Caltech Division of the Humanities and Social Sciences

Ec 121a
Theory of Value

KC Border
Fall 2020

Lecture 16: Key ideas in decision theory

16.1 Basics

The modern approach to uncertainty, as formalized by Kolmogorov [18], 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 usually assume that S is finite.

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

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

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

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

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

16.2 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.

16.3 Subjective probability and betting

16.3.1 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). ∎

16.4 Statisticians' view of the world

Statisticians take a different approach from Kolmogorov. There is still a space S of *samples* or *signals*, but the a state of the world is not a sample, but rather a probability distribution on the set of samples. (In other words, statisticians believe that God does nothing but play dice.) The set of states of the world to a statistician is a set Θ of *hypotheses*. We can think of $Theta$ as a set of urns, each urn θ describes a probability p_θ on S. A particular urn θ_0 is used to choose 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?

16.4.1 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**.

16.4.2 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.

16.4.3 Statistical inference: the game

Freedman and Purves [13] describe statistical inference in terms of the following game.

The Master of Ceremonies chooses an urn, and announces the sample 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.)

16.4.4 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

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

16.4.1 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). ∎

16.5 The Ellsberg Paradox

Daniel Ellsberg [5] (of *Pentagon Papers* [6] 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 strict preference for urn A in each deal. Reportedly these included Jimmy Savage.[1] 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

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

Strictly preferring 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] 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).

16.6 Expected utility hypothesis

The standard model of choice over random variables 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 $\boldsymbol{E}\, u(X) \geqslant \boldsymbol{E}\, u(Y)$.

Two Bernoulli utilities u and v represent the same preference ranking 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.*

16.7 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:

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

16.8 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 $k = 1, \dots, n-1$,

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

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 $u_1 < \cdots < u_n$.

16.8.1 Expected utility theorem *Suppose p and q are distinct probability vectors in $\boldsymbol{R}^n$ (where $n \geqslant 2$). 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 the existence of $u_1, \ldots, u_n$ satisfying

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

Gordan's Alternative 16.13.2 asserts that the alternative is that there exists $y = (y_0, y_1, \ldots, y_{n-1}) > 0$ and

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

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.

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

$$\begin{aligned} p_1 - q_1 & & - y_1 &= 0 \\ p_2 - q_2 &+ y_1 & - y_2 &= 0 \\ &\vdots & \vdots & \\ p_{n-1} - q_{n-1} &+ y_{n-2} & - y_{n-1} &= 0 \\ p_n - q_n &+ y_{n-1} & &= 0. \end{aligned}$$

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

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

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

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

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

16.9.1 Theorem *Let $p^0, p^1, \ldots, p^m$ 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,*

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

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 [9], Ledyard [19], and Border [2]) Condition (i) is equivalent to the following matrix equation, with $m+n-1$ rows and n columns.

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

Gordan's Alternative 16.13.2 assets that the alternative is that there is some semipositive $m+n-1$-vector

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

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 16.8.1, 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. ∎

16.10 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

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

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 [2] and Ledyard [19].

16.11 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.

16.11.1 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 [19] 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.

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

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

Gordan's Alternative 16.13.2 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

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

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 16.9.1, 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 $\sum_{i=1}^m \pi_i c(B_i)$.

16.12 Appendix: Random variables

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

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

is often used to describe events involving X.

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.

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

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

The expectation has the following properties:

$$\boldsymbol{E}(\boldsymbol{E}\, X) = \boldsymbol{E}\, X$$
$$\boldsymbol{E}(aX + bY) = a\, \boldsymbol{E}\, X + b\, \boldsymbol{E}\, Y$$
$$\boldsymbol{E}(X - \boldsymbol{E}\, X) = 0$$
$$X \geqslant Y \implies \boldsymbol{E}\, X \geqslant \boldsymbol{E}\, Y$$

16.12.1 Definition *A set $\{X_i\}$ of random variables is* ***stochastically independent*** *(or simply* ***independent****) if for distinct i and j,*

$$\boldsymbol{E}(X_i X_j) = (\boldsymbol{E}\, X_i)(\boldsymbol{E}\, X_j).$$

A collection of events is independent if their indicator functions are independent. (The ***indicator function*** *$\mathbf{1}_E$ of E is defined by*

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

16.12.2 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,

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

Now let $b \to \infty$ and use the Monotone Convergence Theorem for the unbounded case. The general conclusion (without differentiability of F) uses a more sophisticated integration by parts [3] result based on Fubini's Theorem. ∎

16.12.3 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 $\boldsymbol{E}(X - \boldsymbol{E}\, X)(Y - \boldsymbol{E}\, Y)$.

16.13 Appendix: Theorems of the Alternative

The mathematical tools we shall use are presented here without proof. See Gale [14, Chapter 2] for proofs. Here is the notation I use for vector orders.

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

16.13.1 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*

$$Ax = b \tag{2}$$

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

$$\begin{aligned} pA &= 0 \\ p \cdot b &> 0. \end{aligned} \tag{3}$$

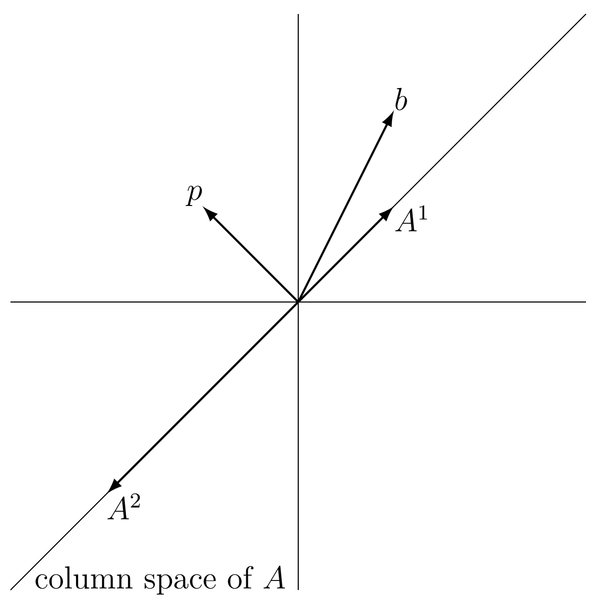

Figure 16.1. Geometry of the Fredholm Alternative

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

$$Ax \gg 0. \tag{4}$$

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

$$\begin{aligned} pA &= 0 \\ p &> 0 \end{aligned} \tag{5}$$

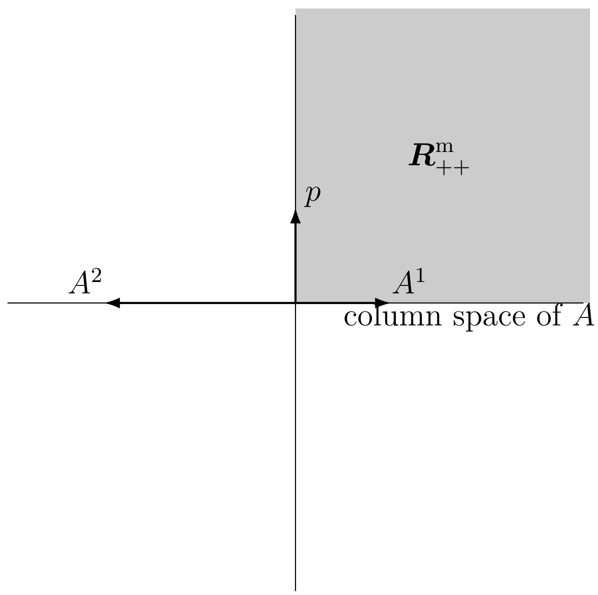

Figure 16.2. Geometry of Gordan's Alternative.

16.13.3 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*

$$Ax > 0 \tag{6}$$

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

$$\begin{aligned} pA &= 0 \\ p &\gg 0. \end{aligned} \tag{7}$$

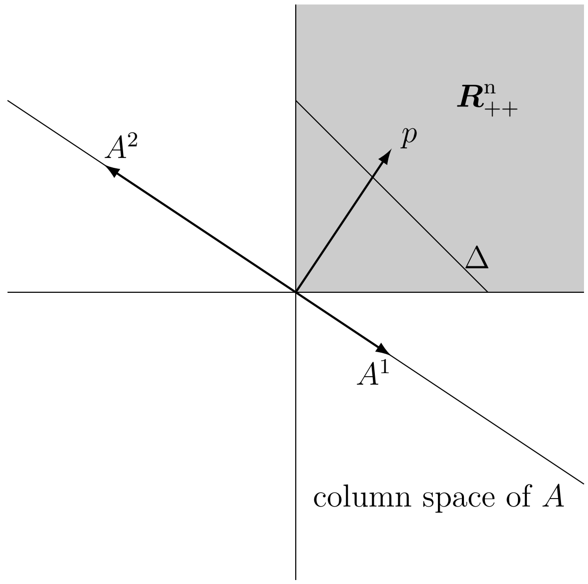

Figure 16.3. Geometry of the Stiemke Alternative

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. http://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] ——— . 1996. Notes on integration by parts. http://www.its.caltech.edu/~kcborder/Notes/IntegrationByParts.pdf

[4] 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

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

[6] ——— . 2002. *Secrets: A memoir of Vietnam and the Pentagon Papers*. New York: Penguin Books.

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

[8] ——— . 1970–72. *Theory of probability*. London: Wiley. 2 vols.

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

[10] ——— . 1983. Ellsberg revisited: A new look at comparative probability. *Annals of Statistics* 11:1047–1059. http://www.jstor.org/stable/2241296

[11] 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 [**?**, pp. 317–319].

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

[13] 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

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

[15] 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

[16] 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

[17] 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

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

[19] 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

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

[21] 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

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

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

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