22. Preferences and Risk

The important distinction between risk and uncertainty in economics is due to Frank Knight (1921). He applied the term *risk* to cases where the actual outcome is unknown, but the probabilities of possible outcomes can be quantified. He reserved the term *uncertainty* for situations where our lack of knowledge about the future could not be quantified. We call this latter case *Knightian uncertainty*, and use the term *uncertainty* to cover any situation where the future is unknown, regardless of its quantifiability.

This chapter and the next are about risk, quantifiable uncertainty. We start with a simple, but influential gambling problem—Nicolas Bernoulli's St. Petersburg paradox in section one. In section two, we set up a framework suitable for studying risk, a world of possible states and known probabilities. We define lotteries that may pay off differently in different states. Preferences and utility are defined over these lotteries. In section three, we present Von Neumann and Morgenstern's charactferization of expected utility, the centerpiece of the entire chapter. Section four examines Herstein and Milnor's (1953) abstract utility representation theorem, a theorem extends von Neumann and Morgenstern's results well beyond their original setting. While sections two and three focused on models with a finite number of states, section five examines distribution functions, allowing us to define lotteries on infinite state spaces. We build up enough probability theory to define expected utility on such a space. Section six investigates some of the technical complications that arise on infinite state spaces. The chapter closes with a short section that uses the Herstein and Milnor result to prove a utility representation theorem for infinite state spaces.

Outline:

1. ✓ Introduction to Risk: The St. Petersburg Paradox
2. ✓ Preferences over Lotteries
3. ✓ The von Neumann-Morgenstern Axioms
4. ✓ *Mixture Spaces
5. ✓ Distribution Functions and Infinite State Spaces
6. **Lebesgue Measure and Integration
7. **Expected Utility on Infinite State Spaces

22.1 Introduction to Risk: The St. Petersburg Paradox

The theory of decision-making under risk and uncertainty has a long history. One of the first serious attempts to address such a problem was that of Daniel Bernoulli in 1738. His solution has had an enormous impact on the theory of risk and uncertainty. Bernoulli introduced what we now call expected utility to resolve the St. Petersburg paradox. The paradox was a gambling problem created by his elder cousin, Nicolas Bernoulli (1687-1759) in a letter to Pierre Raymond de Montmort in 1713.[1]

[1] I include the dates since it is impossible to tell the various Nicolas Bernoulli's apart without them. Our Nicolas Bernoulli was son of Daniel Bernoulli's grandfather, also named Nicolas Bernoulli.

22.1.1 St. Petersburg Paradox

Example 22.1.1: St. Petersburg Paradox. The St. Petersburg paradox, created by Nicolas Bernoulli, involves the following type of gamble. I flip a coin. If it's heads, I pay you one ducat. If it's tails we double the stake and I flip again. We continue this until a head appears.

The payoff from the bet is zero as long as the coin comes up tails, and 2^{n-1} if the first head occurs on the n^{th}coin flip. The coin is presumed to be fair, with equal chances of heads and tails, so there is a 2^{-n} chance that the first head occurs at flip n. The expected payoff is then infinite:

$$\sum_{n=1}^{\infty} 2^{n-1} \times 2^{-n} = \sum_{n=1}^{\infty} \frac{1}{2} = +\infty.$$

The paradox is that people are not willing to pay very much to play this lottery, even though it has an infinite expected value.

22.1.2 Daniel Bernoulli Resolves the Paradox

Daniel Bernoulli (1738) proposed a resolution of the paradox by introducing what we now call an expected utility function. He demonstrated this in detail for the utility function $u(c) = \ln c$. He showed the expected utility was finite, and corresponded to a finite value of consumption.[2] Bernoulli also argued that this was generally true when marginal utility was decreasing.

We compute the expectation of utility, and find it is finite.

$$Eu = \sum_{n=1}^{\infty} 2^{-n} \ln 2^{n-1} = \ln 2 \left(\sum_{n=1}^{\infty} (n-1) 2^{-n} \right) = \ln 2. \tag{22.1.1}$$

In fact, the expectation is the same as the utility obtained from 2 ducats received with certainty. Today, this is called the *certainty equivalent* of the gamble.

[2] He got the idea of utility from Gabriel Cramer, of Cramer's Rule fame.

22.1.3 Computing Infinite Sums

Computation of sums as in equation 22.1.1 is often not taught in calculus classes. Here's a trick that can be helpful in for calculating such sums. We start with a sum that is well-known. For $|a| < 1$, define

$$f(a) = \sum_{n=1}^{\infty} a^n = \frac{a}{1-a}$$

The infinite sum converges both absolutely and uniformly on $\overline{B_r(0)}$ for $r < 1$. Then we compute the derivative term-by-term, obtaining

$$f'(a) = \sum_{n=1}^{\infty} n a^{n-1} = \frac{1}{(1-a)^2}$$

which also converges absolutely and uniformly on $\overline{B_r(0)}$ for $r < 1$. We can combine the previous two equations, writing

$$\sum_{n=1}^{\infty} (n-1)a^n = af'(a) - f(a) = \frac{a}{(1-a)^2} - \frac{a}{1-a} = \frac{a^2}{(1-a)^2}.$$

Applying this to the case above, where $a = 1/2$, we obtain

$$\sum_{n=1}^{\infty} (n-1)2^{-n} = 1.$$

By taking higher derivatives, sums of the form $\sum_{n=1}^{\infty} n^k a^n$ can easily be calculated when $|a| < 1$.

22.1.4 More on the St. Petersburg Paradox

Equation 22.1.1 implies that someone with this utility function would not pay an infinite amount for the gamble. In fact, it is worth the same as 2 ducats with certainty.

We can also calculate the maximum a gambler with wealth w would be willing to pay for this gamble. If the cost of the gamble is $c < w$, then we must compare not taking the gamble, in which case the gambler has utility $\ln w$, and taking the gamble, which yields utility

$$\sum_{n=1}^{\infty} 2^{-n} \ln(w - c + 2^{n-1}).$$

The c that makes both quantities equal is the maximum our gambler would pay to take the St. Petersburg gamble. ◀

Further consideration of the issues raised by the St. Petersburg paradox eventually led to the development of both expected utility theory and decision theory.

22.2 Preferences over Lotteries

In order to model risk, we start with a set of possible states of the world. We don't know what will actually occur, only the range of possibilities. We start with the case where there are finitely many possibilities. Let $\mathbf{S} = \{1, \ldots, S\}$ denote the set of possible states. We will often consider the state to be synonymous with the income received in that state, but this is not required.

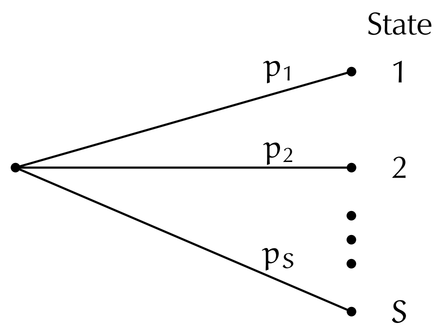

Figure 22.2.1: Tree diagram for a simple lottery.

A *simple lottery* $L = (p_1, \ldots, p_S)$ is an assignment of a probability p_s to each state s such that $0 \le p_s \le 1$ and $\sum_{s=1}^{S} p_s = 1$, as illustrated by the tree diagram in Figure 22.2.1. The probabilities may be objective, or may represent the beliefs of our consumer.[3] Let $\mathfrak{L}(\mathbf{S})$ denote the set of simple lotteries on $\mathbf{S}$.

[3] Savage (1954) showed how both preferences and beliefs about probabilities can be derived from consumer behavior. This is examined in section 24.1.

22.2.1 Compound Lotteries: Mixing Simple Lotteries

One way to combine lotteries (or any probability distributions) is by mixing them. A *compound lottery* is a lottery over lotteries. If $L_k = (p_s^k)$ are K simple lotteries, we can form a compound lottery $L = \big((q_1, L_1), \ldots, (q_K, L_K)\big)$ with $0 \le q_k \le 1$ and $\sum_k q_k = 1$. We interpret this as yielding the lottery L_k with probability q_k, as illustrated in Figure 22.2.2. More generally, when we have two probability distributions, we can always form a new distribution in this way, by taking one distribution with probability a and the other with probability $(1 - a)$ for $a \in [0, 1]$. This is referred to as a *mixture*. By repeating the mixing process on lotteries we can generate any compound lottery.

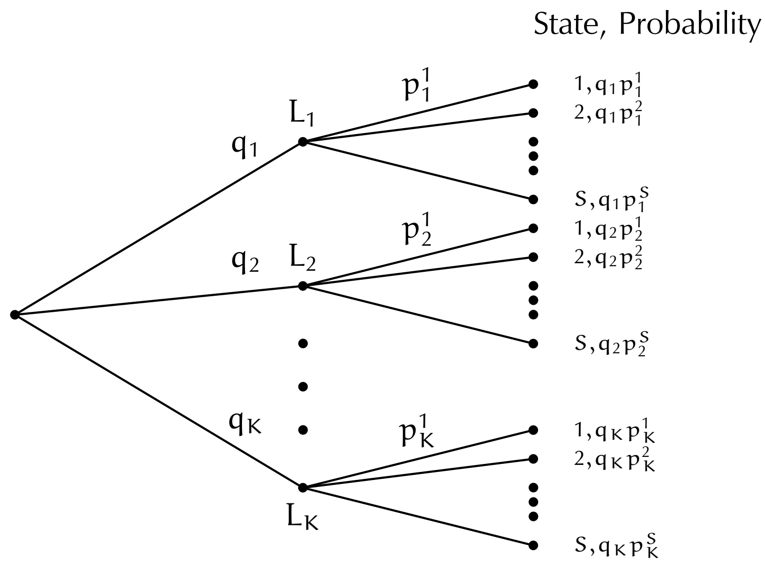

Figure 22.2.2: Tree diagram for a compound lottery. The probabilities for state 1 are shown. We add them to find the probability that state 1 occurs in the compound lottery, which is $\sum_k q_k p_1^k$.

22.2.2 Reduction of Compound Lotteries

When the lotteries L_k are simple lotteries, we can easily compute the probability that each state occurs in the compound lottery $L = \big((q_1, L_1), \dots, (q_K, L_K)\big)$. The probability of state s is $r_s = \sum_k q_k p^k_s$ (state 1 is illustrated on Figure 22.2.2). Note that $0 \le r_s \le 1$ and

$$\sum_s r_s = \sum_s \sum_k q_k p^k_s = \sum_k q_k \left(\sum_s p^k_s \right) = \sum_k q_k = 1,$$

so the r_s are probabilities. The simple lottery $(r_1, \dots, r_S)$ is called the *reduced lottery* associated with L. In fact, any finitely compounded lottery can be reduced to a simple lottery.

For $0 \le \alpha \le 1$ and $L, L' \in \mathfrak{L}(\mathbf{S})$, we use the notation $\alpha L \oplus (1-\alpha)L'$ to denote the *mixture* or *compound lottery* giving lottery L with probability α and lottery L' with probability $(1-\alpha)$. Notice that any mixture of elements of $\mathfrak{L}(\mathbf{S})$ reduces to an element of $\mathfrak{L}(\mathbf{S})$. In other words, $\mathfrak{L}(\mathbf{S})$ is closed under mixtures.

Compound lotteries have three important properties that are used in reduction. For all $L, L', L'' \in \mathfrak{L}(\mathbf{S})$ and $0 \le \alpha, \beta \le 1$.

1. Trivial Mixtures. $(1)L \oplus (0)L' = L$.
2. Symmetry. $\alpha L \oplus (1-\alpha)L' = (1-\alpha)L' \oplus \alpha L$.
3. Mixture Distributive Law.

$$\alpha\big[\beta L \oplus (1-\beta)L'\big] \oplus (1-\alpha)L' = (\alpha\beta)L \oplus (1-\alpha\beta)L'$$

22.2.3 Consequentialism

We will make the *consequentialist* assumption that only the probabilities of the final states matter for utility. The way in which those probabilities are obtained does not affect consumer preference. In particular, it doesn't matter whether the uncertainty is resolved all at once or incrementally, first through one lottery, then through another.[4]

Consequentialism means that all lotteries on $\mathbf{S}$ can be thought of as simple lotteries. And simple lotteries can be represented by points in $\mathbb{R}^S$ if we consider the probability p_s as the s coordinate of the lottery. The restriction that $p_s \geq 0$ and $\sum_s p_s = 1$ amounts to saying we are looking at points in the $(S-1)$-dimensional simplex $\Delta_S = \{\mathbf{p} \in \mathbb{R}^S : p_s \geq 0, \sum_s p_s = 1\}$.[5] When lotteries are thought of as points in the probability simplex Δ_S, the points are convex combination of the states, the corners, with the weights in the convex combination indicating the probabilities.

Preferences will be defined over the lottery space $\mathfrak{L}(\mathbf{S})$, or equivalently, the probability simplex Δ_S.

[4] See Hammond (1988, 1989) for more on the relation between consequentialism and expected utility.

[5] Since Δ_S is $(S-1)$-dimensional, it is often written Δ_{S-1}. In this chapter we use Δ_S instead to emphasize the number of states.

22.2.4 Expected Utility

A utility function has the *expected utility form* if there are $u_1, \dots, u_S$ such that $L = (p_1, \dots, p_s)$ has utility

$$Eu(L) = \sum_s p_s u_s$$

where u_s is the utility in state s. We sometimes write $u(s)$ instead of u_s.

The utility of a lottery L is then the expectation of u under the probability distribution defined by the lottery L. Von Neumann and Morgenstern (1944) characterized preferences with such utility functions. For this reason, they are sometimes called *von Neumann-Morgenstern utility functions*.

22.2.5 Expected Utility: A Simple Example

Example 22.2.3: A Simple Example. Suppose the states are whether or not you get a new job. You earn \$62,500 in your old job (state 1) and would earn \$78,400 if you get the new job (state 2).

If your indirect utility function is $u(m) = \sqrt{m}$, the utilities would be $\sqrt{62500} = 250$ in state 1 and $\sqrt{78400} = 280$ in state 2. If the probability of getting the new job is 25% (state 2), you have a 75% chance of remaining in your old job (state 1). The lottery you face is $(p_1, p_2) = (.75, .25)$ and your expected utility is

$$Eu(L) = .75(250) + .25(280) = 257.5.$$

In comparison, if you knew with certainty that you would stay in your old job your expected utility would be $\sqrt{62500} = 250$. ◀

22.2.6 Transformations of Expected Utility

The expected utility form is not preserved under arbitrary increasing transformations of utility. That is, it is usually false that $f\big(Eu(L)\big) = E\big((f \circ u)(L)\big)$ for all lotteries L. Only increasing affine transformations preserve the expected utility form. These transformations can be written $\varphi(u) = au + b$ for $a > 0$. In other words, expected utility is not ordinal, and can be normalized to be cardinal by picking values for any two lotteries with differing expected utility.

22.2.7 Examples of Expected Utility

Here are some examples of expected or von Neumann-Morgenstern utility functions.

Example 22.2.4: Selected von Neumann-Morgenstern Utility Functions.

1. Let $u(w) = 2w$ and let $L = \left(\frac{5}{12}, \frac{1}{3}, \frac{1}{4}\right)$ be defined on states $w = 1, 2, 3$. Then $Eu(L) = \frac{5}{12}(2) + \frac{1}{3}(4) + \frac{1}{4}(6) = \frac{44}{12} = \frac{11}{3}$. Since $Eu(L) = u(11/6)$, the certainty equivalent of this lottery is a payoff of $11/6$.
2. Let $u(w) = w^2$ and $L = \left(\frac{3}{4}, \frac{1}{12}, \frac{1}{6}\right)$ be defined on states $w = 1, 2, 3$. Then $Eu(L) = \frac{3}{4}(1) + \frac{1}{12}(4) + \frac{1}{6}(9) = \frac{31}{12}$.
3. Let $u(w) = \sqrt{w}$ and let L be a lottery over wealth levels $\{1, \ldots, 50\}$ with probabilities p_w. Then $Eu(L) = \sum_{w=1}^{50} p_w \sqrt{w}$. If $p_w = 1/50$, we can approximate $Eu(L)$ by using integrals.

$$\frac{1}{50}\int_0^{50} w^{1/2}\, dw \leq Eu(L) \leq \frac{1}{50}\int_1^{51} w^{1/2}\, dw.$$

A short calculation shows that $4.71 \leq Eu(L) \leq 4.84$.

◄

22.2.8 Marschak-Machina Triangle I

When preferences are defined by an expected utility function, we can draw indifference curves in the probability simplex Δ_S. Letting $\mathbf{u} = (u_1, \dots, u_S)$, we can write $\mathrm{E}u(L) = \mathbf{p}\cdot\mathbf{u}$. This means that the indifference curves are parallel hyperplanes perpendicular to $\mathbf{u}$.

When there are three states and the preference order is strict, we can visualize this by using a *Marschak-Machina triangle*.[6] Let $\mathbf{S} = \{s_1, s_2, s_3\}$ with $s_1 \prec s_2 \prec s_3$. To represent the lottery with probabilities $p_1 s_1 \oplus p_2 s_2 \oplus p_3 s_3$ in $\mathbb{R}^2$, we put s_1 at $(1, 0)$, s_2 at the origin, and s_3 at $(0, 1)$. Then the point (p_1, p_3) with $p_i \geq 0$ and $p_1 + p_3 \leq 1$ represents the lottery $p_1 s_1 \oplus p_2 s_2 \oplus p_3 s_3$ where $p_2 = 1 - p_1 - p_3$. Marschak-Machina triangles can also be used to represent lotteries over any three states in a larger state space $\mathbf{S}$.

[6] See Marschak (1950). Machina has repeatedly made use of them, e.g. Machina (1987).

22.2.9 Marschak-Machina Triangle II

The parallel indifference hyperplanes translate to parallel lines in $\mathbb{R}^2$. To find them, suppose $(p_1, p_3) \sim (p'_1, p'_3)$. Then expected utility obeys

$$p_1 u_1 + (1 - p_1 - p_3) u_2 + p_3 u_3 = p'_1 u_1 + (1 - p'_1 - p'_3) u_2 + p'_3 u_3$$

where $u_i = u(s_i)$. We then find

$$(p_1 - p'_1)(u_1 - u_2) = (p'_3 - p_3)(u_3 - u_2),$$

so

$$\frac{\Delta p_3}{\Delta p_1} = \frac{p'_3 - p_3}{p'_1 - p_1} = -\frac{u_3 - u_2}{u_1 - u_2}.$$

This shows that the indifference curves are parallel straight lines with slope $(u_2 - u_3)/(u_1 - u_2)$. Moreover, they are positively sloped because $u_1 < u_2 < u_3$. This is illustrated in Figure 22.2.5.

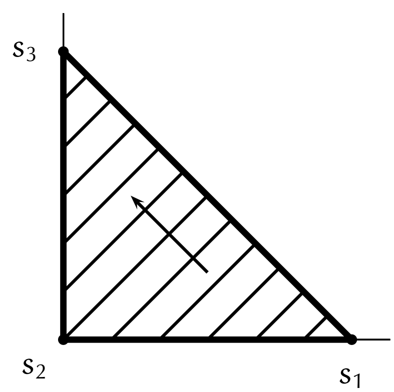

Figure 22.2.5: Expected Utility in a Marschak-Machina Triangle with 3 States. The heavy lines delimit the probability simplex Δ_3. The parallel upward sloping lines are indifference curves for a case with $s_1 \prec s_2 \prec s_3$. The arrow shows the direction of increasing utility. In general, the common slope of the indifference curves depends on the utility of s_1, s_2, and s_3.

22.3 The von Neumann-Morgenstern Axioms 03/24/22

Von Neumann and Morgenstern (1944) characterized expected utility via a set of axioms concerning preferences over $\mathfrak{L}(\mathbf{S})$. Any expected utility function obeys those axioms, and any function obeying those axioms can be written in expected utility form. For this reason, they are often called von Neumann-Morgenstern utility.

von Neumann-Morgenstern Axioms. A binary relation $\succsim$ defined on the lottery space $\mathfrak{L}(\mathbf{S})$ obeys the *von Neumann-Morgenstern Axioms* if

1. Preference Order: $\succsim$ is a preference order on lotteries in $\mathfrak{L}(\mathbf{S})$.
2. Segment Continuity: For all lotteries $L, L', L'' \in \mathfrak{L}(\mathbf{S})$, the sets $\{\alpha \in [0,1] : \alpha L \oplus (1-\alpha)L' \succsim L''\}$ and $\{\alpha \in [0,1] : L'' \succsim \alpha L \oplus (1-\alpha)L'\}$ are closed sets.
3. Independence: For all lotteries $L, L', L'' \in \mathfrak{L}(\mathbf{S})$ and every $0 < \alpha < 1$ we have $L \succsim L'$ if and only if

$$\alpha L \oplus (1-\alpha)L'' \succsim \alpha L' \oplus (1-\alpha)L''.$$

Although segment continuity is weaker than assuming that preferences are continuous, it still allows us to prove that preferences have a continuous utility representation when $\mathfrak{L}(\mathbf{S}) = \Delta_S$.

22.3.1 Consequences of Independence

Independence has some implications.

Proposition 22.3.1. *Suppose $\succsim$ is a preference order on $\mathfrak{L}$ that obeys the Independence Axiom.*

1. *$\succsim$ is convex with respect to mixtures.*
2. *If $0 < \alpha < 1$, then $L \sim L'$ if and only if $\alpha L \oplus (1-\alpha)L'' \sim \alpha L' \oplus (1-\alpha)L''$ for all $L'' \in \mathfrak{L}$.*
3. *If $0 < \alpha < 1$, then $L \succ L'$ if and only if $\alpha L \oplus (1-\alpha)L'' \succ \alpha L' \oplus (1-\alpha)L''$ for all $L'' \in \mathfrak{L}$.*
4. *If $L \sim L'$ and $0 < \alpha < 1$, then $L \sim \alpha L' \oplus (1-\alpha)L$.*

Proof. Recall that Independence says that for all lotteries $L, L', L'' \in \mathfrak{L}(\mathbf{S})$ and every $0 < \alpha < 1$ we have $L \succsim L'$ if and only if

$$\alpha L \oplus (1-\alpha)L'' \succsim \alpha L' \oplus (1-\alpha)L''.$$

For (1), set $L'' = L'$. If $L \succsim L'$, independence implies $\alpha L \oplus (1-\alpha)L' \succsim L'$, which is convexity.

For (2), $L \sim L'$ if and only if $L \succsim L'$ and $L' \succsim L$. Apply independence to both relations to obtain $\alpha L \oplus (1-\alpha)L'' \succsim \alpha L' \oplus (1-\alpha)L''$ and $\alpha L' \oplus (1-\alpha)L'' \succsim \alpha L \oplus (1-\alpha)L''$ for all $L'' \in \mathfrak{L}$. Combining these two relations yields the result.

Item (3) follows by combining Independence and (2) using the fact that $L \succ L'$ if and only if $L \succsim L'$ and not $L \sim L'$.

We obtain (4) by setting $L'' = L$ in (2). $\square$

Item (1) only refers to convexity relative to **mixtures** of lotteries. In other words, this is convexity in terms of probabilities. It does not say anything about convexity in terms of **payoffs**. We will see in Chapter 23 that attitudes toward risk are related to convexity or concavity in terms of payoffs, not probabilities. Since Condition (1) does not address this, it is consistent with many types of attitude toward risk.

Item (4) means that independence implies that indifference curves are flats in lottery space, as illustrated in Figure 22.2.5.

22.3.2 Expected Utility Theorem

Expected Utility Theorem (von Neumann-Morgenstern). *Let the state space* S *be finite. A preference order on* $\mathfrak{L}(S)$ *obeys segment continuity and independence if and only if it has an expected utility representation. Moreover, if* v *is another expected utility representation, then there are* $a > 0$ *and* $b \in \mathbb{R}$ *with* $v = au + b$*.*

As we noted when we introduced the expected utility form, Von Neumann-Morgenstern utility is not preserved under arbitrary increasing transformations. Only increasing affine transformations preserve the expected utility property.

22.3.3 Proof of Expected Utility Theorem, part I

Proof. Part I (if): Suppose preferences have an expected utility representation. This automatically defines a preference order over lotteries.

Now let $L^k = (p^k_s)$ for $k = 1, 2, 3$ be lotteries. The sets

$$\left\{\alpha \in [0,1] : \alpha L^1 \oplus (1-\alpha)L^2 \succsim L^3\right\}$$
$$= \left\{\alpha \in [0,1] : \sum_s [\alpha p^1_s + (1-\alpha)p^2_s]u_s \geq \sum_s p^3_s u_s\right\}$$

and

$$\left\{\alpha \in [0,1] : \alpha L^1 \oplus (1-\alpha)L^2 \precsim L^3\right\}$$
$$= \left\{\alpha \in [0,1] : \sum_s [\alpha p^1_s + (1-\alpha)p^2_s]u_s \leq \sum_s p^3_s u_s\right\}$$

are closed because addition and multiplication are continuous, showing segment continuity.

Further,

$$\alpha L^1 \oplus (1-\alpha)L^3 \succsim \alpha L^2 \oplus (1-\alpha)L^3$$

if and only if

$$\sum_s \left[\alpha p^1_s + (1-\alpha)p^3_s\right]u_s \geq \sum_s \left[\alpha p^2_s + (1-\alpha)p^3_s\right]u_s$$

which holds if and only if

$$\sum_s p^1_s u_s \geq \sum_s p^2_s u_s$$

or equivalently $L^1 \succsim L^2$.

22.3.4 Proof of Expected Utility Theorem, part IIA

Part II (only if): Suppose preferences obey segment continuity and independence. Pick a best state s^* and worst state s_* (ties are okay). Any lottery will be at least as good as receiving the worst state with certainty, and no better than receiving the best state with certainty. Independence implies $s^* \succsim L \succsim s_*$ for all $L \in \mathfrak{L}(\mathbf{S})$ (see Exercise 22.4.5).

If $s^* \sim s_*$, then utility is constant and we are done. Otherwise, $s^* \succ s_*$, in which case $s^* \succ \alpha s^* \oplus (1-\alpha)s_*$ for $0 < \alpha < 1$ by part (3) of Proposition 22.3.1.

For $L \in \mathfrak{L}(\mathbf{S})$, let $A^+ = \{\alpha \in [0,1] : \alpha s^* \oplus (1-\alpha)s_* \succsim L\}$ and $A^- = \{\alpha \in [0,1] : \alpha s^* \oplus (1-\alpha)s_* \precsim L\}$. Both A^+ and A^- are closed subsets of $[0,1]$ by segment continuity. They are non-empty as $1 \in A^+$ and $0 \in A^-$. The interval $[0,1]$ is connected, so $A^+ \cap A^-$ is non-empty.

This tells us that there is an α so that

$$L \sim \alpha s^* \oplus (1-\alpha)s_*.$$

We next show that for every lottery L, that α is unique.

22.3.5 Proof of Expected Utility Theorem, part IIB

Now suppose $L \sim \alpha s^* \oplus (1-\alpha)s_*$ and $L \sim \beta s^* \oplus (1-\beta)s_*$ with $\alpha > \beta$. Then we can write $\beta s^* \oplus (1-\beta)s_*$ as the strict convex combination

$$\beta s^* \oplus (1-\beta)s_* = \left(\frac{\beta}{\alpha}\right)\left(\alpha s^* \oplus (1-\alpha)s_*\right) \oplus \left(1-\frac{\beta}{\alpha}\right)s_*. \tag{22.3.2}$$

Here $\alpha > 0$, and we previously established that $\alpha s^* \oplus (1-\alpha)s_* \succ s_*$. Apply Proposition 22.3.1(3) and symmetry to show that the mixture in the right-hand side of equation 22.3.2 is strictly worse than $\alpha s^* \oplus (1-\alpha)s_*$, contradicting transitivity. Thus there is a unique α with $\alpha s^* \oplus (1-\alpha)s_* \sim L$. We define the utility of L to be that α. In other words, we define

$$u(L) = \alpha$$

to mean

$$\alpha s^* \oplus (1-\alpha)s_* \sim L.$$

22.3.6 Proof of Expected Utility Theorem, part IIC

This definition implies that u is continuous as both

$$u^{-1}[\alpha, \infty) = \left\{L'' : L'' \succsim \alpha s^* \oplus (1-\alpha)s_*\right\}$$

and

$$u^{-1}(-\infty, \alpha] = \left\{L'' : L'' \precsim \alpha s^* \oplus (1-\alpha)s_*\right\}$$

are closed sets by segment continuity.

We complete the main result by showing u has the expected utility form. Suppose $u(L) = \alpha$ and $u(L') = \beta$ and consider $\lambda L \oplus (1-\lambda)L'$ for $0 < \lambda < 1$. By the definition of u,

$$\beta s^* \oplus (1-\beta)s_* \sim L'$$

By symmetry and independence

$$\lambda L \oplus (1-\lambda)L' \sim \lambda L \oplus (1-\lambda)\big(\beta s^* \oplus (1-\beta)s_*\big).$$

Also by symmetry and independence,

$$\begin{aligned}
&\lambda L \oplus (1-\lambda)\big(\beta s^* \oplus (1-\beta)s_*\big) \\
&\quad \sim \lambda(\alpha s^* \oplus (1-\alpha)s_*) \oplus (1-\lambda)(\beta s^* \oplus (1-\beta)s_*) \\
&\quad = (\lambda\alpha + (1-\lambda)\beta)s^* \oplus \big(1 - (\lambda\alpha + (1-\lambda)\beta)\big)s_*.
\end{aligned}$$

where the last line uses the mixture distribution law to reduce the compound lottery. The resulting lottery has utility $\lambda\alpha + (1-\lambda)\beta$, so we conclude

$$u\big(\lambda L \oplus (1-\lambda)L'\big) = \lambda u(L) + (1-\lambda)u(L')$$

for any lotteries L and L'. To show u has the expected utility form, we write $L = \sum_s p_s s$ and repeatedly use the above to get $u(L) = \sum_s p_s u(s)$.

22.3.7 Proof of Expected Utility Theorem, part III

Rest of proof. Part III (moreover): Suppose ν is another expected utility representation and that s^* and s_* are best and worst states in $\mathbf{S}$ For every $L \in \mathfrak{L}(\mathbf{S})$, part II tells us there is a unique $\alpha \in [0, 1]$ such that $L \sim \alpha s^* \oplus (1 - \alpha)s_*$ and that $u(L) = \alpha$.

Clearly $u(s^*) = 1$ and $u(s_*) = 0$.

Since $L \sim \alpha s^* \oplus (1 - \alpha)s_*$, we can write

$$\begin{aligned} \nu(L) &= \alpha\nu(s^*) + (1 - \alpha)\nu(s_*) \\ &= u(L)\nu(s^*) + \big(1 - u(L)\big)\nu(s_*). \end{aligned}$$

Set $a = \nu(s^*) - \nu(s_*)$ and $b = \nu(s_*)$. Notice that a and b are independent of L. This allows us to rewrite $\nu(L)$ as

$$\nu(L) = au(L) + b.$$

□

Since independence and segment continuity imply an expected utility representation, while continuity alone only implies a continuous utility representation, the special properties of expected utility derive from independence. In particular, the Independence Axiom forces the indifference curves to be parallel straight lines, a fact which emphasizes the strength of the Independence Axiom.

22.3.8 Not Expected Utility

To be a von Neumann-Morgenstern utility, a function must be linear in probabilities. To find a utility function on lotteries that fails to be a von Neumann-Morgenstern utility, we simply use a function that is non-linear in probabilities, as in the following example.

Example 22.3.2: Not a von Neumann-Morgenstern Utility Function. Consider $\mathbf{S} = \{1, 2\}$. We will write lotteries as $L = (p_1, p_2)$. Suppose utility is $u(L) = \frac{1}{4}p_1^2 + \frac{1}{4}p_2^2$. Define $L = (1, 0)$ and $L' = (0, 1)$. We compute

$$\begin{aligned} u(L) &= 1/4 \\ u(L') &= 1/4. \end{aligned}$$

Set $L'' = \frac{1}{2}L \oplus \frac{1}{2}L'$. The resulting probabilities are $L'' = \left(\frac{1}{2}, \frac{1}{2}\right)$. Then

$$u(L'') = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}.$$

It follows that $L'' \prec L$. This violates the Independence Axiom which requires that $L'' = \frac{1}{2}L' \oplus \frac{1}{2}L \succsim \frac{1}{2}L \oplus \frac{1}{2}L = L$ because $L' \succsim L$. ◀

22.3.9 Defining Expected Utility via Lotteries I

We normally think of expected utility as being derived from utility in each state. However, the consequentialist assumption that the expected utility of a compound lottery is the same as the utility of the resulting reduced lottery means that we need only know the utility on a spanning set of lotteries in order to derive expected utility. Example 22.3.3 shows how it works.

Example 22.3.3: Lottery-based Expected Utility.

Define lotteries by

$$\begin{aligned} L_1 &= (1/2, 1/3, 1/6), \\ L_2 &= (1/3, 1/6, 1/2), \\ L_3 &= (1/6, 1/2, 1/3). \end{aligned}$$

Suppose $Eu(L_1) = 5$, $Eu(L_2) = 3$, and $Eu(L_3) = 1$.

We now compute the expected utility of the certain lottery $L = (1, 0, 0)$. First write L as a linear combination of the L_k. We do that by solving the equation

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/6 & 1/2 \\ 1/6 & 1/2 & 1/3 \end{pmatrix} x. \tag{22.3.3}$$

This yields $x = (7/3, 1/3, -5/3)^T$. Thus $L = \frac{7}{3}L_1 + \frac{1}{3}L_2 - \frac{5}{3}L_3$.

22.3.10 Defining Expected Utility via Lotteries II

We need a relationship that only involves positive multiples of the lotteries, so we rewrite this as $L + \frac{5}{3}L_3 = \frac{7}{3}L_1 + \frac{1}{3}L_2$.

The next step is to convert this to convex combinations (compound lotteries). We do this by multiplying by 3/8 to write both sides as compound lotteries, obtaining

$$\frac{3}{8}L \oplus \frac{5}{8}L_3 = \frac{7}{8}L_1 \oplus \frac{1}{8}L_2.$$

This trick will not work for arbitrary vectors, but works here because we are dealing with lotteries.[7] Now expected utility obeys

$$\frac{3}{8}Eu(L) + \frac{5}{8}Eu(L_3) = \frac{7}{8}Eu(L_1) + \frac{1}{8}Eu(L_2)$$

or

$$Eu(L) = \frac{7}{3}Eu(L_1) + \frac{1}{3}Eu(L_2) - \frac{5}{3}Eu(L_3) = 35/3 + 1 - 5/3 = 11$$

It follows that $Eu(L) = 11$. Notice that in the end, we used the original linear combination that gave L in terms of the L_i. ◀

[7] The key is that the matrix in equation 22.3.3 is both non-negative and doubly stochastic. Then the inverse is doubly stochastic, so the terms of $\mathbf{x}$ sum to one.

22.3.11 Theorem on Lottery-Defined Utility

The method above can be used to show that if you can write one lottery as a linear combination of other lotteries, then the expected utility of that lottery is the same linear combination of the expected utilities of the other lotteries. That inspires the following theorem.

Theorem 22.3.4. *Suppose* L_k, $k = 0, \dots, K$ *are lotteries in* $\mathfrak{L}(\mathbf{S})$ *and that as vectors,* $\sum_{k=1}^{K} \alpha_k L_k = L_0$. *Then* $\sum_{k=1}^{K} \alpha_k Eu(L_k) = Eu(L_0)$.

Proof. Let $L_k = (p_1^k, \dots, p_S^k)$ for $k = 0, \dots, K$. We write $L_0 + \sum_k^- \alpha_k L_k = \sum_k^+ \alpha_k L_k$ where $\sum^+$ and $\sum^-$ respectively denote the sums over k with $\alpha_k > 0$ and with $\alpha_k < 0$. Since all of the terms on the left-hand side are non-negative, and $L_0 > \mathbf{0}$, we know the right-hand side must be positive.

Let $\alpha = \sum_k^+ \alpha_k > 0$, and $\beta_k = \alpha_k/\alpha$ for $k = 1, \dots, K$ and $\beta_0 = 1/\alpha$. Then $\beta_0 L_0 + \sum_k^- \beta_k L_k = \sum_k^+ \beta_k L_k$. For each s, $\beta_0 p_s^0 + \sum_k^- \beta_k p_s^k = \sum_k^+ \beta_k p_s^k$. Summing over s and interchanging the order of summation, we obtain $\beta_0 + \sum_k^- \beta_k = \sum_k^+ \beta_k = 1$. This means we can consider both sides of the equation as compound lotteries.

As compound lotteries, we can compute expected utility of each. This yields $\beta_0 Eu(L_0) + \sum_k^- \beta_k Eu(L_k) = \sum_k^+ \beta_k Eu(L_k)$. Finally, we multiply by α and reassemble the summation to obtain $Eu(L_0) = \sum_{k=1}^{K} \alpha_k Eu(L_k)$. □

22.4* Mixture Spaces

Although the Expected Utility Theorem of von Neumann and Morgenstern only applies to finite state spaces, we will see in section 22.7 that expected utility can also be used on infinite state spaces. There are representation theorems that handle such cases. Although expected utility cannot be defined when probability cannot be applied, we can get a similar type of utility when it is possible to mix outcomes in a way that looks like a probability mixture, regardless of whether we are dealing with a finite or infinite space.

We start with the generalization of probability mixtures, the abstract concept of a mixture. Its natural home will be a mixture space.

Mixture Space. A pair $(\mathfrak{M}, \oplus)$ where $\mathfrak{M}$ is a set and $\oplus$ is a binary operation on $\mathfrak{M}$ is a *mixture space* if for any $\alpha \in [0,1]$ and $L, L' \in \mathfrak{M}$, we have another element of $\mathfrak{M}$, the *mixture* $\alpha L \oplus (1-\alpha)L'$ obeying:

1. Trivial Mixtures. $(1)L \oplus (0)L' = L$.
2. Symmetry. $\alpha L \oplus (1-\alpha)L' = (1-\alpha)L' \oplus \alpha L$.
3. Mixture Distributive Law.

$$\alpha[\beta L \oplus (1-\beta)L'] \oplus (1-\alpha)L' = (\alpha\beta)L \oplus (1-\alpha\beta)L'.$$

For any $L, L' \in \mathfrak{M}$ and $a, b \in [0,1]$.

We previously noted that conditions (1)–(3) can be used to reduce compound lotteries to simple lotteries.

22.4.1 Mixture Spaces: Examples

Here are some examples of mixture spaces.

Example 22.4.1: Three Examples of Mixture Spaces.

1. Any convex set is a mixture space when $\alpha x \oplus (1-\alpha)y$ is defined as $\alpha x + (1-\alpha)y$. The set of simple lotteries on $\mathbf{S}$ is also a mixture space under consequentialism, when the compound lotteries formed by $\oplus$ are reducible to simple lotteries.
2. We can also form a mixture space from a subset of the lottery space $\mathfrak{L}(\mathbf{S})$. Let $\mathbf{S}' = \{L_1, \ldots, L_T\}$ where each $L_s \in \mathfrak{L}(\mathbf{S})$. Then $\mathfrak{L}(\mathbf{S}')$ with $\oplus$ inherited from $\mathfrak{L}(\mathbf{S})$ is a mixture space.
3. Let $\mathcal{F}$ be a σ-algebra of sets of $\mathbf{S}$. Then define $\mathfrak{L}(\mathbf{S})$ to be the set of probability measures on $(\mathbf{S}, \mathcal{F})$. Now define the mixture $\alpha\mu \oplus (1-\alpha)\nu$ as the probability measure $\alpha\mu + (1-\alpha)\nu$. This makes $(\mathfrak{L}(\mathbf{S}), \oplus)$ a mixture space. Similarly, if $\mathcal{F}$ is an algebra of subsets of $\mathbf{S}$, the space of finitely additive probability measures on $(\mathbf{S}, \mathcal{F})$ is mixture space when mixtures are formed the same way as probability measures.

◀

22.4.2 Basic Properties of Mixtures

The list of properties of $\oplus$ is the minimum required, but it implies other properties. For example, combining symmetry and the trivial mixture property yields the other trivial mixture property, that $(0)L \oplus (1)L' = L'$. Notice that the distributive law shows how to reduce the mixture of a mixture with a lottery. However, these properties will also allow us to reduce the mixture of two mixtures, as shown below in Proposition 22.4.2.

All this becomes simpler when representing lotteries as points in Δ_S. Then mixtures are convex combinations of points in Δ_S and we know how to reduce all types of mixtures.

Abstract mixtures are only required to follow the three rules in the definition. They includes mixture spaces where the points are not probabilities and the mixtures are not convex combinations. This considerably complicates matters. In particular, it makes it harder to reduce mixtures to simpler mixtures in an abstract mixture space.

It is still possible to show that they obey some of the other characteristics of convex combinations. For example, it is possible to show that a mixture of a single lottery, which yields the same lottery no matter what, is the same as the original lottery.

It is also possible to strengthen the distributive law so that it works on mixtures of mixtures. This allows us to reduce mixtures of mixtures to simple mixtures. These and the basic mixture properties will be all we need in subsequent theorems.

Proposition 22.4.2. *If $\mathfrak{M}$ is a mixture space, then for any $L, L' \in \mathfrak{M}$ and $\alpha, \beta, \gamma \in [0,1]$,*

1. *Certain Mixtures.* $\alpha L \oplus (1-\alpha)L = L$,
2. *Mixture Reduction Law.*

$$\begin{aligned}\alpha\big[\beta L \oplus (1-\beta)L'\big] &\oplus (1-\alpha)\big[\gamma L \oplus (1-\gamma)L'\big] \\ &= \big[\alpha\beta + (1-\alpha)\gamma\big]L \oplus \big[\alpha(1-\beta) + (1-\alpha)(1-\gamma)\big]L'.\end{aligned}$$

22.4.3 Proof of Proposition 22.4.2

Proof. For (1),

$$\begin{aligned}\alpha L \oplus (1-\alpha)L &= \alpha(1L \oplus 0L) \oplus (1-\alpha)L \\ &= \alpha(0L \oplus 1L) \oplus (1-\alpha)L \\ &= 0L \oplus 1L = L\end{aligned}$$

by successively using triviality, symmetry, the mixture distributive law (with $L = L'$, $\beta = 0$, and $\gamma = 1$), and triviality.

The second equation easily follows if either $\alpha = 0$ or $\beta = 0$. We will consider the case $\beta \leq \gamma$. The key to the proof is to rewrite the convex combinations in different ways and repeatedly apply the mixture distributive law. Since $\beta/\gamma \leq 1$, we can use convex combinations involving it. We start with the left hand side of the mixture reduction law.

$$\begin{aligned}&\alpha\big[\beta L \oplus (1-\beta)L'\big] \oplus (1-\alpha)\big[\gamma L \oplus (1-\gamma)L'\big] \\ &\quad = \alpha\left\{\left(1-\frac{\beta}{\gamma}\right)L' \oplus \left(\frac{\beta}{\gamma}\right)\big[\gamma L \oplus (1-\gamma)L'\big]\right\} \oplus (1-\alpha)\big[\gamma L \oplus (1-\gamma)L'\big] \\ &\quad = \left[\alpha\left(1-\frac{\beta}{\gamma}\right)\right]L' \oplus \left\{\left[1-\alpha\left(1-\frac{\beta}{\gamma}\right)\right]\big[\gamma L \oplus (1-\gamma)L'\big]\right\} \\ &\quad = \left\{\left[\frac{\alpha\beta}{\gamma} + (1-\alpha)\right]\big[\gamma L \oplus (1-\gamma)L'\big]\right\} \oplus \left[1 - \frac{\alpha\beta}{\gamma} - (1-\alpha)\right]L' \\ &\quad = \left\{\gamma\left[\frac{\alpha\beta}{\gamma} + (1-\alpha)\right]\right\}L \oplus \left\{1-\gamma\left[\frac{\alpha\beta}{\gamma} + (1-\alpha)\right]\right\}L' \\ &\quad = \big[\alpha\beta + (1-\alpha)\gamma\big]L \oplus \big[\alpha(1-\beta) + (1-\alpha)(1-\gamma)\big]L',\end{aligned}$$

establishing the mixture reduction law.

The first equation uses symmetry and the mixture distributive law. The second follows from the mixture distributive law. Symmetry and some rewriting of expressions yield the third. The fourth again uses the mixture distributive law, and the last rewrites it.

This establishes the equality for $\beta \leq \gamma$. If $\beta > \gamma$, we can exchange β for γ and use symmetry to return the expressions to the same form, when the above calculations apply. $\square$

22.4.4 Further Results on Mixtures

If $(\mathfrak{M}, \oplus)$ is a mixture space, we say that a function $f \colon \mathfrak{M} \to \mathbb{R}$ is *mixture linear* if $f\big(\alpha L \oplus (1-\alpha)L'\big) = \alpha f(L) + (1-\alpha)f(L')$ for all $\alpha \in [0,1]$ and $L, L' \in \mathfrak{M}$.

Before we get to the main result on mixture linear representation of preferences in mixture space, we first prove a key lemma. Lemma 22.4.3 shows that use preferences have a representation on any preference interval of the form $\mathfrak{M}_{AB} = \{L : A \precsim L \precsim B\}$. Moreover, this representation is unique up to affine transformations.

Lemma 22.4.3. *Let $\succsim$ be a binary relation on a mixture space $(\mathfrak{M}, \oplus)$ that obeys*

1. *$\succsim$ is a preference order on $\mathfrak{M}$.*
2. *$\succsim$ obeys the Independence Axiom.*
3. *$\succsim$ is segment continuous, meaning that for all lotteries $L, L', L'' \in \mathfrak{M}$, the sets $\{\alpha \in [0,1] : \alpha L \oplus (1-\alpha)L' \succsim L''\}$ and $\{\alpha \in [0,1] : L'' \succsim \alpha L \oplus (1-\alpha)L'\}$ are closed.*

Suppose $A, B \in \mathfrak{M}$ with $A \prec B$. Define $\mathfrak{M}_{AB} = \{L \in \mathfrak{M} : A \precsim L \precsim B\}$ Then there is a mixture linear utility function u that represents $\succsim$ on $\mathfrak{M}_{AB}$. Moreover, if v is another mixture linear function representing $\succsim$ on $\mathfrak{M}_{AB}$, $v(L) = au(L) + c$ for some $a > 0$ and $c \in \mathbb{R}$.

22.4.5 Proof of Lemma 22.4.3

Proof. We will borrow portions of Part II of the proof of the von Neumann-Morgenstern Expected Utility Theorem.

Given $L \in \mathfrak{M}$, we define the sets $A^+ = \{\alpha \in [0,1] : \alpha B \oplus (1-\alpha)A \succsim L\}$ and $A^- = \{\alpha \in [0,1] : \alpha B \oplus (1-\alpha)A \precsim L\}$. Both are closed by segment continuity. Completeness implies $A^+ \cup A^- = [0,1]$, and the connectedness of $[0,1]$ ensures that $A^+ \cap A^-$ is non-empty.

We follow the argument from the von Neumann-Morgenstern Expected Utility Theorem to show that $A^+ \cap A^-$ is a singleton $\{\alpha\}$ and define $u(L) = \alpha$. Thus $u(L) = \alpha$ if and only if $L \sim \alpha B \oplus (1-\alpha)A$.

Continuity of u immediately follows, as the inverse images $u^{-1}[\alpha, \infty) = \{L' \in \mathfrak{M} : L' \succsim \alpha B \oplus (1-\alpha)A\}$ and $u^{-1}(-\infty, \alpha]$ are both closed by segment continuity of $\succsim$.

The arguments in the proof of the von Neumann-Morgenstern Theorem show that u is mixture linear on $\mathfrak{M}_{AB}$. Note that Proposition 22.4.2 is used here to ensure that the compound lotteries resolve as in the von Neumann-Morgenstern Theorem.

Now suppose v is another mixture linear utility representation on $\mathfrak{M}_{AB}$. Take $L \in \mathfrak{M}_{AB}$. Then $L \sim u(L)B \oplus \big(1 - u(L)\big)A$. Mixture linearity then implies $v(L) = u(L)v(B) + \big(1 - u(L)\big)v(A)$, so $v(L) = [v(B) - v(A)]u(L) + v(A)$. This means we can set $a = [v(B) - v(A)]$ and $b = v(A)$ to obtain the required linear transformation. $\square$

Recall that in Example 22.3.3, we did not need to define utility on the underlying states. It was enough to define it on a spanning set of lotteries. It was then uniquely defined on all of $\mathfrak{L}(\mathbf{S})$. In Lemma 22.4.3, the finite set of lotteries we start with (there called $\mathbf{S}$) necessarily spans, so utility is uniquely determined by its value on $\mathbf{S}$ and mixture linearity. Part II of the proof of the von Neumann-Morgenstern Expected Utility Theorem shows that the values on these states are determined by the values on s^* and s_*. Thus only two parameters are required to determine which mixture linear utility representation we have, as in Lemma 22.4.3.

Although these representations are not fully ordinal, they are not entirely cardinal either as different people may consider different utility numbers for lotteries B and A to accurately describe their strength of preference.

22.4.6 Uniqueness of Herstein-Milnor Representations

By normalizing the utility function obtained using Lemma 22.4.3 we can make the representation unique.

Corollary 22.4.4. *Suppose the hypotheses of Lemma 22.4.3 hold and that there are* $L_0, L_1 \in \mathfrak{M}$ *with* $L_1 \succ L_0$. *If* u *is a utility representation of* $\succsim$ *on some* $\mathfrak{M}_{AB}$ *with* $L_1, L_0 \in \mathfrak{M}_{AB}$ *that obeys* $u(L_1) = 1$ *and* $u(L_0) = 0$*, then* u *is unique.*

Proof. Suppose v is another such representation. By Lemma 22.4.3 there are a, b with $v(L) = au(L) + b$ for all $L \in \mathfrak{M}_{AB}$. Then $0 = v(L_0) = au(L_0) + b = b$, so $b = 0$. Then $1 = v(L_1) = au(L_1) = a$, so $a = 1$, showing that $v = u$. $\square$

22.4.7 Herstein-Milnor Theorem

Herstein and Milnor (1953) proved a utility representation theorem for a large class of spaces that includes both countably and uncountably infinite state spaces. We include a version of their theorem here.

Herstein-Milnor Utility Theorem. *Let $\succsim$ be a binary relation on a mixture space $\mathfrak{M}$ that obeys*

1. *$\succsim$ is a preference order on $\mathfrak{M}$*
2. *$\succsim$ obeys the Independence Axiom.*
3. *$\succsim$ is segment continuous, meaning that for all lotteries $L, L', L'' \in \mathfrak{M}$, the sets $\{\alpha \in [0,1] : \alpha L \oplus (1-\alpha)L' \succsim L''\}$ and $\{\alpha \in [0,1] : L'' \succsim \alpha L \oplus (1-\alpha)L'\}$ are closed.*

Then there is a mixture linear function $u\colon \mathfrak{M} \to \mathbb{R}$ that represents preferences on $\mathfrak{M}$. The function u obeys $u(\alpha L \oplus (1-\alpha)L') = \alpha u(L) + (1-\alpha)u(L')$ for any $L, L' \in \mathfrak{M}$ and $\alpha \in [0,1]$. Moreover, if v is another such function, $v(L) = a u(L) + c$ for some $a > 0$ and $c \in \mathbb{R}$.

Proof. If everything in $\mathfrak{M}$ is indifferent, set $u = 0$ or any other constant. Otherwise, there are $L_0, L_1 \in \mathfrak{M}$ with $L_0 \prec L_1$.

If $L \in \mathfrak{M}$, we can choose $A, B \in \mathfrak{M}$ so that $B \succsim L_1, L$ and $L, L_0 \succsim A$. By corollary 22.4.4 there is a unique mixture linear function u that represents $\succsim$ on $\mathfrak{M}_{AB}$ with $u(L_0) = 0$ and $u(L_1) = 1$. This uniquely defines $u(L)$. Since $L \in \mathfrak{M}$ was arbitrary, we have defined u on all of $\mathfrak{M}$.

Although u is defined, we still need to show it represents preferences on all of $\mathfrak{M}$. Take arbitrary $L, L' \in \mathfrak{M}$ with $L \succsim L'$. Choose new A and B so that $B \succsim L, L_1$, $L', L_0 \succsim A$. Now u represents $\succsim$ on $\mathfrak{M}_{AB}$, so $u(L) \geq u(L')$. Further, it is mixture linear for L, L'.

Since L and L' were arbitrary points in $\mathfrak{M}$, u represents $\succsim$ and is mixture linear on all of $\mathfrak{M}$. $\square$

22.4.8 More on Herstein-Milnor Theorem

The Herstein-Milnor Utility Theorem does not say that u is continuous. To get continuity requires stronger assumptions on $\succsim$.

Corollary 22.4.5. *Suppose that the assumptions of the Herstein-Milnor Utility Theorem hold. If preferences are also continuous, that is if* $\{L' \in \mathfrak{M} : L' \succsim L\}$ *and* $\{L' \in \mathfrak{M} : L' \precsim L\}$ *are closed for all* $L \in \mathfrak{M}$*, then any mixture linear utility representation is continuous.*

Proof. The Herstein-Milnor Utility Theorem applies, yielding mixture utility u.

For all $L \in \mathfrak{M}$, $u^{-1}[u(L), \infty) = \{L' \in \mathfrak{M} : L' \succsim L\}$ and $u^{-1}(-\infty, u(L)] = \{L' \in \mathfrak{M} : L' \precsim L\}$ are closed for every $L \in \mathfrak{M}$, establishing continuity of u. $\square$

The corollary has two continuity assumptions! The preferences must be both segment continuous and continuous in the usual sense. Continuity will imply segment continuity if the mapping $\alpha \mapsto \alpha L \oplus (1 - \alpha)L'$ is continuous for every $L, L' \in \mathfrak{M}$, if the mixture operation is continuous in α. That happens when the mixtures are probability mixtures, but is not true for all mixture spaces.

22.5 Distribution Functions and Infinite State Spaces

We are now ready to consider expected utility with an infinite number of states. For simplicity, we focus on expected utility when lotteries are purely monetary. This allows us to describe a lottery's payoffs by a real-valued random variable X, a function that randomly assigns the various possible payoffs. Gains are positive, losses are negative.

With a finite state space, it is enough to describe a lottery by listing the probability of each payoff. Expected values are then calculated by taking sums. If we have a countably infinite state space, the same procedure works, although we must take some care to make sure the sums converge.

When the state space is a continuum, such as the interval $[0,1]$, the sums must be replaced by some sort of integral. There are several ways to do this, involving different levels of difficulty and generality. In this section we use a method that is relatively simple and well-adapted to a state space in $\mathbb{R}$, using integrals involving distribution functions.[8]

[8] The early chapters of Chung (1974) use this method.

22.5.1 Infinite State Space: Distribution Functions

Recall that a function $f\colon \mathbb{R} \to \mathbb{R}$ is *weakly increasing* if for all $x, x' \in \mathbb{R}$ with $x < x'$, $f(x) \leq f(x')$. We will require that distribution functions are weakly increasing. Some important properties of weakly increasing functions are covered in section 30.7.

One important property is that limits from both the right and left exist at every point. When x is finite, we denote the limit from the left by $F(x-) = \lim_{x_n \uparrow x} f(x_n)$ and the limit from the right by $F(x+) = \lim_{x_n \downarrow x} f(x_n)$. Recall that a function F is *continuous from the right* at x if $F(x+) = F(x)$, and *continuous from the left* at if $F(x-) = F(x)$. The limits at infinity also exist, but may be infinite. They are denoted $F(-\infty)$ and $F(+\infty)$.

Distribution Function. A function $F\colon \mathbb{R} \to [0, 1]$ is a *distribution function* if F is weakly increasing and continuous from the right with $F(-\infty) = 0$ and $F(+\infty) = 1$. Distribution functions are sometimes called *cumulative distribution functions*.

22.5.2 Distribution Functions form a Mixture Space

The distribution functions will be our space of lotteries.

For distribution functions F and G, we define their mixture by the corresponding convex combination:

$$\big(\alpha F \oplus (1-\alpha)G\big)(x) = \alpha F(x) + (1-\alpha)G(x) \tag{22.5.4}$$

for any $\alpha \in [0,1]$.

This makes the set of distribution functions on $\mathbb{R}$ a mixture space.

Proposition 22.5.1. *Let $\mathcal{D}$ be the set of distribution functions with mixtures defined by convex combinations using equation 22.5.4, then $\mathcal{D}$ is a mixture space.*

Proof. Given $F, G \in \mathcal{D}$ and $\alpha \in [0,1]$, consider the mixture $H = \alpha F \oplus (1-\alpha)G$. Then $H(x) = \alpha F(x) + (1-\alpha)G(x)$ for all $x \in \mathbb{R}$. Since F and G take values in the convex set $[0,1]$, so does H. Moreover, $H(-\infty) = \alpha F(-\infty) + (1-\alpha)G(-\infty) = 0$ and $H(+\infty) = \alpha F(+\infty) + (1-\alpha)G(+\infty) = 1$.

Also, H is weakly increasing and continuous from the right because F and G are. It follows that H is also a distribution function. □

The set $\mathcal{D}$ will serve as our space of lotteries. Given appropriate preferences, we will be able to use the Herstein-Milnor Utility Theorem to obtain a utility representation on the set of distributions. We will use this in section 22.7 to show that those utilities can be represented as expected utility integrals.

22.5.3 Interpretation of Distribution Functions 03/29/22

NB: Problems 21.2.4, 21.2.7, 22.4.1, 22.6.1, and 22.6.4 are due on Thurs., April 7.

We interpret $F(x)$ as the probability that X takes a value less than or equal to x. In symbols, we write this as $F(x) = \Pr(X \le x)$. The requirement that $F(+\infty) = 1$ means that $\Pr(X < +\infty) = 1$. There is 100% chance that the payoff will be less that $+\infty$. The requirement that $F(-\infty) = 0$ ensures that there is also a 100% chance that the payoff will be greater that $-\infty$. Between them, they ensure that X is finite with probability 100%.

Now consider $a < b$. Then

$$F(b) - F(a) = \Pr(a < X \le b) = \Pr\big(X \in (a, b]\big).$$

We can use Proposition 30.7.2 to add to our interpretation of $F(x)$. The proposition tells us that $F(x-)$ exists. Now

$$F(x-) = \lim_{y \uparrow x} F(y) = \lim_{y \uparrow x} \Pr(X \le y) = \Pr(X < x)$$

because $\cup_{y<x}\{z : z \le y\} = \{z : z < x\}$. Then for $a < b$,

$$F(b) - F(a-) = F(b+) - F(a-) = \Pr\big(X \in [a, b]\big).$$

It follows that

$$F(b) - F(b-) = F(b+) - F(b-) = \Pr(X = b).$$

It F is continuous at b, $\Pr(X = b) = 0$, while if F has a jump at b, $\Pr(X = b)$ is the size of the jump.

22.5.4 Riemann Integrals: Tagged Partitions

We will use distribution functions to form Riemann-Stieltjes integrals. Even if you have not encountered a Riemann-Stieltjes integral before, you should have encountered the ordinary Riemann integral in undergraduate calculus. We will use that as a starting point.

To find the Riemann integral $\int_a^b f(x)\,dx$, we consider a partition of the interval $[a, b]$, that is, we take any x_i with $a = x_1 < x_2 < \cdots < x_{n+1} = b$. This generates a collection of intervals, $[x_i, x_{i+1}]$.

We write $\mathbf{x}$ for the ordered set $(x_1, \ldots, x_{n+1})$ (here n varies). The *norm of a partition* generated by $\mathbf{x}$ is $\max\{(x_{i+1} - x_i) : i = 1, \ldots, n\}$.

A *tagged partition* $\mathcal{P}(\mathbf{x}, \mathbf{t})$ is a partition together with numbers $t_i \in [x_i, x_{i+1}]$ with $t_i < t_{i+1}$ (the last condition ensures that no point is tagged twice).

22.5.5 Riemann Integrals: Refinements of Partitions

A tagged partition $\mathcal{P}(\mathbf{y}, \mathbf{s})$ *refines* $\mathcal{P}(\mathbf{x}, \mathbf{t})$ if $\mathbf{x} \subset \mathbf{y}$ and $\mathbf{t} \subset \mathbf{s}$. That is, a refinement of a tagged partition divides intervals $[x_i, x_{i+1}]$ into smaller subintervals by adding additional partition points and adds additional tags for subintervals that do not contain an existing tag.

As a result, the norm of $\mathcal{P}(\mathbf{y}, \mathbf{s})$ is less than or equal to the norm of $\mathcal{P}(\mathbf{x}, \mathbf{t})$. If we order tagged partitions via refinement, the set of tagged partitions of $[a, b]$ becomes a directed set.[9]

Consider the interval $[a, b] = [0, 5]$ and the tagged partition defined by

$$\begin{aligned}\mathbf{x} &= (x_1, \dots, x_7) = (0, .3, 2, 2.5, 3.8, 4.5, 5) \text{ and tags}\\ \mathbf{t} &= (t_1, \dots, t_7) = (.3, .5, 2.2, 3.6, 4.1, 4.7).\end{aligned}$$

The norm of this partition is 1.7. The partition $\mathcal{P}(\mathbf{y}, \mathbf{s})$ defined by

$$\begin{aligned}\mathbf{y} &= (0, .3, \mathbf{.9}, 2, 2.5, \mathbf{3.2}, 3.8, 4.5, 5) \text{ and}\\ \mathbf{s} &= (.3, .5, \mathbf{1.4}, 2.2, \mathbf{3}, 3.6, 4.1, 4.7)\end{aligned}$$

is a refinement of $\mathcal{P}(\mathbf{x}, \mathbf{t})$ and it has norm 1.1. I've bolded the added numbers in the refinement.

The refinement splits $[.3, 2]$ into $[.3, .9]$ and $[.9, 2]$, adding the tag $1.4 \in [.9, 2]$. It also splits $[2.5, 3.8]$ into $[2.5, 3.2]$ and $[3.2, 3.8]$, adding the tag $3 \in [2.5, 3.2]$.

[9] See section 30.8.4 for information on directed sets and nets (generalized sequences).

22.5.6 The Riemann Integral

Now set $\Delta x_i = x_{i+1} - x_i > 0$. The *Riemann sum* associated with the tagged partition $\mathcal{P}(\mathbf{x}, \mathbf{t})$ is

$$R\big(\mathcal{P}(\mathbf{x}, \mathbf{t})\big) = \sum_{i=1}^{n} f(t_i)\,\Delta x_i.$$

Since the tagged partitions of $[a, b]$ form a directed set, the Riemann sums are a net (generalized sequence) in $\mathbb{R}$. The Riemann integral is the limit of that net, if it exists.

$$\int_a^b f(x)\,dx = \lim R\big(\mathcal{P}(\mathbf{x}, \mathbf{t})\big)$$

If there is no limit, the Riemann integral does not converge.

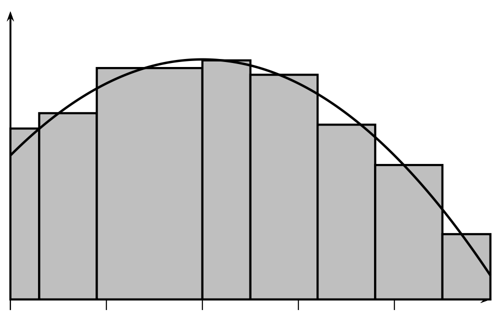

Figure 22.5.2: The gray area illustrates the Riemann sum for $\mathcal{P}(\mathbf{y}, \mathbf{s})$. That is, $\mathbf{y} = (0, .3, .9, 2, 2.5, 3.2, 3.8, 4.5, 5)$ and $\mathbf{s} = (.3, .5, 1.4, 2.2, 3, 3.6, 4.1, 4.7)$.

Riemann provided the first rigorous definition of the integral. Although the limit can be difficult to calculate, the Riemann integral is well-adapted to continuous and piecewise continuous functions. It may not work on badly behaved functions. Further, taking the limit of a sequence of integrable functions can cause problems, even if the sequence is bounded. For more general functions, a more general approach is needed—the Lebesgue integral. We examine it in section 22.6.

22.5.7 The Riemann-Stieltjes Integral

The Riemann-Stieltjes integral is defined the same way, but with the interval lengths Δx_i replaced by $\Delta F_i = F(x_{i+1}) - F(x_i)$. Because F is non-decreasing, $\Delta F_i \geq 0$. In other words, we look at sums of the form $\sum_i f(t_i)\Delta F_i$ and proceed as with the Riemann integral.

So how do we compute $\int_a^b u(x)\,dF(x)$? When F is differentiable, the answer is simple. The integral reduces to the Riemann integral

$$\int_a^b u(x)\,dF(x) = \int_a^b u(x)\,F'(x)\,dx.$$

If F is a distribution function, we refer to $f(x) = F'(x)$ as the *probability density function*. Because F is increasing, f must be non-negative. By the Fundamental Theorem of Calculus,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \tag{22.5.5}$$

It follows that $\int_{\mathbb{R}} f(t)\,dt = 1$.

Probability Density Function. A *probability density function* on $\mathbb{R}$ is a function $f\colon \mathbb{R} \to \mathbb{R}_+$ with $\int_{\mathbb{R}} f(t)\,dt = 1$.

It is easy to see that any probability density function defines a distribution function via equation 22.5.5. However, this is not the only way to get a distribution function.

22.5.8 Beyond Density Functions

Not all distribution functions derive from a density function. For example, consider the *Heaviside step function* at y,

$$H_y(x) = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{if } x < y. \end{cases}$$

The Heaviside step function H_y has a unit step at $x = y$. It is weakly increasing, continuous from the right, and obeys $H_y(-\infty) = 0$ and $H_y(+\infty) = 1$. In other words, the Heaviside step function is a distribution function.

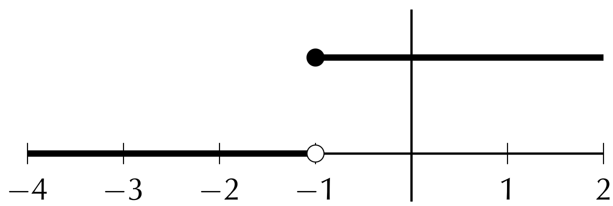

Figure 22.5.3: Heavside Step Function for $y = -1$

The Heaviside function cannot be written in terms of a probability density. For $x \neq y$, $H'_y(x) = 0$. If there was a probability density, it would have to be zero. But the zero function is not a probability density.

22.5.9 Heaviside Functions as Distribution Functions

To calculate integrals where a Heaviside step function is a distribution function, consider a partition $(\mathbf{x}, \mathbf{t})$ and compute the values $\Delta(H_y)_i = H_y(x_{i+1}) - H_x(t_i)$.

The position of y relative to x_i and x_{i+1} is critical. If $y \le x_i$ then $y < x_{i+1}$ and $H(x_{i+1}) - H(x_i) = 1 - 1 = 0$. Also, if $x_{i+1} < y$, so it x_i, and $H(x_{i+1}) - H(x_i) = 0 - 0 = 0$. Such terms will not contribute to the Riemann-Stieltjes sums. The only non-zero terms arise when $x_i < y$ and $y \le x_{i+1}$. Then $H(x_{1+1}) - H(x_i) = 1 - 0 = 1$. This can only happen once.

Given $f(x)$, the Riemann-Stieltjes sum for a partition $(\mathbf{x}, \mathbf{t})$ becomes $f(t_i)$ where $t_i \in [x_i, x_{i+1}]$, which is the only interval with $x_i < y$ and $y \le x_{i+1}$.

Now suppose f is continuous. Let $\varepsilon > 0$. We can find $\delta > 0$ so that $|f(y) - f(y')| < \varepsilon$ for $|y - y'| < \delta$. Now take a refinement $(\mathbf{y}, \mathbf{s})$ of the current partition so that $s_j \in [y_j, y_{j+1}] \subset [y - \delta, y + \delta]$. For any further refinement, f remains within ε of $f(y)$, and so do the Riemann-Stieltjes sums. It follows that their limit is

$$\int_{-\infty}^{+\infty} f(x)\, dH_y(x) = f(y).$$

If $f(x) = 1$, we obtain $\int_{-\infty}^{+\infty} dH_y(x) = 1$.

22.5.10 Expected Values

With some of the basics under control, we are ready to use distribution functions to compute expected values.

Given a distribution function F, the expectation of the associated random variable X is defined by the Lebesgue-Stieltjes integral[10]

$$EX = \int_{-\infty}^{+\infty} x \, dF(x).$$

More generally, the expectation of a function $u(x)$ is

$$Eu(X) = \int_{-\infty}^{+\infty} u(x) \, dF(x),$$

provided the integral exists.

[10] The Lebesgue-Stieltjes is a generalization of the Riemann-Stieltjes integral. When the latter exists, they are the same.

22.5.11 Mean and Variance

We use the Lebesgue-Stieltjes integral to write expectations. The expectation of $u\colon \mathbb{R} \to \mathbb{R}$ when income is given a lottery X with distribution function F is

$$Eu(X) = \int_{-\infty}^{+\infty} u(x)\, dF(x).$$

The expectation of the random variable X itself, the *mean of* X is $EX = \int x\, dF(x)$. The mean is often denoted by μ. We will sometimes use the notation EF rather than EX in order to emphasize the dependence of the mean on the distribution function.

The *variance of* X is defined by

$$\operatorname{var}(X) = E\big((X - EX)^2\big) = \int (x - EX)^2\, dF(x).$$

Since the expectation is a linear operator, we can rewrite the variance:

$$\begin{aligned}
\operatorname{var}(X) &= E\big[X^2 - 2X\,EX + (EX)^2\big] \\
&= E(X^2) - E\big[2X(EX)\big] + E\big[(EX)^2\big] \\
&= E(X^2) - 2(EX)^2 + (EX)^2 \\
&= E(X^2) - (EX)^2.
\end{aligned}$$

The variance is often denoted by σ^2, when the above can be rewritten $\sigma^2 = E(X^2) - \mu^2$. Since x^2 is a convex function, Jensen's Inequality yields $E(X^2) \geq (EX)^2$, implying $\operatorname{var}(X) \geq 0$.

22.5.12 Higher Moments

For $n > 1$, the *n^{th} moment of* F is defined by

$$\int_{-\infty}^{+\infty} (x - \mu)^n \, dF(x).$$

The variance is the second moment of the distribution. Technically, these are moments defined about the mean. The mean itself is the first moment about zero (replace the mean by zero in the formula and set $n = 1$). The third moment is called *skewness*, and the fourth moment is referred to as *kurtosis*.

22.5.13 Uniform Distribution

Uniform distributions provide a class of probability distribution functions.

Example 22.5.4: Uniform Distribution. Let $b > a$. The function

$$f(t) = \begin{cases} \frac{1}{b-a} & \text{when } t \in [a, b] \\ 0 & \text{when } t \notin [a, b] \end{cases}$$

is the density function for

$$F(x) = \begin{cases} 0 & \text{when } x < a \\ \frac{x-a}{b-a} & \text{when } x \in [a, b] \\ 1 & \text{when } x > b. \end{cases}$$

This is a distribution function, the *uniform distribution function* over $[a, b]$. Its associated random variable X assigns probability uniformly over $[a, b]$, with zero probability of either $X < a$ or $X > b$.

22.5.14 Mean and Variance of Uniform Distribution

We can calculate the mean and variance for the uniform distribution. The mean is

$$\mu = \frac{1}{b-a}\int_a^b t\,dt = \frac{1}{b-a}\left.\frac{t^2}{2}\right|_a^b = \frac{1}{b-a}\frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

The variance is

$$\begin{aligned}
\sigma^2 &= \frac{1}{b-a}\int_a^b (t-\mu)^2\,dt = \frac{1}{b-a}\left.\frac{(t-\mu)^3}{3}\right|_a^b \\
&= \frac{1}{b-a}\frac{(b-\mu)^3-(a-\mu)^3}{3} \\
&= \frac{1}{b-a}\frac{(b-a)^3-(a-b)^3}{24} \\
&= \frac{1}{b-a}\frac{(b-a)^3}{12} = \frac{(b-a)^2}{12}.
\end{aligned}$$

◀

22.5.15 Normal Distribution I

Another set of important distribution functions are the normal distributions.

Example 22.5.5: Normal Distribution. We start by making a density function from e^{-x^2}. Since it is non-negative, we need only normalize it so its integral is one. Let

$$A = \int_{-\infty}^{+\infty} e^{-x^2}\, dx. \tag{22.5.6}$$

We will normalize the density by dividing it by A. Squaring, we obtain

$$\begin{aligned} A^2 &= \left(\int_{-\infty}^{+\infty} e^{-x^2}\, dx\right)\left(\int_{-\infty}^{+\infty} e^{-y^2}\, dy\right) \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)}\, dx\, dy \\ &= \int_0^{2\pi}\left(\int_0^{\infty} e^{-r^2} r\, dr\right)\, d\theta \\ &= 2\pi\int_0^{\infty} e^{-r^2} r\, dr = \pi\int_0^{\infty} e^{-u}\, du = \pi. \end{aligned} \tag{22.5.7}$$

In line three, we transformed the (x, y) coordinates to polar coordinates using $x = r\cos\theta$ and $y = r\sin\theta$. Of course, under the integral sign, $dx\, dy$ is another way of writing the 2-form $dx \wedge dy$. We can use the theory of differential forms to do the change of coordinates. We have $dx = \cos\theta\, dr - r\sin\theta\, d\theta$ and $dy = \sin\theta\, dr + r\cos\theta\, d\theta$. Then

$$dx \wedge dy = r\cos^2\theta\, dr \wedge d\theta - r\sin^2\theta\, d\theta \wedge dr = r\, dr \wedge d\theta.$$

Further, we used the substitution $u = r^2$ in equation 22.5.7.

22.5.16 Normal Distribution II

Our calculations showed that

$$1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2}\, dx$$

which means that

$$g(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$$

is a density function. It is a type of *normal distribution*

It has expectation

$$\begin{aligned}
EX &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} xe^{-x^2}\, dx \\
&= \frac{1}{2\pi} \left[\int_{0}^{+\infty} xe^{-x^2}\, dx + \int_{-\infty}^{0} xe^{-x^2}\, dx \right] \\
&= \frac{1}{2\pi} \left[\int_{0}^{+\infty} xe^{-x^2}\, dx - \int_{0}^{+\infty} xe^{-x^2}\, dx \right] \\
&= 0
\end{aligned}$$

because the first integral is the negative of the second. This exploits the fact that $h(x) = xe^{-x^2}$ is an *odd function*, meaning $h(-x) = -h(x)$.

The variance is

$$\int_{-\infty}^{+\infty} x^2 e^{-x^2}\, dx = \left[-\frac{1}{2} x e^{-x^2} \right]_{-\infty}^{+\infty} + \frac{1}{2} \int_{-\infty}^{+\infty} e^{-x^2}\, dx = \frac{1}{2}.$$

Here we have integrated by parts using $v = \frac{1}{2}x$ and $du = 2xe^{-x^2}\, dx$.

22.5.17 Normal Distribution III

We define the normal distributions using density functions:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]\right\}.$$

This density function defines the *normal distribution* with mean μ and variance σ^2. We will reduce this to the previous case by appropriate substitution.

First, we substitute $y = (x-\mu)/\sqrt{2\sigma^2}$ to obtain

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)\,dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left\{-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]\right\}\,dx \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-y^2}\,dy = 1,\end{aligned}$$

showing that $f(x)$ is a density function.

22.5.18 Normal Distribution IV

Now we compute the expectation and variance.

$$EX = \int_{-\infty}^{+\infty} xf(x)\,dx = \int_{-\infty}^{+\infty} (y+\mu)f(y)\,dy = \mu + \int_{-\infty}^{+\infty} yf(y)\,dy = \mu$$

where we made the change of variables $y = x + \mu$ and used the fact that f is a density function and that $yf(y)$ is an odd function to show the integral of $yf(y)$ is zero.

To compute the variance we use the substitution $y = (x-\mu)/\sqrt{2\sigma^2}$. The variance is

$$\begin{aligned}
E\big((X-\mu)^2\big) &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (2\sigma^2 y^2) e^{-y^2} (2^{1/2}\sigma)\,dy \\
&= 2\sigma^2 \int_{-\infty}^{+\infty} y^2 e^{-y^2}\,dy \\
&= \sigma^2.
\end{aligned}$$

It follows that f has mean μ and variance σ^2. ◀

22.5.19 Expected Utility for Uniform Distribution

Example 22.5.6: Expected Utility for Uniform Distribution. Let $u(x) = \sqrt{x}$ and suppose you receive a payoff that is uniformly distributed on the interval $[a, b]$ with $0 \le a < b$. The c.d.f. is

$$F(x) = \begin{cases} 0 & \text{for } x \le a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x \ge b. \end{cases}$$

Then $dF = 1/(b-a)$ on $[a, b]$ and zero elsewhere, so

$$Eu(X) = \frac{1}{b-a}\int_a^b \sqrt{x}\,dx = \frac{2}{3}\cdot\frac{b^{3/2}-a^{3/2}}{b-a}.$$

A second example would be the same distribution with utility $v(x) = \ln x$. Then

$$Ev(X) = \frac{1}{b-a}\Big[x\ln x - x\Big]_a^b = \frac{b\ln b - a\ln a}{b-a} - 1.$$

◀

22.5.20 Multiple Random Variables

We can also consider the case of many random variables, $x_1, \dots, x_N$. When there are many random variables, they can either be inter-related or independent of each other (and everything in between). Roughly speaking, they are independent if you can define a distribution function for each that does not depend on the value of the other random variables.

More precisely, random variables $x_1, \dots, x_N$ with distribution functions $F_1, \dots, F_N$ are *independent random variables* if for any function $u\colon \mathbb{R}^N \to \mathbb{R}$,

$$Eu = \int_{\mathbb{R}^N} u(x_1, \dots, x_N)\, dF(x_1) \cdots dF(x_N).$$

When random variables $X_1, \dots, X_N$ are independent, it is easy to see that the expectation of X_n is $\int x_n \, dF_n(x_n)$ as all of the other variables integrate to 1. Moreover, if $X_1, \dots, X_N$ are independent, then

$$E(X_1 + \cdots + X_N) = EX_1 + \cdots + EX_N$$

and

$$\operatorname{var}(X_1 + \cdots + X_N) = \sum_n \operatorname{var}(X_n).$$

When $F_n = F$ for all n, we say $F_1, \dots, F_N$ are *identically distributed*. Random variables that are both independently and identically distributed are referred to as *iid*.

22.6** Lebesgue Measure and Integration

Skipped Rest of Chapter

Lebesgue approached integration from the opposite direction as Riemann. Riemann partitioned the **domain**, Lebesgue partitioned the **range**. For that to work, Lebesgue needed a way to measure sets of the form

$$\{x : y_i \le f(x) \le y_{i+1}\}.$$

Then he could multiply a value in $[y_i, y_{i+1}]$ by the size of the above set and form sums that would converge to the integral.

It is easy to measure how big intervals are. The length of $(a, b]$ is $(b - a)$. The length of a union of disjoint intervals is just the sum of the lengths of each interval. Lebesgue introduced a generalized notion of length, now called Lebesgue measure, which applies to a large class of sets including all open and closed sets.

The sets with Lebesgue measure zero play a particularly important role. For any finite interval A with endpoints $a < b$, $m(A)$, the *Lebesgue measure* of A is defined as $m(A) = b - a$. We use this to define the *outer Lebesgue measure* m^* on any set $E \subset \mathbb{R}$ by

$$m^*(E) = \inf\left\{ \sum_{n=1}^{\infty} m(A_n) : \text{each } A_n \text{ is a finite interval and } E \subset \bigcup_{n=1}^{\infty} A_n \right\}.$$

A set $E \subset \mathbb{R}$ has *measure zero* if $m^*(E) = 0$. If a set has measure zero, then for every $\varepsilon > 0$, we can find a countable collection of intervals that cover it and have total length less than ε. We will sometimes be interested in properties of functions that hold outside some set of measure zero. In that case we will say the property is true *almost everywhere*, abbreviated *a.e.*

22.6.1 The Cantor Set has Measure Zero

One famous example of a non-trivial set with measure zero is the Cantor set.

Example 22.6.1: The Cantor Set has Measure Zero. To form the Cantor set, start with the closed interval $C_0 = [0,1]$ and remove its open middle third, $(1/3, 2/3)$. That leaves $C_1 = [0,1/3] \cup [2/3,1]$. Then remove the open middle thirds of each part of C_1 to get C_2. In general, C_n is a union of closed intervals, and we obtain C_{n+1} by removing the open middle thirds of each interval. The *Cantor set* is $\mathfrak{C} = \cap_{n=1}^{\infty} C_n$.

As the intersection of non-empty nested compact sets, the Cantor set is compact and non-empty. The Cantor set also has measure zero. To see this, note that $m(C_n) = (2/3)^n$. Since C_n is a finite union of intervals, $0 \le m^*(C) \le (2/3)^n$ for every $n = 1, 2, \ldots$. It follows that $m^*(C) = 0$.

One way to understand the Cantor set a bit better is to write the $x \in [0,1]$ in ternary form (base-3). I.e., we interpret the expression $0.d_1d_2d_3\ldots$ as meaning $\sum_{n=1}^{\infty} d_n/3^n$ rather than $\sum_{n=1}^{\infty} d_n/10^n$. Notice that there may be two ways to write such numbers as $1/3 = 0.1000\ldots = 0.02222\ldots$.

Whenever we *must* use a 1 to write the ternary form of a number in $[0,1]$, we are in one of the middle thirds that is removed to form the Cantor set. That means that the Cantor set consists of all $x \in [0,1]$ that can be written in ternary form using only 0 and 2. Since 1 is not involved, this expression is unique. Numbers starting with 0.0 are in the first third (including $1/3 = 0.0222\ldots$), while numbers starting 0.2 are in the last third (including $2/3 = 0.2$), and so forth.

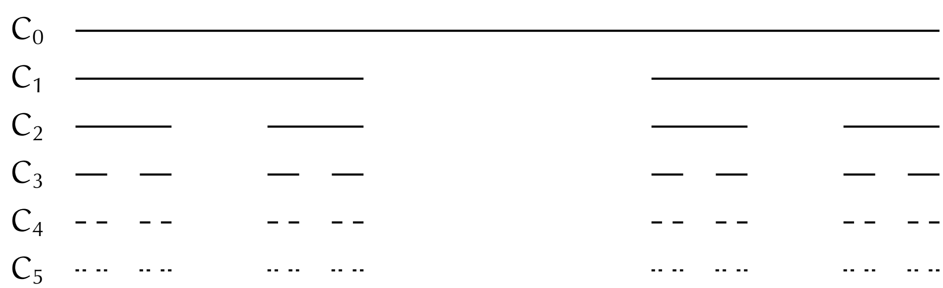

Figure 22.6.2: Here are the $C_0, \ldots, C_5$, the first six sets defined by eliminating middle thirds, and whose intersection comprises the Cantor set $\mathfrak{C}$.

◀

22.6.2 Lebesgue Integrals

We skip over the gory details concerning the construction of Lebesgue measure and concepts such as measurable functions (which include all continuous functions and their pointwise limits) to proceed to construction of the Lebesgue integral.

Simple Function. A function $f : \mathbb{R} \to \mathbb{R}$ is a *simple function* if it is measurable and takes only a finite number of values.

So if f is simple function, we can list its values, $f_1, \dots, f_N$. Then let $E_n = f^{-1}(f_n)$. Since $\{f_1, \dots, f_N\}$ is the range of f, $\cup_n E_n$ is the domain, $\mathbb{R}$. The Lebesgue integral of f is defined by

$$\int_{\mathbb{R}} f(x)\,dx = \sum_n f_n m(E_n)$$

where $m(E_n)$ is the Lebesgue measure of the set E_n. To integrate a more general function g, we first approximate it by simple functions. We take the limit as the approximation converges to g. If the limit of the integrals exists, it defines the Lebesgue integral of g. (Again, more technical details are omitted!)

22.6.3 Problems with Riemann Integrals

Although not all functions are Lebesgue integrable, the Lebesgue integral applies to more functions than the Riemann integral does. Moreover, it is much better behaved when taking pointwise limits. One useful property of Lebesgue integrals is that if two functions agree almost everywhere, they have the same Lebesgue integral.

Example 22.6.3: Lebesgue Integrable, but not Riemann Integrable. Even relatively harmless functions can fail to be Riemann integrable. Consider the function $f\colon [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & \text{when } x \text{ is irrational} \\ 1 & \text{when } x \text{ is rational.} \end{cases}$$

Since $f = 0$, a.e., it is Lebesgue integrable with $\int_0^1 f(x)\, dx = 0$. However, any tagged partition that only has rational tags will have a Riemann sum of one, while tagged partitions that only have irrational tags will have a Riemann sum of zero. The Riemann sums of mixed partitions can be anything in between. More importantly, given a partition, we can always construct a refinement with Riemann sum near one, and another refinement with Riemann sum near zero. This means both one and zero are cluster points of the net of Riemann sums. In fact, so is everything in between. There can be no limit. ◄

22.6.4 Lebesgue-Stieltjes Integrals: Point Masses

The Lebesgue-Stieltjes integral works the same way as the Lebesgue integral, except that we use the distribution function F to define the measure of $(a, b]$ as $F(b) - F(a)$.

Example 22.6.4: Point Mass at y. Recall the *Heaviside step function*

$$H_y(x) = \begin{cases} 1 & \text{if } x \geq y \\ 0 & \text{if } x < y. \end{cases}$$

We saw that when f is continuous, the Riemann-Stieltjes integral is

$$\int_{-\infty}^{+\infty} f(x)\, dH_y(x) = f(y).$$

Of course, this is also the Lebesgue-Stieltjes integral.

The Heaviside distribution is also referred to as a *point mass at* y. The associated measure is the Dirac measure δ_y defined by

$$\delta_y(E) = \begin{cases} 1 & \text{if } y \in E \\ 0 & \text{if } y \notin E. \end{cases}$$

◄

22.6.5 Finite Lotteries as Distribution Functions

We can use point masses to write finite lotteries as distribution functions, provided the payoffs are in $\mathbb{R}$.

Example 22.6.5: Lotteries as Distribution Functions. Consider a lottery X on a finite state space, paying x_i with probability p_i. Of course $p_i \geq 0$ and $\sum_i p_i = 1$. We can construct a distribution function that corresponds to X. Define the step function $F(x) = \sum_i p_i H_{x_i}(x)$. Then F is a distribution function because each H_{x_i} is right continuous, each has limit 0 at $-\infty$, and $F(+\infty) = \sum_i p_i \lim H_{x_i}(x) = \sum_i p_i H_{x_i}(+\infty) = \sum_i p_i = 1$. It follows that

$$Eu(X) = \int_{\mathbb{R}} u(x)\,dF(x) = \sum_i p_i u(x_i)$$

which is the expected value of u under the lottery yielding payoff x_i with probability p_i. Thus a lottery over a finite state space has a distribution function that is a probability mixture of Heaviside functions. ◄

22.6.6 Countably Infinite Lotteries

In fact, we can utilize a similar construction with the jumps of any distribution function F. Let D the set of all jump points of F. Since D is countable (Proposition 30.7.3), we may write $D = \{a_n\}$. Let $b_n = F(a_n+) - F(a_n-) = F(a_n) - F(a_n-) > 0$ and define

$$F_d(x) = \sum_n b_n H_{a_n}(x). \tag{22.6.8}$$

The sum converges uniformly since each $b_n > 0$ and $\sum_n b_n \le F(+\infty) - F(-\infty) = 1$. In fact, $F_d(-\infty) = 0$, $F_d(+\infty) = \sum_n b_n$ and F_d is weakly increasing and right continuous. Such a distribution, consisting entirely of point masses, is called a *discrete distribution*.

Proposition 22.6.6. *Suppose* F *is a distribution function and* F_d *is as defined in Equation 22.6.8. Then* $F = F_c + F_d$ *where* F_c *is a non-negative, weakly increasing, continuous function.*

Proof. By construction, $F_d(x) \le F(x)$ so F_c is non-negative. Moreover, if $x > y$, then

$$\begin{aligned} F_d(x) - F_d(y) &= \sum_{n:y<a_n\le x} b_n \\ &= \sum_{n:y<a_n\le x} [F(a_n) - F(a_n-)] \\ &\le F(x) - F(y). \end{aligned}$$

It follows from this equation that both F_d and F_c are weakly increasing.

Finally, $F_c(x+) = F(x+) - \sum_{n:a_n\le x} b_n$ and $F_c(x-) = F(x-) - \sum_{n:a_n<x} b_n$. Then

$$F_c(x+) - F_c(x-) = \begin{cases} F(x+) - F(x-) = 0 & \text{when } x \ne a_n \text{ for any } n \\ F(x+) - F(x-) - b_n = 0 & \text{when } x = a_n. \end{cases}$$

This shows that F_c is continuous, completing the proof. □

22.6.7 Absolute Continuity

Since F_c is continuous, it assigns probability zero to every point in $\mathbb{R}$. By construction, the term F_d consists purely of point masses. We have shown that any distribution function can be decomposed into a discrete part and a continuous part.

We can further decompose the continuous part of the distribution function. We say that F is *absolutely continuous* with respect to Lebesgue measure if and only if there is an integrable function f such that for every $x < x'$

$$F(x') - F(x) = \int_x^{x'} f(z)\,dz. \tag{22.6.9}$$

Any such absolutely continuous function F, has a derivative almost everywhere. In fact,

$$F' = f, \text{a.e.}$$

If F is a distribution function, we refer to f as the *density function*.

Conversely, if f is a non-negative function obeying $\int_{\mathbb{R}} f(z)\,dz = 1$, then F defined by

$$F(x) = \int_{-\infty}^{x} f(z)\,dz$$

is an absolutely continuous distribution function with $F' = f$.

When F is an absolutely continuous distribution function describing a random variable X,

$$Eu(X) = \int_{-\infty}^{+\infty} u(x)\,f(x)\,dx$$

where $f(x) = F'(x)$ is the density function.

22.6.8 Singular Distributions

We can further decompose the continuous part of a distribution function into an absolutely continuous function, and a function that is the antithesis of absolutely continuous. We call a distribution function F *singular* if F′ exists and equals zero a.e. If such a function were absolutely continuous, it would be zero! Singular functions are as far from absolutely continuous as possible.

22.6.9 The Cantor Function

The Cantor function is a very interesting function that maps $[0,1]$ onto $[0,1]$. Incredibly, all of its increase takes place on the zero-measure Cantor set $\mathfrak{C}$. Not surprisingly, it is a singular distribution function.

Example 22.6.7: Cantor Function. We define the Cantor function $c(x)$ as follows. For $x > 1$, set $c(x) = 1$ and for $x < 0$, set $c(x) = 0$.

Let $x \in [0,1]$. We start by writing x in ternary form as in Example 22.6.1. If a 1 appears in the ternary form, replace all following ternary digits by 0. This ensures that any numbers within a given interval that is removed to construct the Cantor set will be represented by the same ternary form. E.g., anything in $(1/3, 2/3)$ is mapped to $0.1000\cdots = 1/3$ while anything in $(7/9, 8/9)$ is mapped to $0.21000\ldots = 7/9$. Now replace any 2's in the ternary form by 1. Let $0.d_1d_2d_3\ldots$ be the result and define the Cantor function by $c(x) = \sum_{n=1}^{\infty} d_n/2^n$. That is, we treat the resulting $0.d_1d_2d_3\ldots$ as a binary number.

The Cantor function is constant on every removed open interval, so $c' = 0$ there, as well as on $(-\infty, 0)$ and $(1, +\infty)$. The complement of this collection of intervals is the Cantor set. It has measure zero, so $c' = 0$, a.e. In other words, c is a singular function.

The Cantor function is also continuous. Take $x \in [0,1]$ and let $\varepsilon > 0$. Take n such that $1/2^n < \varepsilon$. Set $\delta = 1/3^n$. When $|x - y| < \delta$, the ternary forms are the same in the first n ternary digits. But when transformed as above, the resulting binary forms are the same in the first n binary digits, so $|c(x) - c(y)| < 1/2^n < \varepsilon$, proving continuity.

Since c is continuous and weakly increasing with $c(-\infty) = 0$ and $c(+\infty) = 1$, it is a distribution function. It is continuous, so $c_d = 0$. Because $c' = 0$ a.e., it has no absolutely continuous part, $c_{ac} = 0$. That means that $c = c_s$. It is a purely singular distribution function with all of its mass concentrated on the Cantor set.

The existence of the Cantor functions has an interesting implication for the Cantor set. The complement of the Cantor set is mapped to points of the form $n/2^m$ with $n \le 2^m$. Let $A = \{n/2^m : 0 \le n \le 2^m\}$. Then A is a countable set. That means that the image of the Cantor set is $c(\mathfrak{C}) = [0,1] \setminus A$. The right-hand side has cardinality $\mathfrak{c}$, the cardinality of the continuum. It follows that $\mathfrak{C}$ itself also has cardinality $\mathfrak{c}$. Although the Cantor set is small in size (measure zero), it contains a lot of points! ◄

22.6.10 Completing the Decomposition

In order to extract the absolutely continuous part of F_c, we use Corollary 30.7.7 to see that F_c has a derivative almost everywhere. The existence of such a derivative does not mean that F_c is absolutely continuous. After all, the Cantor function has $c' = 0$ a.e., but c is certainly not the integral of 0.[11]

Notice that this gives 0 is the absolutely continuous part of the Cantor function. The same is true of all singular functions.

Let $f = F_c' = F'$ a.e. We will use f as a density function. Since F is non-decreasing, $f \geq 0$. Define the absolutely continuous part of F, F_{ac} by

$$F_{ac}(x) = \int_{-\infty}^{x} f(z)\,dz.$$

Then $F_s = F_c - F_{ac}$ is singular. We have decomposed any distribution function F into a discrete part F_d, a singular part F_s, and an absolutely continuous part F_{ac} obeying

$$F = F_d + F_s + F_{ac}.$$

In other words, the probability density F has a discrete part F_d that has its mass concentrated on a countable set of points, a singular part F_s that has $F_s' = 0$ a.e., and an absolutely continuous part F_{ac} with density function $f = F' = F_{ac}'$.

Now we let

$$\begin{aligned}
a &= F_d(+\infty) - F_d(-\infty)\\
b &= F_s(+\infty) - F_s(-\infty)\\
c &= F_{ac}(+\infty) - F_{ac}(-\infty)
\end{aligned}$$

Then $a, b, c \geq 0$ and $a + b + c = 1$. If $a > 0$, define $G_d = F_d/a$, if $b > 0$, let $G_s = F_s/b$, and if $c > 0$ set $G_{ac} = F_{ac}/c$ (each is defined as 0 otherwise). Then we can write F as a convex combination of **distribution functions**,

$$F = aG_d + bG_s + cG_{ac}.$$

[11] Our method yields 0 is the absolutely continuous part of the Cantor function. The same is true of all singular functions.

22.7** Expected Utility on Infinite State Spaces

03/23/21

The Herstein-Milnor Utility Theorem does not immediately yield an expected utility representation for infinite state spaces. However, as in the von Neumann-Morgenstern Expected Utility Theorem, it is easy to show that it is an expected utility function if $\mathfrak{L} = \mathfrak{L}(\mathbf{S})$ for some finite state space $\mathbf{S}$. To go further, we first define lotteries of finite support.

Lottery of Finite Support. A lottery has *finite support* if there are only finitely many states x_s, $s = 1, \dots, S$, with $p_s > 0$ and $\sum_{s=1}^{S} p_s = 1$.

Even if $\mathbf{S}$ is infinite, it is still possible to derive an expected utility representation for all lotteries of finite support. To extend the representation further, we must first introduce the *weak topology* on the space of measures. Let $\mathcal{C}_b(\mathbf{S})$ be the set of all bounded and continuous functions on $\mathbf{S}$.

Weak Convergence of Measures. A sequence of measures μ_n on $\mathbf{S}$ *weakly converges* to μ if

$$\int f \, d\mu_n \to \int f \, d\mu \text{ for every } f \in \mathcal{C}_b(\mathbf{S}).$$

Since each distribution function defines a measure on $\mathbb{R}$, and each measure on $\mathbb{R}$ defines a distribution function, we can extend the above definition to weak convergence of distribution functions. A sequence of distribution functions F_n on $\mathbf{S}$ *weakly converges* to a distribution function F if

$$\int f \, dF_n \to \int f \, dF \text{ for every } f \in \mathcal{C}_b(\mathbf{S}).$$

22.7.1 Weak Convergence of Measures

The weak topology can be described by a metric on the space of probability measures. As a result, sequences are sufficient to characterize its topology. The fact that mixtures are just convex combinations of probability measures, and hence continuous functions of mixtures, means that weak continuity implies segment continuity. When μ has finite support, we can write $\mu = \sum_{s=1}^{S} p_s \delta_s$. The Expected Utility Theorems allow us to obtain an expected utility representation on all finitely supported lotteries.

Let $\mathbf{S}$ be a locally compact Hausdorff space and $\mathcal{C}(\mathbf{S})$ denote the space of continuous functions on $\mathbf{S}$. We say $f \in \mathcal{C}(\mathbf{S})$ converges to zero at infinity if for $\varepsilon > 0$ there is a compact set K with $|f(x)| < \varepsilon$ for $x \notin K$. Let $C_0(\mathbf{S})$ denote the set of continuous functions on $\mathbf{S}$ that converge to zero at infinity. This is a normed vector space with $\|f\| = \sup_{s \in \mathbf{S}} |f(s)|$. The fact that f converges to zero at infinity implies the norm is always finite.

The dual space of $C_0(\mathbf{S})$ is the set of regular Borel measures on S, denoted $ca_r(S)$.[12]

[12] See Aliprantis and Border (2006) for explanation of any unfamiliar mathematics in this section, including discussion of the weak topology on $\mathcal{C}_b$, the space ca_r, and the duality between $\mathcal{C}_b$ and ca_r.

22.7.2 Finitely Supported Measures are Dense

Theorem 22.7.1. *Let* F *be the set of finitely supported measures on* $\mathbf{S}$. *Then* F *is dense in* $\mathrm{ca}_r(\mathbf{S})$ *in the* $\sigma\big(\mathrm{ca}_r(\mathbf{S}), \mathcal{C}_0(\mathbf{S})\big)$ *topology.*

Proof. Suppose F is not dense in $\mathrm{ca}_r(\mathbf{S})$, let $\mu \in \mathrm{ca}_r(\mathbf{S})$ with $\mu \notin \bar{F}$. Define a linear form φ on the span of μ and $\bar{F}$ by $\varphi(\alpha\mu + \beta f) = \alpha$ for any $f \in \bar{F}$. The Hahn-Banach Extension Theorem allows us to extend this to a σ-continuous linear functional on $\mathrm{ca}_r(S)$. By the Riesz Representation Theorem, we may regard this as an element g of $\mathcal{C}_0(\mathbf{S})$. But then $g(\delta_s) = g(s) = 0$ for all $s \in S$, implying that $g = 0$. This is impossible due to our definition of φ, and so contradicts our suppose that F is not dense in $\mathrm{ca}_r(\mathbf{S})$. It follows that F is dense in $\mathrm{ca}_r(\mathbf{S})$. □

By Theorem 22.7.1, any probability measure is the limit of finitely supported lotteries supported lotteries. This allows us to obtain an integral representation.[13]

[13] See Fishburn (1970). Kreps (1988) has a nice discussion of these matters.

22.7.3 General Expected Utility Theorem

Finally, the Herstein-Milnor Theorem can be used to prove a general result on expected utility.

General Expected Utility Theorem. *Let $\succsim$ be a binary relation on $\mathfrak{L}(\mathbf{S})$, the set of countably additive Borel probability measures on $\mathbf{S}$. Suppose $\succsim$ obeys:*

1. *$\succsim$ is a preference order on $\mathfrak{L}(\mathbf{S})$.*
2. *$\succsim$ obeys the Independence Axiom.*
3. *The sets $\{L \in \mathfrak{L}(\mathbf{S}) : L \succsim L'\}$ and $\{L \in \mathfrak{L}(\mathbf{S}) : L' \succsim L\}$ are weakly closed.*

Then there is a continuous function $u\colon \mathbf{S} \to \mathbb{R}$ so that

$$L \succsim L' \quad \textit{if and only if} \quad \int u(s)\, dL(s) \ge \int u(s)\, dL'(s).$$

Moreover, if v is any other such function, there are $a > 0$ and $c \in \mathbb{R}$ with $v = au + c$.

April 22, 2022

Copyright ©2022 by John H. Boyd III: Department of Economics, Florida International University, Miami, FL 33199