CALIFORNIA INSTITUTE OF TECHNOLOGY
Division of the Humanities and Social Sciences

Aspects of normative decision theory

KC Border

Spring 1984; Revised Fall 2007
v. 2011.03.29::20.58

Normative decision theory attempts to answer the question, "How can I make good decisions when I don't have good information?" It does not try to describe how real people make real decisions, but we believe that if we do a good job of answering the question, then real decision makers will want to behave in accordance with our theory. In fact, this is a large part of the rationale that good business schools have for including decision theory in their curriculum.

Problem: What do we mean by "good?"

The first example of normative decision theory is the theory of probability that was developed by the Bernoullis (Jacob and Daniel) in the first half of the 18th century for games involving randomizing devices such as cards, dice, and roulette wheels. It used simple combinatorial rules for calculating probabilities and expected values. In the 20th century, a notion of **subjective** probability was developed that did not rely on notions of randomness, but of belief and behavior (Bruno de Finetti and Jimmie Savage). How, you may ask, does it make sense to think of probability without randomness?

1 The probabilists' model of uncertainty

The modern approach to uncertainty, as formalized by Kolmogorov, 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 the purpose of these notes S is assumed to be finite.*

The collection $\mathcal{E}$ of **events** is usually assumed to be an **algebra** of subsets of S. That is, $\mathcal{E}$ satisfies:

i. $S \in \mathcal{E}$, $\varnothing \in \mathcal{E}$.

ii. If $E \in \mathcal{E}$, then $E^c \in \mathcal{E}$.

iii. If $E, F \in \mathcal{E}$, then $E \cap F \in \mathcal{E}$ and $E \cup F \in \mathcal{E}$.

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

i. For each $E \in \mathcal{E}$,

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

ii. If $E \cap F = \varnothing$, then

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

A **probability vector** $p \in \boldsymbol{R}^S$ satisfies

$$p_i \geqslant 0,\ i \in S \qquad \text{and} \qquad \sum_{i \in S} p_i = 1.$$

A probability vector defines a probability p on $\mathcal{E} = 2^S$ via

$$p(E) = \sum_{i \in E} p_i.$$

2 The statisticians' model of uncertainty

The statisticians' approach to the world is sightly different. Its key ingredients are:

S, a **sample space**.
$\mathcal{E} \subset 2^S$, a collection of **sample events**.
$\{p_\theta : \theta \in \Theta\}$, a set of probabilities on $\mathcal{E}$.

The **sample space** is the set of outcomes of a **statistical experiment**. Statisticians tend to regard elements of Θ as **states of the world**. Depending on the school of thought, one probability p_{θ_0} may be regarded as the **true state of the world**. **Bayesian** statisticians also put a probability measure on Θ, which may be either a **prior** or **posterior** depending on the stage of their analysis.

When S is a subset of $\boldsymbol{R}^{\mathrm{n}}$ and each p_θ has a density f_θ, the **likelihood function** is a mapping $L\colon \Theta \times S \to \boldsymbol{R}_+$ defined by

$$L(\theta|s) = f_\theta(s).$$

3 Subjective likelihood

The subjective relative likelihood of an individual is a binary relation on events (subsets of S). We write $E \succcurlyeq F$ to mean that event E is at least as likely as event F. Let us say that the subjective likelihood $\succcurlyeq$ is **represented by a probability measure $\boldsymbol{p}$** if

$$E \succcurlyeq F \iff p(E) \geqslant p(F).$$

The following are some obvious necessary conditions for a subjective likelihood to have a representation by a probability p:

C (Completeness) For all E, F, either $E \succcurlyeq F$ or $F \succcurlyeq E$, or both.

T (Transitivity) For all E, F, G,

$$[E \succcurlyeq F \ \& \ F \succcurlyeq G] \quad \Longrightarrow \quad E \succcurlyeq G.$$

A (Additivity) If $E \cap G = \varnothing$ and $F \cap G = \varnothing$, then

$$E \succcurlyeq F \quad \Longleftrightarrow \quad E \cup G \succcurlyeq F \cup G.$$

N (Nontriviality) For every event E, $E \succcurlyeq \varnothing$ and $S \succ \varnothing$.

A subjective likelihood relation satsifying these properties has been called an **intuitive probability** [13] or **qualitative probability** [16].

While it is clearly necessary that a binary relation be a qualitative probability in order to be representable by a probability, but it is not sufficient, as the following example due to Kraft, Pratt, and Seidenberg [14] shows. (There is an unfortunate typographical error on page 414 of their paper, but it is corrected later on.)

1 Example (Qualitative probability not representable) Partially define $\succcurlyeq$ on the finite set $\{a, b, c, d, e\}$ by

$$\begin{gathered} \{a,b,d\} \succ \{c,e\} \succ \{a,b,c\} \succ \{b,e\} \succ \{a,d\} \succ \{a,c\} \succ \{b,c,d\} \succ \{e\} \\ \succ \{c,d\} \succ \{a,b\} \succ \{a\} \succ \{b,d\} \succ \{b,c\} \succ \{d\} \succ \{c\} \succ \{b\} \succ \varnothing \end{gathered} \tag{1}$$

This orders seventeen of the thirty-two subsets. Each of the remaining fifteen subsets is a complement of one of these, so if we assign a probability to each of these sets, the probability of the remainder is determined. The complements must be ordered in the reverse order. That is, we must have

$$\begin{gathered} \{a,b,c,d,e\} \succ \{a,c,d,e\} \succ \{a,b,d,e\} \succ \{a,d,e\} \succ \{a,c,e\} \succ \{b,c,d,e\} \succ \{a,b,e\} \\ \succ \{a,b,c,d\} \succ \{a,e\} \succ \{b,d,e\} \succ \{b,c,e\} \succ \{a,c,d\} \succ \{d,e\} \succ \{a,b,d\} \succ \{c,e\} \end{gathered}$$

This specifies a linear order on all the subsets. Checking additivity is simple, but tedious. K–P–S prove a little lemma to simplify things a bit, but I leave to you to verify that the additivity condition A is satisfied. (Their lemma is that under a linear order, if the bottom half of the order satisfies additivity, and the top half consists of the complements of the bottom half ordered in reverse, then the entire order satisfies additivity.)

Now to show that this order has no probability representation. From (1) we have

$$\{a\} \succ \{b,d\}, \quad \{c,d\} \succ \{a,b\}, \quad \{b,e\} \succ \{a,d\}$$

so a representation p would have to satisfy

$$p(a) > p(b) + p(d), \quad p(c) + p(d) > p(a) + p(b), \quad p(b) + p(e) > p(a) + p(d).$$

Adding these inequalities, we would have to have

$$p(a) + p(b) + p(c) + p(d) + p(e) > 2p(a) + 2p(b) + 2p(d),$$

or

$$p(c) + p(e) > p(a) + p(b) + p(d),$$

which contradicts $\{a, b, d\} \succ \{c, e\}$. Thus no representation exists. □

K–P–S give a necessary and sufficient condition for a likelihood relation (on a finite set) to be representable by a probability, but their condition is expressed in terms of monomials in the letters representing the elements of the set. The next result translates this into more familiar set-theoretic terms.

2 Theorem (Scott [17]) *Let S be a finite set and let $\mathcal{E}$ be an algebra of subsets of S and let $\succcurlyeq$ be a binary relation on $\mathcal{E}$. For $\succcurlyeq$ to be representable by a probability measure p on $\mathcal{E}$, that is,*

$$A \succcurlyeq B \iff p(A) \geqslant p(B),$$

it is necessary and sufficient that for all events $A, A_0, \dots, A_n, B, B_0, \dots, B_n$ in $\mathcal{E}$ the following conditions hold:

i. $S \succ \varnothing$

ii. $A \succcurlyeq \varnothing$

iii. $A \succcurlyeq B$ *or* $B \succcurlyeq A$

iv. $\left[A_i \succcurlyeq B_i \ i = 1, \dots, n \text{ and } \sum\limits_{i=0}^{n} \mathbf{1}_{A_i} = \sum\limits_{i=0}^{n} \mathbf{1}_{B_i} \right] \implies B_0 \succcurlyeq A_0$

Condition (iv) is somewhat opaque. To understand it, let's see why it is a necessary condition for $\succcurlyeq$ to be represented by a probability.

Suppose by way of contradiction that the antecedent of (iv) holds but $A_0 \succ B_0$. Recall that $\mathbf{1}_E$ is the indicator function of E. That is, $\mathbf{1}_E(s) = 1$ if $s \in E$ and $\mathbf{1}_E(s) = 0$ if $s \notin E$. Thus $\sum_{i=0}^{n} \mathbf{1}_{A_i}(s)$ is the number of the sets $A_0, \dots, A_n$ that contain s. This implies that

$$p\left(\bigcup_{i=0}^{n} A_i\right) = \sum_{i=0}^{n} p(A_i) - \sum_{s \in S} p(s) \left(\sum_{i=0}^{n} \mathbf{1}_{A_i}(s) - 1\right)^{+}, \tag{2}$$

where $x^+ = x$ if $x \geqslant 0$, and $= 0$ otherwise. Now for $\succcurlyeq$ to be representable by a probability p, it would have to satisfy $p(A_i) \geqslant p(B_i)$, for $i = 1, \ldots, n$ and $p(A_0) > p(B_0)$. This in turn implies

$$\sum_{i=0}^{n} p(A_i) > \sum_{i=0}^{n} p(B_i).$$

But then by condition (4), we must also have

$$\sum_{i=0}^{n} p(A_i) - \sum_{s \in S} p(s) \left(\sum_{i=0}^{n} \mathbf{1}_{A_i}(s) - 1 \right)^{+} > \sum_{i=0}^{n} p(B_i) - \sum_{s \in S} p(s) \left(\sum_{i=0}^{n} \mathbf{1}_{B_i}(s) - 1 \right)^{+}.$$

So by (2), we have

$$p\left(\bigcup_{i=0}^{n} A_i \right) > p\left(\bigcup_{i=0}^{n} B_i \right).$$

But $\sum_{i=0}^{n} \mathbf{1}_{A_i} = \sum_{i=0}^{n} \mathbf{1}_{B_i}$ implies $\bigcup_{i=0}^{n} A_i = \bigcup_{i=0}^{n} B_i$. This contradiction shows the necessity of condition (iv).

The sufficiency is a lot subtler, and some day I will write it out for you.

4 Subjective probability and betting

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 given by

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

That is, $q(E)$ is the price of a lottery ticket that pays \$1 in event E. Let's call such a lottery ticket an ***E*-ticket**.[1]

3 Subjective probability theorem *Either*
(i) The function q is a probability and $\lambda(E) = \frac{q(E^c)}{q(E)}$ for each E.
Or else
*(ii) 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: Condition (ii) 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 $x(E)$ is the number of E-tickets).

Gordan's Alternative 11 assets that the alternative is that there is some probability vector $p \in \boldsymbol{R}^S$, such that for each event E,

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

---

[1] Young people think an E-ticket is something that lets you on an airplane, but we older Southern Californians know you it's what lets you get on the Matterhorn.

or

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

which is (i). ∎

5 The Ellsberg Paradox

Daniel Ellsberg [3] (of *Pentagon Papers* [4] 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 Jimmy 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 $p_B(\text{red}) = 0$, bit I doubt that's what Savage had in mind. Later on, I'll describe more satisfactory alternatives that allow for these sorts of preferences.

---

[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 or Gerard Debreu), see pp. 655–656.

6 Statisticians' view of the world

The statistical view of the world can be caricatured as follows: 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. Or as the unofficial motto of the American Statistical Association puts it, "Statistics means never having to say you're certain."[3])

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**.

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.

Statistical inference: the game

Freedman and Purves [8] describe statistical inference in terms of the following game.

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

---

[3] Dave Grether often wears a T-shirt from the ASA with this motto. For those of you who are too young to remember, the motto is a takeoff on the tag line "Love means never having to say you're sorry," from the movie *Love Story* (1970) based on the novel of the same name by Yale professor Erich Segal.

2. A Bookie posts odds against subsets $T \in \mathfrak{T}$ of Θ.
3. Bets are placed.
4. The MC reveals the urn, and bets are settled.

(However, in the real world, the MC never tells.)

Strategies

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

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

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

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

4 Bayesian updating theorem *Either*

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

Or else

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

Proof: (ii) is equivalent to

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

Gordan's Alternative 11 assets that 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 (i). ∎

7 Measurable utility

A **mixture space** M is a set such that for each $x, y \in M$ and $\alpha \in [0,1]$ there is an element $\alpha x + (1-\alpha)y$ in M, where

i. $1x + 0y = x$,

ii. $\alpha x + (1-\alpha)y = (1-\alpha)y + \alpha x$,

iii. $\lambda[\alpha x + (1-\alpha)y] + (1-\lambda)y = (\lambda\alpha)x + (1-\lambda\alpha)y$.

The set of lotteries form a mixture space under the interpretation that $\alpha x + (1-\alpha)y$ is a lottery yielding a ticket to play x with probability α and a ticket to play y with probability $1-\alpha$.

When is there a measurable utility u for a regular preference $\succcurlyeq$ on M? A utility satisfies

$$x \succcurlyeq y \iff u(x) \geqslant u(y)$$

and a **measurable utility** additionally satisfies

$$u(\alpha x + (1-\alpha)y) = \alpha u(x) + (1-\alpha)u(y).$$

5 Theorem (Herstein and Milnor [12]) *Let $\succcurlyeq$ be a regular preference on the mixture space M satisfying*

i. $\{\alpha : \alpha x + (1-\alpha)y \succcurlyeq z\}$ *and* $\{\alpha : \alpha x + (1-\alpha)y \preccurlyeq z\}$ *are closed.*

ii. *(Strong Independence) If $x \sim y$, then $\alpha x + (1-\alpha)z \sim \alpha y + (1-\alpha)z$ for all z and all α.*

Then $\succcurlyeq$ has a measurable utility.

8 Stochastic dominance and expected utility

In this section we consider lotteries over monetary prizes. Let $S = \{x_1 < \cdots < x_n\}$ be a finite set of money prizes. A **lottery** is a probability distribution over the prizes. Lotteries thus correspond to probability vectors in $\boldsymbol{R}^{\mathrm{n}}$. We say that q **stochastically dominates** p if for each $k = 1, \ldots, n-1$,

$$\sum_{i=k}^{n} q_i \geqslant \sum_{i=k}^{n} p_i,$$

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. Intuitively one should prefer a stochastically dominating lottery.

A utility on S can be thought of as vector u in $\boldsymbol{R}^{\mathrm{n}}$, where the j^{th} component is the utility of x_j. It is natural to demand in addition that $u_1 < \cdots < u_n$.

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

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

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

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

$$\begin{bmatrix} p_1-q_1 & p_2-q_2 & p_3-q_3 & \cdots & \cdots & \cdots & p_{n-1}-q_{n-1} & p_n-q_n \\ -1 & +1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & +1 & 0 & \ddots & & & 0 \\ 0 & 0 & -1 & +1 & 0 & & & 0 \\ \vdots & & 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & 0 & \vdots \\ 0 & & & & 0 & -1 & +1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & -1 & +1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ u_{n-1} \\ u_n \end{bmatrix} \gg 0.$$

Gordan's Alternative 11 assets that the the alternative is that there exists $y =$

$(y_0, y_1, \ldots, y_{n-1}) > 0$ and

$$\begin{array}{lcccccc}
y_0(p_1 - q_1) & & & - & y_1 & = & 0 \\
y_0(p_2 - q_2) & + & y_1 & - & y_2 & = & 0 \\
 & \vdots & & \vdots & & & \\
y_0(p_{n-1} - q_{n-1}) & + & y_{n-2} & - & y_{n-1} & = & 0 \\
y_0(p_n - q_n) & + & y_{n-1} & & & = & 0.
\end{array} \tag{1}$$

It is easy to see that $y_0 > 0$, for if $y_0 = 0$, then (1) implies everything unravels and the entire vector $y = 0$, a contradiction.

Write $x_i = \frac{y_i}{y_0} \geqslant 0$, $i = 1, \ldots, n-1$. Then

$$\begin{array}{ccccccc}
p_1 - q_1 & & & - & x_1 & = & 0 \\
p_2 - q_2 & + & x_1 & - & x_2 & = & 0 \\
 & \vdots & & & & \vdots & \\
p_{n-1} - q_{n-1} & + & x_{n-2} & - & x_{n-1} & = & 0 \\
p_n - q_n & + & x_{n-1} & & & = & 0.
\end{array}$$

Since $p \neq q$ we cannot have $x_1 = \cdots = x_{n-1} = 0$, so at least one $x_i > 0$.

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

$$\begin{array}{ccccccc}
p_n & - & q_n & = & -x_n & \leqslant & 0 \\
(p_{n-1} + p_n) & - & (q_{n-1} + q_n) & = & -x_{n-1} & \leqslant & 0 \\
 & \vdots & & & & \vdots & \\
\sum_{i=2}^{n-1} p_i & - & \sum_{i=2}^{n-1} q_i & = & -x_2 & \leqslant & 0
\end{array}$$

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

9 Stochastic dominance and expected utility, *deux*

This generalizes the preceding result to larger collections of vectors $p^0, p^1, \ldots p^m$. 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 $u \cdot p^0 > u \cdot \sum_{j=1}^{m} \lambda_j p^j = u \cdot p^0$.

7 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 $u_1 < \cdots < u_n$ such that p^0 has the highest expected utility, that is,*

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

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 [6], Ledyard [15], and Border [2]) Condition (i) is equivalent to the following matrix equation, with $m + n - 1$ rows and n columns.

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

Gordan's Alternative 11 assets that the alternative is that there is some semipositive $m+n-1$-vector

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

satisfying

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

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 6, 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. ∎

10 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:

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

11 The Allais paradox and stochastic dominance

The Allais paradox above presented subject 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 lottery chosen. 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

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

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.

We could alter say A_2 to be $A_2' = [\$5m, .1; \$1m, .89 + 2\varepsilon; \$0, .01 - 2\varepsilon]$ for some tiny $\varepsilon > 0$. Then if B_2 remained the choice, the compound lottery $[B_1, \frac{1}{2}; A_2', \frac{1}{2}]$ reduces to $C_2' = [\$5m, .05; \$1m, .50 + \varepsilon; \$0, .45 - \varepsilon]$, which strictly stochastically dominates C_1.

The next section shows that this is not an isolated case. It is based on Border [2] and Ledyard [15].

12 Stochastic dominance and expected utility, *trois*

Let $S = \{x_1 < \cdots < x_n\}$ be a finite set of money prizes. Let $B_1, \dots, B_m$ 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 $u_1 < u_2 < \cdots < u_n$ on S such that for each $i = 1, \dots, m$,

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

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.

8 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*

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

Proof: (cf. Ledyard [15] and Border [2]) Let's enumerate each B_i as $p^{i0}, \dots p^{im_i}$ where $p^{i0} = c(B_i)$. Create the matrix A with $\sum_{i=1}^{m} m_i + n - 1$ rows and n columns defined as follows.

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

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

Gordan's Alternative 11 assets that the alternative is that there is some semipositive $\sum_{i=1}^{m} m_i + n - 1$-vector

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

satisfying

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

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

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

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

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

Then as in the proof of Theorem 7, 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 $\sum_{i=1}^m \pi_i c(B_i)$.

A Theorems of the Alternative

The mathematical tools we shall use are presented here without proof. See Gale [9, Chapter 2] for proofs. Here is the notation I use for vector orders.

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

9 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*

$$Ax = b \tag{1}$$

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

$$\begin{aligned} pA &= 0 \\ p \cdot b &> 0. \end{aligned} \tag{2}$$

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

$$Ax > 0 \tag{3}$$

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

$$\begin{aligned} pA &= 0 \\ p &\gg 0. \end{aligned} \tag{4}$$

11 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*

$$Ax \gg 0. \tag{5}$$

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

$$\begin{aligned} pA &= 0 \\ p &> 0. \end{aligned} \tag{6}$$

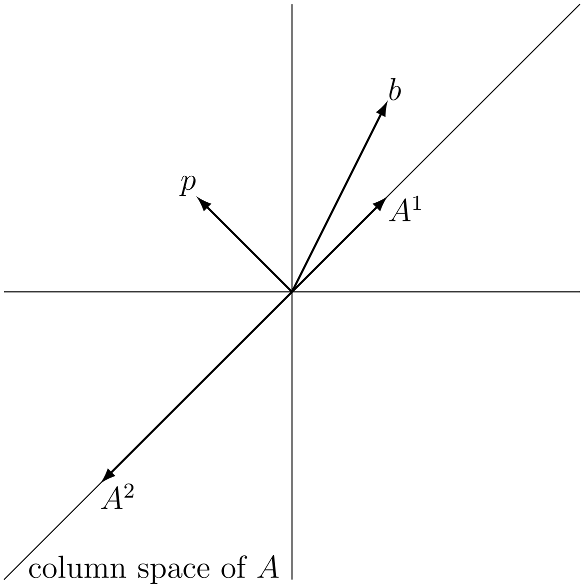

Figure 1. Geometry of the Fredholm Alternative

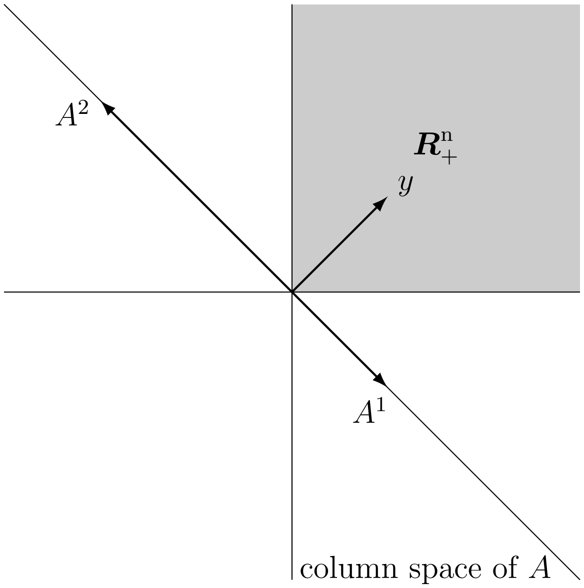

Figure 2. Geometry of the Stiemke Alternative

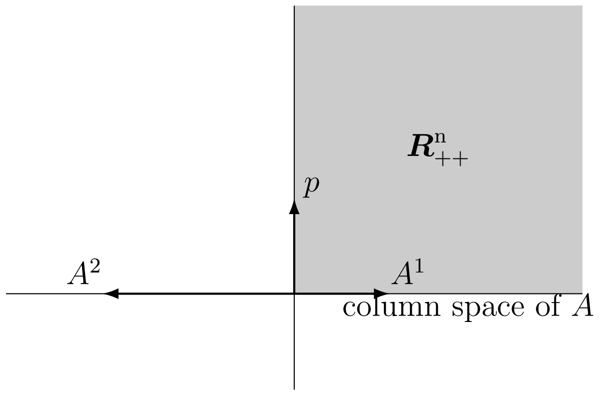

Figure 3. Geometry of the Gordan Alternative

References

[1] M. Allais. 1953. L'extension des theories de l'equilibre economique general et du rendement social au cas du risque. *Econometrica* 21(2):269–290. www.jstor.org/stable/1905539.

[2] K. C. Border. 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.

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

[4] ——— . 2002. *Secrets: A memoir of Vietnam and the Pentagon Papers*. New York: Penguin Books.

[5] B. de Finetti. 1940. Il problema dei 'pieni'. *Giornale dell'Istituto Italiano degli Attuari* 1:1–88.

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

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

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

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

[10] 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.

[11] 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. www.jstor.org/stable/2240227.

[12] I. N. Herstein and J. Milnor. 1953. An axiomatic approach to measurable utility. *Econometrica* 21:291–297. www.jstor.org/stable/1905540.

[13] B. O. Koopman. 1940. The axioms and algebra of intuitive probability. *Annals of Mathematics* 41(2):269–292. DOI: 10.2307/1969003.

[14] C. H. Kraft, J. W. Pratt, and A. Seidenberg. 1959. Intuitive probability on finite sets. *Annals of Mathematical Statistics* 30:408–419. www.jstor.org/stable/2237090.

[15] 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.

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

[17] D. Scott. 1964. Measurement structures and linear inequalities. *Journal of Mathematical Psychology* 1:233–247. DOI: 10.1016/0022-2496(64)90002-1.

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