CHAPTER 7

Stochastic Choice

We now study the empirical content of individual rational choice when choice is stochastic. There are two possible interpretations of this exercise.

The first is that we lack data on individual choices. There is instead a population of agents, and we observe the *distribution* of choices in the population. For example we may know how many people purchased Italian wine in a wine store, and how many purchased French cheese in a cheese store, but we do not know if those who bought the French product in one store are the same people who bought Italian in the other. The theory to be tested is that of rational agents with stable preferences. Thus we want to know when an observed distribution of choices is consistent with a population of rational agents with potentially different, but stable, preferences.

The second interpretation is that we observe an individual who literally randomizes among different alternatives. We might observe this individual agent over time, enough to infer a stochastic rule that he uses to select an element at random when faced with a given set of available choices.

7.1 STOCHASTIC RATIONALITY

The model of stochastic choice can be described as follows. A *system of choice probabilities* is a pair (X,P), where X is a finite set of *alternatives* and P is a function with domain contained in $2^X \backslash \{\varnothing\} \times X$, where $P(A,x) = P_A(x)$ is a non-negative number for each nonempty $A \subseteq X$ and $x \in A$, and such that $\sum_{x \in A} P_A(x) = 1$. In other words, $P_A(x)$ defines a probability distribution over the set A.

In our first interpretation of stochastic choice, there would be an underlying large population of agents: $P_A(x)$ is the fraction of agents who choose x when the set A of possible alternatives is available. In the second interpretation of stochastic choice, an individual agent's choices really are random, and we know that $P_A(x)$ is her probability of choosing x from A.

We assume here that all nonempty $A \subseteq X$ are possible sets of alternatives to choose from (they are budgets in the terminology of Chapters 2 and 3). We explain below (Remark 7.5) which results hold true when this assumption is relaxed.

Given a finite set X, we consider the set Π of all strict preferences on X. We can identify each element of Π with a one-to-one function $\pi : X \to \{1,\ldots,|X|\}$; such a function is a specific utility representation of the preference in question. It is convenient in what follows to describe preferences using utility functions.

We use the following notational simplification. Denote by $\pi(A)$ the set $\{\pi(x) : x \in A\}$ and write $\pi(x) \geq \pi(A)$ to mean that $\pi(x) \geq \pi(y)$ for all $y \in A$. Recall that the preferences in Π are strict, so if $x \in A$ then $\pi(x) \geq \pi(A)$ means that x gives the highest utility in A for utility function π.

A probability distribution on Π is a function $\nu : \Pi \to \mathbf{R}_+$ with $\sum_{\pi\in\Pi} \nu(\pi) = 1$. Denote by $\Delta(\Pi)$ the set of all probability distributions on Π. When $\nu \in \Delta(\Pi)$ and $E \subseteq \Pi$ then we write $\nu(E)$ for $\sum_{\pi\in E} \nu(\pi)$.

A system of choice probabilities (X,P) is *rationalizable* if there is $\nu \in \Delta(\Pi)$ such that for all nonempty $A \subseteq X$ and all $x \in A$,

$$P_A(x) = \sum_{\pi\in\Pi} \nu(\pi)\mathbf{1}_{\{\pi(x)\geq\pi(A)\}} = \nu\left(\{\pi \in \Pi : \pi(x) \geq \pi(A)\}\right).$$

Rationalizability has different interpretations, depending on how we interpret the stochastic choice and the system (X,P). If we interpret stochastic choice as the choices of a population of agents, and $P_A(x)$ as the fraction of agents that choose x from A, then rationalizability means that there is a population distribution ν over the possible preferences that agents can have. Then the fraction of agents choosing x is the fraction of agents for whom x is best in the set A. If, on the other hand, $P_A(x)$ is the result of individual random choices, then rationalizability means that the individual's preferences change. Given a choice problem A, the individual agent draws a utility at random from Π, and chooses x with probability equal to the probability of drawing a utility for which x is best in A. Given this interpretation, the model is often called a model of *random utility*.

Before we go any further, it is worth setting down a very basic implication of rationalizability:

Observation 7.1 *If (X,P) is rationalizable, then for any $x \in X$ and nonempty $A,A' \subseteq X$ such that $x \in A \subseteq A'$, we have*

$$P_A(x) \geq P_{A'}(x).$$

The monotonicity property in Observation 7.1 is the stochastic counterpart of Sen's α (discussed in Chapter 2). The property is called *regularity* in the literature on stochastic choice. Regularity is clearly too weak to characterize rationalizable systems of choice probabilities. We turn instead to two stronger properties: the axiom of revealed stochastic preference and the non-negativity of the Block–Marschak polynomials.

A system of choice probabilities (X,P) satisfies the *axiom of revealed stochastic preference* if, for all sequences $(x_1,A_1),\ldots(x_n,A_n)$, with $x_i \in A_i$ for

$i = 1,\dots,n$, we have that

$$\sum_{i=1}^{n} P_{A_i}(x_i) \leq \max_{\pi \in \Pi} \sum_{i=1}^{n} \mathbf{1}_{\{\pi(x_i) \geq \pi(A_i)\}}.$$

Note that the sequence $(x_1,A_1),\dots(x_n,A_n)$ may repeat the same term (x_i,A_i) many times. As a consequence, testing the satisfaction of the axiom of revealed stochastic preference is problematic because one must verify that an infinite number of sequences have the property above. This is never an issue when using other revealed preference axioms, for example when testing for SARP or GARP. It turns out, however, that there is an algorithm that infinitely many steps determines if the axiom is satisfied. The algorithm amounts to checking the existence of a solution of a system of linear inequalities.[1]

In addition to the axiom of revealed stochastic preference, a certain system of polynomials turns out to characterize stochastic rationality. For all $A \subsetneq X$ and $x \in A^c = X \setminus A$, define the number $K_{x,A}$ by

$$K_{x,A} = \sum_{i=0}^{|A|} (-1)^{|A|-i} \sum_{\{C \subseteq A : |C| = i\}} P_{C^c}(x).$$

The collection of all $K_{x,A}$ comprise the *Block–Marschak polynomials* for the system of choice probabilities (X,P).

The meaning of the Block–Marschak polynomials is made clear by Proposition 7.3 below. They are in principle difficult to interpret. Note, however, that a simple calculation gives:

$$K_{x,\{y\}} = P_{X \setminus \{y\}}(x) - P_X(x),$$

for $x \neq y$. So Observation 7.1 means that $K_{x,\{y\}} \geq 0$ is necessary for rationalizability. It turns out that, not only $K_{x,\{y\}}$, but all the Block–Marschak polynomials must be non-negative for rationalizability; and conversely that if they are all non-negative, then the system of choice probabilities is rationalizable.

The following result collects two theorems, one due to McFadden and Richter, and one due to Falmagne.

Theorem 7.2 *Let (X,P) be a system of choice probabilities, where X is a finite set. The following statements are equivalent:*

I) *(X,P) is rationalizable.*
II) *(X,P) satisfies the axiom of revealed stochastic preference.*
III) *The Block–Marschak polynomials for (X,P) are non-negative.*

The role of the Block-Marschak polynomials may seem obscure. The following result gives them a natural interpretation. We present this result

[1] The existence of such solutions lies at the heart of many problems studied in this book: see the discussion in Chapter 12.

before the proof of Theorem 7.2 because it turns out to play a crucial rule in the proof.

For $C \subseteq X$, and any $x \in C^c$, let

$$M_{x,C} = \{\pi \in \Pi : \pi(C) > \pi(x) \geq \pi(C^c)\}$$

be the set of all utilities in Π that rank any member of C above x, and x at the top of C^c. Proposition 7.3 says that, if (X, P) is rationalizable, then $K_{x,A}$ is the probability that all the elements in A are ranked above x, and that x is at the top of A^c; put differently, $K_{x,A}$ is the probability that A is the upper contour set of x.

Proposition 7.3 *The system of choice probabilities* (X, P) *is rationalized by* $\nu \in \Delta(\Pi)$ *iff* $K_{x,A} = \nu(M_{x,A})$ *for all* (x, A).

Proof. The proof is an application of a combinatorial technique called *Möbius inversion*; this specific type of Möbius inversion is called the *inclusion–exclusion principle*. The technique lets us invert variables which are defined by cumulative sums of real-valued functions defined on a lattice. For a set A and $x \in A$, ν rationalizes the system of choice probabilities iff $P_A(x)$ is the probability that the strict upper contour set of π at x is contained in A^c; formally

$$P_A(x) = \nu(\{\pi \in \Pi : \{y : \pi(y) > \pi(x)\} \subseteq A^c\}).$$

Moreover, $\nu(M_{x,B})$ is by definition the probability that the strict upper contour set of x is exactly B. Consequently, ν rationalizes the system of choice probabilities iff $P_A(x) = \sum_{B \subseteq A^c} \nu(M_{x,B})$, or inverting the role of A and A^c,

$$P_{A^c}(x) = \sum_{B \subseteq A} \nu(M_{x,B}). \tag{7.1}$$

Equation (7.1) shows that P_{A^c} is defined by a cumulative sum; namely, one defines it by summing $\nu(M_{x,B})$ across all $B \subseteq A$. Möbius inversion tells us that, conversely, if we know the value of the left-hand side of equation (7.1) for every A, we can recover $\nu(M_{x,B})$ for all B. An explicit formula for this inversion in this case is well known, and is called the inclusion–exclusion principle. The application of this principle in this environment gives exactly $\nu(M_{x,A}) = K_{x,A}$. (See Proposition 2 of Rota (1964) and the Corollary (Principle of Inclusion–Exclusion) on p. 345.)

7.1.1 Proof of Theorem 7.2

The proof requires the following technical lemma, which we state here without proof. Note that the first part of the lemma is an alternative inductive definition of the Block–Marschak polynomials.

Lemma 7.4 *For all* $A \subseteq X$*:*

I) $K_{x,A} = P_{A^c}(x) - \sum_{C \subsetneq A} K_{x,C}$ *if* $x \in A^c$*;*
II) $\sum_{x \in A^c} K_{x,A} = \sum_{x \in A} K_{x,A \setminus \{x\}}$.

We start by proving the equivalence of (I) and (III). The proof uses Proposition 7.3 in important ways. First, note that the implication (I) $\Longrightarrow$ (III) is immediate from Proposition 7.3. We shall prove that (III) $\Longrightarrow$ (I) by constructing a rationalizing $\nu \in \Delta(\Pi)$.

The structure of this proof is quite involved, so we divide it into steps. Here is the basic structure. Our ultimate goal is to construct $\nu \in \Delta(\Pi)$ which rationalizes P. We do this by first recursively constructing a set function ν^* on a collection of "cylinders" in Π. Let us consider $\pi^* \in \Pi$, and suppose x_1 is the π^*-maximal element, x_2 is the second highest ranked element, and so forth. We will first define ν^*, the probability of the set of all π for which x_1 is maximal. Using this number, we then find the probability of the set of all π for which x_1 is maximal and x_2 comes second. Ultimately, this will allow us to construct the probability of π^*.

We use the notation ν^* simply because the function is not defined on all subsets of Π, but rather on the set of cylinders described in the previous paragraph. But ν^* will then be defined on atoms (on each singleton $\{\pi\}$); and will thus have a natural extension from the set of atoms to a probability measure. This probability measure is ν. Of course, ν and ν^* will coincide on all cylinders. Along the way we will simply need to show that $\nu(\pi) \geq 0$ for every π, that the associated numbers add to one, and that we have rationalization. Note that the cylinders referred to in the previous paragraph will be exactly the type we need to consider in order to discuss rationalization.

Step 1: Defining the cylinders.

We use the term *d-sequence* for a sequence $(x_1, \ldots, x_k)$ such that all its terms are (pairwise) distinct. For any d-sequence $(x_1, \ldots, x_k)$ let

$$S_{(x_1,\ldots,x_k)} = \{\pi \in \Pi : \pi(x_1) > \pi(x_2) > \cdots > \pi(x_k) > \pi(X \setminus \{x_1, \ldots, x_k\})\}.$$

The set $S_{(x_1,\ldots,x_k)}$ is the *cylinder* associated to $(x_1, \ldots, x_k)$. Let $\mathcal{S}$ be the collection of all cylinders: the subsets of Π of the form $S_{(x_1,\ldots,x_k)}$, for some d-sequence $(x_1, \ldots, x_k)$. We shall define a function ν^* on $\mathcal{S}$.

Step 2: Constructing ν^*, and verifying two important additivity properties.

We want to define ν^* on $\mathcal{S}$ by induction. Let $A = \{x_1, \ldots, x_k\}$ for some d-sequence $(x_1, \ldots, x_k)$ and let R_A be the set of d-sequences of length k with elements in A. The key properties required of ν^* will be the following.

$$\sum_{(x'_1,\ldots,x'_k) \in R_A} \nu^*(S_{(x'_1,\ldots,x'_k,x)}) = K_{x,\{x_1,\ldots,x_k\}}. \tag{7.2}$$

and

$$\sum_{x \in A^c} \nu^*(S_{(x_1,\ldots,x_k,x)}) = \nu^*(S_{(x_1,\ldots,x_k)}) \tag{7.3}$$

We proceed to define ν^*. The guiding principle in the definition of ν^* is the interpretation of $K_{x,A}$ obtained from Proposition 7.3. We seek to construct ν so that $K_{x,A}$ is the probability that A is the strict upper contour set of x. The same guiding principle suggests properties (7.2) and (7.3).

First, for every $x \in X$, let $\nu^*(S_{(x)}) = K_{x,\varnothing}$. Second, for every $x, y \in X$, with $x \neq y$, let $\nu^*(S_{(y,x)}) = K_{x,\{y\}}$. Suppose now that $\nu^*(S_{(y_1,\ldots,y_l)})$ has been defined for all d-sequences $(y_1,\ldots,y_l)$ with $l \leq k$. Fix a d-sequence $(x_1,\ldots,x_k,x_{k+1})$. Let $A = \{x_1,\ldots,x_k\}$ and R_A be the set of all d-sequences in A. If $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) = 0$, let $\nu^*(S_{(x_1,\ldots,x_k,x_{k+1})}) = 0$. Otherwise, let

$$\nu^*(S_{(x_1,\ldots,x_k,x_{k+1})}) = \frac{\nu^*(S_{(x_1,\ldots,x_k)})K_{x_{k+1},\{x_1,\ldots,x_k\}}}{\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)})}. \tag{7.4}$$

This defines ν^* on $\mathcal{S}$ by induction.

We shall prove that (7.2) and (7.3) are satisfied. The proof is by induction on the length of the d-sequence defining A. It is easy to see by a direct calculation that (7.2) and (7.3) are satisfied for all sequences of length 0 and 1, where a sequence of length 0 is associated to $A = \varnothing$, and we define $\nu^*(S_\varnothing) = 1$.

Suppose that (7.2) holds for all d-sequences of length $l \leq k-1$. Let $A = \{x_1,\ldots,x_k\}$ for some d-sequence $(x_1,\ldots,x_k)$, of length k. Then we know that

$$\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) = \sum_{x\in A}\left(\sum_{(x'_1,\ldots,x'_{k-1})\in R_{A\setminus\{x\}}} \nu^*(S_{(x'_1,\ldots,x'_{k-1},x)})\right) \tag{7.5}$$

$$= \sum_{x\in A} K_{x,A\setminus\{x\}} \tag{7.6}$$

$$= \sum_{x\in A^c} K_{x,A}. \tag{7.7}$$

Here, (7.5) follows from the inductive hypothesis and (7.3) (by adding over the right-hand-side of (7.3)). Furthermore, (7.6) is a consequence of the inductive hypothesis and (7.2); and (7.7) follows from the second part of Lemma 7.4.

Step 2a: Verifying the two additivity properties ((7.2) and (7.3)) in the case $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) = 0$.

There are two cases to consider. Suppose first that $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) = 0$. Then (7.7) implies that for any $x \in A^c$, we have $K_{x,A} = 0$, as all Block-Marschak polynomials are non-negative (the hypothesis). Therefore, as by definition of ν^*, $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k,x)}) = 0$, we have $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k,x)}) = K_{x,A}$. This verifies (7.2).

Further, Equation (7.3) is clearly satisfied when $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) = 0$, because, by definition, $\nu^*(S_{(x_1,\ldots,x_k,x)}) = 0$.

Step 2b: Verifying the two additivity properties ((7.2) and (7.3)) in the case $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) > 0$.

The second case to consider is when $\sum_{(x'_1,\ldots,x'_k)\in R_A} \nu^*(S_{(x'_1,\ldots,x'_k)}) > 0$. By definition of $\nu^*(S_{(x_1,\ldots,x_k,x)})$, Equation (7.2) is always satisfied when

$\sum_{(x'_1,\dots,x'_k)\in R_A} \nu^*(S_{(x'_1,\dots,x'_k)}) > 0$. To see that (7.3) also holds, note that:

$$\begin{aligned}\sum_{x\in A^c} \nu^*(S_{(x_1,\dots,x_k,x)}) &= \frac{\nu^*(S_{(x_1,\dots,x_k)}) \sum_{x\in A^c} K_{x,\{x_1,\dots,x_k\}}}{\sum_{(x'_1,\dots,x'_k)\in R_A} \nu^*(S_{(x'_1,\dots,x'_k)})} \\ &= \frac{\nu^*(S_{(x_1,\dots,x_k)}) \sum_{x\in A} K_{x,A\setminus\{x\}}}{\sum_{(x'_1,\dots,x'_k)\in R_A} \nu^*(S_{(x'_1,\dots,x'_k)})} \\ &= \nu^*(S_{(x_1,\dots,x_k)}).\end{aligned}$$

The second equality above follows from the second property in Lemma 7.4; the third equality follows from Equation (7.6).

This finishes the proof that ν^* satisfies (7.2) and (7.3).

Step 3: Defining ν from the $\nu^*(\pi)$, and verifying that they coincide on cylinders.

Now, ν can be defined on Π in the following way. If we fix $\pi \in \Pi$ then $S_{(x_1,\dots,x_{|X|})} = \{\pi\}$, where $(x_1,\dots,x_{|X|})$ is the sequence defined by $\pi(x_l) > \pi(x_{l+1})$, for $l = 1\dots,|X|-1$. We write $\nu(\pi)$ for $\nu^*(S_{(x_1,\dots,x_{|X|})})$ and let ν be the obvious extension of this measure to all subsets of Π, namely $\nu(E) = \sum_{\pi\in E}\nu(\pi)$. Equation (7.3) establishes that $\nu^* = \nu$ on $\mathcal{S}$. Moreover, $\nu \geq 0$, and it is easily verified that $\sum_{\pi\in\Pi}\nu(\{\pi\}) = 1$, so it is a probability measure.[2] Finally, the fact that $\nu(M_{x,A}) = K_{x,A}$ is the content of Equation (7.2). By Proposition 7.3, the proof is complete.

We proceed to prove that (I) is equivalent to (II). The proof is a direct application of a version of Farkas' Lemma.

Note first that (I) implies (II) because, for any sequence $(x_1,A_1),\dots,(x_n,A_n)$, rationalizability implies $\sum_{i=1}^n P_{A_i}(x_i) \leq \max_{\nu\in\Delta(\Pi)}\sum_{i=1}^n \nu(\{\pi : \pi(x_i) \geq \pi(A_i)\})$. The latter expression exhibits a linear function being maximized over a compact and convex set. Hence, there is a maximizer at an extreme point; in this case, there is a maximizer at some $\delta_{\pi^*} \in \Delta(\Pi)$, where δ_{π^*} is the point mass on $\{\pi^*\}$. Then $\delta_{\pi^*}(\{\pi : \pi(x_i) \geq \pi(A_i)\}) = \mathbf{1}_{\{\pi^*(x_i)\geq\pi^*(A_i)\}}$, concluding this direction.

Conversely, let W be a matrix that has one column for every $\pi \in \Pi$ and one row for every pair (x,A) with $x \in A$. In the entry corresponding to row (x,A) and column π we have a zero if there is $y \in A$ with $\pi(y) > \pi(x)$ and a one otherwise; so the entry is $\mathbf{1}_{\{\pi(x)\geq\pi(A)\}}$. The matrix W can be represented as follows:

$$\begin{array}{cc} & \begin{array}{ccc} \pi_1 & \cdots & \pi_{|X|!} \end{array} \\ \begin{array}{c} \vdots \\ (x,A) \\ \vdots \end{array} & \left[\begin{array}{ccc} \vdots & \cdots & \vdots \\ \mathbf{1}_{\{\pi_1(x)\geq\pi_1(y)\forall y\in A\}} & \cdots & \mathbf{1}_{\{\pi_{|X|}(x)\geq\pi_{|X|!}(y)\forall y\in A\}} \\ \vdots & \cdots & \vdots \end{array}\right] \end{array}$$

[2] Indeed, using (7.6) we obtain that $(-1)\sum\nu(\pi) = \sum_{i=0}^{|X|-1}(-1)^{|X|-i}\binom{|X|}{i} = -1$, by the Binomial Theorem.

Let p be the vector with as many entries as there are pairs (x,A), whose entries of p are arranged in the same order as the rows of W, so that $P_A(x)$ is the entry in position (x,A) of the vector p.

Then we can represent a probability distribution $\nu \in \Delta(\Pi)$ as a vector with one entry for every $\pi \in \Pi$. The existence of a rationalizing ν is the same as the existence of a solution ν to the system

$$p = W \cdot \nu,$$

such that $\nu \geq 0$ and $\nu \cdot (1,\ldots,1) = 1$.

By Farkas' Lemma (Lemma 1.14), there is no solution to the system iff there is a vector η and a scalar θ such that

$$\eta \cdot W + \theta(1,\ldots,1) \leq 0 \tag{7.8}$$

$$\eta \cdot p + \theta > 0. \tag{7.9}$$

We proceed to show first that a violation of the axiom of revealed stochastic preference follows from the existence of a solution to (7.8)–(7.9) in which the entries of η are non-negative integers (in fact the two statements are equivalent).

Let η and θ be a solution to (7.8)–(7.9) in which the entries of η are non-negative integers. Define a sequence $(x_1,A_1),\ldots(x_n,A_n)$ by including (in any order) $\eta_{(x,A)}$ times the term (x,A); the sequence must have at least one term since a solution η to (7.8)–(7.9) cannot be the null vector. Then $\sum_{i=1}^n P_{A_i}(x_i) = \sum_{(x,A)} \eta_{(x,A)} P_A(x)$ and, for any π, $\sum_{i=1}^n \mathbf{1}_{\{\pi(x_i) \geq \pi(A_i)\}} = \sum_{(x,A)} \eta_{(x,A)} \mathbf{1}_{\{\pi(x) \geq \pi(A)\}}$. Since η and θ solve (7.8)–(7.9) we obtain that

$$\sum_{i=1}^n P_{A_i}(x_i) + \theta > 0 \geq \sum_{i=1}^n \mathbf{1}_{\{\pi(x_i) \geq \pi(A_i)\}} + \theta,$$

for all π. Thus the sequence $(x_1,A_1),\ldots(x_n,A_n)$ presents a violation of the axiom.

Finally, we prove that if there is a solution to (7.8)–(7.9) then we can take the entries of η to be non-negative integers. We show how to reduce the system to a collection of strict inequalities in η, by substituting out θ. Then we can take η to have rational entries and satisfy the system. After multiplying η by a large enough positive integer, we can assume that its entries are integers. We shall prove that they can be assumed to be non-negative.

To see how to substitute out θ, note that Equations (7.8) and (7.9) imply that, for every π,

$$\sum_{(x,A)} \eta_{(x,A)} \mathbf{1}_{\{\pi(x) \geq \pi(A)\}} + \theta \leq 0 < \sum_{(x,A)} \eta_{(x,A)} P_A(x) + \theta.$$

Hence for all π,

$$\sum_{(x,A)} \eta_{(x,A)} \mathbf{1}_{\{\pi(x) \geq \pi(A)\}} < \sum_{(x,A)} \eta_{(x,A)} P_A(x). \tag{7.10}$$

(In fact (7.10) being true for every π is necessary and sufficient for the existence of θ that satisfies Equations (7.8) and (7.9), by setting $\theta = -\max_{\pi\in\Pi}\sum_{(x,A)}\eta_{(x,A)}\mathbf{1}_{\{\pi(x)\geq\pi(A)\}}$ for a solution to (7.10).)

So it is without loss of generality to assume that η is integer-valued.

We show that we can take $\eta \geq 0$ in (7.10) by showing that whenever $\eta_{(\hat{x},\hat{A})} < 0$ then Equation (7.10) holds for some η' with $\eta'_{(\hat{x},\hat{A})} = 0$ and $\eta \leq \eta'$.

Suppose $\eta_{(\hat{x},\hat{A})} < 0$. Note that for any $\pi \in \Pi$, $\mathbf{1}_{\{\pi(\hat{x})\geq\pi(\hat{A})\}} = 1 - \sum_{z\in\hat{A}\setminus\{\hat{x}\}}\mathbf{1}_{\{\pi(z)\geq\pi(\hat{A})\}}$ and $P_{\hat{A}}(\hat{x}) = 1 - \sum_{z\in\hat{A}\setminus\{\hat{x}\}}P_{\hat{A}}(z)$. Consequently we get

$$\eta_{(\hat{x},\hat{A})}\mathbf{1}_{\{\pi(\hat{x})\geq\pi(\hat{A})\}} = \eta_{(\hat{x},\hat{A})} + (-\eta_{(\hat{x},\hat{A})})\sum_{z\in\hat{A}\setminus\{\hat{x}\}}\mathbf{1}_{\{\pi(z)\geq\pi(\hat{A})\}}$$

and

$$\eta_{(\hat{x},\hat{A})}P_{\hat{A}}(\hat{x}) = \eta_{(\hat{x},\hat{A})} + (-\eta_{(\hat{x},\hat{A})})\sum_{z\in\hat{A}\setminus\{\hat{x}\}}P_{\hat{A}}(z).$$

So now find η' as follows. Add $-\eta_{(\hat{x},\hat{A})}$ to both sides of Equation (7.10) for all π and make the preceding substitutions. Hence, η' coincides with η everywhere except that $\eta_{(\hat{x},\hat{A})} = 0$ and for all $z \in \hat{A}\setminus\{\hat{x}\}$ $\eta'_{(z,\hat{A})} = \eta_{(z,\hat{A})} - \eta_{(\hat{x},\hat{A})}$, we obtain that η' satisfies (7.10) for all $\pi \in \Pi$ while $\eta \leq \eta'$ and $\eta'_{(\hat{x},\hat{A})} = 0$.

Remark 7.5 The proof that (I) is equivalent to (II) in the preceding does not rely on the ability to observe the entire system of choice probabilities. In particular, the axiom of revealed stochastic preference is also necessary and sufficient for rationalization by $\nu \in \Delta(\Pi)$ for environments as in Chapter 2, whereby we only observe P_A for A in some set of budgets $\Sigma \subseteq 2^X\setminus\{\varnothing\}$.

7.2 LUCE'S MODEL

We now analyze the special class of systems (X,P) introduced by Duncan Luce. The model is heavily used in applied work, and it lies at the foundation of statistical and econometric studies of discrete choice. We proceed to describe the model, and study its relation to stochastic rationality.

A system of choice probabilities (X,P) satisfies *Luce's independence of irrelevant alternatives (LIIA)* if for any A, and any $x,y \in A$, $P_A(x)P_{\{x,y\}}(y) = P_{\{x,y\}}(x)P_A(y)$. The LIIA axiom is easiest to interpret when the probabilities involved are non-zero, so we can divide and obtain

$$\frac{P_A(x)}{P_A(y)} = \frac{P_{\{x,y\}}(x)}{P_{\{x,y\}}(y)}.$$

So the LIIA axiom says that the likelihood of choosing x relative to y is independent of what other alternatives may be available in A.

An example illustrates that LIIA may be unreasonable. Consider an agent facing the set of alternatives $\{\text{car}, \text{bus}_1\}$, who chooses each with probability $1/2$. If the agent faces instead the set $\{\text{car}, \text{bus}_1, \text{bus}_2\}$, where the two buses only differ in their color (they go to the same place at the same speed), then we might expect him to choose the car or *either* of the two buses with probability $1/2$. LIIA, however, implies that he must choose each alternative with probability $1/3$.

LIIA has a clear implication. For notational simplicity, write $q_{x,y}$ for $P_{\{x,y\}}(x)$. Suppose that $P_A(x) > 0$ for all $x \in A$, and for all nonempty A. Fix an element $z \in X$. Then we can define $u(x) = q_{x,z}/q_{z,x}$.

Note that LIIA implies that $1 = \sum_{x\in A} P_A(x) = P_A(y) \sum_{x\in A} \frac{q_{x,y}}{q_{y,x}}$. Note also that

$$\frac{q_{x,y}}{q_{y,x}} = \frac{P_{x,y,z}(x)}{P_{x,y,z}(y)} = \frac{P_{x,y,z}(z)q_{x,z}/q_{z,x}}{P_{x,y,z}(z)q_{y,z}/q_{z,y}} = \frac{u(x)}{u(y)}.$$

Then,

$$P_A(y) = \frac{u(y)}{\sum_{x\in A} u(x)}. \tag{7.11}$$

We say that (X,P) conforms to the *Luce model* if there is a function $u : X \to \mathbf{R}_+$ such that (7.11) holds for all A and y. We can make the interpretation of u a bit more precise. Say that $x \succeq^* y$ if $q_{x,y} \geq 1/2$. If (X,P) conforms to Luce's model then $x \succeq^* y$ iff $u(x) \geq u(y)$. So $\succeq^*$ is a preference relation represented by u.

The numbers $u(x)$ can be thought of as utility intensities. In previous chapters, the utility functions have played a purely ordinal role in choice. But in the Luce model, an alternative that has a higher utility than another alternative has a higher probability of being chosen. So the utility function u conveys a meaning above and beyond how it orders the different objects of choice.

The "revealed preference" problem of testing whether (X,P) conforms to Luce's model is very simple to solve: set $u(x) = P_X(x)$ and verify whether the resulting u satisfies the definition. Instead of testing Luce's model, we focus on the relation between the model and stochastic rationality, as described in Section 7.1.

Theorem 7.6 *If (X,P) conforms to Luce's model, then it is rationalizable.*

Before proving Theorem 7.6 we show that Luce's model does not exhaust all the rationalizable systems of choice probabilities. Consider the following example.

Let $X = \{a,b,c\}$. Suppose that the utility of a is given by a random variable $\tilde{a}$; and that there are random variables $\tilde{b}$ and $\tilde{c}$ that define the utilities of b and c. Suppose that the three random variables, $\tilde{a}$, $\tilde{b}$ and $\tilde{c}$, are independent and distributed on $\{1,\dots,6\}$ according to the following table:

	1	2	3	4	5	6
$\tilde{a}$	0	0	1/2	1/2	0	0
$\tilde{b}$	0	0.6	0	0	0	0.4
$\tilde{c}$	0.4	0	0	0	0.6	0.

The distributions of $\tilde{a}$, $\tilde{b}$, and $\tilde{c}$ describe a probability distribution on Π, as any specification of random utilities is equivalent to a probability distribution on Π.

Then $q_{a,b} = 0.6$, $q_{b,c} = 0.4 + 0.6 \times 0.4 = 0.64$, and $q_{c,a} = 0.6$. So $a \succ^* b$, $b \succ^* c$ and $c \succ^* a$. Then $\succeq^*$ cannot have any utility representation, let alone one that allows (X,P) to conform with the Luce model.

7.2.1 Proof of Theorem 7.6

Suppose that (X,P) conforms to Luce's model with a corresponding function u; by a normalization we can suppose that $\sum_{x\in X} u(x) = 1$. We use the same notation as in the proof of Theorem 7.2.

For any $\pi \in \Pi$ and j, let x_j^π denote the alternative in X with the jth highest value in π. Thus, $\pi(x_1^\pi) = |X|$, $\pi(x_2^\pi) = |X| - 1, \ldots \pi(x_{|X|}^\pi) = 1$.

The proof proceeds by constructing a probability space in which one can calculate the probability that a sequence of random draws will correspond to a preference π. Consider a probability space defined as follows. Draw infinite sequences in X at random by drawing independently (with replacement) elements from X such that each z is drawn with probability $u(z)$. Let μ be the associated probability measure on X. It should be clear that for any $A \subseteq X$ and $x \in A$, the probability that x is drawn *before* any other element in A is equal to $u(x)/\sum_{y\in A} u(y)$.[3]

For any sequence $j_1 < \ldots < j_{|X|}$ with $j_1 = 1$, let $D_{j_1,\ldots,j_{|X|}}(\pi)$ denote the event that x_k^π is drawn for the first time at draw number j_k. Then $\bigcup_{j_1<\ldots<j_{|X|}} D_{j_1,\ldots,j_{|X|}}(\pi)$ is the event $C(\pi)$ that the draws will conform to π: so $C(\pi) = \bigcup_{j_1<\ldots<j_{|X|}} D_{j_1,\ldots,j_{|X|}}(\pi)$ is the event where x_1^π is drawn first; followed by x_2^π (possibly after several repeated draws of x_1^π); followed by x_3^π (possibly after several repeated draws of x_1^π and x_2^π), and so on. The sets $D_{j_1,\ldots,j_{|X|}}(\pi)$ are disjoint, so the probability that the draws will conform to π is

$$\begin{aligned}\mu(C(\pi)) &= \sum_{j_1<\ldots<j_{|X|}} \mu(D_{j_1,\ldots,j_{|X|}}(\pi)) \\ &= \sum_{j_1<\ldots<j_{|X|}} u(x_1^\pi)^{j_2-j_1} u(x_2^\pi) \left(u(x_1^\pi) + u(x_2^\pi)\right)^{j_3-j_2-1} \\ &\quad \cdot u(x_3^\pi) \left(u(x_1^\pi) + u(x_2^\pi) + u(x_3^\pi)\right)^{j_4-j_3-1} \cdots u(x_{|X|}^\pi).\end{aligned}$$

The events $C(\pi)$ form a partition. Define $\nu(\{\pi\})$ to be the probability of $C(\pi)$. We can explicitly calculate ν as follows. Let $h_k = j_{k+1} - j_k - 1$ and

[3] To see this: Let E be the event that x is drawn before any other element in A. Then E occurs if either x is obtained in the first draw, which has probability $u(x)$, or else an element of A^c is obtained in the first draw (which has probability $1 - \sum_{y\in A} u(y)$) and then E occurs. Then the probability q of E obeys the equation $q = u(x) + (1 - \sum_{y\in A} u(y))q$.

$v_k^\pi = \sum_{l=1}^{k} u(x_l^\pi)$. Then,

$$\begin{aligned}
\nu(\pi) = \mu(C(\pi)) &= \prod_{z\in X} u(z) \sum_{h_1\geq 0,\ldots h_{|X|-1}\geq 0} (v_1^\pi)^{h_1}(v_2^\pi)^{h_2}\cdots(v_{|X|-1}^\pi)^{h_{|X|-1}} \\
&= \prod_{z\in X} u(z) \prod_{k=1}^{|X|-1} \frac{1}{1-v_k^\pi} \\
&= \prod_{z\in X} u(z) \prod_{k=1}^{|X|-1} \frac{1}{\sum_{l=k+1}^{|X|} u(x_l^\pi)} \\
&= \prod_{k=1}^{|X|-1} \frac{u(x_k^\pi)}{\sum_{l=k}^{|X|} u(x_l^\pi)};
\end{aligned}$$

where the next-to-last equality follows by distributing the product $\prod_{z\in X} u(z)$ appropriately, and using that $\sum_{l=1}^{|X|} u(x_l^\pi) = 1$.

Finally, we have already observed that the probability of drawing x before any other alternative in A is equal to $u(x)/\sum_{y\in A} u(y) = P_A(x)$. Note that this probability also equals $\sum_{\{\pi:\pi(x)\geq\pi(A)\}} \mu(C(\pi)) = \nu(\{\pi : \pi(x) \geq \pi(A)\})$. Hence ν rationalizes (X,P).

7.2.2 Luce's model and the logit model

Luce's model can be interpreted as a random utility model in which the "average" utility of x is some known quantity $v(x)$, but where the actual utility is $v(x)+\varepsilon(x)$, where $\varepsilon(x)$ is unknown and random.

The following calculation shows a method of determining v, and motivates why we can think of Luce's model as the "logit model." Suppose that (X,P) conforms to Luce's model, with utility index u. Note that

$$q_{x,y} = \frac{u(x)}{u(x)+u(y)} = \frac{u(x)/u(y)}{u(x)/u(y)+1} = \frac{e^{v(x)-v(y)}}{1+e^{v(x)-v(y)}},$$

where $v(x) = \log(u(x))$. Then the probability of choosing x over y is $q_{x,y} = \varphi(v(x)-v(y))$. The function φ is the logistic distribution function. Then

$$P_A(x) = \frac{e^{v(x)}}{\sum_{y\in A} e^{v(y)}}. \tag{7.12}$$

In particular, Equation (7.12) alternatively derives from a particular random utility model specification, which is called the logit model. Suppose that $P_A(x)$ equals the probability that $v(y)+\varepsilon(y) \leq v(x)+\varepsilon(x)$ for all $y \in A$, $y \neq x$. That is, the probability that $\varepsilon(y)-\varepsilon(x) \leq v(x)-v(y)$ for all $y \in A$, $y \neq x$. Let $v : X \to \mathbf{R}$ be an arbitrary function, and let G be a cdf on $\mathbf{R}$, from which the terms $\varepsilon(x)$ are drawn independently across x. In particular, if G is a Gumbel distribution, then $q_{x,y}$, the probability of choosing x over y, is equal to the probability that a logistic random variable is below $v(x)-v(y)$. The choice of a Gumbel distribution is suggested by the function φ being logistic. The following proposition establishes that this specification implies the choice probabilities in Equation (7.12). It can therefore be viewed as a counterpart to Theorem 7.6.

Proposition 7.7 *Let $v : X \to \mathbf{R}$, and suppose that $G(\alpha) = \exp(-\exp(-\alpha))$. Let ε be a random vector drawn from $\mathbf{R}^X$ according to $|X|$ independent draws of G. Let (X,P) be defined by*

$$P_A(x) = Pr(v(y) + \varepsilon(y) \leq v(x) + \varepsilon(x) \textit{ for all } y \in A^4).$$

Then (X,P) conforms to Luce's model with index $u(x) = e^{v(x)}$.

Proof. Fix A and $x \in A$. Let $u(y) = e^{v(y)}$ and $\delta(y) = e^{-\varepsilon(y)}$. Note that the form of G implies that the distribution function of $\delta(y)$ is $Pr(\delta(y) \leq \alpha) = 1 - e^{-\alpha}$ (the exponential distribution). Then, by the definition of (X,P),

$$\begin{aligned} P_A(x) &= Pr(u(x)/\delta(x) \geq u(y)/\delta(y) \text{ for all } y \in A) \\ &= \int_0^\infty Pr(\delta(y) \geq \bar{\delta}u(y)/u(x) \text{ for all } y \in A)e^{-\bar{\delta}}d\bar{\delta} \\ &= \int_0^\infty \left(\prod_{y \in A \setminus \{x\}} e^{-\bar{\delta}u(y)/u(x)} \right) e^{-\bar{\delta}}d\bar{\delta} \\ &= \int_0^\infty \exp\left\{ -\bar{\delta} \left(1 + \sum_{y \in A \setminus \{x\}} \frac{u(y)}{u(x)} \right) \right\} d\bar{\delta} \\ &= \frac{u(x)}{\sum_{y \in A} u(y)}. \end{aligned}$$

Proposition 7.7 relies on a particular distribution for the random utility term ε. One may ask if there are other distributions for ε that lead to the Luce model. We provide a partial answer in the next result (which is due to McFadden), where we show that if we insist on G being translation complete and the utility index being onto, then the distribution giving rise to the Luce model is unique.

Consider now a system of choice probabilities (X,P), where we depart from the assumptions in this chapter by allowing that X may be infinite. Suppose that $P_A(x)$ is only defined for finite sets A. As before, $x \mapsto P_A(x)$ is a probability distribution on A.

Assume that $P_A(x)$ is obtained from a random utility model, as above. Specifically, suppose that there is a utility v defined on A, and random variables $\varepsilon(x)$ such that x is chosen from A if $v(y) + \varepsilon(y) < v(x) + \varepsilon(x)$ for all $y \in A$. Suppose that the random variables $\varepsilon(x)$ are independent with identical distribution function G on $\mathbf{R}$. Say that (X,P) is *generated* from $v : X \to \mathbf{R}$ and G.

[4] Here, Pr is probability calculated according to independent draws from G.

The distribution function G is *translation complete* if, for any function f such that

$$0 = \lim_{x\to-\infty} f(x) = \lim_{x\to+\infty} f(x),$$

$\int f(x+t)dG(x) = 0$ for all t implies that $f = 0$ a.s. The property of translation completeness is shared by many common distribution functions.

Proposition 7.8 *Suppose that (X,P) is generated from v and G, and that (X,P) conforms to Luce's model with utility index $u(x) = e^{v(x)}$. Suppose that $v : X \to \mathbf{R}$ is such that $v(X) = \mathbf{R}$, and that G is translation complete. Then there is $\alpha > 0$ such that $G(x) = \exp(-\alpha\exp(-x))$.*

Proof. Fix $w \in \mathbf{R}$ and $x \in X$. Let $A = \{x, y_1, \dots, y_K\}$, and let δ_i be such that $v(y_i) = w + \delta_i$ for $i = 1, \dots, K$. Importantly, K is an arbitrary positive integer. Note that

$$P_A(x) = \frac{\exp(v(x))}{\exp(v(x)) + \sum_{i=1}^K \exp(w+\delta_i)} = \int \left[\prod_{i=1}^K (G(v(x)+t-w-\delta_i))\right] dG(t).$$

Because $v(X) = \mathbf{R}$, we can let all the $\delta_i \to 0$ from below to obtain that (using the right continuity of G)

$$\frac{\exp(v(x))}{\exp(v(x)) + K\exp(w)} = \int (G(v(x)+t-w))^K dG(t).$$

On the other hand, if we choose $z \in X$ such that $v(z) = w + \log(K)$ then $P_{\{x,z\}}(x) = \int G(v(x)+t-w-\log(K))dG(t)$. Now,

$$P_{\{x,z\}}(x) = \frac{\exp(v(x))}{\exp(v(x)) + \exp(v(z))} = \frac{\exp(v(x))}{\exp(v(x)) + K\exp(w)},$$

hence

$$\int (G(v(x)+t-w))^K dG(t) = \int (G(v(x)+t-w-\log(K)))dG(t).$$

Since, w and x were arbitrary, we obtain that for all w,

$$\int \left[(G(v(x)+t-w))^K - G(v(x)+t-w-\log(K))dG(t)\right] = 0.$$

Since x was arbitrary, we can choose $v(x) = 0$ and note that by translation completeness, $(G(t))^K = G(t-\log(K))$ for all t. We claim that there is $\alpha > 0$ such that $G(t) = \exp(-\alpha\exp(-t))$.

First, note that $G(t-\log(K)) = (G(t))^K$ implies, setting $t = 0$, that $G(-\log(K)) = (G(0))^K$ for any positive integer K. Likewise, for any positive integer L, we have, by setting $t = \log(K/L)$, $G(-\log(L)) = (G(\log(K/L)))^K$. But $G(-\log(L)) = (G(0))^L$, so $(G(0))^L = (G(\log(K/L)))^K$, or $G(\log(K/L)) = (G(0))^{\frac{L}{K}}$. $G \circ \log$ is therefore continuous on the strictly positive rational numbers, and by definition, it is right-continuous, so that for any strictly positive real number r, $G(\log(r)) = (G(0))^{1/r}$. Now, let $x \in \mathbf{R}$. We have

$G(x) = G(\log(\exp(x))) = (G(0))^{\exp(-x)}$. Finally, since $G(x)$ is a cumulative distribution function, we infer that $0 < G(0) < 1$, so we may set $\alpha = -\log(G(0)) > 0$, and obtain $G(x) = \exp(-\alpha \exp(-x))$.

7.3 RANDOM EXPECTED UTILITY

The previous discussion has not sought to limit the structure of rationalizing preferences in any way. We shall now study the random choice of lotteries, and consider only von Neumann–Morgenstern expected utility preferences.

Let Y be a finite set of "prizes." The objects of choice will be lotteries over Y. A lottery is a probability distribution over Y. Let X be the set of all lotteries over Y. An alternative x in X indicates the probability x_i of the ith element of Y.

In the present setting, a (von Neumann–Morgenstern) utility function is a vector $u \in \mathbf{R}^Y$. An expected utility maximizing agent with utility function u weakly prefers a lottery x over y iff $u \cdot x \geq u \cdot y$.

We have as before a system of choice probabilities, but where X is infinite and we define the choice only over finite nonempty sets. So a system of choice probabilities is a pair (X,P), where P_A is a probability distribution over A, for all finite nonempty sets A. More specifically, P_A is a Borel probability measure on X with $P_A(A) = 1$.

We say that (X,P) is *expected-utility rationalizable* if there is a probability measure μ on $\mathbf{R}^Y$ such that for every finite nonempty A and every $x \in A$,

$$P_A(x) = \mu\left(\{u \in \mathbf{R}^Y : u \cdot x \geq u \cdot y \text{ for all } y \in A\}\right).$$

The domain of μ is required to be the σ-algebra generated by all sets of the form $\{u \in \mathbf{R}^Y : u \cdot x \geq u \cdot y \text{ for all } y \in A\}$. Such sets are discussed in a bit more detail below. Further, μ is required to be *regular*, in the sense that for every possible A, with probability 1, u has a unique maximizer.

The system (X,P) is *monotonic* if $A \subseteq A'$ and $x \in A$, then $P_A(x) \geq P_{A'}(x)$. The property of monotonicity, or *regularity*, was discussed earlier in 7.1. By Observation 7.1, any rationalizable system of choice probabilities must be monotonic.

We say that the system (X,P) is *linear* if

$$P_{\lambda A+(1-\lambda)y}(\lambda x + (1-\lambda)y) = P_A(x).$$

The notion of linearity is analogous to the property of independence in expected-utility theory.[5]

The convex hull of a finite set A, denoted by $\mathrm{conv}(A)$ is the set of all convex combinations of the elements in A. Write $\mathrm{ext}(A)$ for the extreme points of the

[5] If $\succeq$ is a preference relation over lotteries, independence says that $x \succeq y$ iff $\lambda x + (1-\lambda)z \succeq \lambda y + (1-\lambda)z$, for all z and $\lambda \in (0,1)$. In the present setup, the probability of selecting x is unaffected by a mixture with y because any of the intended rationalizing preferences should be unaffected by the mixture.

convex hull of A: these are the points in the convex hull of A which cannot be written as a convex combination of other points in the convex hull of A.

A system (X,P) is *extreme* if $P_A(\text{ext}(A)) = 1$.

Finally, say that (X,P) is *mixture continuous* if $P_{\lambda A+(1-\lambda)A'}$ is continuous as a function of λ, for all finite nonempty sets A and A'.

The next result is due to Gul and Pesendorfer.

Theorem 7.9 *A system of choice probabilities is expected-utility rationalizable iff it is monotone, linear, extreme, and mixture continuous.*

We proceed to discuss the main ideas in the proof of Theorem 7.9. Let

$$N(A,x) = \{u \in \mathbf{R}^Y : u \cdot x \geq u \cdot y \quad \forall y \in A\}.$$

Thus $N(A,x)$ is the set of utilities rationalizing the choice of x from the set A. Note that the rationalizability of (X,P) by a probability measure μ means that $P_A(x) = \mu(N(A,x))$.

The proof of Theorem 7.9 uses basic ideas from convex analysis in Euclidean spaces. Part of the difficulty in the proof is due to the domain of P being subsets of the simplex, not more general subsets of a Euclidean space. So the first step is to recast the problem using finite subsets of $\mathbf{R}^n$ as the domain of P. The way to do that is to assume that Y has $n+1$ elements, and observe that the $(n+1)$-dimensional simplex is contained in an n-dimensional hyperplane. This hyperplane is isomorphic to $\mathbf{R}^n$. Then use linearity to extend P to the domain of all finite subsets of $\mathbf{R}^n$. We omit the details.

Suppose then that we have defined P_A for finite sets $A \subseteq \mathbf{R}^n$. Linearity can be more simply recast in this case as stating that $P_A(x) = P_{A+\{y\}}(x+y)$ for all $x \in A$.

For P to be rationalizable, we need a probability measure μ for which $P_A(x) = \mu(N(A,x))$. So one can define a function μ on all sets of the form $N(A,x)$ by setting $\mu(N(A,x)) = P_A(x)$. The proof proceeds by first showing that μ is well defined, and then that μ is additive on the family of sets of the form $N(A,x)$.

Note that if $u,v \in N(A,x)$ then $\alpha u + \beta v \in N(A,x)$ for any $\alpha,\beta \geq 0$. So $N(A,x)$ is a convex cone. In fact, $0 \in N(A,x)$ so it is a pointed cone, and A is finite so it is a polyhedral cone (one defined by the intersection of a finite number of halfspaces).

To see that μ is well defined on sets of this kind (pointed polyhedral cones), we need to establish two things. The first is a basic result from convex analysis: if K is any pointed cone, then there is always a finite set A and $x \in A$ such that $K = N(A,x)$. The result is geometrically intuitive, and illustrated in Figure 7.1. We do not provide a formal proof of the existence of A and x with $K = N(A,x)$, but hope that the figure suggests one. The dotted lines in the figure are obtained as perpendicular to the extreme rays of the cone.

The second fact we need to establish says that if $K = N(A,x) = N(A',x')$, then $P_A(x) = P_{A'}(x')$. To prove this fact, note first that linearity implies that $P_A(x) = P_{A-\{x\}}(0)$. In an abuse of notation, let A and A' denote $A - \{x\}$ and

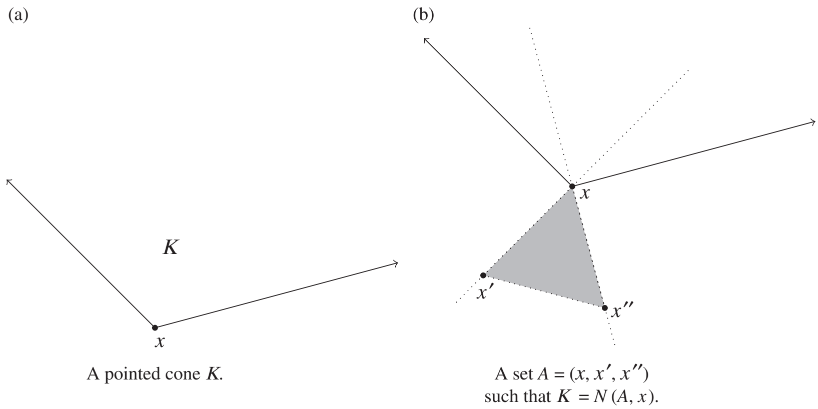

Fig. 7.1 Cones and decision problems.

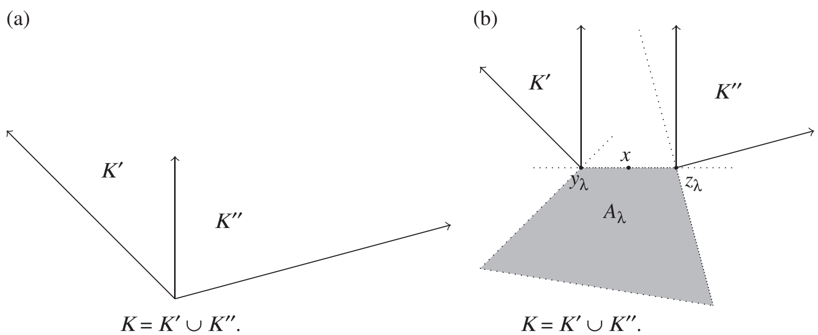

Fig. 7.2 Finite additivity.

$A' - \{x'\}$. Let $K = N(A,0) = N(A',0)$. Then we know that $\text{pos}(A) = N(K,0)$ and $\text{pos}(A') = N(K,0)$, so $A' \subseteq \text{pos}(A)$.[6] Since A' is finite and $0 \in A$, there is $\lambda > 0$ such that $\lambda A' \subseteq \text{conv}(A)$. Then monotonicity and linearity imply that

$$P_{A'}(0) = P_{\lambda A'}(0) \geq P_A(0).$$

The reverse argument establishes $P_{A'}(0) \leq P_A(0)$.

To show that μ is (finitely) additive we need to show that if K, K', and K'' are pointed polyhedral cones (meaning sets of the form $N(A,0)$) such that $K = K' \cup K''$, and the cones have disjoint interior, then $\mu(K) = \mu(K') + \mu(K'')$.[7] We give a sketch of the argument based on Figure 7.2. On the left of the figure we have a cone K that is the union of K' and K'', two cones with disjoint interior. By the previous graphical argument, one can find A' and A'' such that

[6] The notation pos$A$ stands for the set of all positive linear combinations of elements in A.

[7] The intersection of these cones must have probability zero, if we are to obtain a regular probability measure.

$K' = N(y, A')$ and $K'' = N(z, A'')$ for some points y and z. Consider x chosen on the line connecting y and z. Let A be such that $K = K' \cup K'' = N(x, A)$.

Let $A_\lambda = (1-2\lambda)A + \lambda A' + \lambda A''$, and let y_λ and z_λ be as in Figure 7.2(a). By linearity, $P_{A_\lambda}(y_\lambda) = P_{A'}(y) = \mu(K')$ and $P_{A_\lambda}(z_\lambda) = P_{A''}(z) = \mu(K'')$.

Let B be any closed ball such that the intersection of B with the extreme points of A is $\{x\}$. Then by choosing λ small enough we can guarantee that the only extreme points of A_λ in B are y_λ and z_λ. Hence the extremeness property of P implies that

$$P_{A_\lambda}(B) = P_{A_\lambda}(y_\lambda) + P_{A_\lambda}(z_\lambda) = \mu(K') + \mu(K'').$$

By mixture continuity, $P_A(B) = \lim_{\lambda \to 0} P_{A_\lambda}(B)$. Since B was arbitrary, $\mu(K) = P_A(x) = \mu(K') + \mu(K'')$.

The general argument for additivity is more sophisticated. It relies on basic results from convex analysis and the idea that $N(\sum_i \beta_i D_i, \sum_i \beta_i y_i)$ is independent of β_i so long as they are positive (think of the set of all rectangles with sides parallel to the axes).

Once μ has been established to be an additive measure of the set of all pointed cones, an extension argument is required to show that it is a probability measure on vectors u (utility indices).

7.4 CHAPTER REFERENCES

The two interpretations of random choice outlined at the beginning of the chapter are very standard. Many econometric applications of these models assume a population distribution of preferences (so-called individual unobserved heterogeneity). There is also substantial evidence that agents choose different alternatives when faced with the same choice problem: see, for example, Mosteller and Nogee (1951), Papandreou (1953), and Chipman (1960) for early discussions of this fact. Agranov and Ortoleva (2013) conduct an experiment in which subjects seem to deliberately randomize.

Theorem 7.2 is due to McFadden and Richter (1971, 1990) and Falmagne (1978). In particular, the equivalence between (I) and (II) is due to McFadden and Richter, while the equivalence between (I) and (III) is due to Falmagne. The earlier work of Block and Marschak (1960) had established that the non-negativity of the Block-Marschak polynomials was necessary for rationalization. Lemma 7.4 is from Falmagne (1978) (see his Theorem 3). Proposition 7.3 appears in Barberá and Pattanaik (1986). The idea of applying Möbius inversion to these problems first appears in Colonius (1984), see also Fiorini (2004) and Billot and Thisse (2005). Rota (1964) provides a classic discussion of Möbius inversion and the inclusion–exclusion principle.

Luce's model is presented in Luce (1959). The example of the buses and the car is due to Debreu (1960b), while a related example attributed to Savage is presented in Luce and Suppes (1965). This example motivates a model in which objects are first categorized before applying a Luce-style model, see Gul, Natenzon, and Pesendorfer (2014). There is plenty of experimental evidence where subjects exhibit violations of LIIA. One of the best-known such experiments is reported in Huber, Payne, and Puto (1982): it is called the attraction effect.

Theorem 7.6 is due to Block and Marschak (1960). The proof presented here follows Debreu (1960a). The example showing that there are rationalizable systems that do not conform to Luce's model is attributed by Block and Marschak (1960) to Paul Halmos.

Propositions 7.7 and 7.8 appear in McFadden (1974). He attributes Proposition 7.7 to Holman and Marley, who never published their result. Proposition 7.8 is part of a collection of uniqueness results regarding models of random utility: see Yellott (1977). Hausman and Wise (1978) investigate the related random utility model where error terms are normally distributed.

Theorem 7.9 is due to Gul and Pesendorfer (2006). Our informal discussion of the proof is largely taken from their paper.

There are other natural approaches to stochastic choice which we have not discussed here. For example, Machina (1985) suggests that the choice probabilities attributed to any set may be generated by maximizing a convex preference relation on the space of lotteries over that set.