Chapter 4

Expected utility

This chapter introduces the problem of decision-making under uncertainty. We cast decision-making as a computational problem by first introducing the concept of utility*. The notion of utility provides a natural computational description for the more abstract concept of "preferences". We then discuss the problem of uncertainty and present the classical* theory of expected utility*.*

4.1 Preferences and utility

In order to discuss the problem of decision-making from a computational perspective, let us consider the following trivial example.

Example 4.1 A participant in a contest is currently winning 50 EUR. At the last stage of the contest, the contestant is placed before two possibilities:

A: Keep her current prize;

B: Play a game involving two cups, in which the contestant selects one of two cups and keeps the prize in the selected cup.

One of the cups contains a prize of 10 EUR, while the other contains a prize of 75 EUR. The cups are transparent, allowing the contestant to determine the prize under each of the two cups.

The reasoning process of the contestant is along the following lines:

- Keeping the current prize will secure a final prize of 50 EUR.
- Selecting to play the cup game will surely lead to a prize of 75 EUR, since the contestant is able to determine which cup yields the largest prize.

The contestant thus chooses to play the cup game.

Since, in general, contestants will prefer a larger prize to a smaller prize, the decision of the contestant in Example 4.1 is hardly questionable since, indeed, it secures the largest prize. Let us now consider a slight variation of Example 4.1.

Example 4.2 A participant in a contest is currently winning a free meal (valued in 50 EUR) in a steak house. At the last stage of the contest, the contestant is placed before two possibilities:

A: Keep her current prize;

B: Play a game involving two cups, in which the contestant selects one of two cups and keeps the prize in the selected cup.

One of the cups contains a ticket for a free meal in a fast-food chain (valued in 10 EUR), while the other contains a free meal in a gourmet pizza house (valued in 75 EUR). The cups are transparent, allowing the contestant to determine the prize under each cup.

The reasoning process of the contestant is along the following lines:

- Keeping the current prize will secure a free steak meal valued in 50 EUR.
- Since the contestant does not appreciate cheese, between the fast food meal and the pizza meal, she prefers the fast food meal. Therefore, selecting to play the cup game will surely lead to a free fast-food meal valued in 10 EUR.
- However, the contestant prefers the meal in the steak house than the fast-food meal.

The contestant thus chooses to keep her current prize.

The reasoning process in Example 4.2 involves a number of factors other than the sole monetary value of the different options. In that particular example, such factors are specific to the individual contestant, in that different contestants could engage in distinct reasoning processes (for example, a vegetarian contestant may find the pizza meal more appealing than the steak meal).

In general, although it is impractical/infeasible to quantify the subjective factors that guide the decisions of an individual, it is often possible (and, indeed, quite easy) to rank the different possibilities according to their order of preference by the decision-maker. The notion of *preference* can be formalized into a framework that affords a natural computational treatment, and upon which all decision-making discussed throughout the book will be constructed.

◇

We provide a brief overview of binary relations and their main properties, as these provide the formalism necessary to discuss the notion of "preference". Let

$\mathcal{X}$ be a finite set. A *binary relation* R on $\mathcal{X}$ is a subset of $\mathcal{X} \times \mathcal{X}$, i.e., is a set of ordered pairs (x, y), with $x, y \in \mathcal{X}$. We write $x \stackrel{R}{\to} y$ when $(x, y) \in R$ and $x \stackrel{R}{\not\to} y$ when $(x, y) \notin R$.

Binary relations can be:

- *Reflexive*, if $x \stackrel{R}{\to} x$ for every $x \in \mathcal{X}$.
- *Symmetric* if $x \stackrel{R}{\to} y$ implies that $y \stackrel{R}{\to} x$ for all $x, y \in \mathcal{X}$.
- *Asymmetric* if $x \stackrel{R}{\to} y$ implies that $y \stackrel{R}{\not\to} x$ for all $x, y \in \mathcal{X}$.
- *Anti-symmetric* if $x \stackrel{R}{\to} y$ and $y \stackrel{R}{\to} x$ imply that $x = y$ for all $x, y \in \mathcal{X}$.
- *Transitive* if $x \stackrel{R}{\to} y$ and $y \stackrel{R}{\to} z$ imply that $x \stackrel{R}{\to} z$, for all $x, y, z \in \mathcal{X}$.
- *Negative transitive* if $x \stackrel{R}{\not\to} y$ and $y \stackrel{R}{\not\to} z$ imply that $x \stackrel{R}{\not\to} z$, for all $x, y, z \in \mathcal{X}$. Equivalently, R is negative transitive if $x \stackrel{R}{\to} y$ implies that either $x \stackrel{R}{\to} z$ or $z \stackrel{R}{\to} y$.
- *Complete* if $x \stackrel{R}{\to} y$ or $y \stackrel{R}{\to} x$ for every $x, y \in \mathcal{X}$.

A binary relation, being directional, generally induces some kind of *order* among the elements of $\mathcal{X}$. We say that a binary relation R is a *weak order* if it is asymmetric and negative transitive. It is a *strict order* if it is a weak order and, for every $x, y \in \mathcal{X}$ such that $x \neq y$, $x \stackrel{R}{\to} y$ or $y \stackrel{R}{\to} x$.

Finally, we say that a binary relation is an *equivalence relation* if it is reflexive, symmetric and transitive. We encountered an example of an equivalence relation in Chapter 2: *communication* between states of a Markov chain is an equivalence relation.

4.1.1 Preference relations

Let $\mathcal{X}$ denote a (finite) set of outcomes. A *strict preference relation* is a binary relation "$\succ$" defined on $\mathcal{X}$. We write $x \succ y$ when outcome x is preferred to outcome y. We often phrase $x \succ y$ as "x is better than y". For ease of exposition, we sometimes write $y \prec x$ instead of $x \succ y$ with the same meaning, but phrasing it as "y is worse than x". If $x \not\succ y$ and $y \not\succ x$ we say that the two outcomes are *indifferent*, and denote such fact as $x \sim y$. Finally, we write $x \succeq y$ if $x \succ y$ or $x \sim y$, and refer to the relation "$\succeq$" as a *weak preference*; similarly, we write $x \preceq y$ if $x \prec y$ or $x \sim y$.

We require strict preference relations to be *asymmetric* and *negative transitive* (i.e., a weak order on $\mathcal{X}$). Both conditions ensure, in a sense, that the relation is "consistent":

- Asymmetry ensures that if an outcome x is better than an outcome y, then y should not be simultaneously better than x.

- Negative transitivity, on the other hand, ensures that if x is better than y, then an outcome z cannot be, simultaneously, better than x and worst than y.

The following result is a direct consequence of the fact that a strict preference is a weak order on $\mathcal{X}$.

Lemma 4.1. *Let $\mathcal{X}$ be a finite set and $\succ$ a strict preference on $\mathcal{X}$. Then,*

(a) *For any $x, y \in \mathcal{X}$, exactly one of the following holds:*

$$x \succ y; \qquad\qquad y \succ x; \qquad\qquad x \sim y.$$

(b) *The preference $\succ$ is transitive.*

(c) *Indifference is an equivalence relation.*

(d) *If $x \succ y$ and $x \sim z$, then $z \succ y$. Conversely, if $x \sim z$ and $x \prec y$, then $z \prec y$.*

(e) *The relation $\succeq$ is transitive and complete.*

Proof. See Section 4.6. □

The statements in Lemma 4.1, although relatively straightforward, provide the tools necessary to express a preference relation as a numerical function called *utility*, that can be easily used in computation. For that, we require the notion of *rational preference*.

Rational preference

A preference relation over a set $\mathcal{X}$ is a *rational preference* if it is complete and transitive.

From Lemma 4.1 (a), it follows that the weak preference $\succeq$ is complete. Moreover, in virtue of (b) and (c), is also transitive. Therefore, we can conclude that the weak preference $\succeq$ is a rational preference over $\mathcal{X}$.

4.1.2 Existence of utility functions

The importance of rational preferences stems from the fact that such relations can be translated into an *order-preserving function* that can then be used for computational manipulations. This fact is stated in the following result, which is the central result in this section.

Theorem 4.2. *Let $\mathcal{X}$ be a finite set of possible outcomes, and $\succeq$ a rational preference on $\mathcal{X}$. Then, there is a function $u : \mathcal{X} \to \mathbb{R}$ such that $u(x) \geq u(y)$ if and only if $x \succeq y$, for any $x, y \in \mathcal{X}$.*

The function u featured in Theorem 4.2 is known as a *utility function*, a *payoff function* or a *reward function*. In a sense, $u(x)$ provides a numerical value that represents how "advantageous"—in terms of the preference $\succeq$—an outcome x is.

◇

In the previous section, we defined the weak preference $\succeq$ from the strict preference $\succ$. However, the converse is also possible: given a rational preference $\succeq$, the relations $\preceq$, $\succ$, $\prec$ and $\sim$ can be easily defined from $\succeq$. In fact, given $x, y \in \mathcal{X}$,

- $y \preceq x$ corresponds to $x \succeq y$;
- $x \succ y$ corresponds to $y \not\succeq x$;
- $x \prec y$ corresponds to $x \not\succeq y$;
- $x \sim y$ corresponds to the situation where, simultaneously, $x \succeq y$ and $y \succeq x$.

It is immediate to verify that the relations thus obtained verify Lemma 4.1. Then, given the set $\mathcal{X}$ and the rational preference $\succeq$, let $\mathcal{X}_1, \ldots, \mathcal{X}_N$ denote the partition of $\mathcal{X}$ induced by the equivalence relation $\sim$.[1] Since the number of sets in the partition is finite, the index assigned to each subset $\mathcal{X}_n$ can be selected so that, if $m > n$, then $x_m \succ x_n$, with $x_m \in \mathcal{X}_m$ and $x_n \in \mathcal{X}_n$. Then, setting $u(x) = n$ for each $x \in \mathcal{X}_n, n = 1, \ldots, N$, we obtain a function u that verifies the statement of the theorem.

We can extend this result to the situation where $\mathcal{X}$ is countably infinite. As before, let $\mathcal{X}_1, \ldots, \mathcal{X}_n, \ldots$ denote the partition of $\mathcal{X}$ induced by the equivalence relation $\sim$, and let $\{x_1, \ldots, x_n, \ldots\}$ denote a sequence of outcomes selected so that each $x_n \in \mathcal{X}_n$. Set $u(x_1) = 0$. Then, sequentially for each $n > 1$, one of the following holds:

- $x_n \succ x_m$, for all $m < n$. In this case, set $u(x_n) = n$.
- $x_n \prec x_m$, for all $m < n$. In this case, set $u(x_n) = -n$.
- There are x_i and x_j such that $x_i \succ x_n \succ x_j$, $i, j < n$. If this is the case, it is always possible to select x_i and x_j such that no x_m verifies $x_i \succ x_m \succ x_j$, with $m < n$, $m \neq i$ and $m \neq j$. In this case, set $u(x_n)$ to any rational number strictly between $u(x_i)$ and $u(x_j)$.

[1] In other words, each set $\mathcal{X}_n$ contains only outcomes that are pairwise indifferent, i.e., if two outcomes x, y lie in the same $\mathcal{X}_n$, then $x \sim y$.

Note that, by construction, $u(x_n) \neq u(x_m)$ for all $m < n$. Moreover, if $x_m \succ x_n$, $u(x_m) > u(x_n)$ and vice-versa. Finally, by setting $u(x) = u(x_n)$ for all $x \in \mathcal{X}_n$, we again obtain a function u that verifies the statement of the theorem.

Theorem 4.2 allows us to reason about preferences from a computational perspective: a utility function is an object that can be represented in a computer, and our focus in the remainder of the book precisely on automating the process of decision-making, for which utility functions are paramount.

We henceforth refer to a computational decision-maker as an *agent*, interpreted as an abstract (computational) entity that observes its environment and acts upon it. Hence, agents may refer to such distinct entities as office-navigating robots, game playing agents, self-driving cars, or online recommender systems. Our role is that of *agent designers*: we translate our preferences regarding the intended behavior for our agents in terms of an adequate (utility-based) representation.

Utilities for decision-making

Consider an agent facing the problem of selecting one among a set of alternatives, $\mathcal{A}$. Each alternative $a \in \mathcal{A}$ yields an outcome $X(a)$, with $X(a)$ taking values in some set of possible outcomes, $\mathcal{X}$. We call each alternative $a \in \mathcal{A}$ an *action*, and the set $\mathcal{A}$ the *action space* of the agent.

Given a rational preference $\succeq$ over $\mathcal{X}$, let u denote an associated utility function, i.e., a function $u : \mathcal{X} \to \mathbb{R}$ such that if two outcomes $x, y \in \mathcal{X}$ verify $x \succeq y$ then $u(x) \geq u(y)$. We say that the agent is *rational* if it selects its actions so as to maximize the resulting utility. Formally, let

$$Q(a) \stackrel{\text{def}}{=} u(X(a)) = \sum_{x \in \mathcal{X}} \mathbb{I}\left[X(a) = x\right] u(x).$$

A rational agent will select an action $a^* \in \mathcal{A}$ such that

$$a^* \in \operatorname*{argmax}_{a \in \mathcal{A}} Q(a) = \sum_{x \in \mathcal{X}} \mathbb{I}\left[X(a) = x\right] u(x). \tag{4.1}$$

We refer to the function $Q : \mathcal{A} \to \mathbb{R}$ as the *action value function* or, more compactly, as the *Q-function*. We note that, if multiple maximizing actions exist, the agent can select always the same maximizing action, deterministically, or randomize between them. In both situations, the resulting behavior is rational.

From (4.1) it should be clear that it is possible to modify the Q-function without altering the resulting behavior. For example, if we add a constant K to Q,

$$\operatorname*{argmax}_{a \in \mathcal{A}} \{Q(a) + K\} = \operatorname*{argmax}_{a \in \mathcal{A}} Q(a),$$

i.e., the set of maximizing actions remains the same. Similarly, if we multiply Q by a positive constant α,

$$\operatorname*{argmax}_{a \in \mathcal{A}} \{\alpha \times Q(a)\} = \operatorname*{argmax}_{a \in \mathcal{A}} Q(a).$$

We can summarize the observations above in the following more general result, whose proof is left as an exercise (see Exercise 4.1).

Proposition 4.3. *Let $u : \mathcal{X} \to \mathbb{R}$ denote a utility function representing a rational preference $\succeq$ over some finite set of outcomes $\mathcal{X}$, and let $f : \mathbb{R} \to \mathbb{R}$ be a strictly increasing function. Then, the function $u' : \mathcal{X} \to \mathbb{R}$ defined as $u'(x) = f(u(x))$ is also a utility function that represents the agent's preferences over $\mathcal{X}$.*

The framework for decision-making introduced thus far is void of uncertainty: the outcome resulting from each of the agent's actions is determined by the mapping X, and the maximization in (4.1) essentially selects the action leading to the preferred outcome. The next section extends the current paradigm of decision-making to accommodate for actions with uncertain outcomes.

4.2 The theory of expected utility

To extend utility theory to accommodate uncertainty, we start by generalizing the notion of preference from sets of outcomes to sets of *distributions over outcomes*.

4.2.1 Preferences over mixture sets ($\star$)

Let $\mathcal{X}$ be some finite set and $p_1, \ldots, p_N$ arbitrary probability distributions over $\mathcal{X}$. A *convex combination* of $\{p_n, n = 1, \ldots, N\}$ is any linear combination

$$p = \sum_{n=1}^{N} \lambda_n p_n,$$

where $\lambda_1, \ldots, \lambda_N$ are such that $0 \leq \lambda_n \leq 1, n = 1, \ldots, N$, and

$$\sum_{n=1}^{N} \lambda_n = 1.$$

The resulting p is also a probability distribution over $\mathcal{X}$ (see Exercise 4.2) and we henceforth refer to p as a *mixture of* $p_1, \ldots, p_N$. More generally, given a finite set $\mathcal{X}$, let $\Delta(\mathcal{X})$ denote the set of all probability distributions over $\mathcal{X}$. A *mixture set* is any subset of $\Delta(\mathcal{X})$ that is closed under convex combinations.

Mixture set

Let $\mathcal{X}$ be a finite set and $\mathcal{M} \subset \Delta(\mathcal{X})$ an arbitrary set of probability distributions over $\mathcal{X}$. The set $\mathcal{M}$ is a *mixture set over* $\mathcal{X}$ if, given any $p_1, p_2 \in \mathcal{U}$ and any $\lambda \in [0, 1]$, the mixture

$$p(x) = \lambda p_1(x) + (1 - \lambda) p_2(x)$$

is in $\mathcal{M}$.

Given any $p_1, p_2 \in \Delta(\mathcal{X})$ and $\lambda \in [0,1]$, we henceforth write $m_\lambda(p_1, p_2)$ to denote the mixture

$$m_\lambda(p_1, p_2) \stackrel{\text{def}}{=} \lambda p_1 + (1-\lambda)p_2.$$

Given a mixture set $\mathcal{M}$, a function $u : \mathcal{M} \to \mathbb{R}$ is *linear on* $\mathcal{M}$ if

$$u(m_\lambda(p_1, p_2)) \stackrel{\text{def}}{=} u(\lambda p_1 + (1-\lambda)p_2) = \lambda u(p_1) + (1-\lambda)u(p_2),$$

for any $p_1, p_2 \in \mathcal{M}$ and any $\lambda \in [0,1]$. We are now in position to generalize the concept of *preference* to mixture sets.

Preference

A *strict preference* over a mixture set $\mathcal{M}$ is a binary relation $\succ$ on $\mathcal{M}$ such that

(i) The relation $\succ$ is asymmetric and negative transitive on $\mathcal{M}$ (i.e., is a weak order);

(ii) For any $p_1, p_2, p_3 \in \mathcal{M}$ and any $\lambda \in (0,1)$, if $p_1 \succ p_2$,

$$m_\lambda(p_1, p_3) \succ m_\lambda(p_2, p_3). \tag{4.2}$$

(iii) For any $p_1, p_2, p_3 \in \mathcal{M}$, if $p_1 \succ p_2$ and $p_2 \succ p_3$, then there are $\lambda_1, \lambda_2 \in (0,1)$ such that

$$m_{\lambda_1}(p_1, p_3) \succ p_2 \qquad \text{and} \qquad m_{\lambda_2}(p_2, p_3) \succ p_1. \tag{4.3}$$

As in Section 4.1, the requirements above ensure, in a sense, that the preference relation $\succ$ is "consistent":

- Requirement (i) is usually known as the *weak order requirement*, and was already discussed Section 4.1.
- Requirement (ii) is known as the *independence requirement*. It states that, if $p_1 \succ p_2$, then randomly selecting between p_1 and some p_3 should also be preferred to randomly selecting between p_2 and p_3, if the selection probabilities do not change.
- Requirement (iii) is known as the *continuity requirement*, and it states that if $p_1 \succ p_2$, randomizing between p_1 and p_3 should still be preferable to p_2, if the probability of selecting p_1 is very large. Conversely, randomizing between p_2 and p_3 should be preferable to p_1, if the probability of selecting p_3 is very large.

A preference over a mixture set $\mathcal{M}$ introduces an *order* among the probability distributions in $\mathcal{M}$ which can, in turn, be used to extend the decision-making framework from Section 4.1 to accommodate uncertainty.

4.2.2 Existence of utility functions ($\star$)

With the extension of the notion of preference to mixture sets, we now extend Theorem 4.2 accordingly, establishing the existence of an order preserving function u for preferences over mixture sets.

The following is the main result of this chapter and we choose to include its proof in the main text. Although somewhat involved, it provides important insights regarding the structure of a mixture set, and sets the stage for a more intuitive understanding of decision-making in the fact of uncertainty.

Theorem 4.4. *Let $\mathcal{M}$ denote a mixture set over some finite set $\mathcal{X}$, and $\succ$ a strict preference over $\mathcal{M}$. Then, there is a linear function $u : \mathcal{M} \to \mathbb{R}$ such that*

$$u(p_1) > u(p_2) \qquad \text{if and only if} \qquad p_1 \succ p_2, \tag{4.4}$$

for any $p_1, p_2 \in \mathcal{M}$. Moreover, if two linear functions u_1, u_2 exist that verify (4.4) on $\mathcal{M}$, then

$$u_2(p) = \alpha u_1(p) + K,$$

for some constant K and some $\alpha > 0$.

We emphasize that Theorem 4.4 states uniqueness of a *linear* utility function (up to a positive affine transformation), as there are many nonlinear order-preserving utility functions (see Exercise 4.1).

Existence

If there are no distributions $p, q \in \mathcal{M}$ such that $p \succ q$, then the result in Theorem 4.4 is trivially true. Hence, let us consider the case where there are at least two distributions $p_0, q_0 \in \mathcal{M}$ such that $p_0 \succ q_0$, and let us fix p_0 and q_0 for now. We use the following facts.

Lemma 4.5. *Given a mixture set $\mathcal{M}$ over $\mathcal{X}$, let p_1, p_2 and p_3 be arbitrary distributions in $\mathcal{M}$ and λ, λ_1 and λ_2 be arbitrary scalars in $[0, 1]$. Then,*

(a) *If $p_1 \succ p_2$ and $\lambda_1 > \lambda_2$, then $m_{\lambda_1}(p_1, p_2) \succ m_{\lambda_2}(p_1, p_2)$.*

(b) *If $p_1 \succeq p_2 \succeq p_3$ and $p_1 \succ p_3$, then there is a unique λ such that $p_2 \sim m_\lambda(p_1, p_3)$.*

(c) *If $p_1 \sim p_2$, then $m_\lambda(p_1, p_3) \sim m_\lambda(p_2, p_3)$.*

Proof. See Section 4.6. $\square$

Consider the set $\mathcal{M}_0 = \{p \in \mathcal{M} \mid p_0 \succeq p \succeq q_0\}$. For each distribution $p \in \mathcal{M}_0$, Lemma 4.5 (b) implies that there is λ_p such that

$$p \sim m_{\lambda_p}(p_0, q_0),$$

where $\lambda_{p_0} = 1$ and $\lambda_{q_0} = 0$. Moreover, for any $p, q \in \mathcal{M}_0$ such that $\lambda_p > \lambda_q$, Lemma 4.5 (a) implies that

$$m_{\lambda_p}(p_0, q_0) \succ m_{\lambda_q}(p_0, q_0).$$

Since, by construction, $p \sim m_{\lambda_p}(p_0, q_0)$ and $q \sim m_{\lambda_q}(p_0, q_0)$, it follows that $p \succ q$. In contrast, if $p, q \in \mathcal{M}_0$ are such that $\lambda_p = \lambda_q$, a similar reasoning yields that $p \sim q$. Therefore, in $\mathcal{M}_0$, the function $u(p) = \lambda_p$ is order preserving.

To show that it is also a linear function, note that $\mathcal{M}_0$ is a mixture set, i.e., for any $\lambda \in [0, 1]$ and any $p, q \in \mathcal{M}_0$, $m_\lambda(p, q) \in \mathcal{M}_0$. Then, by construction,

$$m_\lambda(p, q) \sim m_{u(m_\lambda(p,q))}(p_0, q_0).$$

On the other hand, by virtue of Lemma 4.5 (c), if $p \sim m_{\lambda_p}(p_0, q_0)$ and $q \sim m_{\lambda_q}(p_0, q_0)$, then

$$m_\lambda(p, q) \sim m_\lambda(m_{\lambda_p}(p_0, q_0), m_{\lambda_q}(p_0, q_0)).$$

Using a useful property of the mixtures m_λ (see Exercise 4.4),

$$m_\lambda(m_{\lambda_p}(p_0, q_0), m_{\lambda_q}(p_0, q_0)) = m_{\lambda\lambda_p + (1-\lambda)\lambda_q}(p_0, q_0),$$

yielding that

$$m_\lambda(p, q) \sim m_{\lambda\lambda_p + (1-\lambda)\lambda_q}(p_0, q_0).$$

Finally, all the above implies that

$$u(m_\lambda(p, q)) = \lambda\lambda_p + (1-\lambda)\lambda_q = \lambda u(p) + (1-\lambda)u(q),$$

and we showed that u is linear in $\mathcal{M}$.

To conclude establishing the first statement of the theorem, it remains to show that u can be extended to all of $\mathcal{M}$. Let $p_1, q_1 \notin \mathcal{M}_0$ be such that $p_1 \succeq p_0$ and $q_0 \succeq q_1$, and define the set

$$\mathcal{M}_1 = \{p \in \mathcal{M} \mid p_1 \succeq p \succeq q_1\}.$$

Clearly, $\mathcal{M}_0 \subset \mathcal{M}_1$. Moreover, $\mathcal{M}_1$ is also a mixture set and it is possible to repeat the process above to construct a linear utility u_1 on $\mathcal{M}_1$ that agrees with u in $\mathcal{M}_0$. In particular, we scale u_1 so that $u_1(p_0) = 1$ and $u_1(q_0) = 0$.

Similarly, let $p_2, q_2 \notin \mathcal{M}_0$ be such that $p_2 \succeq p_0$ and $q_0 \succeq q_2$, and set

$$\mathcal{M}_2 = \{p \in \mathcal{M} \mid p_2 \succeq p \succeq q_2\}.$$

Let u_2 denote the linear utility on $\mathcal{M}_2$ that agrees with u on $\mathcal{M}_0$, again scaled so that $u_1(p_0) = 1$ and $u_1(q_0) = 0$. For any $p \in \mathcal{M}_1 \cap \mathcal{M}_2$, we have that

- If $q_0 \succ p$, then, by Lemma 4.5 (b), there is $\lambda \in [0,1]$ such that $q_0 \sim m_\lambda(p_0, p)$; but then $u_k(q_0) = \lambda u_k(p_0) + (1-\lambda)u_k(p)$, for $k = 1, 2$. By construction, this means that $u_1(p) = u_2(p) = \lambda/(\lambda - 1)$.
- Similarly, if $p \succ p_0$, there is $\lambda \in [0,1]$ such that $p_0 \sim m_\lambda(p, q_0)$ and $u_k(p_0) = \lambda u_k(p) + (1-\lambda)u_k(q_0)$, for $k = 1, 2$. But then $u_1(p) = u_2(p) = 1/\lambda$.
- Finally, if $p_0 \succeq p \succeq q_0$, $p \in \mathcal{M}_0$ and $u_1(p) = u_2(p)$ by construction.

Therefore, the function u can uniquely be extended to arbitrary $\mathcal{M}_k$ such as those defined above. The desired conclusion follows from noting that it is possible to cover all of $\mathcal{M}$ with such sets.

Uniqueness up to affine transformation

We now establish the second statement of the Theorem, i.e., that u is unique up to an affine transformation. Suppose that v is a second linear utility over $\mathcal{M}$. As before, we fix $p_0, q_0 \in \mathcal{M}$ such that $p_0 \succ q_0$ and define, for $p \in \mathcal{M}$,

$$f(p) = \frac{u(p) - u(q_0)}{u(p_0) - u(q_0)} \qquad\qquad g(p) = \frac{v(p) - v(q_0)}{v(p_0) - v(q_0)}.$$

For any $p, q \in \mathcal{M}$, f and g are order preserving, since so are u and v. Moreover, $f(p_0) = g(p_0) = 1$ and $f(q_0) = g(q_0) = 0$ and we can repeat the previous reasoning to conclude that $f(p) = g(p)$ for all $p \in \mathcal{M}$. But then, $u(p) = \alpha v(p) + K$ with

$$\alpha = \frac{u(p_0) - u(q_0)}{v(p_0) - v(q_0)} \qquad\qquad K = u(q_0) - \alpha v(q_0).$$

4.2.3 Decision-making under uncertainty

We return to the situation of an agent facing the problem of selecting one among a set $\mathcal{A}$ of actions, where each action $a \in \mathcal{A}$ is associated with a distribution $P(a)$ over a set of possible outcomes, $\mathcal{X}$. Unlike Section 4.1, we now consider that the action of the agent is actually a r.v. a taking values in $\mathcal{A}$, and the agent is able to specify the distribution π used to sample a.[2]

A probability distribution π over $\mathcal{A}$ is henceforth called a *policy*, and we say that the agent follows policy π if

$$\mathbb{P}\left[\mathrm{a} = a\right] = \pi(a).$$

Note that considering the action of the agent a r.v. does not prevent deterministic action selection: selecting a policy that assigns probability 1 to an action $a \in \mathcal{A}$ is equivalent to deterministically selecting action a.

Let $\mathcal{M}$ denote the *convex hull* of the set $\{P(a), a \in \mathcal{A}\}$, i.e., the set of all convex combinations of the distributions in $\{P(a), a \in \mathcal{A}\}$. There is a one-to-one

[2] Although random action selection was also mentioned in Section 4.1, it was not formally discussed since it was immaterial for the conclusions of that section.

correspondence between the distributions in $\mathcal{M}$ and the policies available to the agent. In fact, any distribution $p \in \mathcal{M}$ can be written as

$$p = \sum_{a \in \mathcal{A}} \lambda_a P(a),$$

for some set of scalars $\{\lambda_a, a \in \mathcal{A}\}$, where each $\lambda_a \in [0,1]$ and

$$\sum_{a \in \mathcal{A}} \lambda_a = 1.$$

Therefore, the policy π_p defined as $\pi_p(a) = \lambda_a, a \in \mathcal{A}$ yields precisely the distribution p. It is straightforward to see that the converse is also true. For convenience of notation, in this section we write π_p to denote the policy corresponding to distribution p and, conversely, write p_π the distribution p corresponding to policy π.

Now given a preference $\succ$ over $\mathcal{M}$, let u denote a corresponding linear utility function, i.e., a linear function $u : \mathcal{M} \to \mathbb{R}$ such that if $p \succ q$ then $u(p) > u(q)$ for any $p, q \in \mathcal{M}$. As in Section 4.1, a rational agent selects its actions so as to maximize the resulting *expected utility*. Formally, let

$$Q(\pi) \stackrel{\text{def}}{=} u(p_\pi).$$

A rational agent thus selects the policy π^* such that

$$\pi^* \in \operatorname*{argmax}_\pi Q(\pi).$$

Alternative interpretation

At first sight, the formulation of expected utility is somewhat unintuitive, as it seems to require the decision-maker to express a preference over all possible probability distributions of outcomes. However, the formulation is both convenient, from a mathematical standpoint, and amenable to a more intuitive interpretation of the concept of expected utility.

We start with the observation that, given an arbitrary set $\mathcal{U}$ and a function $f : \mathcal{U} \to \mathbb{R}$, if p is a distribution p over $\mathcal{U}$, then

$$\sup_{x \in \mathcal{X}} f(x) = \sup_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} p(y) f(x) \geq \sum_{x \in \mathcal{X}} p(x) f(x), \forall x \in \mathcal{X}. \tag{4.5}$$

Since inequality (4.5) will be of use at several points in the book, we state it explicitly as the following lemma.

Lemma 4.6. *Let f be a real-valued function defined over some arbitrary set $\mathcal{U}$, and let p denote a probability distribution over $\mathcal{U}$. Then,*

$$\sup_{x \in \mathcal{X}} f(x) \geq \sum_{x \in \mathcal{X}} p(x) f(x).$$

Recall now that a rational agent selects the policy π^* such that

$$\pi^* \in \operatorname*{argmax}_{\pi} Q(\pi),$$

where

$$Q(\pi) \stackrel{\text{def}}{=} u(p_\pi) = u\left(\sum_{a \in \mathcal{A}} \pi(a) P(a)\right).$$

Due to the linearity of u,

$$Q(\pi) = \sum_{a \in \mathcal{A}} \pi(a) u(P(a)). \tag{4.6}$$

Therefore,

$$\pi^* \in \operatorname*{argmax}_{\pi} \sum_{a \in \mathcal{A}} \pi(a) u(P(a)).$$

Lemma 4.6 thus implies that

$$\max_{\pi} \sum_{a \in \mathcal{A}} \pi(a) u(P(a)) = \max_{a \in \mathcal{A}} u(P(a)).$$

Therefore, the agent needs only to consider *deterministic policies*—i.e., policies that select one action with probability 1. For simplicity of notation, if policy π selects action a with probability 1, we write $Q(a)$ to denote $Q(\pi)$. Using this notation, we again get that a rational agent will select an action a^* such that

$$a^* \in \operatorname*{argmax}_{a \in \mathcal{A}} Q(a) = \operatorname*{argmax}_{a \in \mathcal{A}} u(P(a)).$$

Now, each $P(a)$ is a distribution over some set of outcomes $\mathcal{X}$. We write $P(x \mid a)$ to explicitly denote the probability of outcome x under the distribution $P(a)$. Let $\Delta(\mathcal{X})$ denote the set of all probability distributions over $\mathcal{X}$, and δ_x denote the distribution that assigns probability 1 to outcome x.[3] Clearly, $\mathcal{M} \subset \Delta(\mathcal{X})$ and

$$P(x \mid a) = \sum_{y \in \mathcal{X}} P(y \mid a) \delta_x(y).$$

Let $\hat{u}$ denote the linear extension of u to $\Delta(\mathcal{X})$. The function $\hat{u}$ is a linear utility function that agrees with u on $\mathcal{M}$ and which can be obtained using the procedure described in Section 4.2.2. We write $\hat{u}(x)$ to denote $\hat{u}(\delta_x)$. We have that

$$Q(a) = \hat{u}\left(\sum_{x \in \mathcal{X}} P(x \mid a)\right)$$

which, due to the linearity of $\hat{u}$, can be written as

$$Q(a) = \sum_{x \in \mathcal{X}} P(x \mid a) \hat{u}(x) = \mathbb{E}\left[\hat{u}(\mathrm{x}) \mid \mathrm{a} = a\right],$$

[3]The probability distribution δ_x is the discrete counterpart to the Dirac distribution, and is commonly known as the *Kronecker delta function*.

i.e., $Q(a)$ can be interpreted as the *expected value of $\hat{u}$ when action a is selected.* Finally, the rational agent will select the action a^* such that

$$a^* \in \operatorname*{argmax}_{a \in \mathcal{A}} Q(a) = \operatorname*{argmax}_{a \in \mathcal{A}} \sum_{x \in \mathcal{X}} P(x \mid a)\hat{u}(x). \tag{4.7}$$

We conclude with several remarks:

- In light of (4.7), we can interpret the expected utility framework as follows. Each action $a \in \mathcal{A}$ yields a (random) outcome in $\mathcal{X}$, where the probability of outcome x given action a is given by $P(x \mid a)$. Associated with each outcome $x \in \mathcal{X}$ there is a utility value, $\hat{u}$, and the agent selects its actions so as to maximize the corresponding expected utility.
- It is also illuminating to compare (4.7) with (4.1). Writing (4.1) as

$$a^* \in \operatorname*{argmax}_{a \in \mathcal{A}} \sum_{x \in \mathcal{X}} \delta_x(X(a))u(x),$$

it becomes apparent that both formulations are essentially the same, with (4.1) corresponding to the situation where

$$P(x \mid a) = \delta_x(X(a)).$$

In other words, the decision-making framework discussed in Section 4.1 corresponds to the situation where the actions of the agent result in a deterministic outcome.

- From all of the above, it follows that a (one-shot) decision problem is fully specified under the expected utility framework as long as:
  (a) The outcome utilities, $u(x), x \in \mathcal{X}$, are defined. These utilities implicitly encode the *goal* of the agent, in that they provide the criterion that directs the agent's actions.
  (b) The outcome probabilities, $P(x \mid a), (x, a) \in \mathcal{X} \times \mathcal{A}$, are known. These probabilities, in a sense, model the *environment*, in that they describe how the environment responds to the agent's actions.

4.2.4 Examples

Let us illustrate the application of the expected utility framework in a number of simple examples.

The weather example

An individual must decide, before leaving home in the morning, whether to take her umbrella or not. The weather forecast for the day predicts rain with a probability 0.3; if it rains, there is a 0.5 probability that the person will be outside when it does.

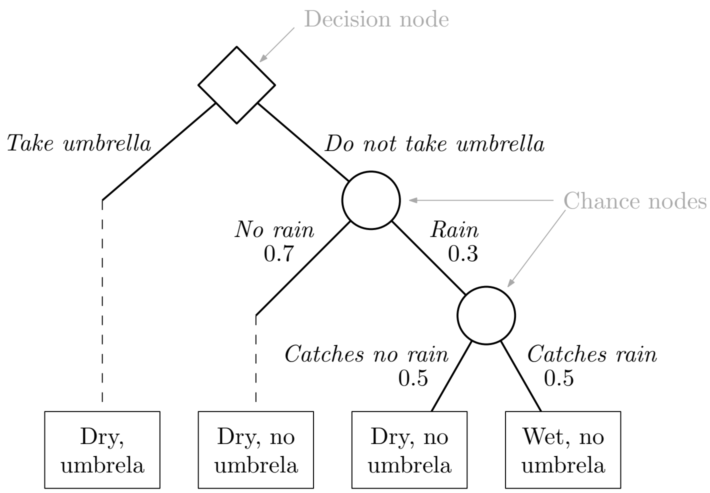

Figure 4.1 Decision tree for the weather example. The tree includes the alternatives/actions available to the agent and the possible outcomes with associated probabilities.

In order to model the decision problem of the individual using expected utility, we must specify the outcome probabilities and the outcome utilities. In order to do so, we describe the decision problem using a visual representation, which is a particular instance of a *decision tree*, from which the probabilities for different outcomes can be easily computed.

The decision-tree for this problem is represented in Fig. 4.1. The only *decision nodes*, at the top, is represented as a rhombus node. There is a branch coming out of the decision node for each action available to the agent. Circular nodes correspond to *chance nodes*, representing random events. Outgoing branches have an associated probability, describing the event associated with the chance node. In our example, the first chance node is concerned with rain—i.e., whether it will rain or not.

The outcomes correspond to the leaves of the tree. In this case, there are four possible outcomes:

A: The agent arrives home dry without an umbrella;

B: The agent arrives home dry with an umbrella;

C: The agent arrives home wet without an umbrella;

D: The agent arrives home wet with an umbrella.

The associated probabilities can be computed by multiplying all probabilities appearing in the branches from the root to the outcome node. An outcome not appearing in the tree has an associated probability of 0. This yields:

$$
\begin{aligned}
&P(A \mid \text{Take umbrella}) = 0; && P(A \mid \text{No umbrella}) = 0.7 + 0.3 \times 0.5 = 0.85;\\
&P(B \mid \text{Take umbrella}) = 1; && P(B \mid \text{No umbrella}) = 0;\\
&P(C \mid \text{Take umbrella}) = 0; && P(C \mid \text{No umbrella}) = 0.3 \times 0.5 = 0.15;\\
&P(D \mid \text{Take umbrella}) = 0; && P(D \mid \text{No umbrella}) = 0.
\end{aligned}
$$

The agent prefers not to carry the umbrella, but above all, it prefers not to get wet. This translates into the preference A $\succ$ B $\succ$ C $\succ$ D. We

We consider two situations.

- We observe that the agent takes the umbrella. If the agent is rational, this must mean that the action-value $Q(\text{Take umbrella})$ is larger than the action value $Q(\text{No umbrella})$. Assuming that $u(A) = 1$ and $u(D) = 0$ and using (4.7), we get
$$u(B) > 0.85 + 0.15u(C).$$
For example, the utility $u(B) = 0.9$ and $u(C) = 0.3$ explains the observed behavior of the agent.
- If, in contrast, the agent does not take the umbrella, $Q(\text{Take umbrella}) < Q(\text{No umbrella})$, which must mean that
$$u(B) < 0.85 + 0.15u(C).$$
For example, the utility $u(B) = 0.9$ and $u(C) = 0.4$ explains the observed behavior of the agent.

The two situations discussed illustrate the use of expected utility theory to *explain* the behavior of rational agents. The following example illustrates the use of expected utility theory to *predict* the behavior of rational agents.

The student example

A first year computer science student finished her final project for a programming course. When on her way to the University to submit the project report, she realizes that her printer failed to print half of the report. As such, she has got two possibilities:

A: Return home and print the remaining pages.

B: Print the remaining pages once she gets to the University.

If she decides to return home, there is a 0.6 probability that she will submit the project late, due to traffic on her way back to the University. Late projects are penalized with 2 values in the final grade.

On the other hand, if she decides to print at the University, there is a 0.3 probability that she may not find a printer at all and, in that case, she will submit

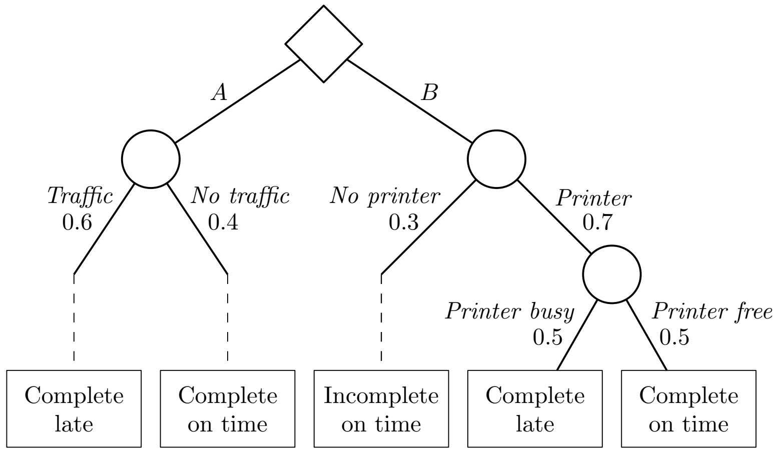

Figure 4.2 Decision tree for the student example.

her project in time but with missing pages. The missing pages will cost her 3 values in the final grade. Even if she manages to find a printer, there is a 0.5 probability that the printer is busy. In that case, she will submit the project late (but complete), incurring the corresponding penalty.

We can translate the preferences of the student by analyzing the impact that the different outcomes have in terms of the students grade. In particular, there are 3 possible outcomes, to which we can associate the following utilities:

- Submit a complete work on time, with an associated utility of 0.
- Submit a complete work but late, with an associated utility of -2.
- Submit an incomplete work, on time (with a utility of flàí3).

The decision tree associated with this problem is depicted in Fig. 4.2, yielding the outcome probabilities

$$
\begin{aligned}
&P(\text{Complete, late} \mid A) = 0.6; && P(\text{Complete, late} \mid B) = 0.35;\\
&P(\text{Complete, on time} \mid A) = 0.4; && P(\text{Complete, on time} \mid B) = 0.35;\\
&P(\text{Incomplete, late} \mid A) = 0; && P(\text{Incomplete, late} \mid B) = 0.3.
\end{aligned}
$$

Computing the expected utility for each of the two actions,

$$
\begin{aligned}
Q(A) &= 0.6 \times (-2) + 0.4 \times 0 + 0 \times (-3) = -1.2\\
Q(B) &= 0.35 \times (-2) + 0.35 \times 0 + 0.3 \times (-3) = -1.6.
\end{aligned}
$$

We can conclude, therefore, that A is the best action, i.e., the agent should return home to print the remaining pages.

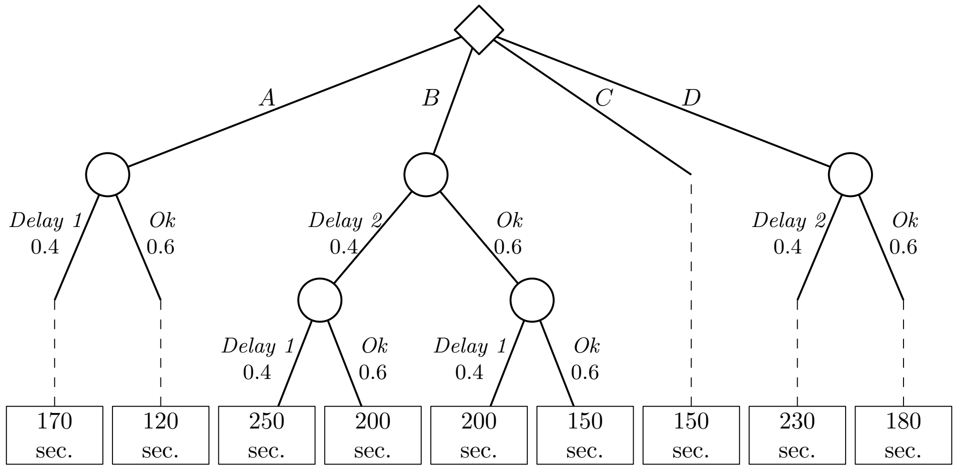

Figure 4.3 Decision tree for the robot example. The chance nodes correspond to the occurrence or not of delays before entering HW 1 (denoted as *Delay 1*) and HW 2 (denoted as *Delay 2*).

The robot example

Consider the household robot example from Chapter 1. A robot moves around the household environment (see Fig. 1.1 in page 4), assisting the human inhabitants in their daily chores. The robot is presently in the Hall and receives a request for assistance in the Kitchen. There are four different paths that the robot may take to reach the kitchen:

A: Hall → Living room → HW 1 → Pantry → Kitchen;

B: Hall → HW 2 → Living room → HW 1 → Pantry → Kitchen;

C: Hall → Living room → Bedroom → HW 1 → Pantry → Kitchen;

D: Hall → HW 2 → Living room → Bedroom → HW 1 → Pantry → Kitchen.

Moving between two rooms takes the robot around 30 seconds. However, there is a step between the Hall and HW 2 as well as between the Living room and HW 1, which the robot has difficulty in crossing. Between these rooms, with a probability 0.4, the robot takes around 80 seconds, instead of 30.

The decision tree for the problem of selecting which path to take is depicted in Fig. 4.3. There are six possible outcomes (corresponding to the different arrival times) and 4 possible actions available to the robot. The robot prefers to arrive as soon as possible, we it is possible to equate the utility with the (negative) travel time. The outcome probabilities and utilities are summarized in Table 4.1.

Table 4.1 Outcome probabilities and utilities for the robot example.

(a) Outcome probabilities.

Action:	A	B	C	D
120 sec	0.6	–	–	–
150 sec	–	0.36	1.0	–
170 sec	0.4	–	–	–
180 sec	–	–	–	0.6
200 sec	–	0.48	–	–
230 sec	–	–	–	0.4
250 sec	–	0.16	–	–

(b) Utilities.

Outcome:	Utility:
120 sec	−120
150 sec	−150
170 sec	−170
180 sec	−180
200 sec	−200
230 sec	−230
250 sec	−250

Computing the expected utility for each action, we get:

$$Q(A) = 0.6 \times (-120) + 0.4 \times (-170) = -140;$$
$$Q(B) = 0.36 \times (-150) + 0.48 \times (-200) + 0.16 \times (-250) = -190;$$
$$Q(C) = -150;$$
$$Q(D) = 0.6 \times (-180) + 0.4 \times (-230) = -200.$$

and the robot should thus take path A.

It should be clear that the navigation problem faced by the mobile robot has been formulated as a single decision problem to fit the expected utility framework discussed so far. However, it is an inherently sequential problem. We postpone to Chapter 5 the extension of the expected utility framework to sequential decision problems.

The two-envelope paradox

This example is sometimes referred as the *two-envelope paradox*. A contestant in a game must select one two envelopes, A and B. The game host reveals that one of the envelopes contains twice as much money as the other. However, it does not tell the contestant which one has more money or the exact amount in each envelope.

With no additional information to guide her actions, the contestant selects envelope A. Before opening the envelope, the contest host offers the contestant the possibility of keeping the prize in the envelope or switching to the other envelope. Given the symmetry of the situation, the contestant decides to keep the envelope, and opens it to reveal a prize of 10 EUR.

At this point, the host again offers the contestant the possibility of keeping the 10 EUR prize or switching envelopes. The corresponding decision tree is depicted in Fig. 4.4. At first sight, and assuming that the utility for each outcome corresponds exactly to the amount of money that the contestant receives, we get that

$$Q(\text{Keep}) = 10; \qquad Q(\text{Switch}) = 0.5 \times 20 + 0.5 \times 5 = 12.5.$$

The best action, then, is to switch. The interesting aspect of the paradox is that the reasoning remains valid whatever value is in the envelope. In fact, letting x

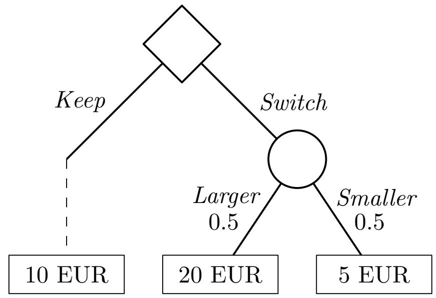

Figure 4.4 Decision tree for the envelope example.

denote the value in the envelope,

$$Q(\text{Keep}) = x \qquad Q(\text{Switch}) = 0.5 \times 2 \times x + 0.5 \times 0.5 \times x = 1.25x.$$

Therefore, *the agent does not actually need to open the envelope to decide that it is better to switch.*

One possible interpretation that solves the paradox is that, in general, the contestant would not be *risk neutral*, i.e., if the amount of money found in envelope A is large enough, the risk of losing half of it outweighs the possibility of doubling it. Formally, if we represent the amount of money in envelopes A and B as two r.v.s a and b, respectively, the above condition

$$\mathbb{E}\left[u(\mathrm{b}) \mid \mathrm{a} = a\right] \leq u(a)$$

for some values of $a \in \mathbb{R}$. We refer to the work of Powers (2015) for a more detailed discussion.

The St. Petersburg paradox

We now describe the widely known *St. Petersburg paradox*, which can be seen as a more elaborate version of the two-envelope paradox discussed above.

A casino in St. Petersburg offers the following game. A fair coin is tossed repeatedly until a tail appears and the game ends. The initial prize is 2 rubles, and the prize doubles for every coin toss. Assuming that the casino has unlimited resources, how much should a player pay to participate so that the game is fair?

The game is fair if the expected gain of the two participants (the casino and the player) is 0, since one player's loss is the other player's gain. We compute the expected winning of a player in the game to determine how much the player should pay to join.

If a tail comes out on the first coin toss, the player wins the initial 2 rubles. Since this can happen with a probability $\frac{1}{2}$, the expected gain in the first coin toss is $\frac{1}{2} \times 2 = 1$ ruble. If not, the game continues (which occurs with a probability $\frac{1}{2}$).

Repeating the reasoning above, the probability of a tail coming out in the second coin toss is also $\frac{1}{2}$, but in this case, the player will win 4 rubles, instead of 2. Therefore, the expected gain in the second coin toss is $\frac{1}{2} \times \frac{1}{2} \times 4 = 1$ ruble. Note that we multiply by $\frac{1}{2}$ twice since the second coin toss only occurs with a $\frac{1}{2}$ probability.

Repeating the process indefinitely, we have

$$\begin{aligned}
\text{Expected gain} &= \tfrac{1}{2} \times 2 + \tfrac{1}{2} \times \left(\tfrac{1}{2} \times 4 + \tfrac{1}{2} \times \left(\tfrac{1}{2} \times 8 + \tfrac{1}{2} \times \ldots\right)\right) \\
&= \tfrac{1}{2} \times 2 + \tfrac{1}{4} \times 4 + \tfrac{1}{8} \times 8 + \ldots \\
&= 1 + 1 + 1 + \ldots = \infty.
\end{aligned}$$

Therefore, no finite amount of money will make this a fair game. This means that—if utility is measured in terms of the winnings in the game—even if a person were asked to gamble all her lifetime savings in this game, it was still a good bet.

The paradox lies in the fact that no player would be willing to take this bet; as with the two-envelope paradox, the "solution" to the paradox is that players are not risk neutral. As such, players are unwilling to risk a large amount of money for a potentially very large prize that is, however, very unlikely to occur.

4.3 Subjective expected utility

According to the framework for decision-making reviewed so far, the decision problem encountered by the agent can be broken down as follows:

1. The agent selects an action $a \in \mathcal{A}$;
2. This action brings forth an outcome $x \in \mathcal{X}$ with a probability $P(x \mid a)$;
3. The agent elicits a utility $u(x)$ from the resulting outcome.

Within this framework, the probabilities $P(x \mid a)$ describe uncertainty regarding the outcome of the agent's action, and the utility u encodes the agent's preferences regarding the outcomes of its actions.

There is, however, a second line of work on expected utility, commonly known as the *subjective expected utility theory*. In this alternative formulation, $\mathcal{S}$ denotes the set of all mappings $S : \mathcal{A} \to \mathcal{X}$, i.e., all mappings from actions to outcomes. We write $S(a)$ to denote the outcome for action $a \in \mathcal{A}$ according to the mapping $S \in \mathcal{S}$,[4] and we write $\mathcal{X}(S)$ to denote the set

$$\mathcal{X}(S) = \{x \in \mathcal{X} \mid S(a) = x, a \in \mathcal{A}\},$$

i.e., $\mathcal{X}(S)$ is the set of all outcomes that may result from the agent's actions according to the mapping S. The set $\mathcal{X}(S)$ corresponds to the set of *relevant outcomes under* S. Each $S \in \mathcal{S}$, maps actions to outcomes and describes, therefore, a possible manner by which the environment responds to the actions of the agent. The mapping $S \in \mathcal{S}$ that actually "describes" the environment is known (according to the theory) as the *state of the world* and is unknown to the agent.

[4]The mapping X used in Section 4.1 was one element of $\mathcal{S}$.

Instead, the agent has access to a probability distribution P over $\mathcal{S}$ that describes the agent's *individual belief* regarding which is the actual state of the world. In other words, the probability $P(S)$ describes the agents (subjective) belief that S is the actual state of the world.

We do not pursue the theory of subjective expected utility any further in this book for two main reasons. First, under certain conditions on the set of states, it can be shown that the subjective expected utility theory and the expected utility theory discussed in Section 4.2 agree—that is the case, for example, if $\mathcal{X}(S) = \mathcal{X}$ for all $S \in \mathcal{S}$ (see Fishburn, 1970, Chapter 13).

Second, the main distinction between the expected utility theory and the subjective expected utility theory lies mostly in the interpretation of the associated probabilities: while in the former the probabilities P describe *uncertainty regarding the "response" of the world*, in the latter P describes *uncertainty regarding the agent's knowledge of the world*. For our purposes in the upcoming chapters it is convenient to explicitly distinguish between uncertainty in the world and uncertainty in the agent's perception of the world. As will soon become apparent, the expected utility theory discussed before provides a more natural framework to reason about the two.

4.4 Utility and cost

Throughout this chapter, we framed the agent's decision as that of maximizing the expected utility. Formally, the agent is provided with a description of the world (in the form of the probabilities P) and a representation of a preference over outcomes (in the form of a utility u), and the agent must determine the action a^* such that

$$a^* \in \underset{a \in \mathcal{A}}{\operatorname{argmax}} \sum_{x \in \mathcal{X}} P(x \mid a) u(x).$$

The utility value $u(x)$ translates the "gain" that the agent elicits from outcome x and our role, as agent designers, is to specify a utility function that reflects our preferences over outcomes and directs the agent's decision-making process. In this sense, the utility function represents the "goal" or "purpose" of our agent.

In many situations, however, it is more natural to express the *cost* of an outcome than the gain. Consider, for instance, the weather example discussed in Section 4.2. In this example, it could be easier to reason about the discomfort that carrying an umbrella or wearing wet clothes causes to the individual than the benefit of not carrying the umbrella and wearing dry clothes. Similarly, in the student example the utility of each outcome was modeled as the corresponding (negative) impact in the student's grade. Finally, in the robot example, the utility of the different routes was modeled as the (negative) corresponding travel time.

In all aforementioned examples, it is possible to interpret the decision problem as that of *minimizing the cost* associated with the outcome resulting from the agent's action. In the weather example, the agent seeks to minimize its discomfort. In the student example, the agent seeks to minimize the impact of its decision on the grade. Finally, in the robot example, the robot seeks to minimize the travel time.

Let $u : \mathcal{X} \to \mathbb{R}$ denote a utility function expressing a preference $\succ$ over a set $\mathcal{X}$ of possible outcomes. As seen before, the utility u is such that $x \succeq y$ if and only if $u(x) \geq u(y)$, for any $x, y \in \mathcal{X}$. The function $c : \mathcal{X} \to \mathbb{R}$ defined as $c(x) = -u(x)$ is such that $x \succeq y$ if and only if $c(x) \leq c(y)$, for any $x, y \in \mathcal{X}$, and we can express (4.7) equivalently as

$$a^* \in \operatorname*{argmin}_{a \in \mathcal{A}} \sum_{x \in \mathcal{X}} P(x \mid a) c(x). \tag{4.8}$$

The function c is called a *cost function* and, in this equivalent formulation, the agent selects the action that minimizes the *expected cost*. Note also that Proposition 4.3 is also valid for cost functions. For this reason—and except where otherwise noted—we can restrict our attention to cost functions taking values in the set $[0, 1]$.

4.5 Bibliographical notes

Expected utility theory is among the most widely used theories for decision-making, for which greatly contributed the classical book of von Neumann and Morgenstern (1944). A good historical overview of this theory can be found in the works of Stigler (1950a,b). The treatment presented herein closely follows the monographs by Fishburn (1970, 1982) although—in order to maintain consistency across the book—we differ significantly in terms of notation. Also, our presentation is necessarily more superficial. Nevertheless, we introduce the main ideas of expected utility and briefly go over the idea of *subjective utilities*, initially formalized in the classical work of Savage (1972) and discussed at length in the aforementioned monographs. The proofs presented in this chapter also follow Fishburn (1982).

The discussion of the two-envelope paradox follows Powers (2015), and we refer to that work for more details. Both the two-envelope paradox as well as the St. Petersburg paradox have been extensively discussed by mathematicians, economists and philosophers. They were originally proposed as arguments against the *expected value theory* in the 18th century by mathematicians such as Daniel Bernoulli (Bernoulli, 1738). The expected value theory is similar in flavor to the expected utility theory, but where wealth/value replaces the notion of utility. Therefore, there is no notion of risk which may explain the behavior that humans exhibit in both the two-envelope and the St. Petersburg paradoxes.

In spite of the key role of the expected utility theory, it is not without criticism. Several studies involving human decisions have shown that human behavior is sometimes in contradiction with expected utility theory, as in the well-known Alais and Ellsberg paradoxes (Alais, 1953; Ellsberg, 1961). Such studies have led to the development of several alternative decision theories, such as *prospect theory*. We refer to the monograph of Gilboa (2009) for an excellent overview of decision-making under uncertainty beyond expected utility theory.

4.6 Proofs

Proof of Lemma 4.1

We prove each statement separately.

(a) Given arbitrary $x, y \in \mathcal{X}$, 4 possible situations may occur:

- $x \succ y$ and $y \succ x$;
- $x \succ y$ and $y \not\succ x$ (or, simply, $x \succ y$);
- $x \not\succ y$ and $y \succ x$ (or, simply, $y \succ x$);
- $x \not\succ y$ and $y \not\succ x$ (or, equivalently, $x \sim y$).

The first situation is not possible, since $\succ$ is asymmetric. The conclusion follows.

(b) Suppose that $x \succ y$. Negative transitivity implies that, for any $z \in \mathcal{X}$, either $x \succ z$ or $z \succ y$ (or both). If, furthermore, we know that $y \succ z$, asymmetry implies that $z \not\succ y$, and we must conclude that $x \succ z$.

(c) Recall that $x \sim y$ means that $x \not\succ y$ and $y \not\succ x$. Then,

- From the definition, $\sim$ is clearly symmetric.
- For any $x \in \mathcal{X}$, $x \not\succ x$ for, otherwise, both symmetry and negative transitivity would not hold for $\succ$. Therefore, indifference is reflexive.
- Finally, to see that is transitive, suppose that $x \sim y$ and $y \sim z$ but $x \not\sim z$. Then, either $x \succ z$ or $z \succ x$. Suppose that $x \succ z$. By negative transitivity, either $y \succ z$ or $x \succ y$—contradicting either $y \sim z$ or $x \sim y$. Then it must be transitive.

(d) Suppose that $x \succ y$. Negative transitivity implies that, for any $z \in \mathcal{X}$, either $z \succ y$ or $x \succ z$. If, furthermore, $x \sim z$, it must hold that $z \succ y$. The second statement is established in exactly the same manner.

(e) Transitivity of $\succeq$ follows from (b), (c) and (d). Completeness follows from (a): for any $x, y \in \mathcal{X}$ either $x \succ y$ (and, thus, $x \succeq y$), or $y \succ x$ (and, thus, $y \succeq x$) or $x \sim y$ (and, thus, both $x \succeq y$ and $y \succeq x$).

Proof of Lemma 4.5

For any $p, q \in \mathcal{M}$ and any $\lambda, \mu, \alpha, \beta, \gamma \in [0, 1]$, we have that (see Exercise 4.4):

$$m_\lambda(p, p) = p; \tag{4.9}$$
$$m_\lambda(p, q) = m_{1-\lambda}(p, q); \tag{4.10}$$
$$m_\lambda(m_\mu(p, q), q) = m_{\lambda\mu}(p, q); \tag{4.11}$$
$$m_\alpha(m_\beta(p, q), m_\gamma(p, q)) = m_{\alpha\beta+(1-\alpha)\gamma}(p, q). \tag{4.12}$$

We can now prove each statement separately.

(a) Taking $\lambda = 1 - \lambda_2$ and $p_3 = p_1$ in (4.2), it follows that $p_1 \succ m_{\lambda_2}(p_1, p_2)$. On the other hand, using (4.11),

$$\begin{aligned} m_{\lambda_1}(p_1, p_2) &= m_{1-\lambda_1}(p_2, p_1) \\ &= m_{(1-\lambda_1)/(1-\lambda_2)}(m_{1-\lambda_2}(p_2, p_1), p_1) \\ &= m_{(1-\lambda_1)/(1-\lambda_2)}(m_{\lambda_2}(p_1, p_2), p_1) \\ &= m_{(\lambda_1-\lambda_2)/(1-\lambda_2)}(p_1, m_{\lambda_2}(p_1, p_2)). \end{aligned}$$

Using the fact that $p_1 \succ m_{\lambda_2}(p_1, p_2)$, it follows that

$$\begin{aligned} m_{\lambda_1}(p_1, p_2) &= m_{(\lambda_1-\lambda_2)/(1-\lambda_2)}(p_1, m_{\lambda_2}(p_1, p_2)) \\ &\succ m_{(\lambda_1-\lambda_2)/(1-\lambda_2)}(m_{\lambda_2}(p_1, p_2), m_{\lambda_2}(p_1, p_2)) \\ &= m_{\lambda_2}(p_1, p_2). \end{aligned}$$

(b) Suppose, first, that $p_1 \sim p_2$ and $p_2 \succ p_3$. Then,

$$p_2 \sim m_1(p_1, p_2) \succ m_\lambda(p_1, p_2),$$

for any $\lambda < 1$, where the last inequality comes from (a). Therefore, $p_2 \succ m_\lambda(p_1, p_2)$ for a single λ (namely, $\lambda = 1$). A similar derivation holds for the case where $p_2 \sim p_3$. Finally, suppose that $p_1 \succ p_2 \succ p_3$. Then, from (4.3), there is $\lambda_0 > 0$ such that $p_2 \succ m_{\lambda_0}(p_1, p_2)$. From ($a$), such inequality holds for all $\lambda > \lambda_0$. However, it does not hold for $\lambda = 0$. Therefore, there must be some $\lambda \in [0, 1]$ such that

$$m_{\lambda_1}(p_1, p_2) \succ p_2 \succ m_{\lambda_2}(p_1, p_2),$$

for all $\lambda_1 > \lambda$ and $\lambda_2 < \lambda$.

(c) The conclusion trivially follows if $\lambda = 0$, $\lambda = 1$ or $p_1 \sim p_3$. We focus, then, on the case where $\lambda \in (0, 1)$ and $p_1 \succ p_3$ (the converse case is identical). From (4.2), it follows that $m_\lambda(p_1, p_3) \succ p_3$.

Suppose now that $m_\lambda(p_2, p_3) \succ m_\lambda(p_1, p_3)$. Since $m_\lambda(p_1, p_3) \succ p_3$, statement ($b$) implies that there is a unique λ_0 such that

$$m_\lambda(p_1, p_3) \sim m_{\lambda_0}(m_\lambda(p_2, p_3), p_3) = m_{\lambda\lambda_0}(p_2, p_3). \tag{4.13}$$

On the other hand, since $p_1 \sim p_2$ and we are considering $p_1 \succ p_3$, $p_2 \succ p_3$. Therefore, $p_2 \succ m_{\lambda_0}(p_2, p_3)$ and $p_1 \succ m_{\lambda_0}(p_2, p_3)$. This finally yields

$$m_\lambda(p_1, p_3) \succ m_\lambda(m_{\lambda_0}(p_2, p_3), p_3) = m_{\lambda\lambda_0}(p_2, p_3),$$

which contradicts (4.13). Then, it must be false that $m_\lambda(p_2, p_3) \succ m_\lambda(p_1, p_3)$. A similar process leads to the conclusion that $m_\lambda(p_1, p_3) \not\succ m_\lambda(p_2, p_3)$, thus establishing that $m_\lambda(p_1, p_3) \sim m_\lambda(p_2, p_3)$.

4.7 Exercises

Exercise 4.1.

(a) Prove that, if $u : \mathcal{X} \to \mathbb{R}$ is a utility function representing a rational preference $\succeq$ over some finite set of outcomes, then so is the function u', defined as $u'(x) = f(u(x))$, where $f : \mathbb{R} \to \mathbb{R}$ is some strictly increasing function.

(b) Repeat the proof in (a) for the case where u is a utility function representing a preference $\succ$ over a mixture set $\mathcal{M}$.

Exercise 4.2.
Given arbitrary probability distributions $p_1, \ldots, p_N$ over some finite set $\mathcal{X}$, show that the mixture

$$p = \sum_{n=1}^{N} \lambda_n p_n$$

is also a probability distribution over $\mathcal{X}$. Recall that $\lambda_1, \ldots, \lambda_N$ are scalars such that, for every $n = 1, \ldots, N$, $0 \leq \lambda_n \leq 1$, and

$$\sum_{n=1}^{N} \lambda_n = 1.$$

Exercise 4.3.
Given any $p_1, p_2 \in \Delta(\mathcal{X})$ and any $\lambda, \lambda_1, \lambda_2, \lambda_3 \in [0, 1]$, show that

- $m_\lambda(p_1, p_1) = p_1$;
- $m_\lambda(p_1, p_2) = m_{1-\lambda}(p_2, p_1)$;
- $m_{\lambda_1}(m_{\lambda_2}(p_1, p_2), p_2) = m_{\lambda_1 \lambda_2}(p_1, p_2)$;
- $m_{\lambda_1}(m_{\lambda_2}(p_1, p_2), m_{\lambda_3}(p_1, p_2)) = m_{\lambda_1 \lambda_2 + (1-\lambda_1)\lambda_3}(p_1, p_2)$.

Exercise 4.4.
Let $\mathcal{M}$ denote a mixture set. Show that:

(a) For any $p \in \mathcal{M}$ and any $\lambda \in [0, 1]$,

$$m_\lambda(p, p) = p.$$

(b) For any $p_1, p_2 \in \mathcal{M}$ and any $\lambda \in [0, 1]$,

$$m_\lambda(p_1, p_2) = m_{1-\lambda}(p_2, p_1).$$

(c) For any $p_1, p_2 \in \mathcal{M}$ and any $\lambda_1, \lambda_2 \in [0, 1]$,

$$m_{\lambda_1}(m_{\lambda_2}(p_1, p_2), p_2) = m_{\lambda 1 \lambda_2}(p_1, p_2).$$

(d) For any $p_1, p_2 \in \mathcal{M}$ and any $\lambda_1, \lambda_2, \lambda_3 \in [0, 1]$,

$$m_{\lambda_1}(m_{\lambda_2}(p_1, p_2), m_{\lambda_3}(p_1, p_2)) = m_{\lambda 1 \lambda_2 + (1-\lambda_1)\lambda_3}(p_1, p_2).$$