Caltech Division of the Humanities and Social Sciences

Notes on selected topics in decision-making under uncertainty

KC Border

Fall 2016

v. 2016.12.07::15.45

Chapter 1

Some key ideas in decision theory

1.1 Basics

The modern approach to uncertainty, as formalized by Kolmogorov [27, 28], has as its fundamentals:

S, a set of **states of the world**.
$\mathcal{E}$, a collection of **events**.
P, a **probability** on $\mathcal{E}$.

The **states** are assumed to be exhaustive and mutually exclusive. What you choose as the set of states is a modeling decision. For today's purposes, we shall mostly assume that S is finite. The reason for this is to avoid any mathematical complications that arise from dealing with with integrals instead of sums. All the result here have generalizations to infinite sets of states, but they may require additional purely technical assumptions that obscure the economic results.

A **probability** P on $\mathcal{E}$ is a function that satisfies the following properties:

Normalization: For each $E \in \mathcal{E}$,

$$0 \leqslant P(E) \leqslant 1, \quad P(S) = 1, \quad \text{and} \quad P(\varnothing) = 0.$$

Additivity: If $E \cap F = \varnothing$, then

$$P(E \cup F) = P(E) + P(F).$$

1.2 Random variables

A **random variable** (or **rv**) X is a real-valued function on S.[1] Notation such as

$$\{X \leqslant t\} \text{ meaning } \{s \in S : X(s) \leqslant t\}$$

is often used to describe events involving X. The **indicator function** of a set E is defined by

$$\mathbf{1}_E(s) = \begin{cases} 1 & \text{if } s \in E \\ 0 & \text{if } s \notin E. \end{cases}$$

[1]There is one additional technical condition, namely that for every interval $I \subset \boldsymbol{R}$, the set $\{s \in S : X(s) \in I\}$ must be an event, that is it it must belong to $\mathcal{E}$. This requirement is known as **measurability of $\boldsymbol{X}$**. It will not be an issue you have to worry about today.

The indicator function of an event is a random variable.

The **cumulative distribution function** (or **cdf**) for X is denoted F_X, and defined by

$$F_X(t) = P\{X \leqslant t\}.$$

If F is differentiable, then F'_X is the **density** of X.

1.2.1 Expectation

The **expectation** of X is denoted $\boldsymbol{E}\,X$. In general it is defined to be

$$\boldsymbol{E}\,X = \int_S X(s)\,dP(s).$$

Let me explain this notation for the special case where X is *simple*, that is S is partitioned into events E_1, E_2,...,E_n, and X is constant on each E_k, say $X(s) = x_k$ for $s \in E_k$. Letting $p_k = P(E_k)$, we have

$$\boldsymbol{E}\,X = \sum_{k=1}^{n} p_k x_k.$$

For the case where X a density f, we have

$$\boldsymbol{E}\,X = \int x f(x)\,dx.$$

1 Proposition *Expectation is a positive linear operator. That is,*

$$\boldsymbol{E}(aX + bY) = a\,\boldsymbol{E}\,X + b\,\boldsymbol{E}\,Y$$

$$X \geqslant 0 \implies \boldsymbol{E}\,X \geqslant 0$$

$$X \geqslant Y \implies \boldsymbol{E}\,X \geqslant \boldsymbol{E}\,Y$$

$$P\{X = c\} = 1 \implies \boldsymbol{E}\,X = c$$

$$\boldsymbol{E}(\boldsymbol{E}\,X) = \boldsymbol{E}\,X$$

$$\boldsymbol{E}(X - \boldsymbol{E}\,X) = 0$$

$$\boldsymbol{E}\,\mathbf{1}_A = P(A)$$

$$\boldsymbol{E}(cX) = c\,\boldsymbol{E}\,X$$

$$\boldsymbol{E}(X + c) = \boldsymbol{E}\,X + c$$

$$\boldsymbol{E}(aX + c) = a\,\boldsymbol{E}\,X + c$$

2 Jensen's inequality *Let u be a concave function defined on an interval that includes the range of X. Assume $\boldsymbol{E}\,|X|$ and $\boldsymbol{E}\,|u(x)|$ are finite. Then*

$$u\big(\boldsymbol{E}\,X\big) \geqslant \boldsymbol{E}\,u(X).$$

Proof: Unless X is degenerate (in which case the conclusion holds trivially) $\boldsymbol{E}\,X$ belongs to the interior of the domain of u, so u has a supergradient there. That is there exists $p \in \boldsymbol{R}$ such that

$$u\big(\boldsymbol{E}\,X\big) + p(x - \boldsymbol{E}\,X) \geqslant u(x)$$

for all x. Thus $u\big(\boldsymbol{E}\,X\big) + p(X - \boldsymbol{E}\,X) \geqslant u(X)$, so taking the expectation on both sides gives

$$u\big(\boldsymbol{E}\,X\big) = \boldsymbol{E}\Big\{u\big(\boldsymbol{E}\,X\big) + p(X - \boldsymbol{E}\,X)\Big\} \geqslant \boldsymbol{E}\,u(X).$$

∎

The **variance** of a random variable is defined to be

$$\operatorname{var} X = \boldsymbol{E}\big((X - \boldsymbol{E}\,X)^2\big) = \boldsymbol{E}\big(X^2 - 2X\,\boldsymbol{E}\,X + (\boldsymbol{E}\,X)^2\big) = (\boldsymbol{E}\,X)^2 - \boldsymbol{E}(X^2).$$

The **covariance** of X and Y is $\boldsymbol{E}(X - \boldsymbol{E}\,X)(Y - \boldsymbol{E}\,Y)$.

3 Proposition *Let X be a non-negative random variable with finite expectation and cdf F_X. Then*

$$\boldsymbol{E}\,X = \int_0^\infty \big(1 - F_X(t)\big)\,dt.$$

Sketch of proof: Assume first that X is bounded above by b and that F is differentiable, so that F' is the density, and that $F(0) = 0$. Then using integration by parts,

$$\begin{aligned}
\boldsymbol{E}\,X &= \int_0^b xF'(x)\,dx \\
&= xF(x)\Big|_0^b - \int_0^b F(x)\,dx \\
&= b - \int_0^b F(x)\,dx \\
&= \int_0^b \big(1 - F_X(x)\big)\,dx.
\end{aligned}$$

The general conclusion uses a more sophisticated theorem on integration by parts based on Fubini's Theorem. ∎

1.3 Odds and prices

The payoffs for betting are usually described in terms of **odds**. If you wager an amount b on the event E and the odds against E are given by $\lambda(E)$, you receive λb if E occurs and lose b if

E fails to occur. We allow λ to take on any value in $[0, \infty]$. The interpretation of $\lambda(E) = \infty$ is that for any positive bet b, if E occurs, then the bettor may name any real number as his payoff. In a frictionless betting market, the odds against E^c are given by

$$\lambda(E^c) = \frac{1}{\lambda(E)},$$

where we use the conventions

$$\frac{1}{\infty} = 0, \quad \frac{1}{0} = \infty.$$

More conveniently, instead of using λ, define

$$q(E) = \frac{1}{1 + \lambda(E)},$$

$$q(E^c) = \frac{1}{1 + \lambda(E^c)} = \frac{1}{1 + \frac{1}{\lambda(E)}} = \frac{\lambda(E)}{1 + \lambda(E)}.$$

Note that

$$q(E) + q(E^c) = 1,$$

and that

$$\lambda(E) = \frac{q(E^c)}{q(E)}.$$

Moreover, if you bet $q(E) = \frac{1}{1+\lambda(E)}$ on E, then your payoff Π in state s is

$$\begin{aligned}
\Pi(s) &= q(E)\left[\lambda(E)\mathbf{1}_E(s) - \mathbf{1}_{E^c}(s)\right] \\
&= q(E)\left[\frac{q(E^c)}{q(E)}\mathbf{1}_E(s) - \mathbf{1}_{E^c}(s)\right] \\
&= q(E^c)\mathbf{1}_E(s) - q(E)\mathbf{1}_{E^c}(s) \\
&= \big(1 - q(E)\big)\mathbf{1}_E(s) - q(E)\big(1 - \mathbf{1}_E(s)\big) \\
&= \mathbf{1}_E(s) - q(E).
\end{aligned}$$

That is, $q(E)$ is the price of a \$1 bet on E. We shall call q the **price function** for bets.

1.4 Subjective probability and betting

4 Subjective probability theorem *Either*
(1) The price function q for bets is a probability and $\lambda(E) = \frac{q(E^c)}{q(E)}$ for each E.
Or else
*(2) The odds are **incoherent**, that is, there is a combination of bets that guarantees the bettor will win a positive amount regardless of which state s occurs.*

A set of incoherent odds is also known as a **Dutch book**.

Proof: (2) is equivalent to

$$S\left\{\ \overbrace{\begin{bmatrix} \vdots \\ \mathbf{1}_E(s) - q(E) \\ \vdots \end{bmatrix}}^{\mathcal{E}} \right. \begin{bmatrix} \vdots \\ x(E) \\ \vdots \end{bmatrix} \gg 0$$

(where $x(E)q(E)$ is the amount bet on E).

The alternative is that there is some probability vector $p \in \boldsymbol{R}^S$, such that for each event E,

$$\sum_{s \in S} p(s)\mathbf{1}_E(s) - q(E) = 0,$$

or

$$q(E) = \sum_{s \in E} p(s) = p(E),$$

which is (1). ∎

1.5 Statisticians' view of the world

Θ is a set of urns, each urn θ describes a probability p_θ on S. A particular urn θ_0 is used to choose signal $s \in S$ according to probability p_{θ_0}. We observe signal $s \in S$. What information does this convey about θ_0? (Statisticians don't call elements of Θ urns, they call them states of the world. In other words, statisticians believe that God does nothing but play dice.)

1.5.1 Conditional probability

The **conditional probability** of event E given event F is

$$p(E|F) = \frac{p(E \cap F)}{p(F)}.$$

Thus

$$p(E|F)p(F) = p(E \cap F) = p(F|E)p(E),$$

Or

$$p(E|F) = \frac{p(E)}{p(F)} \cdot p(F|E),$$

which is known as **Bayes' Law**.

1.5.2 Bayesian updating

Select urn θ_0 according to probability P on Θ, and select s according to p_{θ_0}. Then the probability that $\theta_0 \in T$, given s is

$$P(T|s) = \frac{\sum_{\theta \in T} p_\theta(s)P(\theta)}{\sum_{\theta \in \Theta} p_\theta(s)P(\theta)}.$$

P is known as a **prior**, and $P(\cdot|s)$ is the corresponding **posterior**.

Should Bayes' Law govern our betting behavior? Let's see.

1.5.3 Statistical inference: the game

Freedman and Purves [17] describe statistical inference in terms of the following game.

The Master of Ceremonies chooses an urn, and announces the signal s.

A Bookie posts odds λ against subsets $T \in \mathcal{T}$ of Θ.

Bets are placed.

The MC reveals the urn, and bets are settled.

(In the real world, the MC never tells.)

1.5.4 Strategies

Bookie chooses $q \geqq 0 \in \boldsymbol{R}^{\mathcal{T}\times S}$. For each $s \in S$,

$$q(T,s) + q(T^c,s) = 1.$$

Bettor then chooses $x \in \boldsymbol{R}^{\mathcal{T}\times S}$, and bets

$$x(T,s)q(T,s)$$

on T when s occurs.

Under these strategies, the expected payoff to the bettor when θ is the selected urn is just

$$\sum_{s\in S}\left(\sum_{T\in\mathcal{T}}\big(\mathbf{1}_T(\theta) - q(T,s)\big)x(T,s)\right)p_\theta(s).$$

5 Bayesian updating theorem *Either*

(1) The Bookie chooses some prior P and posts odds according to the posterior $P(\cdot|s)$

Or else

(2) There is a betting strategy that gives the bettor a positive expected payoff regardless of which urn θ is selected.

Proof: (2) is equivalent to

$$\Theta\left\{\ \overbrace{\begin{bmatrix} \\ \big(\mathbf{1}_T(\theta) - q(T,s)\big)p_\theta(s) \\ \\ \end{bmatrix}}^{\mathcal{T}\times S}\right. \begin{bmatrix} \vdots \\ x(T,s) \\ \vdots \end{bmatrix} \gg 0,$$

The alternative is the existence of a probability vector $P \in \boldsymbol{R}^{\Theta}$ such that for each (T,s),

$$\sum_{\theta\in\Theta}\big(\mathbf{1}_T(\theta) - q(T,s)\big)p_\theta(s)P(\theta) = 0.$$

In other words,

$$\sum_{\theta\in T} p_\theta(s)P(\theta) = \sum_{\theta\in\Theta} q(T,s)p_\theta(s)P(\theta),$$

or

$$q(T,s) = \frac{\sum_{\theta\in T} p_\theta(s)P(\theta)}{\sum_{\theta\in\Theta} p_\theta(s)P(\theta)} = P(T|s),$$

which is (1). ∎

1.6 The Ellsberg Paradox

Daniel Ellsberg [9] (of *Pentagon Papers* [10] fame) proposed the following example to test the intuitiveness of the subjective probability model.

There are two urns.

- Urn A contains 30 red balls, 30 black balls, and 30 yellow balls.
- Urn B contains 30 red balls, 60 balls that are either black or yellow.

Ellsberg asked a number of people to respond to the following two kinds of deals.

Deal 1: You will receive \$100 if a red or black ball is drawn from the urn. Which urn do you want to draw from?

Deal 2: You will receive \$100 if a red or yellow ball is drawn from the urn. Which urn do you want to draw from?

Many subjects indicate a preference for urn A in each deal. Reportedly these included L. J. Savage.[2] But such preferences are inconsistent with reasonable subjective probability and certainly with Savage's independence axiom: Let $p_A(\text{red})$ denote the probability of drawing a red ball from urn A, etc. A reasonable requirement is that

$$p_A(\text{red}) = p_B(\text{red}).$$

Choosing urn A in deal 1 implies

$$p_A(\text{red}) + p_A(\text{black}) > p_B(\text{red}) + p_B(\text{black})$$

and in deal 2 implies

$$p_A(\text{red}) + p_A(\text{yellow}) > p_B(\text{red}) + p_B(\text{yellow})$$

Assuming $p_A(\text{red}) = p_B(\text{red})$, this implies

$$p(\text{red}) + p_A(\text{black}) + p_A(\text{yellow}) > p(\text{red}) + p_B(\text{black}) + p_B(\text{yellow}),$$

when both sides are equal to 1.

Of course, if we are completely subjective, we could believe $p_A(\text{red}) = 1$ and $p_B(\text{red}) = 0$, but I doubt that's what Savage had in mind. Later on, I'll describe more satisfactory alternatives that allow for these sorts of preferences.

1.7 Expected utility model

The standard model of choice over random variables or lotteries is the expected utility (EU) model, which posits that a decision maker (dm) ranks random variables according to the expected value of their **Bernoulli utility** function u. That is, X is preferred to Y if $\boldsymbol{E}\,u(X) \geqslant \boldsymbol{E}\,u(Y)$.

6 Theorem *Two Bernoulli utilities u and v represent the same preference ranking over the set of random variables if and only if there are real numbers $a > 0$ and b satisfying $u(x) = av(x) + b$. That is,* Bernoulli utilities are unique up to positive affine transformation.

[2] Ellsberg presents a number of examples and it is not clear if it is this particular example or some other one that tripped up Savage (and Jacob Marshak and Norman Dalkey, but not Paul Samuelson nor Gerard Debreu, see pp. 655–656).

1.8 Alternative models

Multiple probability (MP) models typically rank random variables according to a function of the form

$$V(X) = \min_{P \in \mathcal{P}} \int_S u\big(X(s)\big)\, dP(s),$$

where $\mathcal{P}$ is a set of probabilities. If $\mathcal{P}$ includes all the degenerate probabilities ($\delta_s(\{s\}) = 1$), then this reduces to the **maximin** criterion, which ranks according to $\min_s X(s)$.

Another model is the **Choquet expected utility** (CEU) model, which uses a function of the form

$$V(X) = \int \nu\,\{X > t\}\; dt,$$

where ν is a *Choquet capacity* (a function on events satisfying $E \subset F \implies \nu(E) \leqslant \nu(F)$, but is not necessarily additive). By Proposition 3, if ν is a probability, this agrees with the usual expected utility. It is designed to explain the *Ellsberg paradox* and capture *ambiguity aversion*.

1.9 Allais Paradox

This example is due more-or-less to Allais [1]. Consider the lotteries

$$A_1 = [\$5m, .1; \$0, .9] \quad B_1 = [\$1m, .11; \$0, .89]$$

and

$$A_2 = [\$5m, .1; \$1m, .89; \$0, .01] \quad B_2 = [\$1m, 1]$$

(The notation means that A_1 pays \$5m with probability .1, and nothing with probability .9, etc.) Many real people report $B_2 \succ A_2$ and $A_1 \succ B_1$, which violates EUH:

$$\begin{array}{llll}
B_2 \succ A_2 & \implies & u(1m) > .1u(5m) + .89u(1m) + .01u(0) & \\
& \implies & .11u(1m) > .1u(5m) + .01u(0) & \text{(subtract } .89u(1m) \text{ from each side)} \\
& \implies & .11u(1m) + .89u(0) > .1u(5m) + .9u(0) & \text{(add } .89u(0) \text{ to each side)} \\
& \implies & B_1 \succ A_1. &
\end{array}$$

1.10 Lotteries, stochastic dominance and expected utility

In this section we consider lotteries over monetary prizes. Let

$$S = \{x_1 < \cdots < x_S\}$$

be a finite set of monetary prizes, listed in ascending order. (Note that I use the same symbol, S, to denote both the set of prizes and the number of prizes. It should not confuse you too much.) A **lottery** is a probability distribution over the prizes. Lotteries thus correspond to probability vectors in $\boldsymbol{R}^S$.

7 Definition *We say that lottery q* ***stochastically dominates*** *lottery p if*

$$\sum_{s=k}^{S} q_s \geqslant \sum_{s=k}^{S} p_s, \quad \text{for each } k = 2, \dots, S, \tag{1.1}$$

and $p \neq q$ (so that there is strict inequality for at least one k). That is, q always assigns higher probability than p to larger prizes.

Note that in the definition we require (1.1) only for $k = 2, \dots, S$. Since p and q are both lotteries the case $k = 1$ is an automatic equality, since both sum to one.

Intuitively, one should prefer a stochastically dominating lottery.

A utility $u \colon S \to \boldsymbol{R}$ on S can be thought of as vector u in $\boldsymbol{R}^S$, where the s^{th} component is the utility x_s. Since the prizes are listed in ascending order, it is natural to demand in addition that $u_1 < \cdots < u_S$. That is, we consider only strictly increasing utility functions for money.

Thus the expected utility of lottery $p \in \boldsymbol{R}^S$ with utility $u \in \boldsymbol{R}^S$ is simply

$$E_p u = \sum_{s=1}^{S} u_s p_s = u \cdot p.$$

8 Proposition *Let p and q be distinct lotteries. The following are equivalent:*

1. *Lottery q **stochastically dominates** lottery p.*
2. *For every nondecreasing u (i.e., $u_1 \leqslant \cdots \leqslant u_S$) we have*

$$u \cdot q \geqslant u \cdot p.$$

3. *For every strictly increasing u (i.e., $u_1 < \cdots < u_S$) we have*

$$u \cdot q \geqslant u \cdot p.$$

Proof: (1) $\implies$ (2): Assume q stochastically dominates p. Abel's famous formula for "summation by parts" van be written as follows:

$$\begin{aligned} u \cdot p = u_1 p_1 + u_2 p_2 + \cdots + u_S p_S = \\ & p_S (u_S - u_{S-1}) \\ & + (p_S + p_{S-1})(u_{S-1} - u_{S-2}) \\ & + (p_S + p_{S-1} + p_{S-2})(u_{S-2} - u_{S-3}) \\ & \quad \vdots \\ & + (p_S + p_{S-1} + \cdots + p_2)(u_2 - u_1) \\ & + \underbrace{(p_S + p_{S-1} + \cdots + p_1)}_{=1} u_1 \\ & = u_1 + \sum_{k=2}^{S} \left((u_k - u_{k-1}) \sum_{s=k}^{S} p_s \right). \end{aligned}$$

Likewise the expected utility of u under q is

$$u \cdot q = u_1 + \sum_{k=2}^{S} \left((u_k - u_{k-1}) \sum_{s=k}^{S} q_s \right).$$

Since u is nondecreasing, each $u_k - u_{k-1} \geqslant 0$, and since q dominates p, we have each $\sum_{s=k}^{S} q_s \geqslant \sum_{s=k}^{S} p_s$, so

$$u \cdot q \geqslant u \cdot p.$$

(2) $\implies$ (1): Assume that for every nondecreasing u (i.e., $u_1 \leqslant \cdots \leqslant u_S$) we have

$$u \cdot q \geqslant u \cdot p.$$

Given $k \in \{2, \dots, S\}$, consider u given by $u_s = 0$ for $s < k$ and $u_s = 1$ for $s \geqslant k$. Then

$$\sum_{s=k}^{S} q_s = u \cdot q \geqslant u \cdot p = \sum_{s=k}^{S} p_s.$$

(2) $\implies$ (3): This is obvious, as every strictly increasing u is also nondecreasing.

(3) $\implies$ (2): This follows from the fact that every nondecreasing u is the limit of a sequence of sequence of strictly increasing us. ∎

This theorem says that q dominates p if and only if very expected utility decision makes with an in increasing utility agrees that p is not better than q.

1.11 Choice and stochastic dominance

The next result asks, when is your choice of p rather than q rationalized by a strictly increasing utility u?

9 Expected utility theorem *Suppose p and q are distinct probability vectors. Either*

(i) There are $u_1 < \cdots < u_S$ such that

$$\sum_{i=1}^{S} u_i p_i > \sum_{i=1}^{S} u_i q_i$$

Or else

(ii) q stochastically dominates p.

That is, as long as your choice is not dominated, you act as if you maximize the expected utility of some strictly increasing utility.

Proof: (i) is equivalent to

$$\begin{bmatrix} p_1 - q_1 & p_2 - q_2 & p_3 - q_3 & \cdots & \cdots & \cdots & p_{S-1} - q_{S-1} & p_S - q_S \\ -1 & +1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & +1 & 0 & \ddots & & & 0 \\ 0 & 0 & -1 & +1 & 0 & & & 0 \\ \vdots & & 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & 0 & \vdots \\ 0 & & & & 0 & -1 & +1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & -1 & +1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_{S-1} \\ u_S \end{bmatrix} \gg 0.$$

Note that the matrix is $S \times S$. Gordan's Alternative 13 asserts that the alternative to (i) is that there exists $y = (y_0, y_1, \dots, y_{S-1}) > 0$ satisfying

$$\begin{array}{lclclcl}
y_0(p_1 - q_1) & & & - & y_1 & = & 0 \\
y_0(p_2 - q_2) & + & y_1 & - & y_2 & = & 0 \\
 & \vdots & & \vdots & & & \\
y_0(p_{S-1} - q_{S-1}) & + & y_{S-2} & - & y_{S-1} & = & 0 \\
y_0(p_S - q_S) & + & y_{S-1} & & & = & 0.
\end{array} \tag{1.2}$$

It is easy to see that $y_0 > 0$, for if $y_0 = 0$, then (1.2) implies everything unravels and the entire vector $y = 0$, a contradiction.

So without loss of generality we may set $y_0 = 1$. Then

$$\begin{array}{rclclcl}
p_1 - q_1 & & & - & y_1 & = & 0 \\
p_2 - q_2 & + & y_1 & - & y_2 & = & 0 \\
 & \vdots & & & & \vdots & \\
p_{S-1} - q_{S-1} & + & y_{S-2} & - & y_{S-1} & = & 0 \\
p_S - q_S & + & y_{S-1} & & & = & 0.
\end{array}$$

Since $p \neq q$ we cannot have $y_1 = \cdots = y_{S-1} = 0$, so for at least one $i \geqslant 1$ we have $y_i > 0$.

In other words, starting from the end, and adding up the last k inequalities, we have

$$\begin{array}{ccccccc}
p_n & - & q_n & = & -y_{n-1} & \leqslant & 0 \\
(p_{n-1} + p_n) & - & (q_{n-1} + q_n) & = & -y_{n-2} & \leqslant & 0 \\
\vdots & & & & \vdots & & \\
\displaystyle\sum_{i=2}^{n} p_i & - & \displaystyle\sum_{i=2}^{n} q_i & = & -y_1 & \leqslant & 0
\end{array}$$

and, since the y_is are not all zero, this is just (ii). ∎

1.12 Stochastic dominance and expected utility, *deux*

This generalizes the preceding result to larger collections of vectors $p^0, p^1, \dots, p^m$. We say that p^0 is an **extreme point** of this collection if it *cannot* be written as a convex combination of the others. That is, it is never true that $p^0 = \sum_{j=1}^m \lambda_j p^j$, where the λs are convex weights. In order to stand a chance of p^0 being the unique maximizer of any vector u, we must assume that it is an extreme point, otherwise we would have the contradiction $u \cdot p^0 > u \cdot \sum_{j=1}^m \lambda_j p^j = u \cdot p$.

10 Theorem *Let $p^0, p^1, \dots, p^m$ be probability vectors on S, and assume that p^0 is an extreme point. Then either*

i. *there is a utility u satisfying $u_1 < \cdots < u_S$ such that p^0 has the highest expected utility, that is,*

$$u \cdot p^0 > u \cdot p^i, \quad i = 1, \ldots, m;$$

or else

ii. *there is a probability vector $\pi \in \boldsymbol{R}^{\mathrm{m}}$ such that the mixture*

$$\sum_{i=1}^{m} \pi_i p^i \text{ stochastically dominates } p^0.$$

Proof: (cf. Fishburn [12], Ledyard [30], and Border [3, 4]) Condition (i) is equivalent to the following matrix equation, with $m + S - 1$ rows and S columns.

$$\left[\begin{array}{cccccccc}
p_1^0 - p_1^1 & p_2^0 - p_2^1 & p_3^0 - p_3^1 & \ldots & \ldots & \ldots & p_{S-1}^0 - p_{S-1}^1 & p_S^0 - p_S^1 \\
p_1^0 - p_1^2 & p_2^0 - p_2^2 & p_3^0 - p_3^2 & \ldots & \ldots & \ldots & p_{S-1}^0 - p_{S-1}^2 & p_S^0 - p_S^2 \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
p_1^0 - p_1^m & p_2^0 - p_2^m & p_3^0 - p_3^m & \ldots & \ldots & \ldots & p_{S-1}^0 - p_{S-1}^m & p_S^0 - p_S^m \\
\hline
-1 & +1 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & -1 & +1 & 0 & \ddots & & & 0 \\
0 & 0 & -1 & +1 & 0 & & & 0 \\
\vdots & & 0 & \ddots & \ddots & \ddots & & \vdots \\
\vdots & & & \ddots & \ddots & \ddots & 0 & \vdots \\
0 & & & & 0 & -1 & +1 & 0 \\
0 & 0 & \ldots & \ldots & 0 & 0 & -1 & +1
\end{array}\right]
\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_{S-1} \\ u_S \end{bmatrix} \gg 0.$$

Gordan's Alternative 13 assets that the alternative is that there is some semipositive $m+S-1$-vector

$$(\pi, y) = (\pi_1, \ldots, \pi_m, y_1, \ldots, y_{S-1}) > 0$$

satisfying

$$\begin{array}{lcccccc}
\sum_{i=1}^m \pi_i (p_1^0 - p_1^i) & & & - & y_1 & = & 0 \\
\sum_{i=1}^m \pi_i (p_2^0 - p_2^i) & + & y_1 & - & y_2 & = & 0 \\
 & & \vdots & & \vdots & & \\
\sum_{i=1}^m \pi_i (p_{S-1}^0 - p_{S-1}^i) & + & y_{S-2} & - & y_{S-1} & = & 0 \\
\sum_{i=1}^m \pi_i (p_S^0 - p_S^i) & + & y_{S-1} & & & = & 0.
\end{array}$$

It is easy to see that $\sum_{i=1}^m \pi_i > 0$, for if $\sum_{i=1}^m \pi_i = 0$, then $\pi = 0$, and everything unravels, so $(\pi, y) = 0$, a contradiction. Therefore we may renormalize, and assume without loss of generality that $\sum_{i=1}^m \pi_i = 1$.

Then just as in the proof of Theorem 9, we see that $\sum_{i=1}^m \pi_i p^i$ is either equal to or stochastically dominates $\sum_{i=1}^m \pi_i p^0 = p^0$. But our extremity hypothesis rules out their equality. That is, condition (ii) holds. ∎

1.13 The Allais paradox and stochastic dominance

The Allais paradox above presented a decision maker with two choice problems: Choose a lottery from the pair $\{A_1, B_1\}$ and choose a lottery from the pair $\{A_2, B_2\}$. The "paradoxical" choice is A_1 from the first pair and B_2 from the second pair.

Consider the following two-stage procedure: choose a pair, where each pair is is equally likely, and then play the chosen lottery. Compare that to the two-stage lottery involving the lotteries not chosen. This amounts to the choice problem of choosing a compound lottery from the pair of compound lotteries

$$C_1 = \left[A_1, \frac{1}{2}; B_2, \frac{1}{2}\right] \qquad C_2 = \left[B_1, \frac{1}{2}; A_2, \frac{1}{2}\right]$$

The compound lotteries reduce to

$$C_1 = [\$5m, .05; \$1m, .50; \$0, .45] \qquad C_2 = [\$5m, .05; \$1m, .50; \$0, .45]\,.$$

That is, the compound lotteries reduce to the identical single-stage lottery, yet the paradoxical choices indicate a strict preference for the first. The next theorem shows that this is not an isolated case. It is based on Border [4] and Ledyard [30].

1.14 Stochastic dominance and expected utility, *trois*

Let $S = \{x_1 < \cdots < x_S\}$ be a finite set of money prizes. Let $B_1, \ldots, B_m$ be **lottery budgets**, that is, each is a finite set of lotteries on S. A **choice function** c assigns to each budget B a single lottery $c(B)$ from the budget. Since the choice function selects a single element from budget we shall assume that it is the unique best element. So we shall say that the choice function is **EU-rational** if there is a utility function $u_1 < u_2 < \cdots < u_S$ on S such that for each $i = 1, \ldots, m$,

$$c(B_i) \cdot u > p \cdot u \text{ for all } p \in B_i \setminus c(B_i).$$

The paradoxical choices in the Allais example were not EU-rational, and we showed the existence of a probability measure over the budgets and an alternative choice function such that compound procedure of drawing a budget at random and then making the paradoxical choice is stochastically dominated.

A **mixed choice** assigns to each budget B_i a mixture (convex combination) $\sum_{j=0}^{m_i} \lambda_{ij} p^{ij}$ of the elements of B_i.

11 Theorem i. *The choice c is EU-rational, or else*

ii. *there is a probability vector $\pi \in \boldsymbol{R}^{\mathrm{m}}$, and a mixed choice d, where $d(B_i)$ does not put any weight on $c(B_i)$ for each i, such that the mixture*

$$\sum_{i=1}^m \pi_i d(B_i) \text{ stochastically dominates or equals } \sum_{i=1}^m \pi_i c(B_i).$$

Proof: (cf. Ledyard [30] and Border [4]) Let's enumerate each B_i as $p^{i0}, \dots, p^{im_i}$ where $p^{i0} = c(B_i)$. Create the matrix A with $\sum_{i=1}^m m_i + S - 1$ rows and n columns defined as follows.

$$
\left[\begin{array}{cccccccc}
p_1^{10} - p_1^{11} & p_2^{10} - p_2^{11} & p_3^{10} - p_3^{11} & \cdots & \cdots & \cdots & p_{S-1}^{10} - p_{S-1}^{11} & p_S^{10} - p_S^{11} \\
p_1^{10} - p_1^{12} & p_2^{10} - p_2^{12} & p_3^{10} - p_3^{12} & \cdots & \cdots & \cdots & p_{S-1}^{10} - p_{S-1}^{12} & p_S^{10} - p_S^{20} \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
p_1^{10} - p_1^{1m_1} & p_2^{10} - p_2^{1m_1} & p_3^{10} - p_3^{1m_1} & \cdots & \cdots & \cdots & p_{S-1}^{10} - p_{S-1}^{1m_1} & p_S^{10} - p_S^{1m_1} \\
\hline
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
\hline
p_1^{m0} - p_1^{m1} & p_2^{m0} - p_2^{m1} & p_3^{m0} - p_3^{m1} & \cdots & \cdots & \cdots & p_{S-1}^{m0} - p_{S-1}^{m1} & p_S^{m0} - p_S^{m1} \\
p_1^{m0} - p_1^{m2} & p_2^{m0} - p_2^{m2} & p_3^{m0} - p_3^{m2} & \cdots & \cdots & \cdots & p_{S-1}^{m0} - p_{S-1}^{m2} & p_S^{m0} - p_S^{20} \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
p_1^{m0} - p_1^{mm_m} & p_2^{m0} - p_2^{mm_m} & p_3^{m0} - p_3^{mm_m} & \cdots & \cdots & \cdots & p_{S-1}^{m0} - p_{S-1}^{mm_m} & p_S^{m0} - p_S^{mm_m} \\
\hline
-1 & +1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & -1 & +1 & 0 & \ddots & & & 0 \\
0 & 0 & -1 & +1 & 0 & & & 0 \\
\vdots & & 0 & \ddots & \ddots & \ddots & & \vdots \\
\vdots & & & \ddots & \ddots & \ddots & 0 & \vdots \\
0 & & & & 0 & -1 & +1 & 0 \\
0 & 0 & \cdots & \cdots & 0 & 0 & -1 & +1
\end{array}\right]
$$

Condition (i) is equivalent to the existence of a vector $u \in \boldsymbol{R}^S$ satisfying $Au \gg 0$.

Gordan's Alternative 13 assets that the alternative is that there is some semipositive $\sum_{i=1}^m m_i + S - 1$-vector

$$(\delta, y) = (\delta_{11}, \dots, \delta_{1m_1}, \dots, \delta_{m1}, \dots, \delta_{mm_m}, y_1, \dots, y_{S-1}) > 0$$

satisfying

$$
\begin{array}{lcccccc}
\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij}(p_1^{0j} - p_1^{ij}) & & & - & y_1 & = & 0 \\
\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij}(p_2^{0j} - p_2^{ij}) & + & y_1 & - & y_2 & = & 0 \\
 & \vdots & & \vdots & & & \\
\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij}(p_{S-1}^{0j} - p_{S-1}^{ij}) & + & y_{S-2} & - & y_{S-1} & = & 0 \\
\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij}(p_S^{0j} - p_S^{ij}) & + & y_{S-1} & & & = & 0.
\end{array}
$$

It is easy to see that $\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij} > 0$, otherwise everything unravels, so $(\delta, y) = 0$, a contradiction. Therefore we may renormalize and assume that $\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij} = 1$. Now for each i set

$$\pi_i = \sum_{j=1}^{m_i} \delta_{ij} \qquad i = 1, \dots, m$$

and

$$\lambda_{ij} = \begin{cases} \frac{\delta_{ij}}{\pi_i} & \pi_i > 0 \\ 0 & \pi_i = 0, \end{cases}$$

so $\sum_{i=1}^m \sum_{j=1}^{m_i} \delta_{ij} = \sum_{i=1}^m \pi_i \sum_{j=1}^{m_i} \lambda_{ij}$. Define the random choice d by

$$d(B_i) = \sum_{j=1}^{m_i} \lambda_{ij} p^{ij}, \qquad i = 1, \dots, m.$$

Then as in the proof of Theorem 10, we see that $\sum_{i=1}^m \pi_i d(B_i)$ stochastically dominates or equals $\sum_{i=1}^m \pi_i p^{i0} = \sum_{i=1}^m \pi_i c(B_i)$. ∎

I assert without proof that if $\sum_{i=1}^m \pi_i d(B_i) = \sum_{i=1}^m \pi_i c(B_i)$, then an arbitrarily small perturbation of the p^{ij}s will lead to $\sum_{i=1}^m \pi_i d(B_i)$ strictly dominating $\sum_{i=1}^m \pi_i c(B_i)$.

1.15 Appendix: Theorems of the Alternative

The mathematical tools we shall use are presented here without proof. See Gale [18, Chapter 2] or Border [2] for proofs. Here is the notation I use for vector orders.

$$\begin{aligned} x \geqq y &\iff x_i \geqslant y_i,\ i = 1, \dots, n \\ x > y &\iff x_i \geqslant y_i,\ i = 1, \dots, n \text{ and } x \neq y \\ x \gg y &\iff x_i > y_i,\ i = 1, \dots, n \end{aligned}$$

12 Theorem (Fredholm Alternative) *Let A be an $m \times n$ matrix and let $b \in \boldsymbol{R}^{\mathrm{m}}$. Exactly one of the following alternatives holds. Either there exists an $x \in \boldsymbol{R}^{\mathrm{n}}$ satisfying*

$$Ax = b \tag{1.3}$$

or else there exists $p \in \boldsymbol{R}^{\mathrm{m}}$ satisfying

$$\begin{aligned} pA &= 0 \\ p \cdot b &> 0. \end{aligned} \tag{1.4}$$

13 Gordan's Alternative *Let A be an $m \times n$ matrix. Exactly one of the following alternatives holds. Either there exists $x \in \boldsymbol{R}^{\mathrm{n}}$ satisfying*

$$Ax \gg 0. \tag{1.5}$$

or else there exists $p \in \boldsymbol{R}^{\mathrm{m}}$ satisfying

$$\begin{aligned} pA &= 0 \\ p &> 0 \end{aligned} \tag{1.6}$$

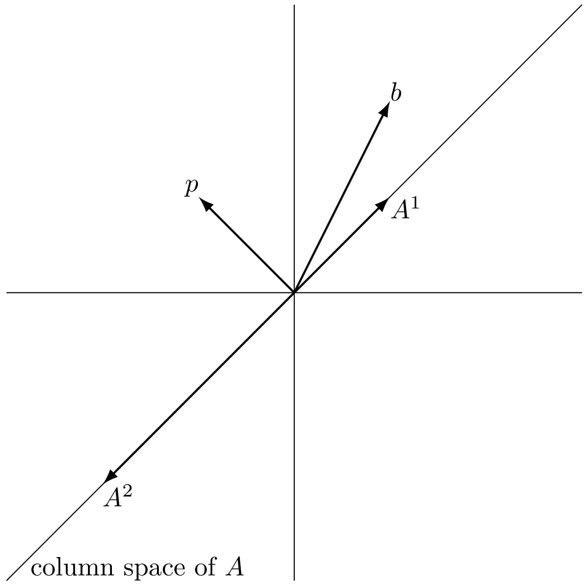

Figure 1.1. Geometry of the Fredholm Alternative

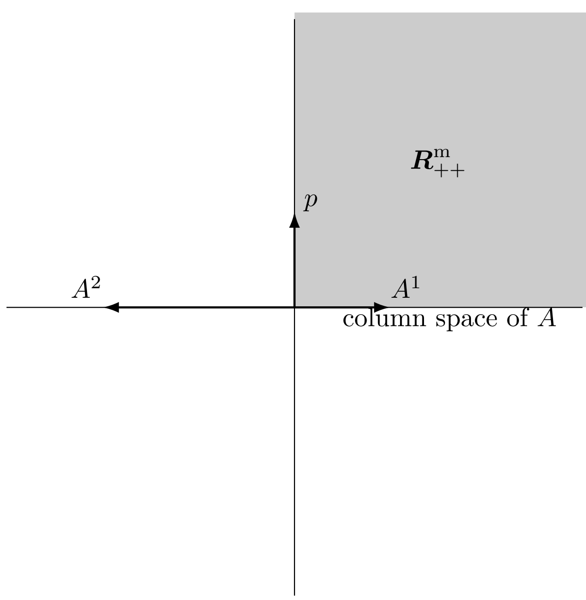

Figure 1.2. Geometry of Gordan's Alternative.

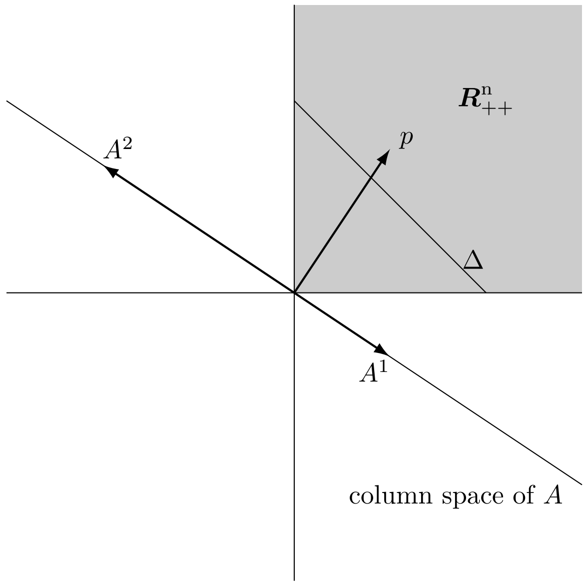

Figure 1.3. Geometry of the Stiemke Alternative

14 Stiemke's Alternative *Let A be an $m \times n$ matrix. Exactly one of the following alternatives holds. Either there exists $x \in \boldsymbol{R}^n$ satisfying*

$$Ax > 0 \tag{1.7}$$

or else there exists $p \in \boldsymbol{R}^m$ satisfying

$$\begin{aligned} pA &= 0 \\ p &\gg 0. \end{aligned} \tag{1.8}$$

Chapter 2

The EUH model

2.1 Risk aversion in the EU model

Risk aversion is the (weak) preference for $\boldsymbol{E}\,X$ for sure over X for all nondegenerate random variables X. That is,
$$U(\boldsymbol{E}\,X) \geqslant \boldsymbol{E}\,u(X).$$
In particular, if an EU dm with utility u is risk averse, and X assumes the values x and y with probabilities p and $1-p$ respectively, then
$$u\big(px + (1-p)y\big) \geqslant pu(x) + (1-p)u(y).$$
In other words, u is **concave**. Conversely if u is concave, then the dm is risk averse, which is a mathematical result known as *Jensen's inequality*.

In practice, it is easiest to identify concave functions by their derivatives. A differentiable utility u is concave if and only $u'(x)$ is a monotone decreasing function of x. A twice-differentiable utility u is concave if and only $u''(x) \leqslant 0$ for all x. Note that linear functions are concave. A dm with a linear utility is **risk neutral** and ranks random variables according to their expectation.

2.2 Stochastic dominance

The rv X **stochastically dominates** Y if
$$\boldsymbol{E}\,u(X) \geqslant \boldsymbol{E}\,u(Y) \text{ for every monotone nondecreasing function } u.$$

15 Theorem *X stochastically dominates Y if and only if*
$$F_X(t) \leqslant F_Y(t) \text{ for all } t.$$

2.3 Riskiness

The rv X is **riskier** than Y if
$$\boldsymbol{E}\,u(X) \leqslant \boldsymbol{E}\,u(Y) \text{ for every concave function } u.$$

16 Theorem *Assume the supports of F_X and F_Y satisfy $F(a) = F_Y(a) = 0$ and $F(b) = F_Y(b) = 1$.*

Then the following are equivalent.

$$\forall s \in [a,b] \quad \int_a^s F_X(t)\,dt \geqslant \int_a^s F_Y(t)\,dt \quad \& \int_a^b F_X(t)\,dt = \int_a^b F_Y(t)\,dt \tag{2.1}$$

$$\boldsymbol{E}\,u(X) \leqslant \boldsymbol{E}\,u(Y) \text{ for every concave function } u. \tag{2.2}$$

$$X = Y + Z \qquad \text{where } \boldsymbol{E}(Z|Y) = 0. \tag{2.3}$$

Proof that (2.2) implies (2.1): Let $s \in [a,b]$. Integrating by parts,

$$\begin{aligned}\int_a^s F_X(t)\,dt &= tF_X(t)\Big|_a^s - \int_a^s t\,dF_X(t).\\ &= sF_X(s) - \int_a^s t\,dF_X(t)\\ &= \int_a^s (s-t)\,dF_X(t)\\ &= \int_a^b (s-t)^+\,dF_X(t).\end{aligned}$$

Similarly

$$\int_a^s F_Y(t)\,dt = \int_a^b (s-t)^+\,dF_Y(t).$$

Since $(s-t)^+$ is a convex function of t, (2.2) implies

$$\int_a^s F_X(t)\,dt = \int_a^b (s-t)^+\,dF_X(t) \geqslant \int_a^b (s-t)^+\,dF_Y(t) = \int_a^s F_Y(t)\,dt.$$

When $s = b$, this becomes $\int_a^b F_X(t)\,dt = \int_a^b (b-t)\,dF_X(t)$. Now $b-t$ is both convex and concave in t, so we must have $\int_a^b F_X(t)\,dt = \int_a^b F_Y(t)\,dt$. ∎

2.4 Comparative risk aversion

A risk averse dm will pay to eliminate risk. We will say that one dm is **more risk averse** than another if his willingness to pay is always higher. Specifically, define **risk premium** $\pi_u(w,Z)$ by the equation

$$u\big(w + \boldsymbol{E}\,Z - \pi_u(w,Z)\big) = \boldsymbol{E}\,u(w+Z). \tag{$\star$}$$

It is the most that an EU decision maker with Bernoulli utility function u would be willing to pay to completely insure against the risk Z to his initial wealth w.

When u is twice differentiable, the **(Arrow–Pratt–de Finetti) coefficient of risk aversion** r_u is defined by

$$r_u(w) = -\frac{u''(w)}{u'(w)}.$$

Note that this coefficient is invariant under positive affine transformations of u, so it really is a property of the preferences.

17 Theorem *Let u and v be continuous strictly increasing functions that are twice differentiable with strictly positive derivatives. Then the following statements are equivalent.*

1. *For all w and all random variables Z that satisfy $\boldsymbol{E}\,Z = 0$,*
$$\pi_u(w, Z) \geqslant \pi_v(w, Z).$$

2. *There exists a concave strictly increasing function g defined on the range of v satisfying*
$$u = g \circ v.$$

3. *For all w,*
$$-\frac{u''(w)}{u'(w)} \geqslant -\frac{v''(w)}{v'(w)}.$$

2.4.1 Interpretation of the Arrow–Pratt-deFinetti coefficient

For each ε small enough let Z_ε be a random variable that takes on each of the values ε and $-\varepsilon$ with probability $\frac{1}{2}$. Then $\boldsymbol{E}\,Z_\varepsilon = 0$ and Z_ε is admissible for u at w. To simplify notation, define the real function p on A by $p(\varepsilon) = \pi_u(w, Z_\varepsilon)$. Note that $p(0) = 0$, $p(\varepsilon) = p(-\varepsilon)$, and by definition,
$$u\big(w - p(\varepsilon)\big) = \tfrac{1}{2}u(w+\varepsilon) + \tfrac{1}{2}u(w-\varepsilon). \tag{2.4}$$

Note that (2.4) implies that the function p is twice differentiable on A.[1] Since (2.4) holds for all small ε, we may differentiate both sides to get
$$-u'\big(w - p(\varepsilon)\big)p'(\varepsilon) = \tfrac{1}{2}u'(w+\varepsilon) - \tfrac{1}{2}u'(w-\varepsilon).$$
In particular, $p'(0) = 0$. Differentiating a second time yields
$$u''\big(w - p(\varepsilon)\big)\big(p'(\varepsilon)\big)^2 - p''(\varepsilon)u'\big(w - p(\varepsilon)\big) = \tfrac{1}{2}u''(w+\varepsilon) + \tfrac{1}{2}u''(w-\varepsilon).$$
In particular, using $p(0) = p'(0) = 0$, we have
$$p''(0) = -\frac{u''(w)}{u'(w)} = r_u(w).$$

We can apply Taylor's Theorem [23, p. 290] to write
$$\begin{aligned} p(\varepsilon) &= p(0) + \varepsilon p'(0) + \frac{\varepsilon^2}{2}\big(p''(0) + R(\varepsilon)\big) \\ &= \frac{\varepsilon^2}{2}\big(p''(0) + R(\varepsilon)\big), \end{aligned} \tag{2.6}$$
where $\lim_{\varepsilon\to 0} R(\varepsilon) = 0$.[2]

Now the variance of Z_ε is ε^2. So $\frac{p(\varepsilon)}{\varepsilon^2}$ is the fraction of the variance that someone with utility u would be willing to pay to insure against Z_ε. The limit of this fraction as $\varepsilon \to 0$ is then $\frac{1}{2}r_u(w)$. In fact, this generalizes to more general admissible small random variables with variance $\varepsilon > 0$

[1] To see this, define the function $f\colon A \times (D - w) \to \boldsymbol{R}$ by
$$f(\varepsilon, \eta) = u\big(w - \eta\big) - \tfrac{1}{2}u(w+\varepsilon) - \tfrac{1}{2}u(w-\varepsilon) \tag{2.5}$$
and note that f is twice differentiable, $f(0,0) = 0$, and $\dfrac{\partial f(0,0)}{\partial \eta} = -u'(w) < 0$. Therefore by the Implicit Function Theorem (see, e.g., [33, Theorem 2, p. 235]) there is a unique twice differentiable function defined on a neighborhood of zero giving η as a function of ε to satisfy equation (2.5).

[2] The form of Taylor's Theorem given by Hardy [23] requires only twice differentiability at 0, not twice continuous differentiability on a neighborhood of 0.

Chapter 3

More topics in uncertainty

3.1 Investment

There are two assets, a safe asset that returns $(1 + r_0)$ for each dollar invested and risky asset that returns $(1 + \boldsymbol{r})$ for each dollar invested, where $\boldsymbol{r}$ is a nondegenerate random variable.

If his wealth is $\hat{w}$, an expected utility maximizing investor will choose the amount x to invest in the risky asset to maximize

$$\boldsymbol{E}\, u\big((\hat{w} - x)(1 + r_0) + x(1 + \boldsymbol{r})\big).$$

The difference $\boldsymbol{r} - r_0$ is the excess of $\boldsymbol{r}$ over the safe return, so for convenience, let us call it $\boldsymbol{q}$, i.e., $\boldsymbol{q} = \boldsymbol{r} - r_0$, and set $w = (1 + r_0)\hat{w}$. Thus x is chosen to maximize

$$\boldsymbol{E}\, u(w + x\boldsymbol{q}),$$

which is a prettier problem.

There are some questions that are frequently glossed over in the literature. One is whether we want to restrict x to lie in the interval $[0, w]$. If so, we have to worry about boundary conditions. We also have to worry whether $w + x\boldsymbol{q}$ lies in the domain of the utility function with probability one. For instance, a utility function that is commonly studied is the logarithmic utility $u(w) = \ln w$ (where $u(0) = -\infty$ is allowed). If we make the limited liability assumption that $1 + \boldsymbol{r} \geqslant 0$ a.s., and also restrict x to $[0, w]$, then we have no problems in that regard. On the other hand, we may actually want to allow borrowing $(x > w)$ and/or short selling $(x < 0)$. In that case, we probably need to have the utility defined on the whole real line, which rules out the logarithmic utility, among others.

In what follows, I shall assume that utilities are defined on an interval D of the real line, are continuous strictly increasing functions on D that are twice continuously differentiable, with strictly positive derivatives everywhere on the interior of D, and that a solution exists and is interior to the domain.

The first order necessary condition for an interior maximum is

$$\boldsymbol{E}\, u'\big(w + x^*\boldsymbol{q}\big)\boldsymbol{q} = 0. \tag{$\star$}$$

Observe that $(\star)$ has a solution only if $\boldsymbol{q} < 0$ with positive probability, which makes perfect economic sense. (Otherwise there would be an arbitrage opportunity: borrow at r_0 and invest at $\boldsymbol{r}$, earning a riskless profit.)

The second order necessary condition is

$$\boldsymbol{E}\, u''\big(w + x^*\boldsymbol{q}\big)\boldsymbol{q}^2 \leqslant 0.$$

If u is concave, then $u'' \leqslant 0$, so the second order condition is automatically satisfied. I may also assume that the strong second order condition

$$\boldsymbol{E}\, u''\big(w + x^*\boldsymbol{q}\big)\boldsymbol{q}^2 < 0$$

holds at a particular solution. This is usually necessary to make the solution a differentiable function of the parameters.

3.1.1 A trivial lemma

18 Lemma *Let f be a nondecreasing real function on an interval I, let x belong to I, and let $\alpha > 0$. Then for any v for which $x + \alpha v \in I$, we have*

$$f(x + \alpha v)v \geqslant f(x)v.$$

This equality is reversed if $\alpha < 0$ or if f is nonincreasing. The inequality is strict provided $v \neq 0$ and f is not constant on the interval from x to $x + \alpha v$.

Proof: We prove the claim for $\alpha > 0$, the others are obvious from its proof. There are two interesting cases: $v > 0$ and $v < 0$. When $v > 0$, then the monotonicity of f implies $f(x+\alpha v) \geqslant f(x)$, so $f(x + \alpha v)v \geqslant f(x)v$. And when $v < 0$, then $f(x + \alpha v) \leqslant f(x)$, but multiplying by the negative quantity v reverses the inequality, so again $f(x + \alpha v)v \geqslant f(x)v$. ∎

3.1.2 Decreasing risk aversion

A natural comparative statics question is: What happens to x^* as a function of w?

19 Proposition *Assume u is C^2 and $u' > 0$, and define the Arrow–Pratt coefficient of risk aversion $r(w) = \dfrac{-u''(w)}{u'(w)}$. Fix w_0, and assume that x_0^* satisfies the strong second order condition. Then there is a neighborhood of w_0 on which x^* is a C^1 function of w.*

Moreover, if r is decreasing at w_0, then x^* is increasing at w_0 if x_0^* is positive and decreasing if x_0^* is negative. If, on the other hand, r is increasing at w_0, then $x^*(w)$ is decreasing when x_0^* is positive and increasing when x_0^* is negative.

Proof: Now x_0^* satisfies the first order condition

$$\boldsymbol{E}\, u'(w_0 + x_0^*\boldsymbol{q})\boldsymbol{q} = 0.$$

By the strong second order condition, the Implicit Function Theorem implies that x^* is a C^1 function of w on an appropriate neighborhood of w_0. Thus differentiating the first order condition with respect to w gives

$$\boldsymbol{E}\, u''(w_0 + x_0^*\boldsymbol{q})\boldsymbol{q}\left(1 + \boldsymbol{q}\frac{dx^*(w_0)}{dw}\right) = 0$$

or

$$\frac{dx^*(w_0)}{dw} = -\frac{\boldsymbol{E}\, u''(w_0 + x_0^*\boldsymbol{q})\boldsymbol{q}}{\boldsymbol{E}\, u''(w + x_0^*\boldsymbol{q})\boldsymbol{q}^2}.$$

The strong second order condition implies that the denominator is negative so the sign of $\dfrac{dx^*(w_0)}{dw}$ is the sign of $\boldsymbol{E}\, u''(w_0 + x_0^*\boldsymbol{q})\boldsymbol{q}$.

Now suppose $r(w)$ is decreasing at w_0. Consider first the case $x_0^* > 0$. By Lemma 18,

$$r(w_0 + x_0^*\boldsymbol{q})\boldsymbol{q} \leqslant r(w_0)\boldsymbol{q}.$$

Therefore, recalling the definition of r and multiplying by the negative quantity $-u'(w_0 + x_0^* \boldsymbol{q})$, we have

$$u''(w_0 + x_0^* \boldsymbol{q})\boldsymbol{q} \geqslant -r(w_0)u'(w_0 + x_0^* \boldsymbol{q})\boldsymbol{q}.$$

Taking the expectation of both sides gives

$$\boldsymbol{E}\, u''(w_0 + x_0^* \boldsymbol{q})\boldsymbol{q} \geqslant -r(w_0)\, \boldsymbol{E}\, u'(w_0 + x_0^* \boldsymbol{q})\boldsymbol{q} = 0$$

where the equality follows from the first order condition ($\star$). Thus

$$\operatorname{sign} \frac{dx^*(w_0)}{dw} = \operatorname{sign}\, Eu''(w + x_0^* \boldsymbol{q})\boldsymbol{q} \geqslant 0$$

when r is decreasing at w_0. Similarly, $\dfrac{dx^*(w_0)}{dw} \leqslant 0$ when r is increasing at w_0.

These conclusions are reversed if $x_0^* < 0$. ∎

3.1.3 What if u is more risk averse than v?

20 Proposition *Assume u is more risk averse than v. If v is risk averse or the two preferences are "sufficiently close" (in a sense to be made precise in the proof), then*

$$0 \leqslant x_u^* \leqslant x_v^* \qquad \text{or} \qquad x_v^* \leqslant x_u^* \leqslant 0.$$

That is, the more risk averse utility adopts the more conservative portfolio.

Proof: We prove only the case $x_u^* \geqslant 0$. The other follows *mutatis mutandis*. Write $u = G \circ v$, where G is strictly increasing and concave. Then ($\star$) becomes

$$\boldsymbol{E}\, G'\big(v(w + x_u^* \boldsymbol{q})\big)v'(w + x_u^* \boldsymbol{q})\boldsymbol{q} = 0.$$

Since G is concave, G' is nonincreasing, and thus so is $G' \circ v$. By Lemma 18,

$$G'\big(v(w + x_u^* \boldsymbol{q})\big)\boldsymbol{q} \leqslant G'\big(v(w)\big)\boldsymbol{q}.$$

Since $v' > 0$, we have

$$G'\big(v(w + x_u^* \boldsymbol{q})\big)v'(w + x_u^* \boldsymbol{q})\boldsymbol{q} \leqslant G'\big(v(w)\big)v'(w + x_u^* \boldsymbol{q})\boldsymbol{q},$$

and taking expectations yields

$$\underbrace{\boldsymbol{E}\, G'(v(w + x_u^* \boldsymbol{q}))v'(w + x_u^* \boldsymbol{q})\boldsymbol{q}}_{=0 \text{ by } (\star)} \leqslant G'(v(w))\, \boldsymbol{E}\, v'(w + x_u^* \boldsymbol{q})\boldsymbol{q}.$$

That is,

$$\boldsymbol{E}\, v'(w + x_u^* \boldsymbol{q})\boldsymbol{q} \geqslant 0.$$

But the first order condition for x_v^* is

$$\boldsymbol{E}\, v'(w + x_v^* \boldsymbol{q})\boldsymbol{q} = 0.$$

Now set $h(x) = \boldsymbol{E}\, v'(w + x\boldsymbol{q})\boldsymbol{q}$. Then $h(x_u^*) \geqslant 0 = h(x_v^*)$. But $h'(x_v^*) = \boldsymbol{E}\, v''(w + x_v^* \boldsymbol{q})\boldsymbol{q}^2 \leqslant 0$ by the second order condition for x_v^*. If u and v are close enough so that $h'(x) \leqslant 0$ on the interval between x_v^* and x_u^*, then $x_u^* \leqslant x_v^*$. (If v is concave, then $h' \leqslant 0$ and no closeness assumption is needed.) ∎

3.2 Deductibles vs. Coinsurance

You are subject to two kinds of risk. With probability $p_1 > 0$ you lose an amount x_1, and with probability $p_2 > 0$ you lose x_2. Assume $x_2 > x_1 > 0$ and $1 - p_1 - p_2 > 0$.

An insurance company offers two kinds of policies. The deductible policy reimburses you for all but d of your loss. The coinsurance policy reimburses you a fraction $1 - \rho$ of your loss. Suppose $0 < d < x_1 < x_2$ and that both policies have the same premium $\pi > 0$, and that both policies have the same expected value.

Suppose you are a risk averse expected utility maximizer and face no other risks. Which policy do you prefer?

Answer: Let w denote your initial wealth. There are three states of the world $\{0, 1, 2\}$. The random variables representing your wealth under the two policies are:

state	deductible X_d	coinsurance X_c	difference Z
0	$w - \pi$	$w - \pi$	0
1	$w - \pi - d$	$w - \pi - \rho x_1$	$d - \rho x_1$
2	$w - \pi - d$	$w - \pi - \rho x_2$	$d - \rho x_2$

That is,

$$X_c = X_d + Z.$$

Now observe that $Z = 0$ in the event $X_d = w - \pi$, and conditional on the event $X_d = w - \pi - d$ the expectation of Z is $\big((p_1 + p_2)d - \rho(p_1 x_1 + p_2 x_2)\big)/(p_1 + p_2)$. But both policies have the same expected value, $(p_1 + p_2)d = \rho(p_1 x_1 + p_2 x_2)$. Therefore

$$E(Z|X_d) = 0.$$

Then X_c is riskier than X_d, so a risk averse expected utility prefers X_d to X_c.

A less elegant but more elementary argument runs like this: Let U denote your utility and w denote your wealth. The expected utilities of the policies are:

$$\begin{aligned}
EU_{\text{deductible}} &= (1 - p_1 - p_2)U(w - \pi) + (p_1 + p_2)U(w - \pi - d) \\
EU_{\text{coinsurance}} &= (1 - p_1 - p_2)U(w - \pi) \\
&\quad + p_1 U(w - \pi - \rho x_1) + p_2 U(w - \pi - \rho x_2)
\end{aligned}$$

Since the policies have the same expected value,

$$p_1(x_1 - d) + p_2(x_2 - d) = (1 - \rho)(p_1 x_1 + p_2 x_2).$$

Rearranging,

$$-d = -\rho \left(\frac{p_1}{p_1 + p_2} x_1 + \frac{p_2}{p_1 + p_2} x_2 \right),$$

so

$$\begin{aligned}
w - \pi - d &= w - \pi - \rho \left(\frac{p_1}{p_1 + p_2} x_1 + \frac{p_2}{p_1 + p_2} x_2 \right) \\
&= \frac{p_1}{p_1 + p_2}(w - \pi - \rho x_1) + \frac{p_2}{p_1 + p_2}(w - \pi - \rho x_2).
\end{aligned}$$

Since U is concave,

$$\begin{aligned}
U(w - \pi - d) &= U\left(\frac{p_1}{p_1 + p_2}(w - \pi - \rho x_1) + \frac{p_2}{p_1 + p_2}(w - \pi - \rho x_2) \right) \\
&\geqslant \frac{p_1}{p_1 + p_2} U(w - \pi - \rho x_1) + \frac{p_2}{p_1 + p_2} U(w - \pi - \rho x_2)
\end{aligned}$$

Multiply both sides by $p_1 + p_2$ and add $(1 - p_1 - p_2)U(w - \pi)$ to conclude that

$$EU_{\text{deductible}} \geqslant EU_{\text{coinsurance}}.$$

3.3 Alternative models

Multiple probability (MP) models typically rank random variables according to a function of the form

$$V(X) = \min_{P \in \mathcal{P}} \int_S u\big(X(s)\big)\, dP(s),$$

where $\mathcal{P}$ is a set of probabilities. If $\mathcal{P}$ includes all the degenerate probabilities $(\delta_s(\{s\}) = 1)$, then this reduces to the **maximin** criterion, which ranks according to $\min_s X(s)$.

Another model is the **Choquet expected utility** (CEU) model, which uses a function of the form

$$V(X) = \int \nu P[X > t]\, dt,$$

where ν is a *Choquet capacity* (a function on events satisfying $E \subset F \implies \nu(E) \leqslant \nu(F)$, but is not necessarily additive). By Proposition 3, if ν is a probability, this agrees with the usual expected utility. It is designed to explain the *Ellsberg paradox* and capture *ambiguity aversion*.

3.4 State preference diagrams

A two-valued random variable can be represented as a point (x_a, x_b) in $\boldsymbol{R}^2$ (the value in event a is x_a and in event b is x_b). The forty-five degree line $\{(x, x) : x \in \boldsymbol{R}\}$ is called the **certainty line**, the value of X is the same in either event.

An **indifference curve** is a set of random variables with the same expected utility. That is, the set of pairs (x, y) such that

$$p_a u(x) + p_b u(y) = \text{constant},$$

where p_a is the probability of event a, etc. For each x, let $\hat{y}(x)$ satisfy

$$p_a u(x) + p_b u\big(\hat{y}(x)\big) = \text{constant}.$$

Since the rhs is independent of x, its derivative wrt x must be zero. That is,

$$p_a u'(x) + p_b u'\big(\hat{y}(x)\big)\hat{y}'(x) = 0,$$

so the slope $\hat{y}'$ of the indifference curve is

$$\hat{y}'(x) = -\frac{p_a u'(x)}{p_b u'\big(\hat{y}(x)\big)}.$$

Along the certainty line we have $\hat{y}(x) = x$, so the slope there is just $-p_a/p_b$.

3.4.1 Bets on a

A **bet on *a*** is a random variable with $x_a > 0$ and $x_b < 0$. A bet is **fair** if its expectation is zero, which entails

$$p_a x_a + p_b x_b = 0, \qquad \text{or} \qquad -\frac{p_a}{p_b} = \frac{x_b}{x_a}.$$

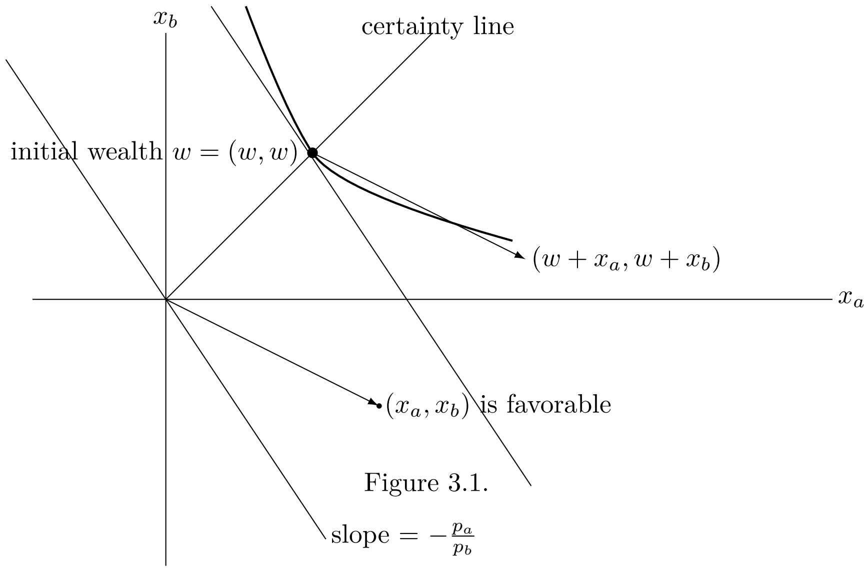

Figure 3.1.

A bet is **favorable** if

$$p_a x_a + p_b x_b > 0, \qquad \text{or} \qquad -\frac{p_a}{p_b} < \frac{x_b}{x_a}.$$

Suppose a risk averse EU dm with wealth w (the point $w = (w, w)$ on the certainty line) is offered the favorable bet x. See Figure 3.1. If his indifference curve is as drawn, he will not want to take the bet, since it would put him on a lower indifference curve. But since his indifference curve has slope $-p_a/p_b$ at w, the line segment joining w and $w + x$ crosses higher indifference curves so for small $\lambda > 0$, the point $w + \lambda x$ is preferred to w. So the dm would prefer to be able to take the bet λx.

21 Proposition *A risk averse EU dm with a smooth Bernoulli utility will prefer to take a small part of any favorable bet.*

Chapter 4

Adverse selection in insurance markets

This section is based on Michael Rothschild and Joseph Stiglitz [36], who argued that in the presence of *adverse selection*, markets for insurance were not guaranteed to deliver efficient outcomes, nor even to have equilibria.

4.1 Consumer types

We use a highly stylized model to starkly illustrate some of the key ideas. There are two **types** of insurance customers who are identical except for one trait—the probability that they will experience a loss. We assume that customers know their own type, but there is no way the insurance company can verify the type of a customer. This **asymmetric private information** is a source of problems in this market.

We consider only two states of the world, state 1 in which no loss occurs, so the wealth is w, and state 2, in which a loss of size c is suffered. Customers of type H are high-risk customers and have a probability p_H of a loss. Customers of type L are low-risk customers and have a probability p_L of a loss. Naturally,

$$1 > p_H > p_L > 0.$$

Assume the customers are EU decision makers with Bernoulli utility u. In the absence of insurance the expected utility of a type θ customer is

$$(1 - p_\theta)u(w) + p_\theta u(w - c),$$

where $\theta \in \Theta = \{L, H\}$.

A *state-preference diagram* is shown in Figure 4.1. Points in the plane represent random variables, that is, they represent the wealth in the two states of the world. The black dot is the *endowment point* $(w, w - c)$, so it lies below the *certainty line*. The red curve in the figure is an the indifference curve of the High-risk type, and its slope at the certainty line is $-(1 - p_H)/p_H$.

4.2 Insurance policies

An **insurance policy** Q is characterized by two parameters, the premium π and the benefit b that is payed in case of a loss. Since in our simple model all consumers are identical in terms of

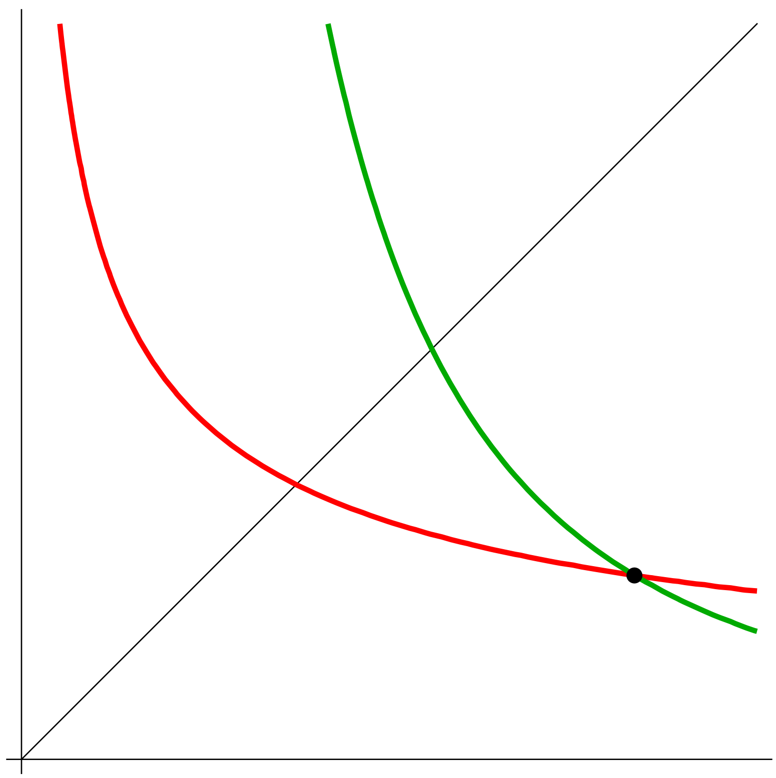

Figure 4.1. The black dot is the initial endowment absent insurance; the red indifference curve is for the High-risk type; the green indifference curve is for the Low-risk type.

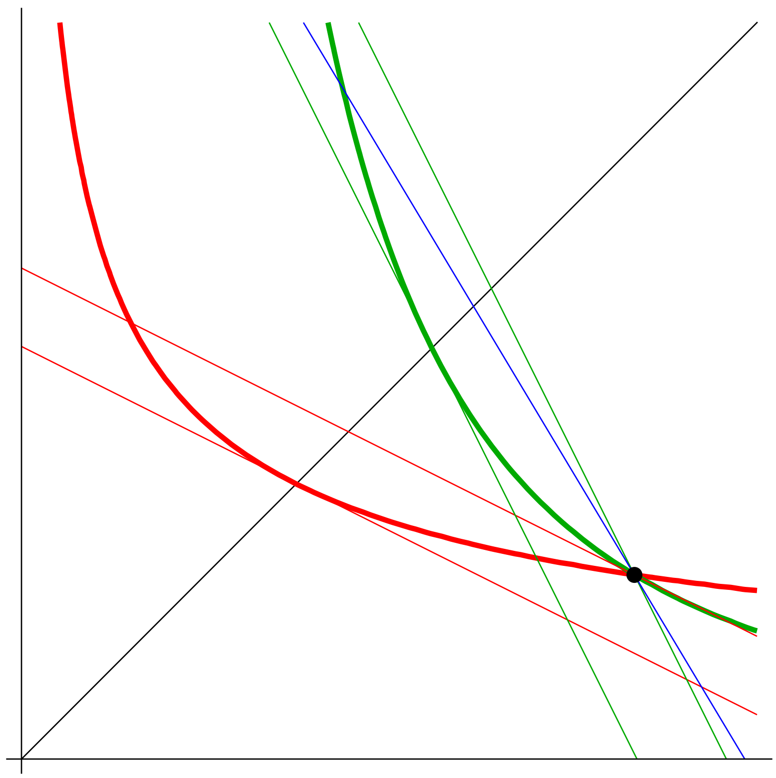

Figure 4.2. The black dot is the initial endowment absent insurance; the red lines are lines of equal expected value for p_H; the green lines are lines of equal expected value for p_L. The blue line is an iso-expected value line for p_A

their initial wealth and size of the loss, it is more convenient to represent a policy by its **result**,

$$X = (w - \pi, w - \pi + b - c).$$

The slope of the line segment connecting this point to the initial endowment is thus $-(b-\pi)/\pi$.

If p is the probability that a policyholder experiences a loss, the expected profit of a policy $Q = (\pi, b)$ to the insurance company is

$$\pi - pb.$$

The expected profit is nonnegative if and only if

$$\frac{1-p}{p} \geqslant \frac{b-\pi}{\pi}.$$

Thus a policy Q has a positive expected profit if and only if its result lies below the line through the endowment having slope $-(1-p)/p$), where p is the probability of a policyholder loss.

Figure 4.2 adds lines of equal expected value for the two types through the endowment. These lines indicate indifference curves for a risk-neutral insurance company.

Let λ denote the fraction of the population that is High-risk. The **average probability of a loss** is then

$$p_A = \lambda p_H + (1-\lambda)p_L.$$

The iso-expected valued line for the average probability of loss is shown in Figure 4.3. Note that in this example the full-insurance policy for the average customer (**FIPAC**), whose result is represented by the blue dot, is preferred to the initial endowment by both types H and L.

4.3 Equilibrium concept

An **equilibrium** in this market consists of a partition $\mathcal{T}$ of the type set Θ, and a list of pairs

$$(Q_T, T), \quad T \in \mathcal{T},$$

where Q_T is the policy purchased by consumers with type $\theta \in T$, such that

Self-selection Each consumer with type θ in T prefers Q_T to any other policy. (Note that $Q = (0,0)$, i.e., no insurance, is allowed to be one of the policies.)

Zero profit Each policy Q_T has expected profit zero, when the probability of a loss is the average probability of a loss for the set T.

Policy stability An insurer cannot make a positive expected profit by introducing a new policy. That is, there does not exist a policy Q' and set S of types such that S is the set of types θ who prefer Q' to Q_T, where $\theta \in T$, and Q' has positive expected profit when purchased by members of S.

In our simple model, there are two types of equilibria. A **separating equilibrium** has two policies Q_H and Q_L, where type H buys Q_H and type L buys Q_L. The policy Q_H has zero expected profit if the probability of loss is p_H and policy Q_L has zero expected profit if the probability of loss is p_L. The second kind of equilibrium is a **pooling equilibrium** with a single policy Q that is purchased by all consumers and has zero expected profit when the probability of loss is $p_A = \lambda p_H + (1-\lambda)p_L$.

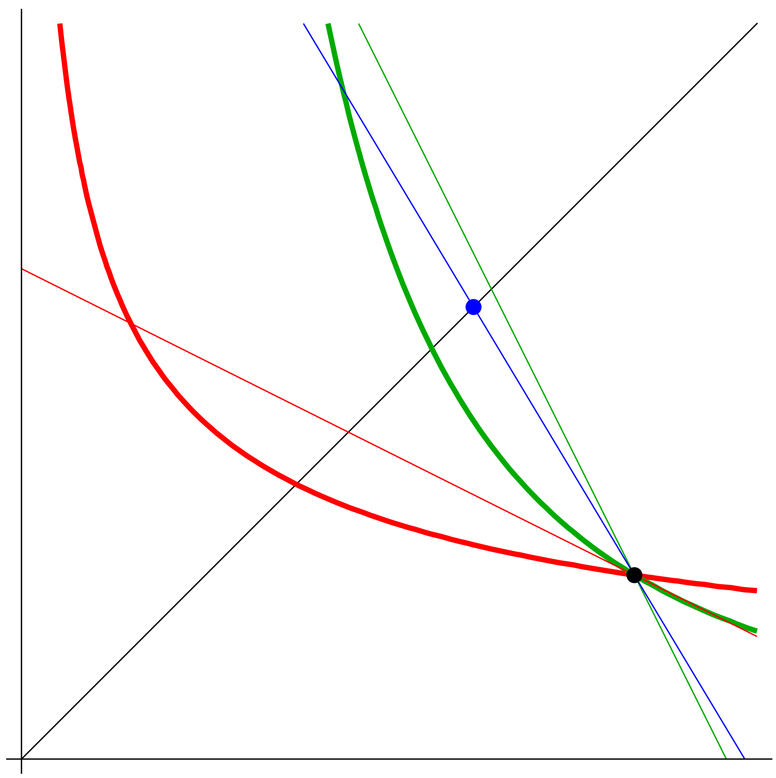

Figure 4.3. The black dot is the initial endowment absent insurance; the red line is an iso-expected value line for p_H; the green line is an iso-expected value line for p_L; and the blue line is an iso-expected value line for p_A.

4.4 Non-existence of pooling equilibrium

Call the pooling policy **FIPAC** (for full insurance policy for average customer). One might be tempted to think that competition among risk-neutral insurers would lead to the pooling policy as the market equilibrium. After all, risk-averse customers prefer full insurance, and the insurance company breaks even in expected value. As Rothschild and Stiglitz pointed out, the problem with this is that it is possible to offer a new policy that will make money by siphoning off the Low-risk customers from the **FIPAC**. That is, there is a policy (many, in fact) that is preferred to **FIPAC** by the Low-risk types, but is not preferred by the High-risk types, and has positive expected value for the insurance company when purchased only by Low-risk types. The orange region in Figure 4.4 shows the set of results of such policies. This siphoning-off of the

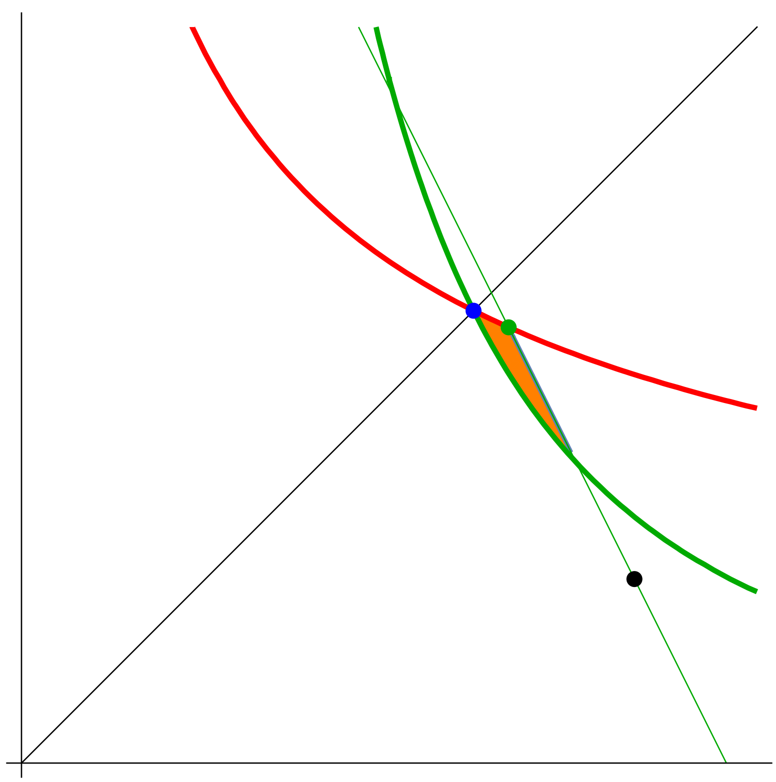

Figure 4.4. The orange region is preferred by type L to the result of **FIPAC**, the blue dot. It is not preferred by type H, and lies below the green line so it is profitable to sell to type L.

Low-risk types leaves, only the High-risk types purchasing **FIPAC**, which now has a negative expected value to the insurance company. This is known as **adverse selection** in the insurance industry.

4.5 Separating equilibrium

So what kind of policies can be supported? Figure 4.5 shows a separating market equilibrium in which the insurance industry offers two policies. The red dot is the result of full insurance

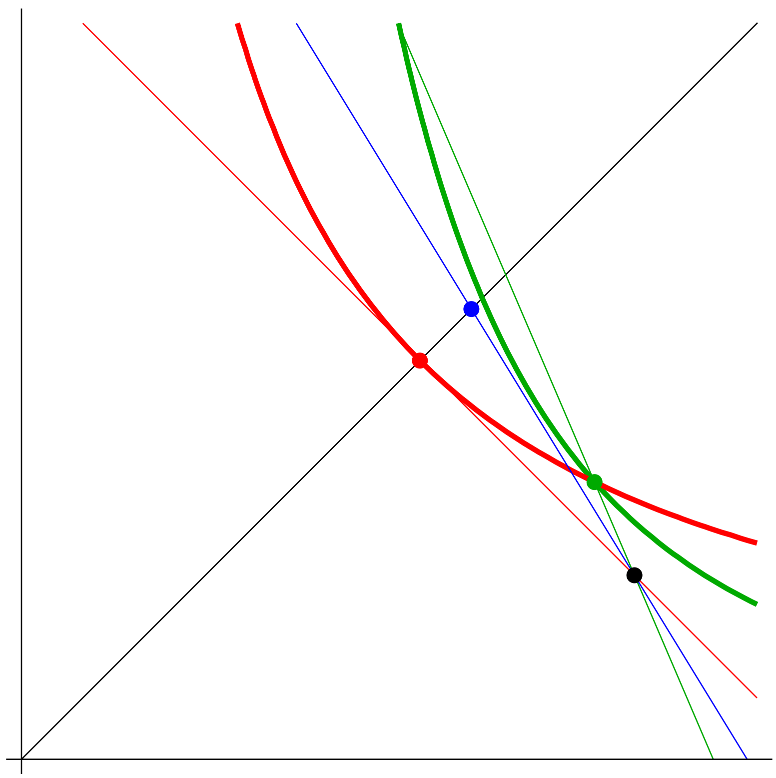

Figure 4.5. Separating Equilibrium

to the High-risk types (**FIH**) and has expected value zero at p_H. The green dot is the result of partial insurance to the Low-risk types (**PIL**) and has expected value zero. In this example, the **PIL** result is preferred to any result on the blue line, which would pool High and Low risks into an average risk. The **PIL** is the best policy the market can deliver to the Low-risk types, so the policy offerings are stable.

4.6 Non-existence of any equilibrium

Figure 4.6 shows a market in which the separating equilibrium described above does not exist. The red dot is again the result of full insurance to the High-risk types (FIH) and has expected value zero at p_H. The green dot is the result of the most favorable partial insurance to the Low-risk types (**PIL**) and has expected value zero. In this example, the **PIL** result is inferior

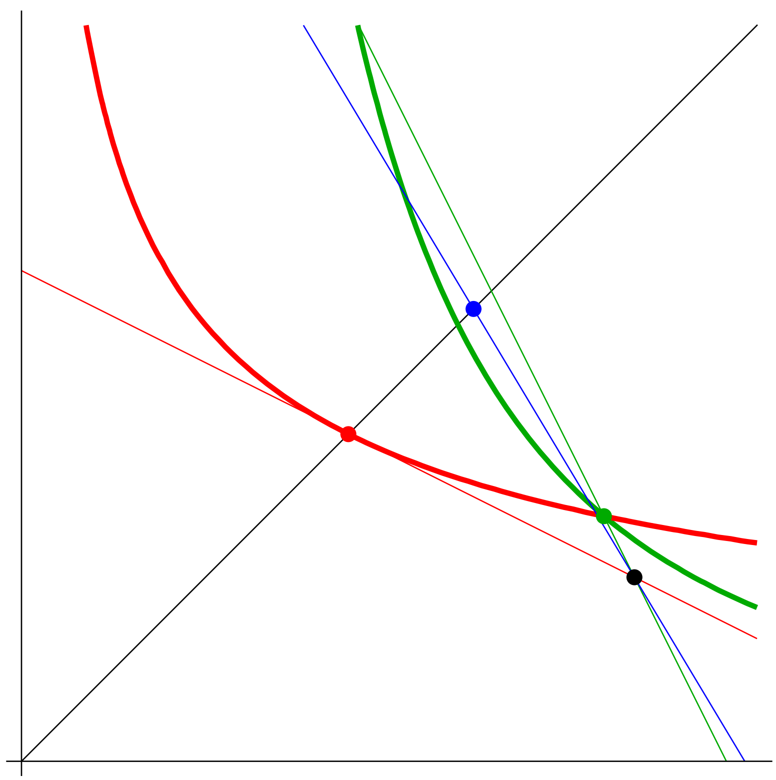

Figure 4.6. Failure of separating equilibrium.

to the blue point, which would pool High and Low risks into an average risk. This means that the blue policy would be bought by everyone if it were offered, so the policy offerings are not stable—a minor perturbation of the **FIPAC** will earn strictly positive profits and siphon off both types.

4.7 Appendix: Parameters for the examples

The parameters for the examples were chosen to yield legible figures, not for "realism."

Example	Utility	p_H	p_L	λ	w	c
Section 4.5	$u(x) = \ln x$	1/2	3/10	2/5	10	7
Section 4.6	$u(x) = \ln x$	2/3	1/3	1/8	10	7

Bibliography

[1] M. Allais. 1953. Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica* 21(4):503–546.
http://www.jstor.org/stable/1907921

[2] K. C. Border. Alternative linear inequalities.
http://www.hss.caltech.edu/~kcb/Notes/Alternative.pdf

[3] ——— . 1991. Functional analytic tools for expected utility theory. In C. D. Aliprantis, K. C. Border, and W. A. J. Luxemburg, eds., *Positive Operators, Riesz Spaces, and Economics*, number 2 in Studies in Economic Theory, pages 69–88. Berlin: Springer–Verlag.
http://www.hss.caltech.edu/~kcb/Courses/Ec181/pdf/Border1991-Dominance.pdf

[4] ——— . 1992. Revealed preference, stochastic dominance, and the expected utility hypothesis. *Journal of Economic Theory* 56(1):20–42. DOI: 10.1016/0022-0531(92)90067-R

[5] K. C. Border and U. Segal. 2002. Coherent odds and subjective probability. *Journal of Mathematical Psychology* 46(3):253–268. DOI: 10.1006/jmps.2001.1387

[6] S. H. Chew and L. G. Epstein. 1990. A unifying approach to axiomatic non-expected utility theories. *Journal of Economic Theory* 49:207–240.

[7] G. Darboux, ed. 1890. *Oeuvres de Jean-Baptiste Joseph Fourier*, volume 2. Paris: Gauthier–Villars.

[8] B. de Finetti. 1952. Sulla preferibilità. *Giornale degli Economisti e Annali di Economia* 11(11/12):685–709. http://www.jstor.org/stable/23236169

[9] D. Ellsberg. 1961. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics* 75(4):643–669. http://www.jstor.org/stable/1884324

[10] ——— . 2002. *Secrets: A memoir of Vietnam and the Pentagon Papers*. New York: Penguin Books.

[11] B. de Finetti. 1970–72. *Theory of probability*. London: Wiley. 2 vols.

[12] P. C. Fishburn. 1975. Separation theorems and expected utility. *Journal of Economic Theory* 11:16–34. DOI: 10.1016/0022-0531(75)90036-8

[13] ——— . 1983. Ellsberg revisited: A new look at comparative probability. *Annals of Statistics* 11:1047–1059. http://www.jstor.org/stable/2241296

[14] ——— . 1983. Transitive measurable utility. *Journal of Economic Theory* 31:293–317.

[15] J.-B. J. Fourier. 1826. Solution d'une question particuliére du calcul des inégalités [Solution of a particular problem in the calculus of inequalities]. *Nouveau Bulletin des Sciences par la Société Philomathique de Paris* pages 99–101. Reprinted in [7, pp. 317–319].

[16] I. Fredholm. 1903. Sur une classe d'équations fonctionelles. *Acta Mathematica* 27(1):365–390. DOI: 10.1007/BF02421317

[17] D. A. Freedman and R. A. Purves. 1969. Bayes' method for bookies. *Annals of Mathematical Statistics* 40(4):1177–1186. http://www.jstor.org/stable/2239586

[18] D. Gale. 1989. *Theory of linear economic models*. Chicago: University of Chicago Press. Reprint of the 1960 edition published by McGraw-Hill.

[19] I. Gilboa and D. Schmeidler. 1989. Maximin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2):141–153. DOI: 10.1016/0304-4068(89)90018-9

[20] P. Gordan. 1873. Über die auflösung linearer Gleichungen mit reelen Coefficienten [On the solution of linear inequalities with real coefficients]. *Mathematische Annalen* 6(1):23–28. DOI: 10.1007/BF01442864

[21] J. Hadar and W. R. Russell. 1969. Rules for ordering uncertain prospects. *American Economic Review* 59(1):25–34. http://www.jstor.org/stable/1811090

[22] ——— . 1978. Applications in economic theory and analysis. In Whitmore and Findlay [41], chapter 7, pages 293–333.

[23] G. H. Hardy. 1952. *A course of pure mathematics*, 10th. ed. Cambridge: Cambridge University Press.

[24] D. C. Heath and W. D. Sudderth. 1972. On a theorem of de Finetti, oddsmaking, and game theory. *Annals of Mathematical Statistics* 43:2072–2077. http://www.jstor.org/stable/2240227

[25] C. Hildreth. 1974. Expected utility of uncertain ventures. *Journal of the American Statistical Association* 69(345):9–17. http://www.jstor.org/stable/2285494.pdf

[26] ——— . 1974. Ventures, bets and initial prospects. In M. Balch, D. L. McFadden, and S. Wu, eds., *Decision Rules and Uncertainty: NSF–NBER Proceedings*, pages 99–131. Amsterdam: North Holland.

[27] A. N. Kolmogorov. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer. A Russian translation by G. M. Bavli, appeared in 1936, with a second edition, slightly expanded by Kolmogorov with the assistance of A. N. Shiryaev, in 1974, and a third edition in 1998. An English translation by N. Morrison appeared under the title *Foundations of the Theory of Probability* (Chelsea, New York) in 1950, with a second edition in 1956.

[28] ——— . 1956. *Foundations of the theory of probability*. New York: Chelsea. Translated from the 1933 German edition by N. Morrison.

[29] D. M. Kreps. 1988. *Notes on the theory of choice*. Underground Classics in Economics. Boulder, Colorado: Westview Press.

[30] J. O. Ledyard. 1986. The scope of the hypothesis of Bayesian equilibrium. *Journal of Economic Theory* 39(1):59–82. DOI: 10.1016/0022-0531(86)90020-7

[31] E. L. Lehmann. 1955. Ordered families of distributions. *Annals of Mathematical Statistics* 26:399–419. http://www.jstor.org/stable/2236469

[32] R. D. Luce and H. Raiffa. 1957. *Games and decisions*. New York: Wiley.

[33] J. E. Marsden. 1974. *Elementary classical analysis*. San Francisco: W. H. Freeman and Company.

[34] J. W. Pratt. 1964. Risk aversion in the small and in the large. *Econometrica* 32(1/2):122–136. http://www.jstor.org/stable/1913738

[35] J. P. Quirk and R. Saposnik. 1962. Admissibility and measurable utility functions. *Review of Economic Studies* 29(2):140–146. http://www.jstor.org/stable/2295819

[36] M. Rothschild and J. Stiglitz. 1976. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics* 90:629–649. http://www.jstor.org/stable/1885326

[37] T. J. Sargent. 1979. *Macroeconomic theory*. New York: Academic Press.

[38] L. J. Savage. 1954. *Foundations of statistics*. New York: John Wiley and Sons. Reprint. New York: Dover, 1972.

[39] D. Schmeidler. 1989. Subjective probability and expected utility without additivity. *Econometrica* 57:571–587. http://www.jstor.org/stable/1911053

[40] E. Stiemke. 1915. Über positive Lösungen homogener linearer Gleichungen. *Mathematische Annalen* 76(2–3):340–342. DOI: 10.1007/BF01458147

[41] G. A. Whitmore and M. Findlay, eds. 1978. *Stochastic dominance: An approach to decision making under risk*. Lexington, Mass.: Heath.