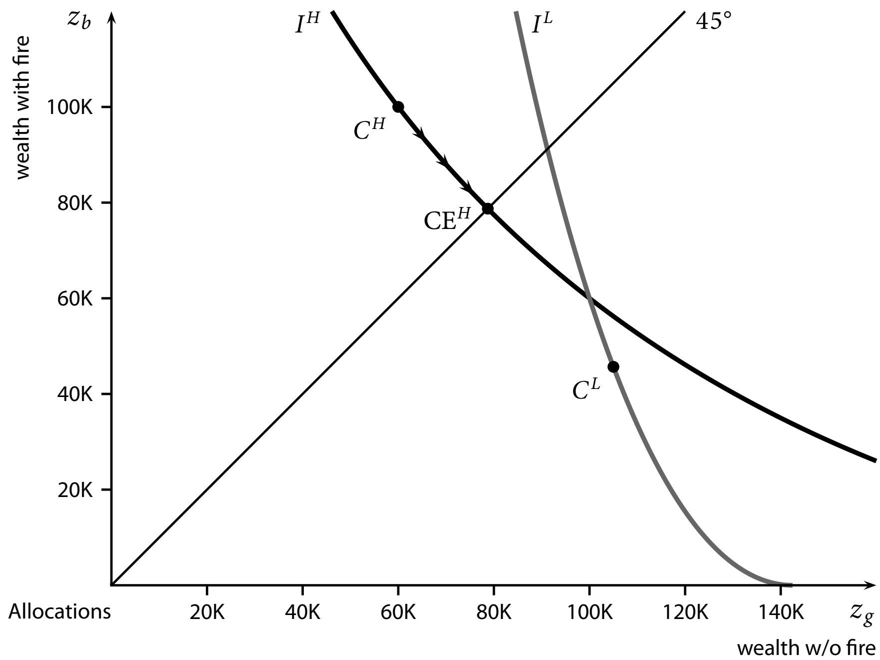
Contracts in this region have less income in the bad state and more in the good state, compared to C^H, and hence have less coverage. Therefore, we have shown:

PROPOSITION 2. *To satisfy the self-selection constraints, the high-risk firms must get more (or as much) coverage as low-risk firms.*

9.3 Efficient contracts with adverse selection

All pooling contracts are second-best (since they are first-best and satisfy the self-selection constraints). The second-best separating menus are more interesting. In this section, we will come up with the following characterization of an efficient separating

Figure 9.6



One type of firm gets full coverage: $\{C^L, C^H\}$ satisfies self-selection, but neither type of firm has full coverage. $\{C^L, \mathrm{CE}^H\}$ also satisfies the self-selection constraints. Both types of firms are indifferent between $\{C^L, \mathrm{CE}^H\}$ and $\{C^L, C^H\}$. The government's expected profits from high-risk contracts increase by the risk premium $E[C^H] - \mathrm{CE}^H$. Hence, $\{C^L, \mathrm{CE}^H\}$ Pareto dominates $\{C^L, C^H\}$.

menu:

> One firm has full coverage and a binding self-selection constraint.
> This firm is the one that is worse off than under the pooling contract that gives the government the same expected profits (which we call the *revenue-equivalent* pooling contract).

(When we say that a self-selection constraint is *binding*, we mean that the party is indifferent between the two contracts. For example, if $C^L \sim^L C^H$, the low-risk self-selection constraint is binding. If $C^L \succ^L C^H$, it is satisfied but not binding.)

Proposition 3. *If a separating menu is efficient, then either low-risk or high-risk firms get exactly full coverage.*

As is typical for proving this kind of result, to show that an efficient menu has a particular property, we show that if a menu does not have the property then it is not efficient (we prove the "contrapositive"). To show that a menu is not efficient, we construct a Pareto-dominating menu of contracts.

Proof. Suppose $\{C^L, C^H\}$ satisfies the self-selection constraints but neither type of firm has full coverage. Such a situation is shown in Figure 9.6. Pick the type of firm with the lower certainty equivalent, that is, the type whose indifference curve intersects the 45° line at a lower point. In Figure 9.6, this is the high-risk firm: $\mathrm{CE}^H \leq \mathrm{CE}^L$. Then we can replace C^H by CE^H without violating the self-selection constraints:

- The high-risk firms weakly prefer CE^H to C^L since they weakly prefer C^H to C^L and they are indifferent between CE^H and C^H.
- The low-risk firms weakly prefer C^L to CE^H since they are indifferent between C^L and CE^L and $\mathrm{CE}^L \geq \mathrm{CE}^H$.

Both types of firms are indifferent between $\{C^L, \mathrm{CE}^H\}$ and $\{C^L, C^H\}$, the government's expected profits increase by the high-risk firms' risk premium: $E[C^H] - \mathrm{CE}^H$. Hence, the menu $\{C^L, \mathrm{CE}^H\}$ Pareto dominates $\{C^L, C^H\}$. Similarly, if $\mathrm{CE}^L \leq \mathrm{CE}^H$, then the menu $\{\mathrm{CE}^L, C^H\}$ Pareto dominates $\{C^L, C^H\}$. □

Proposition 4. *If a separating menu is efficient, then the self-selection constraint of the firm with full coverage is binding.*

Proof. The idea is that a firm that does not have full coverage bears risk in order to keep the other type of firm from choosing its contract. We want to reduce this risk as much as possible, which means up to where the other firm is just indifferent between switching. For example, suppose the high-risk firm has full coverage, but its self-selection constraint is not binding. This means that $C^H \succ^H C^L$, and C^L lies strictly below the high-risk indifference curve through C^H. This situation is shown in Figure 9.7. If we move the low-risk contract in the direction of the arrows along the low-risk indifference curve (toward the 45° line), we are reducing the risk and (since the low-risk firm is neither better or worse off) reducing the expected value of the low-risk firms wealth. This increases the government's expected profits. We can move C^L all the way up to $\hat{C}^L$, on the high-risk indifference curve, without violating the self-selection constraints. Hence, $\{\hat{C}^L, C^H\}$ Pareto dominates $\{C^L, C^H\}$. □

Proposition 5. *If a separating menu is efficient, then the firms with full coverage and binding self-selection constraints are the ones that are worse off than under the revenue-equivalent pooling equilibrium.*

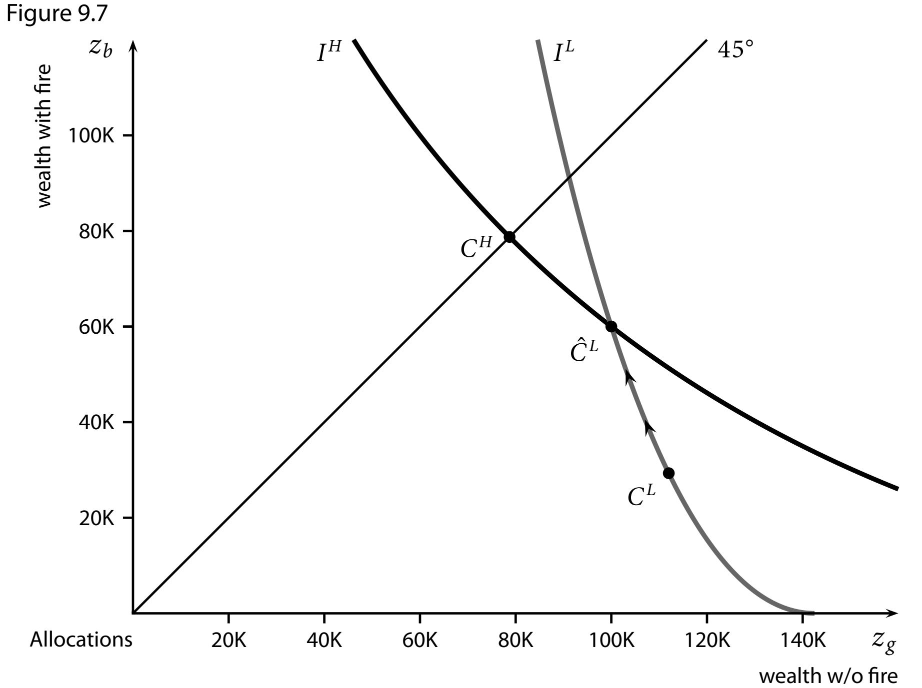
Proof. The revenue-equivalent pooling contract for a menu $\{C^L, C^H\}$ gives each firms wealth between $E[C^L]$ and $E[C^H]$. If the high-risk firms have full coverage, then low-risk firms prefer C^L to the constant amount C^H, and so $E[C^L] > C^H$. Hence, the high-risk firms have lower wealth in C^H than in the revenue-equivalent pooling contract. □

9.4 The efficient contracts with zero expected profits

Let's consider what a "slice" of the Pareto frontier looks like with adverse selection. I will show how to calculate it for the case where the government has zero expected profits.

Starting from the pooling contract, we can make the low-risk firms better off and

Figure 9.7



The self-selection constraint of firms with full coverage is binding: $\{C^L, C^H\}$ and satisfies self-selection and the high-risk firm get full coverage, but they strictly prefer C^L to C^H. $\hat{C}^L$ is closer to 45° line, and hence government appropriates greater risk premium from low-risk firms. Hence, $\{\hat{C}^L, C^H\}$ Pareto dominates $\{C^L, C^H\}$.

the high-risk firms worse off by increasing the premium of the high-risk types but continuing to provide them with full coverage. The premium *rate* for low-risk types falls (since the government earns zero expected profits), but to keep the high-risk firms from choosing the low-risk contract, the level of coverage must also fall (just enough so that the high-risk self-selection constraint is binding). Specifically:

1. High-risk firms get full coverage at premium p^H.
2. Low-risk firms get coverage x^L at premium p^L such that
   (a) The government has zero expected profits:

$$\alpha \pi^L x^L + (1-\alpha)\pi^H L \;=\; \alpha p^L + (1-\alpha) p^H. \tag{9.1}$$

   (b) The high-risk firms self-selection constraint is satisfied with equality:

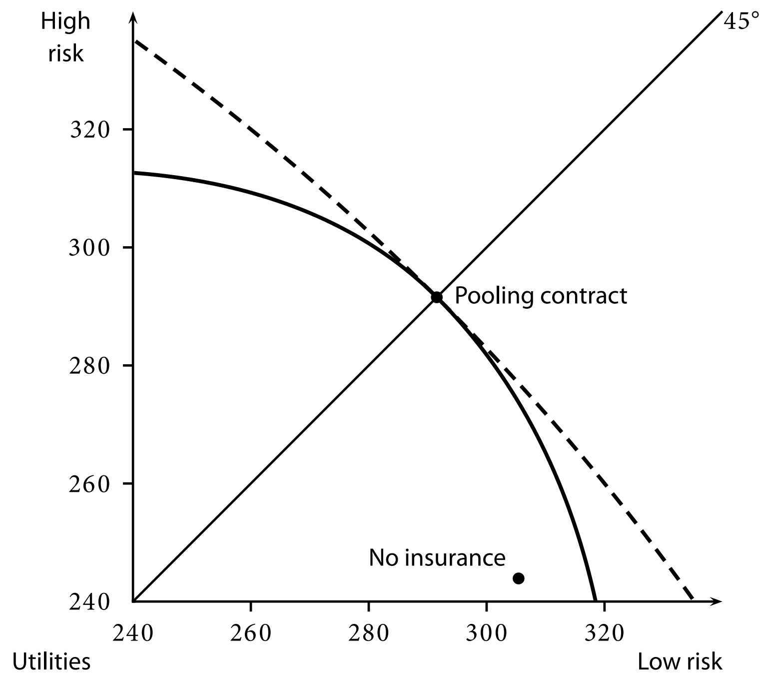
$$u(w - p^H) \;=\; (1-\pi^H) u(w_g - p^L) + \pi^H u(w_b - p^L + x^L) \tag{9.2}$$

(Utility of C^H equals expected utility of C^L for high-risk firms.)

For each p^H, we can solve the system of equations (9.1–9.2) for p^L and x^L. By varying p^H, we can trace out the entire efficiency frontier. In a similar way, we can derive the frontier for the contracts where the low-risk firms get full coverage and the high-risk firms get partial coverage. This slice of the efficiency frontier is plotted in Figure 9.8 for the case where $\alpha = .5$ (the solid line).

For comparison, the efficiency frontier without adverse selection is drawn. Observe that the two frontiers coincide at the pooling contract, but there is a loss in welfare due

Figure 9.8



Efficiency frontiers with zero expected profits, for $\alpha = .5$. Solid line is frontier with adverse selection, and dashed line is frontier without adverse selection.

to the self-selection constraints when we try to give different allocations to the two types of firms.

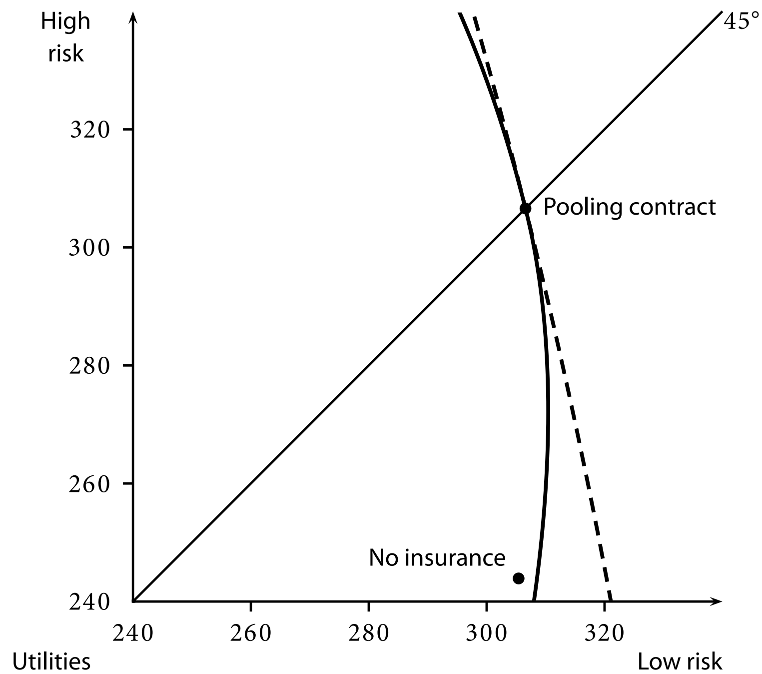
Consider how the efficiency frontier changes when α is higher. The locus of utilities for the solution to equations (9.1) and (9.2) are shown in Figure 9.9 for $\alpha = .8$. Here we can observe something that was not apparent in Figure 9.8. The curve bends around, which means that some of the solutions are not efficient. Here is an explanation. As we increase the premiums for high-risk firms and redistribute expected wealth to low-risk firms, the low-risk firms have to bear more risk so that high-risk firms do not choose the low-risk contract. For high p^H, the extra risk that low-risk firms must bear outweighs the extra wealth they obtain in expected value. Then both low-risk and high-risk firms are made worse off by the attempt to redistribute wealth from high-risk firms. Hence, when calculating the true efficiency frontier, we have to check that solutions to equations (9.1) and (9.2) are not dominated by other solutions to these equations.

9.4.1 Competitive insurance markets with adverse selection

We don't have to do much more work to figure out what the competitive outcome will be when there are many (potential) insurance companies acting competitively. It will be the *actuarially fair, second-best* menu, if there is such a thing.

We need to be precise about what we mean by a competitive equilibrium. We imagine that there are many (potential) insurers. The market is in equilibrium when no insurer can increase its (expected) profits by changing its contract offering, *given that*

Figure 9.9



Solid line is the locus of solutions to equations (9.1) and (9.2), which are the potential points on the efficiency frontier for zero expected profits, for $\alpha = .8$. Not all points are efficient. Dashed line is the efficiency frontier without adverse selection.

the remaining insurers do not change their offerings. In particular, no potential entrant should be able to make a strictly positive expected profit by offering one or more contracts, and all active insurers should be making zero or positive profits. This is a pure-strategy Nash equilibrium in a simultaneous-move game in which the players are the insurance companies and the strategies are contract offerings.

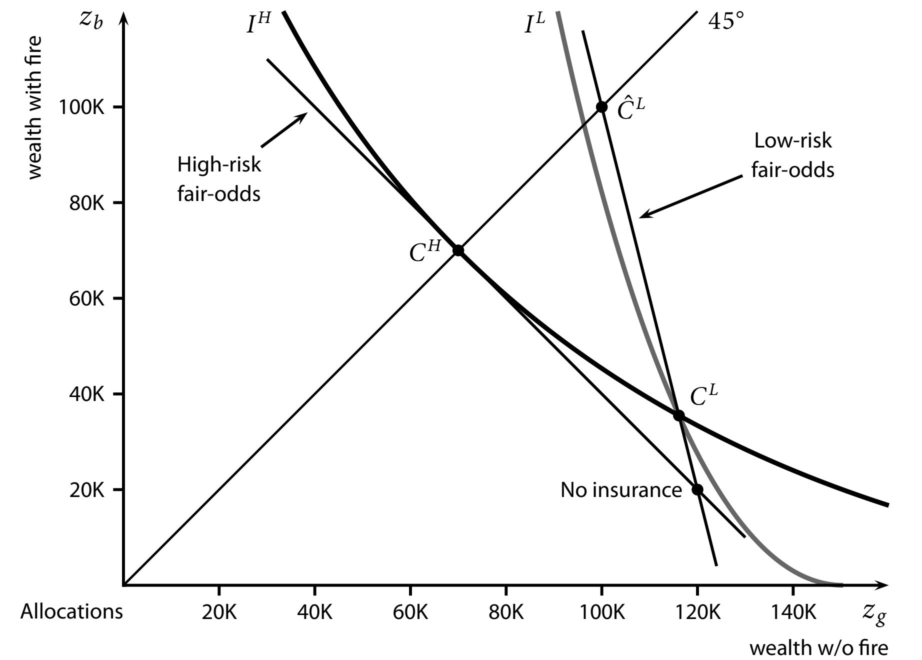
Proposition 6. *Each contract in the competitive equilibrium is actuarially fair for the designated type (and hence the equilibrium is separating).*

Proof. Given any separating menu of contracts, you can always shift one type's contract a small amount and make that type strictly better off without violating the self-selection constraints. E.g., for low-risk types, move their contract to the region down and to the right of the contract that is bounded by the indifference curves through that types' contract, and for high-risk firms move their contract into the similar region that is up and to the left. Thus, if either type's contract is not actuarially fair, a company can come along and offer a similar contract that attracts all the firms' of that type but no firms of the other type, and that is still unfair, i.e., that generates a profit for the company. □

Proposition 7. *The competitive equilibrium is (constrained) efficient.*

Proof. If the current market menu is not constrained efficient, it is possible to come up with a menu that makes all three groups better off (not just one of them). An entrant

Figure 9.10



The competitive equilibrium with adverse selection is $\{C^L, C^H\}$. C^H is the intersection of the high-risk fair-odds line with the 45° line. C^L is the intersection of the low-risk fair-odds line with the high-risk firms' indifference curve through C^H. For comparison, $\{C^L, C^H\}$ is the competitive equilibrium without adverse selection.

can offer such a menu, attract all the firms, and make a positive profit. □

Using the two properties of the competitive equilibrium—actuarial fairness and efficiency—we can calculate and illustrate graphically the equilibrium contracts. First the high-risk contract: From fairness, we know that the high-risk firms are worse off than under the revenue equivalent polling contract. Combining this with efficiency, we know that the high-risk firms have full coverage. Hence, high-risk firms have coverage L of a premium $\pi^H L$. Now the low-risk contract:

- From fairness:
$$p^L = \pi^L x^L \tag{9.3}$$
- From efficiency, the high-risk self-selection constraint is binding:
$$u(w_g - \pi^H L) = (1 - \pi^H)u(w_g - p^L) + \pi^H u(w_b - p^L + x^L) \tag{9.4}$$
(The left-hand side is a high-risk firm's utility when it chooses C^H. The right-hand side is a high-risk firm's expected utility when it chooses C^L.)

Graphically, the low-risk allocation is the intersection of the low-risk firms' fair-odds line and the high-risk firms' indifference curve through the high-risk firms' allocation. This is shown in Figure 9.10. Arithmetically, the low-risk firm's contract $\langle p^L, x^L \rangle$ is the solution to equations (9.3) and (9.4). In our example, the high-risk firms get coverage of \$100,000 at a premium of \$50,000, and they end up with \$70,000 for sure. The low-risk firms' actuarially fair premium is $p^L = .2x^L$, and so the high-risk firms' binding self-selection constraint, equation (9.4), becomes:

$$\begin{aligned} u(70000) &= .5u(120000 - .2x) + .5u(20000 + .8x) \\ 70000^{1/2} &= .5(120000 - .2x)^{1/2} + .5(20000 + .8x)^{1/2} \end{aligned}$$

The solution is $X^L = 19,356$. Then the high-risk firms' utility is $u(70,000) = 265$, and the low-risk firms' utility is

$$.8(120000 - .2(19356))^{1/2} + .2(20000 + .8(19536))^{1/2} = 310$$

There is one important caveat. The menu of contracts we have calculated is certainly the Pareto superior menu out of all actuarially fair menus, and hence is the only candidate for an actuarially fair and efficient menu. However, it is possible that this menu is Pareto dominated by some other menu in which the high-risk firms have lower premiums and the low-risk firms have actuarially unfair insurance but are better off because they have more coverage. In fact, it may even be Pareto dominated by the pooling contract that has zero expected profits for the insurance companies. Observe that the actuarially fair, "efficient" menu we have found does not depend on the fraction α of firms that are low risk. On the other hand, the zero-profit pooling contract does depend on α. As α approaches 1, the premium for the pooling contracts approach the actuarially fair premium for low-risk firms, and the low-risk firms are better off than with the partial coverage they get in the actuarially fair menu. Hence, for large enough value sof α, the actuarially fair, "efficient" menu we have calculated is not efficient at all.

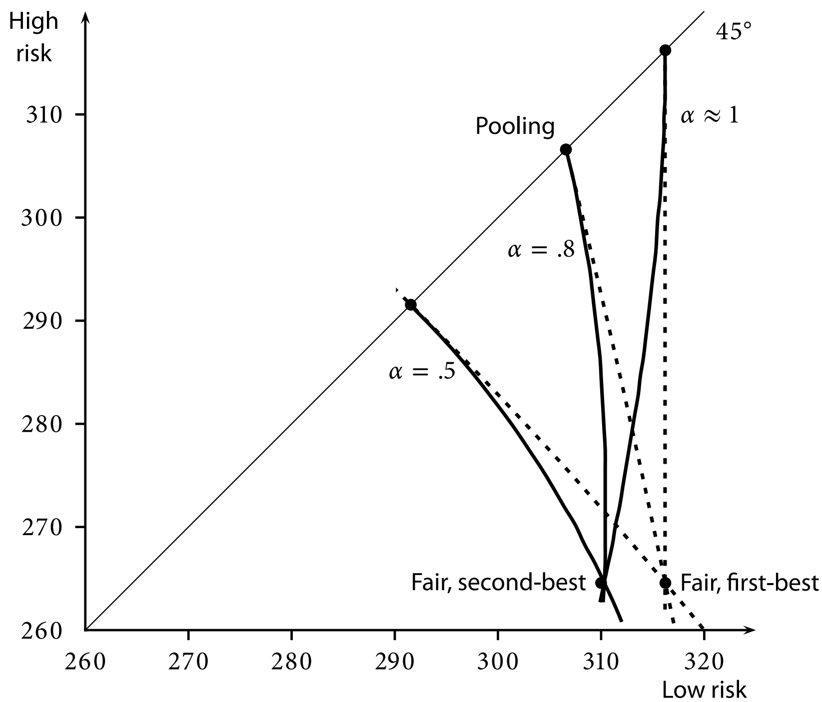
This is illustrated in Figure 9.11, which shows the solutions to equations (9.1) and (9.2) for $\alpha = .5, \alpha = .8$ and $\alpha \approx 1$. When $\alpha = .5$, the fair, second-best menu is efficient. When $\alpha = .8$, the actuarially fair menu is not Pareto dominated by the pooling contract, but it is Pareto dominated by an intermediate contract. When $\alpha \approx 1$, the fair, second-best menu is Pareto dominated by the pooling contract. For example, this is true when $\alpha = .9$:

Utilities	*low risk*	*high risk*
no insurance	305	244
fair, first-best	316	265
fair, second-best	310	265
pooling $\alpha = .5$	292	292
pooling $\alpha = .9$	311	311

What happens when the fair, self-selection menu is not efficient? There is no menu that satisfies the conditions for an equilibrium that were derived above. Of course, the world doesn't stop turning as a result. This simply means that our model no longer makes a prediction, and we need a better model. Several alternative models of competition have been devised all of which are beyond the scope of this course.[4] Some models predict that the equilibrium is separating, but not necessarily actuarially fair

4. One technical extension to our model is to allow *mixed strategies*. Other models, unlike the one we studied, allow firms to consider the long-term reactions of other firms to their contract offerings. One such equilibrium concept is called the *reactive equilibrium*, and it predicts that the outcome is always separating.

Figure 9.11



The solid lines are the loci of utilities for potentially efficient contracts (solutions to equation (9.1) and (9.2)) when high-risk firms get full coverage, for α equal to .5, .8 and ≈ 1. For comparison, the dashed lines are portions of the efficient frontiers without adverse selection. For α equal to .8 and ≈ 1, the menu of actuarially fair contracts that satisfies equations (9.1) and (9.2) is not constrained efficient, and hence cannot be the competitive equilibrium.

for each type of firm. What is important is that you understand that our sharp results about the competitive equilibrium hold when there are not too few high-risk firms, and that otherwise high-risk firms may have actuarially favorable insurance and low-risk firms may have actuarially unfavorable insurance but more coverage than in the actuarially fair menu.

Exercise 9.1. Consider the following problem of adverse selection in insurance markets. There are two types of people looking for automobile liability insurance, good drivers and bad drivers. For simplicity, suppose the state's torte law says that for any accident, no matter what kind, the party at fault pays the other party \$90,000. Bad drivers have an accident at which they are at fault with probability 1/3. Good drivers have an accident at which they are at fault with probability 1/10. The utility over money for both types is $u(m) = m^{1/2}$, and the initial wealth for both types is $\$250,000$.

The market for insurance is perfectly competitive and there are no administrative costs.

a. Suppose that the insurance companies can observe which people are high risk and which are low risk. Describe the market equilibrium. (I.e., what kind of contract will each type get?) Explain.

b. What does it mean for there to be adverse selection in this problem? If there is adverse selection, what would happen if the insurance companies offered the two contracts you found in the previous part?

c. Suppose that there is adverse selection. What is the best menu of contracts (from the insured's point of view), with a distinct contract for each type of driver, such that (i) each driver chooses his or her corresponding contract and (ii) each contract is actuarially fair for the designated type. (You can stop at the point where you have given an equation that determines the key provision of the contract, or you can solve the equation and give the exact contract.)

d. Explain the meaning of a menu of contracts and self-selection in the context of this problem.

Exercise 9.2. Consider the following problem of adverse selection in insurance markets. There are two types of people looking for health insurance, high risk and low risk. The insurance is to cover the cost of back surgery, which each type will have performed no matter what, if the need arises. The operation costs \$12,000. High risk people will need the surgery over the year-long life of the policy with probability 1/2. Low risk people will need it with probability 1/10. The fraction α of the population that is low risk is 1/2. The utility over money for both types is $u(m) = m^{1/2}$, and the initial wealth for both types is $\$24,000$ (i.e., without insurance, they have \$24,000 when they do not need surgery, and they have \$12,000 when they need surgery).

The market for insurance is perfectly competitive and there are no administrative costs.

When answering the questions below, explain the steps you are taking in finding the solution.

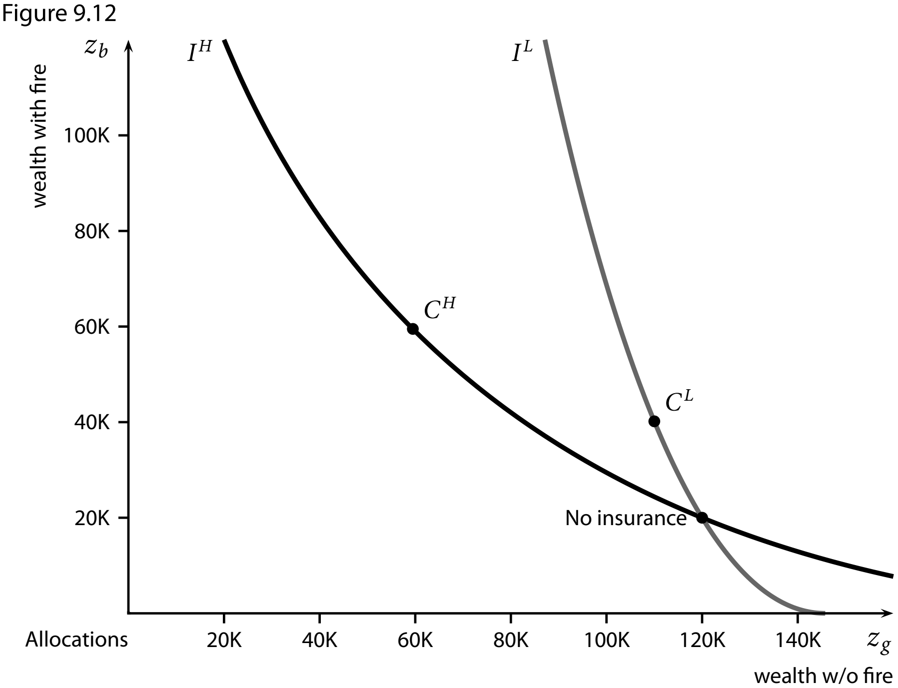
a. Suppose that the insurance companies can observe which people are high risk and which are low risk. Describe the market equilibrium. (I.e., what type of contracts will each type get?)

b. Now find the separating market equilibrium if the insurers *cannot* observe who is high or low risk. You should specify the high-risk contract, and find the low-risk contract as a solution to the self-selection constraint. (You will have to solve an equation with square roots on both sides. You can solve it numerically, or you can solve it by rearranging, squaring both sides, rearranging, squaring both sides again, rearranging, and solving the resulting quadratic equation.)

c. Show that if the contracts you found above are in the market, then the fair pooling contract does not attract all the consumers.

d. Let α be the fraction of people in the market that are low-risk. We set this to 1/2 before, but now we want to treat it as a parameter. How does your answer to the previous problem depend on α?

Figure 9.12



The menu $\{C^L, C^H\}$ is such that both types of firms are indifferent between their contracts and no insurance, and the low-risk firms get positive coverage. However, $\{C^L, C^H\}$ violates the high-risk firms' self-selection constraint.

9.5 Monopolistic screening

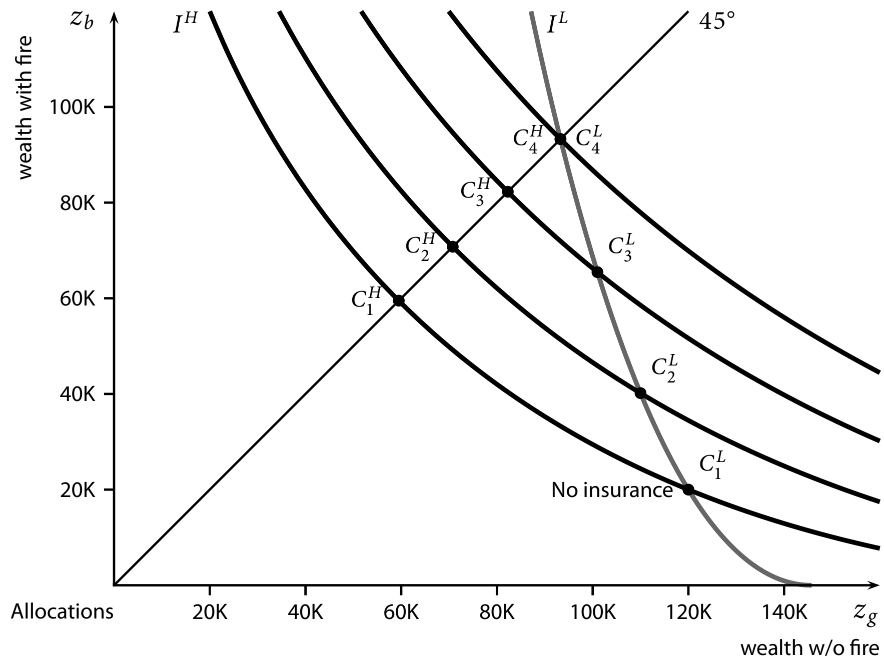
In the competitive equilibrium, the insurance companies get zero expected profits, and the insured firms get all the gains from trade. Our characterization of the efficient menus is also useful for understanding the outcomes of other market structures.

For example, suppose that there is a single model insurance company that acts as a monopolist to maximize its expected profits. The menu the monopolist chooses will be constrained-efficient, but the monopolist will drive the customers down to their reservation utility—to the extent possible—in order to extract the greatest surplus.

The outside option for the firms is to get no insurance. If the monopolist could observe each firms type, it would provide full coverage to all firms at premium that kept firms with their certainty equivalent (plus perhaps a penny). The monopolist gets each firm's risk premium in expected profits. Each firm is roughly indifference between accepting the monopolist's insurance or opting for no insurance.

However, this menu does not satisfy the self-selection constraints, because the firms get full coverage at different premia. In fact, there is no menu that satisfies the high-risk firms' self-selection constraint in which both firms are indifferent between their insurance and no insurance and in which the low-risk firms get positive coverage. This is illustrated in Figure 9.12. The single-crossing properties implies that the portion of the low-risk firms' indifference curve through the no-insurance allocation lies above the high-risk firms' indifference curve through this allocation.

Figure 9.13



Menus of contracts that satisfy the low-risk participation constraint and the high-risk self-selection constraints with equality. Moving from menu $\{C_1^L, C_1^H\}$ to $\{C_4^L, C_4^H\}$, the monopolist extracts a greater risk premium from the low-risk firms but the high-risk firms get larger informational rents.

The monopolist can opt to provide no insurance to low-risk firms and to give the high-risk firms their certainty equivalent, or it can give more wealth to the high-risk firms so that it can provide partial coverage to the low-risk firms. In the second type of menu, the high-risk firms are better off than without insurance, and are better off than when the monopolist can observe their type. Such surplus due to private information is called *informational rents*.

Menus of this type are shown in Figure 9.13. Given C^H, the low-risk contract C^L that maximizes the monopolist's expected wealth is the one for which the high-risk firms' self-selection constraint and the low-risk firms' participation constraint is binding. That is, it lies at the intersection of the high-risk firms' indifference curve through C^H and the low-risk firms' indifference curve through C^L.

As the monopolist increases the high-risk firms wealth, starting at the high-risk firms' certainty equivalent and ranging up to the low-risk firms' certainty equivalent, the coverage of the low-risk firms' contract increases up to full coverage. The expected profits form the high-risk firms' contract decreases and may even be negative, while the expected profits from the low-risk firms' contract increases. Which of these menus is the best for the monopolist depends on the fraction α of firms that are low risk. When α is small, it can be optimal to only insure the high-risk firms. The larger is α, the more coverage the low-risk firms get. However, as long as the low-risk firms have differentiable utility, full-coverage (pooling) is never optimal for the monopolist (if $\alpha < 1$). Because of local risk neutrality, decreasing coverage for the low-risk firms by a small amount has a negligible effect on the profits from the low-risk firms' contract, but it allows the monopolist to increase the premiums for high-risk firms by a non-negligible amount.

Let's conclude by writing the monopolist's maximization problem. The monopolist chooses two contracts, C^L and C^H, to maximize its expected profits, subject to *four* constraints: the low and high-risk firms' self-selection and participation constraints. However, we have shown, using the single-crossing property of the indifference curves along with the properties of efficient menus, that only the high-risk firms' self-selection constraint and the low-risk firms' participation constraint are binding. Furthermore, the high-risk firms get full coverage. Hence, the monopolist's maximization problem is

$$\max_{p^L, x^L, p^H} \quad \alpha(p^L - \pi^L x^L) + (1-\alpha)(p^H - \pi^H L)$$

$$\text{subject to:} \quad u(w_g - p^H) = (1-\pi^H)u(w_g - p^L) + \pi^H u(w_b - p^L + x^L)$$

$$(1-\pi^L)u(w_g - p^L) + \pi^L u(w_g - p^L + x^L) = (1-\pi^L)u(w_g) + \pi^L u(w_b).$$

Chapter 10

Signaling

10.1 The difference between screening and signaling

10.2 Signaling in labor markets

Chapter 11

Long-term versus short-term contracting

11.1 Ex-ante versus ex-post efficiency

One feasible social outcome is *Pareto superior* to another feasible social outcome if every involved party weakly prefers the former and some party strictly prefers it. A social outcome is *Pareto optimal* if there is no Pareto superior outcome. Pareto optimality and Pareto superiority are weak normative criteria; it is hard to argue against them. Once consequence is that they are not complete. Most pairs of social outcomes cannot be compared this way, because there is no unanimity among the involved parties about which is better.

Another term for Pareto optimal is Pareto efficient, or just plain efficient if you are talking among economists.

In applying these Pareto criteria, one must decide what is meant by an outcome, when an outcome is feasible, who the involved parties are, and what preferences over outcomes are. Economists take preferences as given, as if they were biologically inherited. This is a weak point in economic theory, but there is one way in which changing preferences sneak in. Because preferences over uncertain prospects (state-dependent plans) depend on beliefs, the revelation of information can change preferences. For example, your demand for health insurance will change drastically after you learn that you are or are not sick.

The Pareto criterion can still be applied when beliefs, and hence preferences, change over time due to new information. However, one has to specify what preferences are being used when applying the Pareto criterion. In fact, it is interesting to compare Pareto optimality before and after the revelation of information. The term *ex-ante* means before information is observed, and *ex-post* means after information is observed. Thus, one has the terms ex-ante and ex-post efficiency or Pareto optimality.

Here an example. Suppose there are two farmers living on two neighboring islands. One grows red potatoes and the other grows white potatoes. These potatoes are perfect substitutes and there are no other goods, and hence there are no reasons to trades.

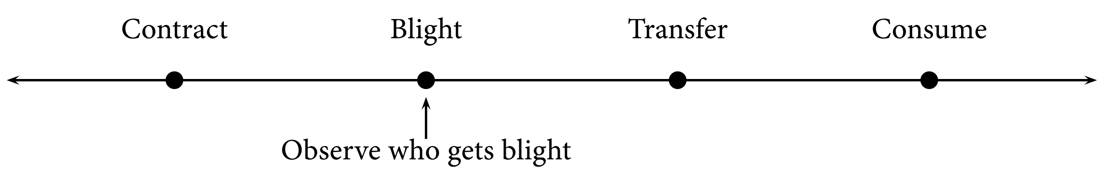
But now suppose that the farmers learn that one (and only one) of their crops will be wiped out by a blight. Each farmer will be hit with equal probability. The farmers then might sign a state-contingent contract which specifies potato transfers from one farmer to the other, depending on who gets hit by the blight. Such a contract, together with the state-dependent endowment of potatoes (the state here is which farmers gets the blight), determine the state-dependent allocations of potatoes. The farmers have preferences over these state-dependent allocations, and hence over contracts, that depend on their information and risk preferences.

Here is what the time line looks like:

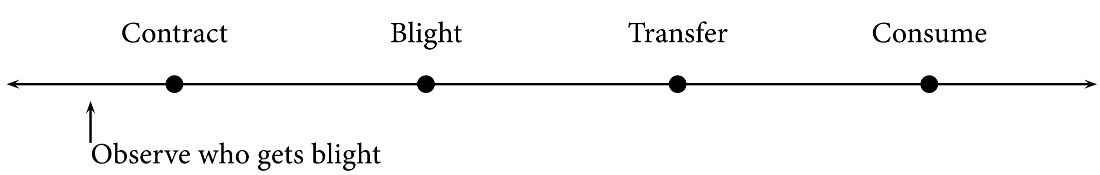
Ex-post means after the farmers know who is hit by the blight, and ex-ante means before they know this. In this simple example, *any allocation of potatoes is ex-post efficient:* The only way to make the farmer who is hit by the blight better off is to give her some of the other farmer's potatoes, but this will make that farmer worse off, since he know his potatoes are OK.

Ex-ante, however, efficient contracts involve some sharing of risk. In fact, if the farmers are risk averse, then any ex-ante efficient contract will have no risk—it will specify simply what fraction of good potatoes each farmer gets. (This is possible because there is no aggregate risk, i.e., the total amount of potatoes in the economy is not random.) For example, the farmers might agree to divide the good potatoes evenly.

Information can have "negative" social value (it can make everyone worse off), in ex-ante terms, because it destroys the possibility of insuring against risk when it is observed before contracts can be signed. If contracting takes place before the farmers know who is hit by the blight:



then presumably the farmers would agree on an ex-ante efficient contract. However, if the farmers learn who gets hit before they are able to contract:



then they will not agree to any transfer at all.. (The same breakdown of trade will occur if only one party knows who gets hit by the blight, as long as the other party knows she knows.) Although this outcome is ex-post efficient, it is not ex-ante efficient.

Before considering an application of ex-ante and ex-post efficiency, let's look at the definitions a little more formally. We need the Savage framework for uncertainty and information. This begins with a set S of states of the world, representing possible descriptions of the world. That there is uncertainty means that it is not known what the true state is. Observing formation might mean that one observes exactly what the true state is, or it might mean only that one is able to exclude some probabilities, e.g., one learns not the entire weather forecast, but only whether it is going to be sunny. For simplicity, we will only consider the case where the state of the world is fully revealed.

It will be more convenient to use state-dependent preferences. Hence, an outcome is not a description of everything that matters to each decision maker, but perhaps only those factors that are under the control of the decision makers or some other authority that could intervene, such as a government. Let Z be the set of feasible outcomes. A contingent outcome is a function $f : S \to Z$. This is called an act in the Savage model, but we will not use this term because preferences over outcomes are not necessarily state independent. For person i there is a probability measure π^i on S and a utility function $u^i : Z \times S \to \mathbb{R}$ such that

$$f \succsim^i g \quad \iff \quad \sum_{s \in S} \pi^i(s) u^i(f(s), s) \geq \sum_{s \in S} \pi^i(s) u^i(g(s), s)$$

$u(z, s)$ is the utility of outcome z in state s, $f(s)$ is the outcome in state s for the contingent outcome f, and $u(f(s), s)$ is thus the utility in state s for f. Then we would say that f is preferred to g if only only if the expected utility of f is greater than the expected utility of g.

Suppose that everyone observes that the true state is s. Then each person's preferences over outcomes are represented by the utility function $u^i(\cdot, s)$. An outcome z is ex-post efficient in state s if it is efficient for these utilities. i.e., if there is no other feasible outcome y such that

$$u^i(y, s) \geq u^i(z, s)$$

for all i, and with strict inequality for some i. A contingent outcome f is said to be ex-post efficient if $f(s)$ is ex-post efficient for each s.

On the other hand, a contingent outcome is ex-ante efficient if there is no other Pareto superior contingent outcome in terms of expected utility, i.e., if there is no other feasible contingent outcome g such that

$$\sum_{s \in S} \pi^i(s) u^i(g(s), s) \ \geq \ \sum_{s \in S} \pi^i(s) u^i(f(s), s)$$

for each i, and with strict inequality for some i.

Here is an important observation:

A contingent outcome is ex-post efficient if it is ex-ante efficient.

In other words, ex-post efficiency is a necessary condition of ex-ante efficiency. For if a contingent outcome is not ex-post efficient, there is another contingent outcome that gives the same outcome or an ex-post Pareto superior outcome in each state, and hence the latter contingent outcome is ex-ante preferred by each individual.

Note that the converse is not true. I.e., a contingent outcome can be ex-post efficient, but not ex-ante efficient. The potato example above showed this. Whenever there is a single good (e.g., "money") and hence no ex-post gains from trade, every allocation is ex-post efficient. Due to risk, however, not every contingent outcome is ex-ante efficient.

11.2 Long-term and short-term insurance contracts