Caltech Division of the Humanities and Social Sciences

Aspects of normative decision theory

KC Border

Spring 1984; Revised Fall 2013
v. 2016.05.04::13.25

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 Blaise Pascal, Pierre de Fermat, and Jacob and Daniel Bernoulli for games involving randomizing devices such as cards, dice, and roulette wheels. 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 L. J. 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 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*. As usual, we write $E \succ F$ to mean $E \succcurlyeq F$ & $F \not\succcurlyeq E$, and $E \sim F$ to mean $E \succcurlyeq F$ & $F \succcurlyeq E$. The **graph** of $\succcurlyeq$ is
$$\operatorname{gr} \succcurlyeq = \{(E, F) : E \succcurlyeq F\}.$$

Let us say that the subjective likelihood relation $\succcurlyeq$ is **represented by a probability measure $\boldsymbol{p}$** if
$$E \succcurlyeq F \iff p(E) \geqslant p(F).$$

Savage [21, p. 32] calls such subjective likelihood relation a **qualitative probability** if it satisfies the following obvious necessary conditions 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) $S \succ \varnothing$, and for every event E, $E \succcurlyeq \varnothing$.

Bruno de Finetti [7] posed the question of whether these conditions were sufficient to guarantee that $\succcurlyeq$ was representable by a probability. The following example due to Kraft, Pratt, and Seidenberg [17] shows that is not the case. (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

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

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

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

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, due to Dana Scott [22, Theorem 4.1] gives a friendlier set-theoretic statement. I have replaced Scott's condition (4_B) with a similar condition that is perhaps more transparent. I refer to it as Condition **S**, but there should be a better name. The proof is also mine.

2 Theorem *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,*

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

it is necessary and sufficient that $\succcurlyeq$ satisfy the following three conditions:

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

C *(Completeness) For all $E, F \in \mathcal{E}$, either $E \succcurlyeq F$ or $F \succcurlyeq E$, or both. (Or equivalently, for all $E, F \in \mathcal{E}$, exactly one of $E \succ F$, $F \succ E$, or $E \sim F$ holds.)*

S *(Condition **S**) For every finite list $(E_1, F_1), \dots, (E_n, F_n)$ of pairs of events (where repetitions are allowed),*

$$\left[(E_i \succcurlyeq F_i,\ i = 1, \dots, n) \ \&\ \sum_{i=1}^{n} \mathbf{1}_{E_i} = \sum_{i=1}^{n} \mathbf{1}_{F_i} \right] \implies E_i \sim F_i,\ i = 1, \dots, n.$$

Proof: ($\Longrightarrow$) Assume that $\succcurlyeq$ is representable by p. Then it is obvious that Nontriviality and Completeness must be satisfied.

To see that Condition **S** is also necessary, 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 $\sum_{i=1}^{n} \mathbf{1}_{E_i}(s)$ is the count of the events $E_1, \dots, E_n$ that contain s. Also observe that for any event E,

$$p(E) = \sum_{s \in E} p(s) = \sum_{s \in S} p(s) \mathbf{1}_E(s).$$

Thus for events $E_1, \dots, E_n$, we have

$$\sum_{i=1}^{n} p(E_i) = \sum_{i=1}^{n} \left(\sum_{s \in S} p(s) \mathbf{1}_{E_i}(s) \right) = \sum_{s \in S} p(s) \left(\sum_{i=1}^{n} \mathbf{1}_{E_i}(s) \right). \tag{2}$$

In other words, the function $\sum_{i=1}^{n} \mathbf{1}_{E_i}$ is a random variable whose expected value is the sum of probabilities $\sum_{i=1}^{n} p(E_i)$.

Let $(E_1, F_1), \dots, (E_n, F_n)$ be a list of pairs of events satisfying (i) $E_i \succcurlyeq F_i$, $i = 1, \dots, n$ and (ii) $\sum_{i=1}^{n} \mathbf{1}_{E_i} = \sum_{i=1}^{n} \mathbf{1}_{F_i}$. By (ii) and (2), we have that

$$\sum_{i=1}^{n} p(E_i) = \sum_{i=1}^{n} p(F_i).$$

From (i), we have $p(E_i) \geqslant p(F_i)$ for each i. Therefore we must actually have $p(E_i) = p(F_i)$, or $E_i \sim F_i$, for each i.

($\Longleftarrow$) We prove the converse by proving its contrapositive. That is, we shall show that if $\succcurlyeq$ is not representable, but satisfies Nontriviality and Completeness, then it must violate Condition **S**.

Consider the following system of inequalities, where the rows of the first matrix are indexed by the graph of $\succ$ and rows are of the second matrix are indexed by the graph of $\succcurlyeq$, and the columns are indexed by the states S.

$$\begin{array}{c}
\begin{array}{cc} & s \\ E \succ F & \begin{bmatrix} & \vdots & \\ \cdots & \mathbf{1}_E(s) - \mathbf{1}_F(s) & \cdots \\ & \vdots & \end{bmatrix} \end{array} \begin{bmatrix} \vdots \\ p(s) \\ \vdots \end{bmatrix} \gg 0 \\[2ex]
\begin{array}{cc} & s \\ E \succcurlyeq F & \begin{bmatrix} & \vdots & \\ \cdots & \mathbf{1}_E(s) - \mathbf{1}_F(s) & \cdots \\ & \vdots & \end{bmatrix} \end{array} \begin{bmatrix} \vdots \\ p(s) \\ \vdots \end{bmatrix} \geqq 0
\end{array} \tag{3}$$

If the system (3) has a solution p, then the row corresponding to $\{s\} \succcurlyeq \varnothing$ implies $p(s) \geqslant 0$. The row corresponding to $S \succ \varnothing$ implies $\sum_{s \in S} p(s) > 0$. We may normalize p so that it is indeed a probability measure on S. Thus $\succcurlyeq$ is representable if and only if (3) has a solution. We now show that if no solution exists, then Condition **S** is violated.

So suppose (3) does not have a solution. Then by Motzkin's Rational Transposition Theorem 17 there exist integer-valued nonnegative vectors $k^{\succ}$ (indexed by the graph of $\succ$) and $k^{\succcurlyeq}$ (indexed by the graph of $\succcurlyeq$) such that for each column $s \in S$,

$$\sum_{(E,F):E \succ F} k^{\succ}_{(E,F)} \big(\mathbf{1}_E(s) - \mathbf{1}_F(s)\big) + \sum_{(E,F):E \succcurlyeq F} k^{\succcurlyeq}_{(E,F)} \big(\mathbf{1}_E(s) - \mathbf{1}_F(s)\big) = 0. \tag{4}$$

Moreover, Motzkin's Theorem guarantees that $k^{\succ}$ is nonzero.

Construct a list of pairs by taking $k^{\succ}_{(E,F)}$ copies of (E, F) for each (E, F) with $E \succ F$ and $k^{\succcurlyeq}_{(E,F)}$ copies of (E, F) for (E, F) with $E \succcurlyeq F$, and enumerate it as $(E_1, F_1), \dots, (E_n, F_n)$.

By construction, for each (E_i, F_i), we have $E_i \succcurlyeq F_i$ and by (4) we have

$$\sum_{i=1}^{n} \mathbf{1}_{E_i} = \sum_{i=1}^{n} \mathbf{1}_{F_i}.$$

But since $k^{\succ}$ is nonzero, for at least one pair we have $E_i \succ F_i$, which violates Condition **S**.

This completes the proof. ∎

3 Remark Note that Condition **S** and Completeness imply Transitivity. We proceed by contraposition. Assume Completeness and that Transitivity fails. That is, there are A, B, C with

$A \succcurlyeq B$, $B \succcurlyeq C$, and $C \succ A$. Set

$$E_1 = A, \qquad F_1 = B,$$

$$E_2 = B, \qquad F_2 = C,$$

$$E_3 = C, \qquad F_3 = A.$$

Then $E_i \succcurlyeq F_i$ for all i and

$$\sum_{i=1}^{3} \mathbf{1}_{E_i} = \sum_{i=1}^{3} \mathbf{1}_{F_i} = \mathbf{1}_A + \mathbf{1}_B + \mathbf{1}_C.$$

But $E_3 \succ F_3$, which violates Condition **S**.

4 Remark Now let's see that Condition **S** and Completeness imply Additivity. So assume $A \cap C = \varnothing$ and $C \cap C = \varnothing$, then we want to show that

$$A \succcurlyeq B \quad \iff \quad A \cup C \succcurlyeq B \cup C.$$

First assume $A \succcurlyeq B$, and suppose $A \cup C \succcurlyeq B \cup C$ fails. Then $B \cup C \succ A \cup C$. Define

$$E_1 = A, \qquad F_1 = B,$$

$$E_2 = B \cup C, \qquad F_2 = A \cup C.$$

Since $A \cap C = \varnothing$ we have that $\mathbf{1}_{A \cup C} = \mathbf{1}_A + \mathbf{1}_C$. Similarly, $\mathbf{1}_{B \cup C} = \mathbf{1}_B + \mathbf{1}_C$. So now observe that

$$\sum_{i=1}^{2} \mathbf{1}_{E_i} = \mathbf{1}_A + \mathbf{1}_B + \mathbf{1}_C = \mathbf{1}_B + \mathbf{1}_A + \mathbf{1}_C = \sum_{i=1}^{2} \mathbf{1}_{F_i}.$$

This violates Condition **S**.

For the converse, assume $A \cup C \succcurlyeq B \cup C$, but that $A \succcurlyeq B$ fails, so that $B \succ C$ and define

$$E_1 = A \cup C, \qquad F_1 = B \cup C,$$

$$E_2 = B, \qquad F_2 = C.$$

This violates Condition **S**.

This finishes the proof of Additivity.

5 Remark We now note that the Kraft–Pratt–Seidenberg example violates Condition **S**. The following list of pairs will do. (These are the same pair we used above to show that the relation was not representable.)

$$E_1 = \{a\}, \qquad F_1 = \{b, d\},$$

$$E_2 = \{c, d\}, \qquad F_2 = \{a, b\}.$$

$$E_3 = \{b, e\}, \qquad F_3 = \{a, d\}.$$

$$E_4 = \{a, b, d\}, \qquad F_4 = \{c, e\}.$$

6 Remark I mentioned above that what I call Condition **S** is not Condition (4_B) of his Theorem 4.1, [22, p.246]. In the notation of this note, condition (4_B) is:

For every finite list $(E_0, F_0), \dots, (E_n, F_n)$ of pairs of events (where repetitions are allowed),

$$\left[(E_i \succcurlyeq F_i,\ i = 1, \dots, n) \ \& \ \sum_{i=0}^{n} \mathbf{1}_{E_i} = \sum_{i=0}^{n} \mathbf{1}_{F_i} \right] \implies F_0 \succcurlyeq E_0.$$

(Pay attention to the fact that his indices run from 1 to n in one place and from 0 to n in another place.)

My Condition **S** does not imply the conclusion $F_0 \succcurlyeq E_0$ in the situation described—it only implies the weaker $E_0 \not\succ F_0$. But in the presence of Completeness, $F_0 \succcurlyeq E_0$ is equivalent to $E_0 \not\succ F_0$.

4 Subjective probability and betting

This section is based on de Finetti [6] as exposited by Heath and Sudderth [14].

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

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

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]

7 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: Let $x(E)$ denote the number of E-tickets bought by the Bettor. Condition (ii) is equivalent to the matrix inequality

$$\begin{array}{cc} & E \\ s & \begin{bmatrix} & \vdots & \\ \cdots & \mathbf{1}_E(s) - q(E) & \cdots \\ & \vdots & \end{bmatrix} \end{array} \begin{bmatrix} \vdots \\ x(E) \\ \vdots \end{bmatrix} \gg 0$$

where rows are indexed by S and columns are indexed by $\mathcal{E}$.

Gordan's Alternative 15 asserts that the alternative to (ii) 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 (i). ∎

[1] Young people think an E-ticket is something that lets you board an airplane, but older Southern Californians remember when it let you board the Matterhorn.

5 The Ellsberg Paradox

Daniel Ellsberg [4] (of *Pentagon Papers* [5] 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.

[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 [4, pp. 655–656].

6 Statisticians' view of the world

The statistical view of the world can be caricatured as follows: Θ is a finite set of urns, each urn θ describes a probability p_θ on the finite set S of *samples*. A particular urn θ_0 is used to choose a sample $s \in S$ according to probability p_{θ_0}. We observe the sample $s \in S$. 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

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

provided $p(F) > 0$. 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

The Bayesian approach to statistics turns Θ into a probability space, with an algebra $\mathcal{T}$ of events, and a probability measure P on $\mathcal{T}$, called the **prior**. The prior represents our beliefs about the urn θ_0 that is used to select the sample select s according to p_{θ_0}. Then the **posterior probability** that $\theta_0 \in T$, given sample s, is calculated according to Bayes' Law as

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

$P(\cdot|s)$ represents our belief about θ_0 after the experiment.

Should Bayes' Law govern our betting behavior regarding θ_0? That is, should we use it to set prices for lottery tickets?

Statistical inference: the game

Freedman and Purves [11] caricature statistical inference in terms of the following game.

1. The Master of Ceremonies chooses an urn θ_0 from a set Θ of urns, draws a sample s from the urn, and exhibits the sample to the Bettor and the Bookie.

[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 prices for lottery tickets corresponding to subsets $T \in \mathcal{T}$ of Θ.

3. The Bettor buys a portfolio of lottery tickets.

4. The MC reveals the urn θ_0, and the tickets are payed off.

The reason this is a caricature is that in the real world of statistical inference, there is never an MC to reveal θ_0.

Strategies

The Bettor and the Bookie choose their strategies for the game in advance of playing, so they must decide what to do for each possible sample that could be observed.

The Bookie chooses $q \geqq 0 \in \boldsymbol{R}^{\mathcal{T}\times S}$. For each $s \in S$ and $T \in \mathcal{T}$, $q(T,s)$ is the price he sets for a lottery ticket that pays \$1 if the urn belongs to the set T, after having seen the sample s.

Then Bettor then chooses $x \in \boldsymbol{R}^{\mathcal{T}\times S}$, and buys $x(T,s)$ T-tickets, after having the seen the sample s.

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

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

8 Bayesian updating theorem *Either*
(i) *The Bookie chooses some prior P on Θ and sets prices 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 by the MC.*

Note that this result does not say that the MC actually selected the urn at random according to P—it is merely a device to calculate the prices to avoid (ii).

Proof: Condition (ii) is equivalent to the matrix inequality

$$\begin{array}{c c}
 & (T,s) \\
\theta & \begin{bmatrix} & \vdots & \\ \cdots & (\mathbf{1}_T(\theta)-q(T,s))p_\theta(s) & \cdots \\ & \vdots & \end{bmatrix}
\end{array}
\begin{bmatrix} \vdots \\ x(T,s) \\ \vdots \end{bmatrix} \gg 0,$$

where rows are indexed by Θ and columns are indexed by $\mathcal{T}\times S$.

Gordan's Alternative 15 asserts that the alternative to (ii) is the existence of a probability vector $P \in \boldsymbol{R}^{\Theta}$ such that for each column $(T,s) \in \mathcal{T}\times S$,

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

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 $1-\alpha$.

A utility satisfies

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

and a **measurable utility** or a **von Neumann–Morgenstern utility** additionally satisfies

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

9 Theorem (Herstein and Milnor [15]) *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.

The Independence axiom is often used as a normative justification for measurable, utility, but it is not clear why a decision maker *ought* to want to obey it.

8 Stochastic dominance and expected utility

In this section we consider lotteries over monetary prizes. Let me abuse notation and use S to denote both the set and number of prizes, so $S = \{x_1 < \cdots < x_S\}$ is a finite set of money prizes. A **lottery** is a probability distribution over the prizes. Lotteries thus correspond to probability vectors in $\boldsymbol{R}^{\mathrm{S}}$. We say that q **stochastically dominates** p if for each $k = 0, \ldots, S-1$,

$$\sum_{s=S-k}^{S} q_s \geqslant \sum_{s=S-k}^{S} p_s,$$

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 the set of the k largest prizes. Intuitively, one should prefer a stochastically dominating lottery, (assuming larger prizes are better).

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

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

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

$$\sum_{s=1}^{S} u_s p_s > \sum_{s=1}^{S} u_s q_s$$

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

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

Gordan's Alternative 15 asserts that the alternative is that there exists a vector $y = (y_0, y_1, \dots, y_{S-1}) > 0$ such that

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

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 $x_s = \frac{y_s}{y_0} \geqslant 0$, $i = 1, \dots, S-1$. Then

$$\begin{array}{cccccccc} p_1 - q_1 & & & - & x_1 & = & 0 \\ p_2 - q_2 & + & x_1 & - & x_2 & = & 0 \\ p_3 - q_3 & + & x_2 & - & x_3 & = & 0 \\ & \vdots & & & & \vdots & \\ p_{S-1} - q_{S-1} & + & x_{S-2} & - & x_{S-1} & = & 0 \\ p_S - q_S & + & x_{S-1} & & & = & 0. \end{array} \tag{1'}$$

Start from the end, and add up the last k inequalities to get

$$\begin{array}{ccccccc} p_S & - & q_S & = & -x_{S-1} & \leqslant & 0 \\ (p_{S-1} + p_S) & - & (q_{S-1} + q_S) & = & -x_{S-2} & \leqslant & 0 \\ \vdots & & & & \vdots & & \\ \sum_{s=1}^{S} p_s & - & \sum_{s=1}^{S} q_s & = & -x_1 & \leqslant & 0 \end{array}$$

which asserts that q stochastically dominates p. ∎

9 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 $u \cdot p^0 > u \cdot \sum_{j=1}^m \lambda_j p^j = u \cdot p^0$.

11 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,*

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

or else

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

$$\sum_{j=1}^{m} \pi_j p^j \text{ stochastically dominates } p^0.$$

Proof: (cf. Fishburn [9], Ledyard [18], and Border [3]) Condition (i) is equivalent to the following matrix equation, with $m+S-1$ rows and S columns.

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

Gordan's Alternative 15 asserts 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

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

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

Then just as in the proof of Theorem 10, we see that $\sum_{j=1}^m \pi_j p^j$ is either equal to or stochastically dominates $\sum_{j=1}^m \pi_j p^0 = p^0$. But our extremity hypothesis rules out their equality. That is, condition (ii) holds. ∎

10 Allais Paradox

This example is due more-or-less to Maurice 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 people report $B_2 \succ A_2$ and $A_1 \succ B_1$, which violates EUH:

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

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 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 [3] and Ledyard [18].

12 Stochastic dominance and expected utility, *trois*

Let $S = \{x_1 < \cdots < x_S\}$ be a finite set of money prizes. Let $B_1, \ldots, B_K$ be **lottery budgets**. That is, each B_k is a finite set $\{p^{k0}, \ldots, p^{km_k}\}$ of $m_k + 1$ 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 strictly increasing utility function $u_1 < u_2 < \cdots < u_S$ on S such that for each $k = 1, \ldots, K$,

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

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 a budget B_k a mixture (convex combination) $\sum_{j=0}^{m_k} \lambda_{kj} p^{kj}$ of the elements of B_k. (For each k, we have $\sum_{j=0}^{m_k} \lambda_{kj} = 1$.)

12 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 $d(B_k)$ does not put any weight on $c(B_k)$ for each k, such that the mixture*

$$\sum_{k=1}^{K} \pi_k d(B_k) \textit{ stochastically dominates or equals } \sum_{k=1}^{K} \pi_k c(B_k).$$

Proof: (cf. Ledyard [18] and Border [3]) Assume without loss of generality that p^{k0} is the choice for B_k. Create the matrix A with $\sum_{k=1}^{K} m_k + S - 1$ rows and S columns defined as follows.

$$
\left[\begin{array}{cccccccc}
p_1^{10}-p_1^{11} & p_2^{10}-p_2^{11} & p_3^{10}-p_3^{11} & \dots & \dots & \dots & p_{S-1}^{10}-p_{S-1}^{11} & p_S^{10}-p_S^{11} \\
p_1^{10}-p_1^{12} & p_2^{10}-p_2^{12} & p_3^{10}-p_3^{12} & \dots & \dots & \dots & p_{S-1}^{10}-p_{S-1}^{12} & p_S^{10}-p_S^{20} \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
p_1^{10}-p_1^{1m_1} & p_2^{10}-p_2^{1m_1} & p_3^{10}-p_3^{1m_1} & \dots & \dots & \dots & p_{S-1}^{10}-p_{S-1}^{1m_1} & p_S^{10}-p_S^{1m_1} \\
\hline
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
\hline
p_1^{K0}-p_1^{K1} & p_2^{K0}-p_2^{K1} & p_3^{K0}-p_3^{K1} & \dots & \dots & \dots & p_{S-1}^{K0}-p_{S-1}^{K1} & p_S^{K0}-p_S^{K1} \\
p_1^{K0}-p_1^{K2} & p_2^{K0}-p_2^{K2} & p_3^{K0}-p_3^{K2} & \dots & \dots & \dots & p_{S-1}^{K0}-p_{S-1}^{K2} & p_S^{K0}-p_S^{20} \\
\vdots & \vdots & \vdots & & & & \vdots & \vdots \\
p_1^{K0}-p_1^{Km_K} & p_2^{K0}-p_2^{Km_K} & p_3^{K0}-p_3^{Km_K} & \dots & \dots & \dots & p_{S-1}^{K0}-p_{S-1}^{Km_K} & p_S^{K0}-p_S^{Km_K} \\
\hline
-1 & +1 & 0 & 0 & 0 & \dots & 0 & 0 \\
0 & -1 & +1 & 0 & \ddots & & & 0 \\
0 & 0 & -1 & +1 & 0 & & & 0 \\
\vdots & & 0 & \ddots & \ddots & \ddots & & \vdots \\
\vdots & & & \ddots & \ddots & \ddots & 0 & \vdots \\
0 & & & & 0 & -1 & +1 & 0 \\
0 & 0 & \dots & \dots & 0 & 0 & -1 & +1
\end{array}\right]
$$

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

Gordan's Alternative 15 asserts that the alternative is that there is some semipositive $\sum_{k=1}^{K} m_k + S - 1$-vector

$$(z, y) = (z_{11}, \dots, z_{1m_1}, \dots, z_{K1}, \dots, z_{Km_K}, y_1, \dots, y_{S-1}) > 0$$

satisfying

$$
\begin{array}{lcccccc}
\sum_{k=1}^{K} \sum_{j=1}^{m_k} z_{kj}(p_1^{k0} - p_1^{kj}) & & & - & y_1 & = & 0 \\
\sum_{k=1}^{K} \sum_{j=1}^{m_k} z_{kj}(p_2^{k0} - p_2^{kj}) & + & y_1 & - & y_2 & = & 0 \\
 & \vdots & & \vdots & & & \\
\sum_{k=1}^{K} \sum_{j=1}^{m_k} z_{kj}(p_{S-1}^{k0} - p_{S-1}^{kj}) & + & y_{S-2} & - & y_{S-1} & = & 0 \\
\sum_{k=1}^{K} \sum_{j=1}^{m_k} z_{kj}(p_S^{k0} - p_S^{kj}) & + & y_{S-1} & & & = & 0.
\end{array}
$$

It is easy to see that $\sum_{k=1}^K \sum_{j=1}^{m_k} z_{kj} > 0$, otherwise everything unravels, so $(z, y) = 0$, a contradiction. Therefore we may renormalize and assume that $\sum_{k=1}^K \sum_{j=1}^{m_k} z_{kj} = 1$. Now for each k set

$$\pi_k = \sum_{j=1}^{m_k} z_{kj} \qquad k = 1, \dots, K$$

and

$$\lambda_{kj} = \begin{cases} \frac{z_{kj}}{\pi_k} & \pi_k > 0 \\ 0 & \pi_k = 0, \end{cases}$$

so $\sum_{k=1}^K \sum_{j=1}^{m_k} z_{kj} = \sum_{k=1}^K \pi_k \sum_{j=1}^{m_k} \lambda_{kj}$ and for each k with $\pi_k > 0$, we have $\sum_{j=1}^{m_k} \lambda_{kj} = 1$. Define the random choice d by

$$d(B_k) = \sum_{j=1}^{m_k} \lambda_{kj} p^{kj}, \qquad k = 1, \dots, K.$$

Then just as in the proof of Theorem 11, we see that $\sum_{k=1}^K \pi_k d(B_k)$ stochastically dominates or equals $\sum_{k=1}^K \pi_k p^{k0} = \sum_{k=1}^K \pi_k c(B_k)$. ∎

I assert without proof that if $\sum_{k=1}^K \pi_k d(B_k) = \sum_{k=1}^K \pi_k c(B_k)$, then an arbitrarily small perturbation of the p^{kj}s will lead to $\sum_{k=1}^K \pi_k d(B_k)$ strictly dominating $\sum_{k=1}^K \pi_k c(B_k)$.

A Theorems of the Alternative

The mathematical tools we shall use are presented here without proof. See Gale [12, Chapter 2] or my on-line notes for proofs. Here is the notation I use for vector orders.

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

I call the next result Fredholm's Alternative, as Fredholm [10] contains a version of it.

13 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

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

The next result is due to Stiemke [23].

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

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

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

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

A different version of the alternative is due to Gordan [13].

15 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

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

Another alternative is Motzkin's Transposition Theorem [19], proven in his 1934 Ph.D. thesis. This statement is taken from his 1951 paper [20].[4]

[4] Motzkin [20] contains an unfortunate typo. The condition $Ax \gg 0$ is erroneously given as $Ax \ll 0$.

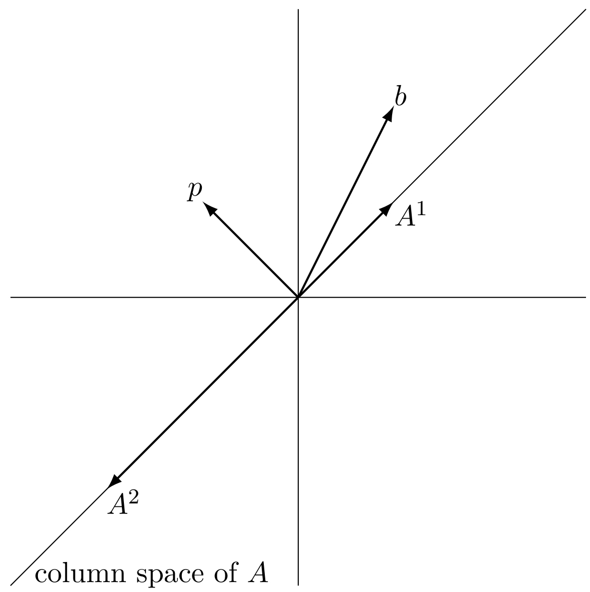

Figure 1. Geometry of the Fredholm Alternative

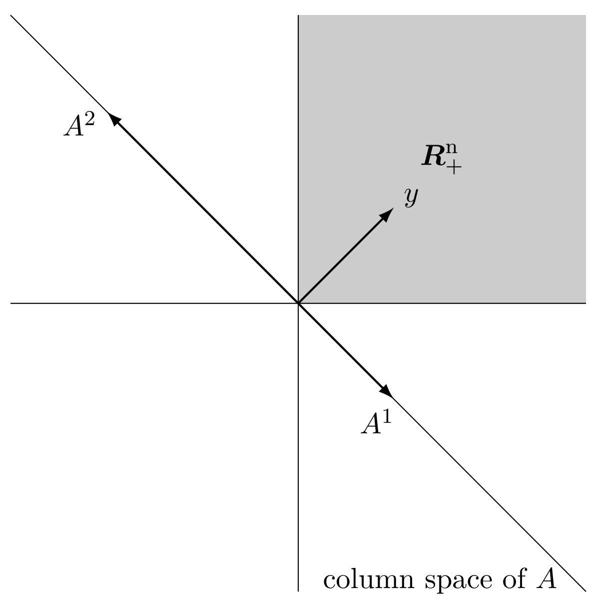

Figure 2. Geometry of the Stiemke Alternative

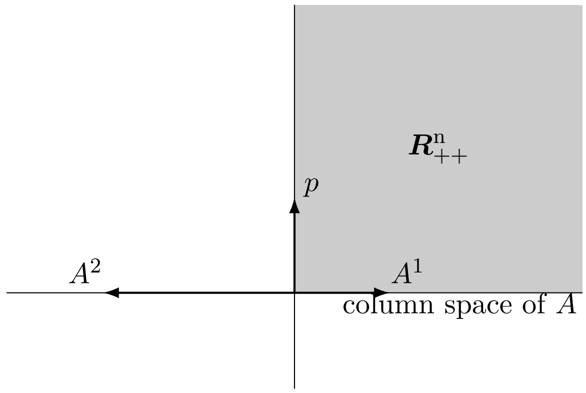

Figure 3. Geometry of the Gordan Alternative

16 Motzkin's Transposition Theorem *Let A be an $m \times n$ matrix, let B be an $\ell \times n$ matrix, and let C be an $r \times n$ matrix, where B or C may be omitted (but not A). Exactly one of the following alternatives holds. Either there exists $x \in \boldsymbol{R}^{\mathrm{n}}$ satisfying*

$$\begin{aligned} Ax &\gg 0 \\ Bx &\geqq 0 \\ Cx &= 0 \end{aligned} \tag{7}$$

or else there exist $p^A \in \boldsymbol{R}^{\mathrm{m}}$, $p^B \in \boldsymbol{R}^{\ell}$, and $p^C \in \boldsymbol{R}^{\mathrm{r}}$ satisfying

$$\begin{aligned} p^A A + p^B B + p^C C &= 0 \\ p^A &> 0 \\ p^B &\geqq 0. \end{aligned} \tag{8}$$

Motzkin expressed (8) in terms of the transpositions of A, B, and C. The reason the matrix A may not be omitted is that without A, the vectors $x = 0$, $p^B = 0$, $p^C = 0$ solve both systems (7) and (8). Note that Gordan's Alternative is the case of Motzkin's Theorem where B and C are both omitted. Stoer and Witzgall [24, Theorem 1.4.4, p. 18] and Fishburn [8, Theorem 3.2, pp. 31–32] also provide a rational version of Motzkin's theorem, which can be used to prove the following result.[5]

17 Motzkin's Rational Transposition Theorem *Let A be an $m \times n$ rational matrix, let B be an $\ell \times n$ rational matrix, and let C be an $r \times n$ rational matrix, where B or C may be omitted (but not A). Exactly one of the following alternatives holds. Either there exists $x \in \boldsymbol{R}^{\mathrm{n}}$*

[5]Stoer and Witzgall [24] use terminology that makes it difficult to realize that Theorem 1.4.4 implies what follows. It is also stated for a general commutative ordered field, which they denote by R, which is not to be confused with the real numbers $\boldsymbol{R}$. Both the real field and the rational field are covered by their results. Fishburn [8] presents an all-integer version of Motzkin's Theorem. The integer version can derived from the rational version by multiplying by a common denominator.

satisfying

$$\begin{aligned} Ax &\gg 0 \\ Bx &\geqq 0 \\ Cx &= 0 \end{aligned} \tag{9}$$

or else there exist $p^A \in \mathbb{Z}^{\mathrm{m}}$*,* $p^B \in \mathbb{Z}^{\ell}$*, and* $p^C \in \mathbb{Z}^{\mathrm{r}}$ *satisfying*

$$\begin{aligned} p^A A + p^B B + p^C C &= 0 \\ p^A &> 0 \\ p^B &\geqq 0. \end{aligned} \tag{10}$$

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. http://www.jstor.org/stable/1905539

[2] K. C. Border. 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/Ec181/pdf/Border1991-Dominance.pdf

[3] ——— . 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

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

[5] ——— . 2002. *Secrets: A memoir of Vietnam and the Pentagon Papers*. New York: Penguin Books.

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

[7] ——— . 1951. La 'logico del plausible' secondo la concezione di Polya. In *Atti della XLII Riunione della Società Italiana per il Progresso delle Scienze (Novembre 1949)*, pages 1–10. Rome: Società Italiana per il Progresso delle Scienze.

[8] P. C. Fishburn. 1973. *The theory of social choice*. Princeton: Princeton University Press.

[9] ——— . 1975. Separation theorems and expected utility. *Journal of Economic Theory* 11:16–34. DOI: 10.1016/0022-0531(75)90036-8

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

[11] 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

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

[13] 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

[14] 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

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

[16] B. O. Koopman. 1940. The axioms and algebra of intuitive probability. *Annals of Mathematics* 41(2):269–292. http://www.jstor.org/stable/1969003.pdf

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

[18] 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

[19] T. S. Motzkin. 1934. *Beiträge zur Theorie der linearen Ungleichungen*. PhD thesis, Universität Basel. It appears from citations I have seen that this was published in Jerusalem in 1936. Motzkin [20] gives 1934 as the date.

[20] ——— . 1951. Two consequences of the transposition theorem on linear inequalities. *Econometrica* 19(2):184–185. http://www.jstor.org/stable/1905733

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

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

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

[24] J. Stoer and C. Witzgall. 1970. *Convexity and optimization in finite dimensions I*. Number 163 in Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berüsichtigung der Anwendungsgebiete. Berlin: Springer–Verlag.