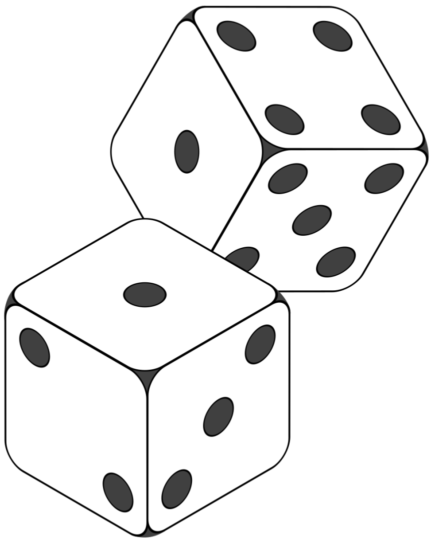

Decision Theory and Human Behaviour

Rational Choice, Utility, Risk, Uncertainty and Ambiguity

Giuseppe Bassi

Academic Year 2024/2025

An introductory course

Abstract

These notes originate from a personal undertaking to consolidate and summarize the core material of a Decision Theory course taught by Professors Massimo Marinacci and Fabio Angelo Maccheroni at Bocconi University. They offer a concise re-elaboration of the main topics covered, rather than an exhaustive, textbook-level treatment. Hence, they are not intended to serve as an official reference and do not provide comprehensive coverage of the entire syllabus. I encourage readers to suggest improvements or report any errors they may encounter so that future editions can be improved.

Version 1.0 Date: September 23, 2025

Contents

1 Preliminary Results

In this section, we introduce key concepts and a taxonomy that, while not typically asked directly, establish the foundational knowledge required for a comprehensive understanding of the program. Feel free to skip this information initially and return to it as necessary. More advanced topics which are not essential but still included in the syllabus are discussed in the appendix.

1.1 What are we studying?

Rational choice theory, also known as neoclassical economics, examines the decision-making process of rational beings, referred to here as **decision makers** (DMs). It is assumed that these individuals make optimal choices among alternatives.
We assume that there is a **choice space $\mathbf{X}$** of conceivable alternatives and a collection $\mathcal{X}$ of nonempty subsets X of $\mathbf{X}$, called **choice sets**. Each choice set X is a collection of distinct alternatives the DM has to choose from. We call the pair $(\mathbf{X}, \mathcal{X})$ a **decision framework**.
In our formulation, the decision framework is considered convex, meaning that the choice space $\mathbf{X}$ is a subset of a locally convex topologically ordered vector space, and $\mathcal{X}$ includes convex choice sets.
For example, in consumer theory, $\mathbf{X}$ is a convex subset of $\mathbb{R}^n_+$ consisting of consumption bundles $x = (x_1, \ldots, x_n)$, and $\mathcal{X}$ is the collection of budget sets X that a consumer may face according to their wealth and market prices.

1.2 Logic

Definition 1.1 (Proposition). Statement that can be either true or false. For example, *"4 is an even number"*.

Definition 1.2 (Predicate). An expression $p(x)$ whose truth value depends on the value of the argument x, with x belonging to a domain X. For example, if $p(x) =$ *"x is an even number"*, then $p(4)$ is a true proposition, while $p(3)$ is a false proposition. In our analysis, we will focus on binary predicates $p(x, y)$, depending on two arguments.

Definition 1.3 (Extension of a predicate). The subset A of X for which $p(x)$ is true, that is, $A = \{x \in X : p(x) \text{ true}\}$. For binary predicates, the extension R is defined as $R = \{(x, y) \in X \times Y : p(x, y) \text{ true}\}$.

Definition 1.4 (Binary relation). Subset R of $X \times X$. If $(a, b) \in R$, we simply write aRb, which reads "a is in the relation R with b". Alternatively, a binary relation can be defined as the extension R of a binary predicate $p(x, y)$, with $(x, y) \in X \times X$.

Definition 1.5 (Properties of binary relations). A binary relation R on X can be:

- **Reflexive:** if xRx for all $x \in X$;
- **Transitive:** if xRy and yRz implies xRz for all $x, y, z \in X$;
- **Complete:** if for every $x, y \in X$, either xRy or yRx or both; for example, $>$ is complete in $\mathbb{R}$ but not in $\mathbb{R}^2$. Completeness implies reflexivity;
- **Symmetric:** if xRy implies yRx for all $x, y \in X$;
- **Asymmetric:** if xRy implies not yRx for all $x, y \in X$;
- **Antisymmetric:** if xRy and yRx implies $x = y$ for all $x, y \in X$. For example, $\geq$ is an antisymmetric relation.

Definition 1.6 (Classification of relations). We have the following taxonomy:

- **Preorder:** reflexivity + transitivity.
- **Weak order:** transitivity + completeness.
- **Partial order:** reflexivity + transitivity + antisymmetry.
- **Total order:** transitivity + antisymmetry + completeness.
- **Equivalence relation:** reflexivity + transitivity + symmetry.

Definition 1.7 (Preference). In decision theory, binary relations are called *preferences* and delineate how a decision maker ranks alternatives within a choice set X. These are represented by $\succsim_X$, or more succinctly as $\succsim$. The preference $\succsim$ induces other binary relations:

i) $x \succsim y$ if the DM either strictly prefers x to y or is indifferent between x and y;

ii) $x \succ y$ if the DM strictly prefers x to y. Formally, $x \succ y$ if $x \succsim y$ but not $y \succsim x$;

iii) $x \sim y$ if the DM is indifferent between x and y. Formally, $x \sim y$ if both $x \succsim y$ and $y \succsim x$;

iv) $x \| y$ if x and y are not comparable in X, since the alternatives are too unfamiliar for the DM (usually excluded by assuming completeness).

Lemma 1.8. *If $\succsim$ is reflexive and transitive, then $\sim$ is an equivalence relation.*

Proof.

i) Reflexivity: Let $x, y \in X$ and $x = y$. By reflexivity of $\succsim$, we have $x \succsim y$ and $y \succsim x$. By definition, this implies $x \sim y$.

ii) Symmetry: Let $x, y \in X$ and $x \sim y$. By definition, we have $x \succsim y$ and $y \succsim x$, which implies, again by definition, $y \sim x$.

iii) Transitivity: let $x, y, z \in X$ with $x \sim y$ and $y \sim z$. By definition, we have $x \succsim y$, $y \succsim x$, $y \succsim z$ and $z \succsim y$. By transitivity applied on the first and third relation, and on the second and fourth, we get $x \succsim z$ and $z \succsim x$, that is, $x \sim z$.

□

Definition 1.9 (Indifference curves). If $\succsim$ is reflexive and transitive, we call *indifference curves* the equivalence classes

$$[x] = \{y \in X : y \sim x\}, \qquad x \in X.$$

We also denote by

$$X/\sim \; = \; \{[x] : x \in X\}$$

the *quotient space* of X, i.e. the collection of all indifference curves.[1]

1.3 Some Topological Notions

Definition 1.10 (Topology). Topology denotes the field of mathematics studying *proximity* in general spaces, dealing with concepts such as continuity and convergence. We will focus on metric spaces, where we can define the notion of distance between elements. For our purposes, a prototypical environment is $\mathbb{R}^n$ with its natural order $\geq$ and Euclidean topology generated by the Euclidean norm $\|x\| = \sqrt{\sum x_i^2}$.

Definition 1.11 (Ordered vector space). A set endowed with a partial order $\leq$ constitutes an ordered space. Examples include $(\mathbb{R}^n, \leq)$ and $(2^X, \subseteq)$, where 2^X is the power set of a set X. Furthermore, if $x \leq y \implies x + z \leq y + z$ for all $x, y, z \in V$, and $x \leq y \implies \alpha x \leq \alpha y$ for all $\alpha \geq 0$, then V is an ordered vector space.

Definition 1.12 (Space for Order Axioms). Order axioms, such as monotonicity (axiom A.3), apply in ordered topological vector spaces with a partial order $\geq$ and topology τ. For our purposes, imagine simply $\mathbb{R}^n$. Here, $x > y$ means $x_i \geq y_i$ for each $i = 1, \dots, n$, with at least one strict inequality, and $x \gg y$ indicates $x_i > y_i$ for each $i = 1, \dots, n$. For instance, in $\mathbb{R}^2$, we have $(1, 2) \gg (0, 1) \geq (0, 0)$. Note that $\gg$ does not compare $(1, 1)$ and $(2, 1)$, hence it is not a complete relation.

Definition 1.13 (Order interval). In an Archimedean M-space $(V, \geq, \|\cdot\|)$, such as $(\mathbb{R}^n, \geq, \|\cdot\|)$, $X \subseteq V$ is an *order interval* if, for all $y, z \in X$ with $y \leq z$, we have

$$x \in X \iff y \leq x \leq z.$$

It can be open, closed, or half-open. We call **interval body** an order interval that is either closed, open, or half-open.[2]

Example 1.14. Some examples include:

[1] The quotient space is a partition of X, since each alternative $x \in X$ belongs to one, and only one, indifference curve.

[2] In economics, it is commonly assumed that utility functions are strongly monotone and quasi-concave on an ordered interval.

Table 1: Axioms of Preferences

Set-theoretic Axioms	**Order Axioms**	**Algebraic Axioms**
Reflexivity	Monotonicity	Archimedean
Transitivity	Strict monotonicity	Convexity
Strict transitivity	Strong monotonicity	Strict convexity
Completeness		Affinity

i) $\mathbb{R}^n_+ = [\mathbf{0}, \infty)$ is a closed order interval, containing all the points with all coordinates ≥ 0;

ii) $[(1,1),(2,3)] = \{(x,y) \in \mathbb{R}^2 : 1 \leq x \leq 2,\ 1 \leq y \leq 3\}$ is a closed order interval;

iii) $\mathbb{R}^n_{++} = (\mathbf{0}, \infty)$ is an open order interval, containing all the points with all coordinates > 0;

iv) $((1,1),(2,3)] = \{(x,y) \in \mathbb{R}^2 : 1 < x \leq 2,\ 1 < y \leq 3\}$ is a half-open order interval.

1.4 Quasi-Concavity

Definition 1.15 (Concavity). A function $f : C \to \mathbb{R}$ is *concave* if

$$f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha) f(y), \qquad \forall x, y \in V,\ \alpha \in [0,1].$$

Definition 1.16 (Quasi-concavity). A function $f : C \to \mathbb{R}$ is *quasi-concave* if

$$f(\alpha x + (1-\alpha)y) \geq \min\{f(x), f(y)\}, \qquad \forall x, y \in V,\ \alpha \in [0,1].$$

We have:

i) f is quasi-concave iff all upper level sets $(f \geq t) = \{x \in A : f(x) \geq t\}$ are convex for all $t \in \mathbb{R}$;

ii) Convex functions are quasi-convex, concave functions are quasi-concave, affine functions are quasi-affine (the converse does not hold);

iii) Monotone functions are quasi-affine, thus quasi-concave and quasi-convex. In particular, $f(x) = e^x$ is an example of a quasi-concave (since increasing) function which is strictly convex.

Remark 1.17 (Ordinal invariance of quasi-concave functions). If f is quasi-concave and φ is increasing, then $\varphi \circ f$ is quasi-concave for any increasing φ.[3]
In fact,

$$(\varphi \circ f)(\alpha x + (1-\alpha)y) \geq \varphi(\min\{f(x), f(y)\}) = \min\{(\varphi \circ f)(x), (\varphi \circ f)(y)\}.$$

Remark 1.18 (Irreducibility of quasi-concave functions). Quasi-concave functions are not ordinally concave, meaning that there are quasi-concave functions f such that, for any increasing φ, $\varphi \circ f$ is not concave. Consequently, some quasi-concave utility functions cannot be ordinally reduced to concave functions. This is a significant limitation because concave utility functions are much more manageable in optimisation problems.

2 Utility Theory

2.1 Axioms of Preferences

In this section we will analyze the axioms summarized in Table 1, which formalize some important economic ideas.

Extensionality: if $X = X'$, then $x \succsim_X y$ if and only if $x \succsim_{X'} y$ for all $x, y \in X$.

Comment: This assumption presupposes that preferences are determined solely by the available alternatives, **not by the manner in which they are described**. For instance, suppose our choices on goods and services are based on budget sets, namely pairs of prices and wealth, i.e. $X = B(p, w)$. Now, if both prices and wealth were to double, resulting in $X' = B(2p, 2w)$, the description of the budget set changes. However, extensionality requires that the elements of the

[3] This property does not hold for concave functions, which are preserved only if φ is both monotone increasing and concave. Quasi-concavity is thus an ordinal property, while concavity is not.

set do not change.[4] Such property may be falsiable if the DM is somehow biased -for instance, it suffers from *money illusion*-, but we abstract from such *framing effects*.

(A.0) Reflexivity: for all $x \in X$, it holds $x \succsim x$.

(A.1) Transitivity: for all $x, y, z \in X$, $x \succsim y$ and $y \succsim z$ imply $x \succsim z$.

Comment: transitivity is the foundational formalization of the concept of "*rationality*" in choice behaviour, and it's essential for to consistently defining rational choice problems as optimization problems. If transitivity is violated, cycles such as $x \succ y \succ z \succ x$ can occur. The presence of such cycles indicates that there may be no optimal solution, leading to pairwise inconsistencies in evaluations. A stronger version of transitivity is discussed in the appendix.

(A.2) Completeness: for all $x, y \in X$, $x \succsim y$ or $y \succsim x$.

Comment: completeness is not a rationality assumption like transitivity, but rather an *information assumption* associated with the decision maker's (DM's) familiarity with alternatives.

(A.3) Strict Monotonicity: for all, $x, y \in X$, $x > y \implies x \succ y$.[5]

Comment: Strict monotonicity essentially says that "the more of any good, the better", which assumes that all goods are essential, the DM cares about each of them. However, there are also weaker forms of monotonicity:

(A.3bis) Strong Monotonicity: for all, $x, y \in X$, $x \gg y \implies x \succ y$, and $x \geq y \implies x \succsim y$.[6]

(A.3ter) (weak) Monotonicity: for all, $x, y \in X$, $x \geq y \implies x \succsim y$.

Comment: (weak) monotonicity models **satiation** of goods, that is, a maximal quantity $\overline{x}$ above which our utility does not increase (but doesn't decrease either).

Notice that **(A.3)** $\implies$ **(A.3bis)** $\implies$ **(A.3ter)** (prove it as an exercise).

The next two axioms require X to be a convex set.

(A.4) Archimedean: for all, $x, y, z \in X$, with $x \succ y \succ z$, there exists $\alpha, \beta \in (0, 1)$ such that

$$\alpha x + (1 - \alpha)z \succ y \succ \beta x + (1 - \beta)z$$

Comment: According to this axiom, there are no infinitely preferred or infinitely despised alternatives. Specifically, given any pairs $x \succ y$ and $y \succ z$, there always exists a mixture of x and z that is preferred to y, and a mixture of x and z that is less preferred than y. This is the only axiom that cannot be easily behaviorally falsified, thus it is typically assumed to hold, since it's a form of continuity of $\succsim$ on X.

(A.5) Convexity: for all, $x, y \in X$ and $\alpha \in [0, 1]$, $x \sim y \implies \alpha x + (1 - \alpha)y \succsim x$.

Comment: Convexity formalizes the *principle of diversification*, which can be interpreted differently depending on the context. When alternatives are bundles of goods, it reflects a preference for variety, suggesting that a mixture of goods is preferred over having a high quantity of one good and a low quantity of another. When alternatives are financial assets, convexity indicates a preference for hedging; when alternatives are strategies, it models a preference for mixed strategies, that is, for randomization. The dual notion of convexity is concavity, which indicates a dislike for diversification.

(A.5) Strict convexity: for all, $x, y \in X$ and $\alpha \in (0, 1)$, $x \sim y \implies \alpha x + (1 - \alpha)y \succ x$.

(A.5) Affinity: affinity expresses indifference to diversification: for all, $x, y \in X$ and $\alpha \in [0, 1]$, $x \sim y \implies \alpha x + (1 - \alpha)y \sim x$.

[4]This property will be clear once defined the menu correspondence, and will imply the positive homogeneity of degree zero of the demand correspondence and of the indirect utility.

[5]Monotonicity axioms require X to be a subset of an ordered topological vector space V, characterized by a partial order $\geq$ and topology τ. Therefore, we regard $(V, \geq, \tau)$ as the space of our analysis. For simplicity, imagine simply $\mathbb{R}^n$. Additional details on this topic, including the notations $x > y$ and $x \gg y$, are discussed in chapter 1.

[6]Strong monotonicity allows goods to be **perfect complements**. For example, consider the preference on $\mathbb{R}^2_+$ defined by $x \succsim y \Leftrightarrow \min\{x_1, x_2\} \geq \min\{y_1, y_2\}$. This is a strongly monotone preference which is not strictly monotone, and can model the need of x_1 right shoes and x_2 left shoes.

2.2 Ordinal Utility Theory

In this section we will study utility functions uniquely based on an ordinal, rather than cardinal, analysis. Utility functions are helpful because they translate preference relations $\succsim$ ingo numerical terms. As we will see, utility functions are **unique up to strictly increasing transformations**, meaning that distinct utility functions may characterize the same preference, and thus, the same choice behaviour. Therefore, they can be treated only in ordinal terms, that is, comparing the utilities of two different alternatives, but not as a cardinal measurement where a real value has inherent meaning beyond being greater or smaller than another, since the actual difference $u(x) - u(y)$ is meaningless, except for its sign. In particular, comparisons like $u(x) - u(y) \geq u(z) - u(w)$ are irrelevant, since it's sufficient to take a strictly increasing transformation to reverse the inequality (try to make an example). Consequently, in ordinal utility theory, difference quotients and derivatives, which require cardinality, cannot be analyzed. We will conclude this section with some theoretical results regarding the existence of a utility function given a preference relation $\succsim$ and a set X of alternatives.

Definition 2.1 (Utility function). Given a preference relation $\succsim$ on X, a utility function is a function $u : X \to \mathbb{R}$ which satisfies

$$u(x) \geq u(y) \iff x \succsim y, \qquad u(x) > u(y) \iff x \succ y, \qquad u(x) = u(y) \iff x \sim y,$$

where the first equivalence is the definition of u and the others are trivial consequences.

Definition 2.2 (Indifference curves). These definitions are equivalent:

$$[x] = \{y \in X : y \sim x\} \;=\; \{y \in X : u(y) = u(x)\}.$$

Theorem 2.3 (Ordinal invariance of utility functions). *If $u : X \to \mathbb{R}$ is a utility function for $\succsim$, then a function $u' : X \to \mathbb{R}$ is also a utility function for $\succsim$ if and only if there exists a strictly increasing function $f : Im\,u \to \mathbb{R}$ such that $u' = f \circ u$.*

Proof.

We only prove the "if" direction. We need a way to characterize strictly increasing functions. We know that a strictly increasing function satisfies $a > b \implies f(a) > f(b)$, but this is a sufficient condition, what about a necessary condition? The necessary (and sufficient) condition we are looking for is:[a]

$$f \text{ strictly increasing if and only if } \quad f(a) \geq f(b) \Leftrightarrow a \geq b.$$

If f is strictly increasing and u is a utility function, then

$$f(u(x)) \geq f(u(y)) \iff u(x) \geq u(y) \iff x \succsim y, \quad \forall x, y \in X,$$

where the first equivalence uses the characterization of strictly increasing function, and the second the definition of u. So $f \circ u$ is a utility function. □

[a] f is strictly increasing if $a > b$ implies $f(a) > f(b)$. From this definition, you can derive the following result, which is left as a trivial exercise, as it's easier to perform than to explain.

Example 2.4. Utility functions are unique (invariant) up to strictly increasing transformations, even featuring highly different marginal utility patterns. For instance, there are many versions of the classic *Cobb–Douglas* utility function $u : \mathbb{R}^2_+ \to \mathbb{R}$:

i) Cobb–Douglas: $u(x_1, x_2) = x_1^a x_2^{1-a}$, with $0 \leq a \leq 1$;

ii) Log-linear utility: $\log\big(u(x_1, x_2)\big) = a \log x_1 + (1 - a) \log x_2$;

iii) Quadratic Cobb–Douglas: $u^2(x_1, x_2) = x_1^{2a} x_2^{2-2a}$.

Luckily, the optimal bundles which maximise the utility u subject to some constraints are invariant to the utility function chosen.

2.3 Existence of Utility Functions

Proposition 2.5 (First necessary condition). *A preference relation has a utility representation only if it is a weak order, i.e. it is complete and transitive.*

Proof. The proof relies on the fact that utility translate the preference $\succsim$ to an inequality between real numbers. Thus, both transitivity and completeness follow from the completeness and transitivity of $\geq$ on $\mathbb{R}$.[7] $\square$

Proposition 2.6 (Second necessary condition). *A preference relation has a utility representation only if $X/\sim$ has at most the power of the continuum.*[8]

Definition 2.7 ($\succsim$-order density). $Z \subseteq X$ is said to be $\succsim$-order dense in X if, for each $x, y \in X$ with $x \succ y$, there exists $z \in Z$ such that $x \succsim z \succsim y$. For instance, $\mathbb{Q}$ is $\geq$-dense in $\mathbb{R}$.

Theorem 2.8 (Cantor–Debreu). *If $\succsim$ is complete and transitive on X, then it has a utility representation if and only if there exists an at most countable $\succsim$-order dense subset Z in X.*

Corollary 2.9. *If $X/\sim$ is at most countable, then $\succsim$ has a utility representation if and only if it is a weak order, i.e. it is complete and transitive.*

Proof.
If $\succsim$ has a utility representation, then $\succsim$ is a weak order. We need to prove that, adding the hypothesis $|X/\sim| \leq |\mathbb{N}|$, the converse is also true.
A first proof comes from Cantor Theorem: If $X/\sim$ is at most countable, then the set Z formed by the representative x of the indifference curves $[x]$ is countable and, trivially, $\succsim$-order dense in X.
A second proof is independent from it, but we discuss only the case where X is finite. The proof relies on defining $u(x)$ as the number of alternatives strictly dispreferred to x. Specifically, $u(x) := |\{z \in X : z \prec x\}|$. Now

$$\begin{aligned} y \precsim x &\iff \{z \in X : z \prec y\} \subseteq \{z \in X : z \prec x\} \\ &\iff |\{z \in X : z \prec y\}| \leq |\{z \in X : z \prec x\}| \iff u(y) \leq u(x). \end{aligned}$$

$\square$

Corollary 2.10. *If X is at most countable, $\succsim$ has a utility representation if and only if it is a weak order, i.e. it is complete and transitive.*[9]

Lemma 2.11. *For each $x \in X \subseteq \mathbb{R}^n_+$ there exists a unique constant $k_x \geq 0$ such that $x \sim k_x e$, where e denotes the unit vector $e = (1, \dots, 1)$.*

Theorem 2.12 (Necessary and sufficient condition in $\mathbb{R}^n_+$). *A preference relation on $\mathbb{R}^n_+$ has a strongly monotone and continuous utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ if and only if it is transitive, complete, strongly monotone and Archimedean.*

Proof.
Here, we prove that an strongly monotone, Archimedean weak order admits a strongly monotone and continuous utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$. As for the converse, refer to the appendix.
Defining $u(x) := k_x$ for every $x \in X$, we employ the previous lemma, obtaining the chain of equivalences

$$x \succsim y \quad \overset{1}{\iff} \quad k_x e \succsim k_y e \quad \overset{2}{\iff} \quad k_x \geq k_y \quad \overset{3}{\iff} \quad u(x) \geq u(y)$$

where 1 is the previous lemma, 2 is the strong monotonicity of $\succsim$ and 3 is the definition of u. Thus, $u(x)$ is a proper utility function. We won't prove its continuity. As for strong monotonicity of u, it directly follows from the strong monotonicity of $\succsim$: if $x \gg y$, then, by strong monotonicity of $\succsim$, $x \succ y$, implying $u(x) > u(y)$; and if $x \geq y$, then, by strong monotonicity of $\succsim$, $x \succsim y$, implying $u(x) \geq u(y)$. $\square$

Theorem 2.13 (Quasi-concavity). *In the same setting of the previous lemma, $\succsim$ is (strictly) convex if and only if u is (strictly) quasi-concave.*[10]

[7] If $x \succsim y$ and $y \succsim z$, then $u(x) \geq u(y)$ and $u(y) \geq u(z)$. By transitivity of $\geq$ on $\mathbb{R}$, $u(x) \geq u(z)$, deducing $x \succsim z$. Similarly, either $u(x) \geq u(y)$ or $u(y) \geq u(x)$ implying completeness of $\succsim$.

[8] $X/\sim$ is the quotient vector space representing the collection of the indifference curves. The power of the continuum is the cardinality of $\mathbb{R}$. Intuitively, if $X/\sim$ had a cardinality higher than $\mathbb{R}$, we wouldn't be able to associate a real number $u(x)$ to $[x]$.

[9] Since X is at most countable, then the collection of indifference curves $X/\sim$ is at most countable, and the previous corollary yields the conclusion.

[10] For the definition of quasi-concavity and some basic results, refer to the section "Preliminary Results".

Proof. Suppose u is quasi-concave, $x \sim y$ and $z = \alpha x + (1-\alpha)y$ with $\alpha \in [0,1]$. Then, by quasi-concavity,

$$u(z) = u\big(\alpha x + (1-\alpha)y\big) \geq \alpha u(x) + (1-\alpha)u(y) = \alpha u(x) + (1-\alpha)u(x) = u(x) \implies z \succsim x.$$

The strictly quasi-concave case is analogous. The converse is omitted. □

Example 2.14 (Lexicographic Preference). A classic example of a preference without a utility representation. In $\mathbb{R}^2$, it is defined as

$$x \succsim y \iff x_1 > y_1 \quad \text{or} \quad (x_1 = y_1 \text{ and } x_2 \geq y_2).$$

i) This preference is complete, transitive, antisymmetric and strongly monotone (the proof is an easy exercise);

ii) The indifference curves are singletons due to antisymmetry, thus they are uncountably many;

iii) This preference does not satisfy the Archimedean axiom (take $x = (1,0), y = (0,1)$ and $z = (0,0)$). Thus, the earlier existence theorem does not apply;

iv) It has no utility representation. Suppose $u : \mathbb{R}^2 \to \mathbb{R}$ exists. If $x < y$ and $a < b$, then $u(x,a) < u(x,b) < u(y,a) < u(y,b)$. By density of $\mathbb{Q}$ in $\mathbb{R}$, we can find q_x and q_y such that

$$u(x,a) < q_x < u(x,b) < u(y,a) < q_y < u(y,b).$$

Therefore, $x \neq y$ implies $q_x \neq q_y$, meaning that $q : \mathbb{R} \to \mathbb{Q}$ is an injective map from $\mathbb{R}$ to $\mathbb{Q}$, a contradiction, since $\mathbb{Q}$ is countable and $\mathbb{R}$ is uncountable.

3 Rational Choice

In this section, we formalize the decision-making setting in which the decision maker (DM) selects among various alternatives. We will begin by defining foundational concepts such as **decision framework**, **decision environment** and **decision problem**. Next, we will adapt these definitions according to the chosen perspective, modifying them in four distinct ways:

i) **Universal analysis**: the broadest model assumes that the preferences depend on the choice set we are considering, highlighting the significance of **context effects**. In universal analysis, it is assumed that a universal preference exists, valid across all choice sets.

ii) **Contextualized analysis**: this model reintroduces context effects but establishes a new preference for comparing alternatives across different contexts, referred to as contextualized alternatives.

iii) **Parametric analysis**: in this model, choice sets are parameterized through a menu correspondence, necessitating slight modifications to every definition accordingly.

iv) **Utility analysis**: this approach assumes the existence of a utility representation $u_X : X \to \mathbb{R}$ for $\succsim_X$, called **decision criterion**.

3.1 Optimal choice

Definition 3.1 (Decision framework). The pair $(\mathbf{X}, \mathcal{X})$ made of a **choice space** $\mathbf{X}$ of conceivable alternatives and a collection $\mathcal{X}$ of non-empty subsets X of $\mathbf{X}$, called **choice sets**.

Definition 3.2 (Preference map). $P : X \mapsto \succsim_X$ associates, to each $X \in \mathcal{X}$, a preorder $\succsim_X$ on X. This map is well defined thanks to extensionality.

Definition 3.3 (Contextualized alternative). The pair (x, X) with $x \in X$; let $\mathcal{C}$ be the set of all contextualized alternatives. Changing the choice set X may change the preference $\succsim_X$, modelling **context effects.**[11]

Definition 3.4 (Decision environment). The triple $(\mathbf{X}, \mathcal{X}, P)$ made of a decision framework endowed with a preference map P.

Definition 3.5 (Decision problem). For the DM (or choice problem) the pair $(X, \succsim_X)$. The aim is to find the alternatives in X that are optimal according to $\succsim_X$.

[11] Some marketing techniques may be able to reverse a consumer's ranking between a basic option x and a richer one x', for instance, by adding a top option x''.

Definition 3.6 (Optimal alternative). Given a decision problem, $\hat{x} \in X$ is optimal if there is no $x \in X$ such that $x \succ_X \hat{x}$. Equivalently, $\hat{x}$ is optimal iff $\hat{x} \succsim_X x$ for all alternatives in X comparable with $\hat{x}$.

Proposition 3.7 (Ascent Algorithm). *In a decision problem $(X, \succsim_X)$, if X is finite, optimal alternatives exist.*[12]

Proof.

We prove this by induction on the cardinality $N := |X|$ of X. If $N = 1$, the property clearly holds. Supposing there exists an optimal for any set of $N \leq k$ elements, we prove that this implies there exists an optimal for a set of $k+1$ elements, concluding the proof. Take a set of $k+1$ elements $x_1, \ldots, x_{k+1}$. Let $\hat{x}$ be the optimal of the first k elements. If it doesn't happen that $x_{k+1} \succ \hat{x}$, then $\hat{x}$ is optimal and we are done. If $x_{k+1} \succ \hat{x}$, we know that there is no $i \in \{1, \ldots, k\}$ such that $x_i \succ \hat{x}$, and, by transitivity, there's no i such that $x_i \succ x_{k+1}$. Thus x_{k+1} is an optimal alternative.

$\square$

3.2 Correspondences

In this section, we introduce two primary multivalued functions: the **rational correspondence**, which maps each choice set to its set of optimal alternatives, and the **menu correspondence**, which parameterizes choice sets based on a parameter θ.

Definition 3.8 (Correspondence). A correspondence is a multivalued function $\varphi : D \subseteq X \rightrightarrows Y$ which associates a non-empty subset of Y to each element of D. If $\varphi(x)$ is a singleton for every $x \in D$, then φ reduces to a function.[13] A main instance is the budget correspondence $B : \mathbb{R}^{n+1}_+ \rightrightarrows \mathbb{R}^n_+$ given by $B(p, w) = \{x \in \mathbb{R}^n_+ : p \cdot x \leq w\}$.

Definition 3.9 (Rational correspondence). The rational choice correspondence $\sigma : \mathcal{D} \rightrightarrows \mathbf{X}$ is the subset of $\mathbf{X}$ that associates to each $X \in \mathcal{D}$ the set of optimal alternatives of X:

$$\sigma(X) = \{\hat{x} \in X : \nexists x \in X,\ x \succ_X \hat{x}\},$$

where $\mathcal{D} \subseteq \mathcal{X}$ is the collection of choice sets in which optimal alternatives exist (*a correspondence cannot map to an empty set*). It follows that:

i) If $x \in \sigma(X)$ and x and y are comparable, then $x \succsim_X y$.

ii) If $\succsim$ is a weak order, then $\sigma(X) = \{\hat{x} \in X : \hat{x} \succsim x,\ \forall x \in X\}$.

iii) If $\succsim_X$ is a preorder, $\sigma(X)$ is made of alternatives pairwise either incomparable or indifferent, so $\sigma(X)$ can be partitioned into indifference curves $[\hat{x}]_X$.

iv) If $\succsim_X$ is a weak order, then optimal alternatives are indifferent.

Definition 3.10 (Value function and rational correspondence). If there exists a utility representation u_X of $\succsim_X$, we can define the rational correspondence and the value function $v : \mathcal{D} \to \mathbb{R}$ as

$$\sigma(X) = \arg\max_{x \in X}\ u(x) \qquad \text{and} \qquad v(X) = \max_{x \in X}\ u(x),$$

where $\arg\max_{x \in X} u(x)$ is the set of maximizers of u in X.

Definition 3.11 (Menu correspondence). The menu correspondence $\varphi : \Theta \rightrightarrows \mathbf{X}$ parametrises the choice sets, associating to each parameter θ a set $\varphi(\theta)$.

Example 3.12. The prototypical example of menu correspondence is the budget correspondence used in the consumer problem, where $\mathcal{X}$ is made of budget sets $B(p, w) = \{x \in \mathbb{R}^n_+ : p \cdot x \leq w\}$, depending on two pairs $p \in \mathbb{R}^n_+$ and $w \in R$, prices and wealth respectively, forming the parametric set $\mathbb{R}^n_+ \times \mathbb{R}_+ = \Theta$. Therefore,

$$\mathcal{X} = \{B(p, w) : (p, w) \in \mathbb{R}^n_+ \times \mathbb{R}_+\} \quad \iff \quad \mathcal{X} = \{\varphi(\theta) : \theta \in \Theta\}.$$

Note how the dependence of $X = \varphi(\theta)$ on θ allows us to directly relate any function of X as a function of θ:

- The preference map $P(X) = \succsim_X$ becomes a function of θ, as $P(\theta) = \succsim_\theta$.[14]
- The contextualized alternative $(x, X) = (x, \varphi(\theta))$ becomes (x, θ).

[12] It is clear that the converse is false. For example, if $X = \mathbb{R}_+$ and $\succsim_X$ is strictly monotone, there is no optimal. This fact formalizes the idea that there is no optimal amount of money: more is always better.

[13] Depending on whether the image of φ consists of closed, compact or convex subsets, we say that φ is closed-valued, compact-valued or convex-valued.

[14] Preferences do not change across equivalent parametrizations, since it may happen that, if φ is not injective, then the same set X may be parametrized in different ways.

3.3 Universal Analysis

We assume there exists a universal preference $\succsim$ on the choice space $\mathbf{X}$ such that, for all $X \in \mathcal{X}$, it holds

$$a \succsim b \iff a \succsim_X b \quad \forall a, b \in X.$$

The decision problems now take the form $(X, \succsim)$ for each $X \in \mathcal{X}$, and the decision environment is $(\mathbf{X}, \mathcal{X}, \succsim)$ rather than $(\mathbf{X}, \mathcal{X}, P)$. With this structure, we can more deeply explore the properties of the rational correspondence, which is now based on $\succsim$, and further develop the concept of a **menu preference**, that establishes a preference over choice sets themselves.
The following propositions list the main properties of the rational correspondence:

Proposition 3.13. *If $\succsim$ is a preorder, then for each $X, Y \in \mathcal{D}$:*

i) $X \subseteq Y$ implies $X \setminus \sigma(X) \subseteq Y \setminus \sigma(Y)$;

ii) $X \subseteq Y$ implies $\sigma(Y) \cap X \subseteq \sigma(X)$.

Proof. If $x \in X \setminus \sigma(X)$, there exists $x' \in X$ such that $x' \succ x$. Because $X \subseteq Y$, we have $x', x \in Y$, hence $x \in Y \setminus \sigma(Y)$. For (ii), if $x \in \sigma(Y) \cap X$, then $x \in X$ and no $x' \in Y$ satisfies $x' \succ x$; in particular, no $x' \in X$ does, so $x \in \sigma(X)$. □

Proposition 3.14. *If $\succsim$ is a weak order, then for each $X, Y \in \mathcal{D}$:*

*i) Independence of irrelevant alternatives **(IIR)**: $X \subseteq Y$ and $\sigma(Y) \cap X \neq \emptyset$ imply $\sigma(Y) \cap X = \sigma(X)$.*[15]

*ii) Weak axiom of revealed preference **(WARP)**: If $x \in \sigma(X)$ and $y \in X \setminus \sigma(X)$, then there is no $Y \in \mathcal{D}$ such that $x \in Y$ and $y \in \sigma(Y)$.*[16]

Proof. (i). $\sigma(Y) \cap X \subseteq \sigma(X)$ holds by the previous theorem. We only need to prove that $\sigma(X) \subseteq \sigma(Y) \cap X$. Let $x \in \sigma(X)$ and $z \in \sigma(Y) \cap X \subseteq \sigma(X)$. By completeness, $x \succsim z$, and, by transitivity, $x \succsim y$ for every $y \in Y$. In particular, $x \in \sigma(Y) \cap X$.
(ii). Let $x \in \sigma(X)$ and $y \in X \setminus \sigma(X)$. Then $x \succ y$ by completeness. If $y \in \sigma(Y)$ and $x \in Y$, then $y \succsim x$, contradicting $\succ y$. □

Definition 3.15 (Choice function). If $\sigma : \mathcal{D} \rightrightarrows \mathbf{X}$ is single-valued, it is a choice function $\sigma : \mathcal{D} \to \mathbf{X}$. Such a function satisfies **WARP** iff, for all $X, Y \in \mathcal{D}$,

$$\sigma(Y) \in X \text{ and } \sigma(X) \neq \sigma(Y) \implies \sigma(X) \notin Y.$$

Definition 3.16 (Menu preference). The menu preference $\succeq$ over the choice sets in $\mathcal{D}$, induced by $\succsim$, is defined by

$$X \succeq Y \iff \forall y \in Y, \exists x \in X, \ x \succsim y, \qquad X, Y \in \mathcal{D}.$$

The following propositions list the main properties of the menu preference:

Proposition 3.17. *If $\succsim$ is a preorder, then for each $X, Y \in \mathcal{D}$:*

i) $Y \subseteq X$ implies $X \succeq Y$: more options are always welcomed. This directly follows from reflexivity.

ii) $X \succeq Y$ implies $X \sim X \cup Y$.[17]

Proposition 3.18. *If $\succsim$ is a weak order, then:*

i) $\succeq$ is a weak order;

ii) For each $X \in \mathcal{D}$, $X \sim \sigma(X)$;

iii) $\{\hat{x}\} \sim X$ for each optimal alternative $\hat{x} \in X$. In particular, $\sigma(X) = \{\hat{x} \in X : \hat{x} \sim X\}$.

Proof.
(i) We first prove transitivity: let $X \succeq Y$ and $Y \succeq Z$. Since $Y \succeq Z$, for every $z \in Z$, there exists

[15] This asserts that enlarging a choice set by adding alternatives that are not superior does not affect the optimal choice within the subset.
[16] This means an optimal choice in one context cannot become suboptimal in another when both choices are present.
[17] From the previous result, $X \preceq X \cup Y$. If $z \in X \cup Y$, then either $z \in X$ or $z \in Y$. In both cases, you can easily conclude that $X \succeq X \cup Y$.

$y_z \in Y$ such that $y_z \succsim z$. Similarly you find $x_{y_z} \in X$ such that $x_{y_z} \succsim y_z$ and then you conclude $x_{y_z} \succsim z$ using the transitivity of $\succsim$. As to completeness, let $X, Y \in \mathcal{D}$ and let $\hat{x} \in \sigma(X)$, $\hat{y} \in \sigma(Y)$. By the completeness of $\succsim$, we have $\hat{x} \succsim \hat{y}$ or $\hat{y} \succsim \hat{x}$. By the transitivity of $\succsim$, this implies either $X \succeq Y$ or $Y \succeq X$.
(ii) We already know $\sigma(X) \preceq X$ since $\sigma(X) \subseteq X$. By completeness of $\succsim$, it holds $\hat{x} \succsim x$ for all $x \in X$ and $\hat{x} \in \sigma(X)$, implying $\sigma(X) \succeq X$. Point 3 is an obvious corollary of this property.
(iv) Since $X \sim \sigma(X)$, $Y \sim \sigma(Y)$ and $\succeq$ is a weak order, the conclusion holds. $\square$

3.4 Variations

Preference: we have the following:

- **Universal preference**: $a \succsim b \iff a \succsim_X b$.
- **Contextualized preference**: $(x, X) \succsim (x', X) \iff x \succsim_X x'$ for every $x, x' \in X$. Of course, we can also rank alternatives across context, like $(x, X) \succsim (y, Y)$ or $(x, X) \succsim (x, Y)$, if $x \in X$ and $x \in Y$.
- **Parametric contextualised preference**: $(x, \theta) \succsim (x', \theta) \iff x \succsim_\theta x'$; in general we may compare (x, θ) with (x', θ').

Optimal alternatives: in contextualised analysis, $\hat{x} \in X$ is optimal whenever $(\hat{x}, X) \succsim (x, X)$ for every $x \in X$ (if $\succsim$ is a weak order).

Menu preference: we have the following:

- **Contextualised analysis**: $X \succeq Y$ if, for every $y \in Y$, there exists $x \in X$ such that $(x, X) \succsim (y, Y)$.
- **Parametric analysis**: $\theta \succeq_\Theta \theta'$ whenever, for every $y \in \varphi(\theta')$, there exists $x \in \varphi(\theta)$ with $(x, \theta) \succsim (y, \theta')$.

3.5 Utility Analysis

We now assume that for every choice set X, there exists a utility representation $u_X : X \to \mathbb{R}$, representing $\succsim_X$, which we call the **Decision Criterion**, and then we will adjust this definition to align with various assumptions.
Utility analysis: the decision problem $(X, \succsim_X)$ takes the form (X, u_X); the optimal alternatives are the solutions of the optimisation problem

$$\max_x u_X(x) \quad \text{sub } x \in X.$$

In this case,

$$\sigma(X) = \{\hat{x} \in X : \forall x \in X,\ u_X(\hat{x}) \geq u_X(x)\}.^{18}$$

Universal analysis: if $\succsim$ is a universal weak-order preference and $u : \mathbf{X} \to \mathbb{R}$ its utility representation, the decision environment and the decision problem take the forms $(\mathbf{X}, \mathcal{X}, u)$ and (X, u), and:

- the equivalent optimisation problem becomes

$$\max_x u(x) \quad \text{sub } x \in X;$$

- in this case,

$$\sigma(X) = \{\hat{x} \in X : \forall x \in X,\ u(\hat{x}) \geq u(x)\} = \arg\max_{x \in X} u(x);$$

- the value function is the utility function for the menu preference. In fact, for every $X, Y \in \mathcal{D}$,

$$X \succeq Y \iff \max_{x \in X} u(x) \geq \max_{y \in Y} u(y) \iff v(X) \geq v(Y).$$

Parametric analysis if $\succsim$ is a contextualised preference and $u : \operatorname{Gr}\varphi \to \mathbb{R}$ its utility representation, then:

- $\operatorname{Gr}\varphi$ is the graph of φ, i.e. the set of pairs (x, θ) with $x \in \varphi(\theta)$;

[18] Note that here we are assuming context independence, since u does not depend on θ. A main instance of this context-independent problem is the **consumer problem**.

- u must satisfy
$$u(x,\theta) \geq u(x',\theta') \iff (x,\theta) \succsim (x',\theta')$$
for each $(x,\theta),(x',\theta') \in \operatorname{Gr}\varphi$ and, for each $\theta \in \Theta$,
$$u(x,\theta) \geq u(x',\theta) \iff x \succsim_\theta x', \quad \forall x,x' \in \varphi(\theta);$$

- the optimisation problem becomes
$$\max_x u(x,\theta) \quad \text{sub } x \in X_\theta = \varphi(\theta);$$

- we have
$$\sigma(\theta) = \{\hat{x} \in \varphi(\theta) : \forall x \in \varphi(\theta),\, u(\hat{x},\theta) \geq u(x,\theta)\} = \arg\max_{x\in\varphi(\theta)} u(x,\theta), \quad v(\theta) = \max_{x\in\varphi(\theta)} u(x,\theta).$$

3.6 Convexity

In this section, we explore the link between the convexity of the preference relation $\succsim$ and the convexity and uniqueness of solutions in a decision problem. It is a crucial result which will be employed several times throughout these notes.

Theorem 3.19. *If $\succsim_X$ is convex, then $\sigma(X)$ is convex. If $\succsim_X$ is strictly convex, then $\sigma(X)$ is a singleton.*[19]

Proof.
If $\succsim_X$ is convex, $a,b \in \sigma(X)$ and $\lambda \in [0,1]$ imply $a \sim b$; by convexity, $\lambda a + (1-\lambda)b \succsim_X a$, hence $\lambda a + (1-\lambda)b \in \sigma(X)$. If $\succsim_X$ is strictly convex and $a,b \in \sigma(X)$ with $a \neq b$, then $a \sim b$ and, by strict convexity, $c = \frac{1}{2}a + \frac{1}{2}b \succ a$, contradicting the optimality of a.

□

Theorem 3.20. *Let $\succsim$ be a weak order.*

i) If $\succsim$ is convex, then $\sigma : \mathcal{D} \rightrightarrows \mathbf{X}$ is convex-valued. In particular, optimal alternatives are either unique or uncountably many.

ii) If $\succsim$ is strictly convex, then σ is single-valued.

Theorem 3.21. *Let $\succsim_X$ be transitive, complete, strongly monotone and Archimedean.*

i) If u_X is quasi-concave, then $\sigma(X)$ is convex (recall that u_X is quasi-concave iff $\succsim_X$ is convex).

ii) If u_X is strictly quasi-concave, then $\sigma(X)$ is a singleton.

4 Decisions Under Certainty

In this section, we address decision problems where the *material consequences* of an action are well-defined and certain, such as bundles of goods or monetary amounts. The framework is the quartet
$$(\mathbf{A}, \mathcal{A}, \mathbf{C}, \rho),$$
where $\mathbf{A}$ is the space of actions, $\mathbf{C}$ is the space of consequences, $\mathcal{A}$ is a collection of action sets of $\mathbf{A}$ and $\rho : \mathbf{A} \to \mathbf{C}$ is a consequence function, assumed surjective, mapping each action to its outcome.

4.1 Outcome Consequentialism

Our first assumption is *outcome consequentialism*: if two actions lead to the same consequence, then the two actions are indifferent to a DM:
$$\rho(a) = \rho(b) \implies a \sim b.$$
With this perspective, if we define $\dot{\succsim}$ as a preference over consequences, then
$$a \succsim b \iff \rho(a) \,\dot{\succsim}\, \rho(b),$$

[19] The converse is false, since the lexicographic preference has singleton indifference curves, but it is not strictly convex.

Table 2: Decision environments and decision problems

Rational choice	Choice under certainty EF	Choice under certainty RF
$(\mathbf{X}, \mathcal{X}, u)$	$(\mathbf{A}, \mathcal{A}, \dot{u} \circ \rho)$	$(\mathbf{C}, \mathcal{C}, \dot{u})$
(X, u)	$(A, \dot{u} \circ \rho)$	$(C, \dot{u})$

meaning action a is preferred to action b if and only if the consequence of a is preferred over that of b. This reflects that while actions are evaluated, what really matters to the DM are the material consequences that they deliver.
In this context, the decision environment is the sextet $(\mathbf{A}, \mathcal{A}, \mathbf{C}, \rho, \dot{\succsim}, \succsim)$.

Choice under certainty, extended form: Assume $\dot{\succsim}$ can be represented by a utility function $\dot{u} : \mathbf{C} \to \mathbb{R}$. In terms of utility functions we have

$$a \succsim b \iff \rho(a) \mathrel{\dot{\succsim}} \rho(b) \iff \dot{u}(\rho(a)) \geq \dot{u}(\rho(b)),$$

so that $u = \dot{u} \circ \rho$ is a proper utility function $u : \mathbf{A} \to \mathbb{R}$ for $\succsim$. Now the DM confronts the decision problem under certainty $(A, \dot{u} \circ \rho)$, equivalent to the optimization problem under certainty

$$\max_a \ \dot{u}(\rho(a)) \quad \text{sub } a \in A.$$

The rational choice correspondence and the value function are given by

$$\sigma(A) = \arg\max_{a \in A} \ \dot{u}(\rho(a)) \quad \text{and} \quad v(A) = \max_{a \in A} \dot{u}(\rho(a))$$

Choice under certainty, reduced form: If $\succsim$ and $\dot{\succsim}$ are both preorders or both weak orders, and $\succsim$ satisfies the outcome consequentialist relation, then we can directly link actions to their consequences:

$$c \mathrel{\dot{\succsim}} d \iff a \succsim b, \quad \forall a \in \rho^{-1}(c), b \in \rho^{-1}(d),$$

meaning that the consequence c is preferred to d if and only if any action delivering c is preferred to any action delivering d. There is a perfect symmetry between the action-based decision environment and a consequence-based decision environment, shown in table 2, with $\mathcal{C} = \{\rho(A) : A \in \mathcal{A}\}$.
The equivalent optimization problem is

$$\max_c \dot{u}(c) \quad \text{sub } c \in C.$$

4.2 Example: Firm's Problem

Consider a firm that must decide its production level to maximize profit, denoted as $\rho(a) = r(a) - c(a) \in [0, +\infty)$, where r and c represent the revenue and cost functions, respectively.
There are three important market forms:

i) **Perfect competitive output market:** the firm is a price-taker, with revenue $r(a) = pa$.

ii) **Monopolistic output market:** as the sole producer, revenue is $r(a) = aD(a)$, with $D : [0, +\infty) \to \mathbb{R}$ indicating the price chosen by the demand.

iii) **Competitive inputs market:** To produce a, the firm may need to buy a vector of inputs $x = (x_1, \ldots, x_n)$ at unit prices $w = (w_1, \ldots, w_n)$. The cost for x is $w_1 x_1 + \cdots + w_n x_n = x \cdot w$.

The **production function** $f : \mathbb{R}^n_+ \to [0, +\infty)$ transforms input x with cost $w \cdot x$ into output $f(x)$. For a specific output a, we define the **isoquant** as the counterimage of a through f, i.e. $f^{-1}(a) = \{x \in \mathbb{R}^n_+ : f(x) = a\}$, the set of x which can output a through f. Thus the cost function is

$$c(a) = \min_{x \in f^{-1}(a)} w \cdot x.$$

Assuming w is given and p varies as a parameter, the firm's profit function transforms to:

$$\rho_p(a) = pa - c(a) = pa - \min_{x \in f^{-1}(a)} w \cdot x.$$

The firm's **production decision problem** is to maximize this profit (we are assuming that the utility function of money is simply $\dot{u}(x) = x$):

$$\max_{\rho(a)} \dot{u}(\rho(a)) \quad \text{sub } a \in A \iff \max_{a} r(a) - c(a) \quad \text{sub } a \in A,$$

where $A(k) = [0, k]$ denotes the firm's production possibilities, with k representing full capacity.
Assuming the continuity and strict concavity of ρ, and considering the firm's production constraints and the parameter $\theta = (k, p)$, there exists a unique optimal production level $\sigma(k, p)$ and corresponding profit $v(k, p)$. This guarantees a singular, optimal decision for each set of market conditions.

5 Consumer Problem

5.1 Description of the problem

Here, the DM is a consumer who has to choose between bundles of goods. These are the assumptions and the basic objects of the model:

i) It's a decision problem under certainty in reduced form, assuming outcome consequentialism, since the DM chooses shopping actions which deliver bundles but identifies them with the delivered bundle itself.

ii) Goods are infinitely divisible and a consumer can buy any quantity of them. Thus, the consequence space $\mathbf{C}$ is the **consumption set** $\mathbb{R}^n_+$.

iii) The market is competitive, so that prices are fixed *a priori*.

iv) The choice sets are the **budget sets**, parametrized by

$$B(p, w) = \{c \in \mathbb{R}^n_+ : p \cdot c \leq w\}.$$

Note that such set is closed and convex.[20] This implies that the budget correspondence $B : \mathbb{R}^{n+1}_+ \rightrightarrows \mathbb{R}^n_+$, which associates a budget set to any pair (p, w), is closed-valued and convex-valued. The budget correspondence plays the role of the menu correspondence.

v) We typically assume $\succsim$ being represented by a utility function, but that's not always the case.[21]

With such assumption, the consumer decision environment becomes

$$(\mathbb{R}^n_+, \boldsymbol{B}, \succsim) \implies (\mathbb{R}^n_+, \boldsymbol{B}, u),$$

where $\succsim$ is a weak order on $\mathbf{C}$ and $\boldsymbol{B}$ is the collection of budget sets, and the consumer problem becomes

$$(B(p, w), \succsim) \implies (B(p, w, u)) \implies \max_{c} u(c) \quad \text{sub } c \in B(p, w)$$

5.2 Optimal bundles

Here we study the set D of pairs (p, w) for which there exists an optimal solution to the consumer problem.

Proposition 5.1. *Budget sets are compact if and only if there are no free goods.*

Proof. We know that budget sets are closed and convex. To establish compactness, we need to show they are bounded under the condition that there are no free goods, i.e., when $p \gg \mathbf{0}$. If $p \gg \mathbf{0}$, then for any $x \in \mathbb{R}^n_+$ that belongs to the budget set, we have $w \geq p \cdot x \geq p_i x_i$ for each i. This implies that each component x_i of the vector x must satisfy $x_i \leq \frac{w}{p_i}$, establishing a finite upper bound $\frac{w+1}{p_i}$ for each x_i. Thus, the set is bounded in every dimension. Conversely, if there exists at least one free good, say $p_j = 0$, the consumer can afford an unbounded quantity of this good without affecting their ability to buy other goods. This means the budget set is unbounded in the direction of x_j, and therefore, not compact. $\square$

Proposition 5.2. *If u is upper semicontinuous, then an optimal solution exists for every $p \gg \mathbf{0}$ and $w \geq 0$. If u is also strictly monotone, then there exists an optimal solution only when there are no free goods.*

[20] It's convex since if $p \cdot a \leq w$ and $p \cdot b \leq w$, then $p \cdot (\alpha a + (1 - \alpha)b) = \alpha p \cdot a + (1 - \alpha)p \cdot b \leq \alpha w + (1 - \alpha)w = w$ and it's closed since limits preserve non-strict inequalities, so that if $p \cdot c_n \leq w$, then $p \cdot c \leq w$, where $c_n \to c$.

[21] The lexicographic preference has no utility function, and still is a valid consumer problem. The unique optimal bundle is $\hat{c} = (w, p_1, 0)$, when $p \gg \mathbf{0}$, since you can easily show that any bundle with $c_2 \neq 0$ is not optimal. Using the strong monotonicity and the completeness of $\succsim$, we conclude that $(w/p_1, 0)$ is the unique optimal bundle.

Proof. With $p \gg \mathbf{0}$, budget sets are compact, ensuring the existence of optimal solutions because u is upper semicontinuous, a generalization of Weierstrass' Theorem (which doesn't appear on the syllabus). For strict monotonicity, consider u strictly monotone and an optimal solution $\hat{c}$ where $p_i = 0$ for some i, making good e_i free. Purchasing $\hat{c} + e_i$ stays within budget ($p \cdot (\hat{c} + e_i) = p \cdot \hat{c} \leq w$), yet $u(\hat{c} + e_i) > u(\hat{c})$ contradicts the optimality of $\hat{c}$, proving no free goods exist when u is strictly monotone. □
The next theorem provides a necessary condition for optimal bundles:

Theorem 5.3. *If u is differentiable and $\hat{c} \gg \mathbf{0}$ is an optimal bundle, then*[22]

$$\frac{\frac{\partial u(\hat{c})}{\partial c_k}}{p_k} = \frac{\frac{\partial u(\hat{c})}{\partial c_j}}{p_j} \implies MRS_{k,j}(\hat{c}) = \frac{\frac{\partial u(\hat{c})}{\partial c_k}}{\frac{\partial u(\hat{c})}{\partial c_j}} = \frac{p_k}{p_j}.$$

5.3 Demand Correspondence

Here we present the basic properties of demand correspondences.

Definition 5.4 (Demand correspondence (DC)). The demand correspondence associates to the pair (p, w) the set of optimal solutions. If the optimal is unique, the demand correspondence is a (single–valued) demand function:

$$d(p, w) = \arg \max_{c \in B(p,w)} u(c).$$

Remark 5.5 (Positive homogeneity of degree 0). It holds $d(p, w) = d(\alpha p, \alpha w)$ for all $\alpha > 0$. This directly stems from the equivalence $B(p, w) = B(\alpha p, \alpha w)$ for $\alpha > 0$.

Remark 5.6. the economic interpretation is straightforward: what matters are *relative* prices p_i/p_j and real income. It allows one to normalise the price of a non-free good, termed the **numeraire**, to 1, becoming the **real unit of account**: relative prices and **real income** are scaled by $1/p_j$ without changing the budget set. Usually, however, we employ a **nominal unit of account**, called **fiat money**, which is used to express **nominal prices** p and **nominal incomes** w.

Theorem 5.7 (Walras' Law). *If u is strongly monotone, then the demand correspondence satisfies Walras' Law, i.e. $p \cdot \hat{c} = w$ for every $\hat{c} \in d(p, w)$.*

Proof.
If $w = 0$ there is nothing to prove. Otherwise, suppose by contradiction $p \cdot \hat{c} < w$. There exists $\varepsilon > 0$ such that

$$\sum_{i=1}^{n} p_i(\hat{c}_i + \varepsilon) = \sum_{i=1}^{n} p_i \hat{c}_i + \varepsilon \sum_{i=1}^{n} p_i \leq w.$$

Setting $z = \hat{c} + \varepsilon(1, \ldots, 1)$, we see that $z \gg \hat{c}$; thus, by strong monotonicity, $u(z) > u(\hat{c})$, a contradiction. □

Remark 5.8. This implies that the optimal solutions must lie on the budget line

$$\Delta(p, w) = \{c \in \mathbb{R}^n_+ : p \cdot c = w\} \subseteq \partial B(p, w),$$

the boundary of $B(p, w)$.

Theorem 5.9. *If u is quasi-concave, then $d(p, w)$ is convex-valued. If u is strictly quasi-concave, $d(p, w)$ is a (single-valued) demand function.*

Proof. The proof relies on a property proven in Chapter "Utility Theory": u is quasi-concave if and only if $\succsim$ is convex. In such case, the optimal–choice correspondence σ is convex-valued, hence so is d. If u is strictly quasi-concave, then $\succsim$ is strictly convex, σ is single-valued, and therefore d is also single-valued, i.e. a function. □

5.4 Indirect Utility

Here we present the basic properties of the indirect utility. The indirect utility function represents the utility level attained at the optimal consumption bundles in a given budget set:

$$v(p, w) = \max_{c \in B(p,w)} u(c).$$

[22] The first is a cardinal notion which relates relative prices to marginal utilities. The second is an ordinal notion, since marginal rate of substitution is invariant under strictly monotone transformations, and thus it does not depend on which utility representation we choose.

Indirect utility functions are:

i) **Positively homogeneous of degree zero**; this directly follows from $B(p, w) = B(\alpha p, \alpha w)$ for $\alpha > 0$.

ii) **Monotone in wealth and in prices**; it is increasing in w, since

$$w \geq w' \implies B(p, w') \subseteq B(p, w) \implies v(p, w') \leq v(p, w),$$

and decreasing in p, since

$$p \geq p' \implies B(p, w) \subseteq B(p', w) \implies v(p, w) \leq v(p', w).$$

iii) **Strictly monotone in wealth and prices if u is strongly monotone**; it is strictly increasing in w and, when $w > 0$, strictly decreasing in p.

iv) **Quasi-convex**; in particular, it is quasi-affine in wealth.

v) **(Strictly) concave in wealth if u is (strictly) concave**.

Proof.
We focus on property (v), proving it for concave u (analogous arguments apply in the strict case). Let $(p, w), (p, w') \in D$ and $\alpha \in [0, 1]$. We aim to prove that

$$v(p, \alpha w + (1 - \alpha)w') \geq \alpha v(p, w) + (1 - \alpha)v(p, w'). \tag{5.1}$$

Let $\hat{c} \in d(p, w)$ and $\hat{c}' \in d(p, w')$. Then

$$(\alpha\hat{c} + (1 - \alpha)\hat{c}') \cdot p = \alpha(\hat{c} \cdot p) + (1 - \alpha)(\hat{c}' \cdot p) \leq \alpha w + (1 - \alpha)w'.$$

Therefore $\alpha\hat{c} + (1 - \alpha)\hat{c}' \in B(p, \alpha w + (1 - \alpha)w')$, and thus

$$v(p, \alpha w + (1 - \alpha)w') \geq u(\alpha\hat{c} + (1 - \alpha)\hat{c}') \geq \alpha u(\hat{c}) + (1 - \alpha)u(\hat{c}') = \alpha v(p, w) + (1 - \alpha)v(p, w'),$$

which proves (5.1). $\square$

6 Comparative Statistics

Comparative statistics exercises explore the effects of changes in parameters on choice variables, holding all other factors constant, *ceteris paribus*.
We will assume continuous, strongly monotone and strictly quasi-concave utility functions. These properties ensure the presence of an individual demand function $d(p, w)$, which obeys Walras' law and indicates how much of each good the DM will purchase given p and w.

6.1 Slutsky's Theorems

For the purposes of our analysis, we denote $\Delta := p' - p$ and use this notation interchangeably to signify changes in prices, with $\Delta > 0$ unless stated otherwise. Specifically, $\Delta_k = p'_k - p_k$ indicates price changes for the k-th good only, *ceteris paribus*.

Definition 6.1 (Price effects). Price effects arise from two distinct forces:

i) a **substitution effect**, which shifts optimal consumption away from goods that have become relatively more expensive;

ii) a **wealth effect**, which reduces the overall budget set, lowering the real income, the purchasing power.

Definition 6.2 (Slutsky wealth adjustment). Let $\hat{c}$ be the optimal bundle in $B(p, w)$. If p changes to p', define the wealth level w' as the adjusted wealth such that $\hat{c}$ is just affordable under the new price p', i.e. $w' = p' \cdot \hat{c}$ (by Walras' law). The adjustment

$$w' - w = p' \cdot \hat{c} - p \cdot \hat{c} = \Delta \cdot \hat{c},$$

is called the **Slutsky wealth adjustment**. It neutralises wealth effects by holding real income constant, allowing a clearer examination of the substitution effect alone.

Theorem 6.3 (Slutsky law of demand). *Let $(p, w), (p', w') \in D$ be such that $p' \cdot \hat{c} = w'$, with $\hat{c} = d(p, w)$ and $\hat{c}' = d(p', w')$. Then*

$$\boxed{(p' - p) \cdot (\hat{c}' - \hat{c}) \leq 0},$$

with strict inequality when $\hat{c} \neq \hat{c}'$.

Proof.
Applying Walras' law we derive:

$$(p' - p) \cdot (\hat{c}' - \hat{c}) = \underbrace{p' \cdot \hat{c}'}_{w'} - \underbrace{p' \cdot \hat{c}}_{w'} - p \cdot \hat{c}' + \underbrace{p \cdot \hat{c}}_{w} = w - p \cdot \hat{c}'.$$

We need to demonstrate that $p \cdot \hat{c}' \geq w$, which holds due to WARP. Otherwise, we can argue by contradiction. If $\hat{c} = \hat{c}'$, clearly $p \cdot \hat{c}' = p \cdot \hat{c} = w$. If not, we have $\hat{c} \in B(p', w')$ but $\hat{c} \neq \hat{c}'$, meaning $\hat{c}$ is still affordable yet not optimal, this implies $u(\hat{c}') > u(\hat{c})$. If $\hat{c}' \in B(p, w)$, it would contradict the optimality of $\hat{c}$. Hence, $\hat{c}' \notin B(p, w)$, and $p \cdot \hat{c}' > w$. □

Remark 6.4. if $p'_k > p_k$ and $p'_i = p_i$ for $i \neq k$, which is the typical ceteris paribus assumption, then the inequality indicates that, under a Slutsky wealth adjustment, an increase in the price of some good leads to a decrease in its demand.

Slutsky difference identity: Given $\hat{c} = d(p, w)$ and $\hat{c}' = d(p', w')$:

$$d(p', w) - d(p, w) = \underbrace{d(p', w) - d(p', w')}_{\text{wealth effect}} + \underbrace{d(p', w') - d(p, w)}_{\text{substitution effect}}.$$

Setting $\Delta = p' - p$ and $w' - w = \Delta \cdot \hat{c}$:

$$d(p + \Delta, w) - d(p, w) = \underbrace{d(p + \Delta, w) - d(p + \Delta, w + \Delta \cdot \hat{c})}_{\text{wealth term}} + \underbrace{d(p + \Delta, w + \Delta \cdot \hat{c}) - d(p, w)}_{\text{substitution term}}.$$

If only the price of the k-th component changes and $\Delta \to 0$, we derive:

$$\frac{\partial d_k(p, w)}{\partial p_k} = \underbrace{-\frac{\partial d_k(p, w)}{\partial w} d_k(p, w)}_{\text{wealth term}} + \underbrace{\frac{\partial d_k(p, w)}{\partial w} d_k(p, w) + \frac{\partial d_k(p, w)}{\partial p_k}}_{\text{substitution term}} = s_{kk}(p, w) - \frac{\partial d_k(p, w)}{\partial w} d_k(p, w),$$

where $s_{kk}(p, w)$ denotes the substitution term.
This result, however, is more general, since it holds also for local cross-price substitution terms, defined as

$$s_{kj}(p, w) \equiv \lim_{\Delta_j \to 0} \frac{d_k(p + \Delta_j, w + \hat{c}_j \Delta_j) - d_k(p, w)}{\Delta_j}.$$

Here, s_{kj} is the cross-price substitution effect on good k due to a price change $\Delta_j = p'_j - p_j$ in good j. If the demand function is differential, the Slutsky Differential Identity is:

$$\boxed{\frac{\partial d_k(p, w)}{\partial p_j} = s_{kj}(p, w) - \frac{\partial d_k(p, w)}{\partial w} d_j(p, w)}.$$

6.2 Types of Goods

Definition 6.5 (Normal good). if $w > w' \implies d_k(p, w) \geq d_k(p, w')$, thus the wealth effect is positive (in the local case, $\partial d_k(p, w)/\partial w \geq 0$).

Definition 6.6 (Inferior good). if $w > w' \implies d_k(p, w) \leq d_k(p, w')$, thus the wealth effect is negative.

Definition 6.7 (Giffen good). If, for some $(p, w), (p', w) \in D$, it holds $p'_k > p_k \implies d_k(p', w) > d_k(p, w)$, meaning that an increase in price increases the demand, ceteris paribus. They are a special class of inferior goods, where wealth effects are strong.

Theorem 6.8 (Normal law of demand). *An increase (decrease) in the price of a normal good decreases (increases) its demand, ceteris paribus. Moreover, if we add the condition that k has an initial positive demand, a strict increase (decrease) in its price strictly decreases (increases) its demand, ceteris paribus.*

Table 3: Cross-price effects

	Normal	**Inferior**
Substitutes	Ambiguous	Positive
Complements	Negative	Ambiguous

Proof.
We prove the statement in the case $p'_k > p_k$. For ease of notation, we omit to specify we are dealing with k-th components. We only need to prove that $d(p', w) - d(p, w) \leq 0$. Let w' be the wealth level that would make the previous optimal bundle $\hat{c}$ just affordable given the price p'.
Starting from Slutsky difference identity and using the normality of k:

$$d(p', w) - d(p, w) = \underbrace{d(p', w) - d(p', w')}_{\leq 0} + \underbrace{d(p', w') - d(p, w)}_{\leq 0} \leq 0$$

where the first inequality is the normality of k, since $w = p' \cdot d(p, w) \leq p' \cdot d(p, w) = w'$, and the second is the Slutsky Law of demand. □

Theorem 6.9 (Local Normal Law of Demand). *If the demand function is differentiable at $(p, w) \gg \mathbf{0}$ and k is normal at (p, w), then*

$$\frac{\partial d_k(p, w)}{\partial p_k} \leq 0.$$

Proof. Since good k is normal at (p, w), we have $\partial d_k(p, w)/\partial w \geq 0$. On the other hand, by the Slutsky law of demand $s_{kk}(p, w) \leq 0$ (use the difference quotient definition). Thus the conclusion holds. □

Proposition 6.10 (Law of Reversibility). *If the indirect utility function v is twice differentiable at $(p, w) \gg \mathbf{0}$, with $\partial v(p, w)/\partial w \neq 0$, then $s_{kj}(p, w) = s_{jk}(p, w)$.*

Definition 6.11 (Cross-price relations). If the demand function is differentiable at $(p, w) \gg \mathbf{0}$, two goods k and j, with $s_{kj} = s_{jk}$, are

i) substitutes at (p, w) if $s_{kj}(p, w) \geq 0$; coffee and tea are substitutes: decreasing the price of tea decreases the demand of coffee, since the consumer replaces coffee with tea.

ii) complements at (p, w) if $s_{kj}(p, w) \leq 0$; coffee and sugar are complements: decreasing the price of one decreases the demand of the other.

iii) independent at (p, w) if $s_{kj}(p, w) = 0$; coffee and salt are independent.

Remark 6.12 (Cross-Price Effects). The effect of a change of the price of some good j on the consumption of some other good k is determined by the combination of income and substitution effects, which can either reinforce or offset each other. There are four cases:

i) Goods k and j are substitutes at (p, w) and k is normal at w: the effect is **ambiguous**, since substitution and wealth effects go in opposite directions.

ii) Goods k and j are substitutes at (p, w) and k is inferior at w: the effect is **positive**. For example, a fall in p_j will cause a fall in the demand for k.

iii) Goods k and j are complements at (p, w) and k is normal at w: the effect is **negative**. For example, a fall in p_j will cause a rise in the demand for k.

iv) Goods k and j are complements at (p, w) and k is inferior at w: the effect is **ambiguous**, since substitution and wealth effects go in opposite directions.

7 Decisions Under Risk

Rational choice under risk broadens the domain of rational choice theory from certainty to uncertainty by considering random outcomes. It models scenarios such as games of chance or decisions where probabilities are precise and known, yet the final results remain uncertain. This approach essentially extends the concepts of decisions under certainty.

7.1 Modelling random consequences

Definition 7.1 (Prize space). The set $\mathbf{C}$ of all possible consequences, referred to here as prizes.

Definition 7.2 (Lottery). A lottery l is a density function $l : \mathbf{C} \to [0,1]$ with a finite support, meaning it satisfies:

i) $l(c) \geq 0$ for every $c \in \mathbf{C}$;

ii) $l(c) = 0$ unless $c \in \operatorname{supp} l$, where $\operatorname{supp} l = \{c \in \mathbf{C} : l(c) > 0\}$ is a finite set;

iii) $\sum_{c \in \mathbf{C}} l(c) = \sum_{c \in \operatorname{supp} l} l(c) = 1$.

Definition 7.3 (Risky prospect). A risky prospect is the set of pairs $\{c_i, l(c_i)\} = \{c_i, p_i\}$, defined as

$$\{c_1, p_1; \ldots; c_n, p_n\} = \{(c,p) \in \operatorname{Gr} l : p > 0\},$$

where $\operatorname{Gr} l = \{(c,p) \in \mathbf{C} \times [0,1] : p = l(c)\}$.

Example 7.4. Consider a lottery where you win 1 if you flip a coin and get heads, and lose 1 if you get tails. The risky prospect can be represented as $\{-1, 1/2; 1, 1/2\}$.

Definition 7.5 (Lottery space). The set of lotteries, denoted $\mathbf{L}$. It is convex, indicating that any mixture of lotteries $ql + (1-q)l'$ for $q \in [0,1]$ is also a lottery.

Proof.

Let $l, l' \in \mathbf{L}$ and $q \in [0,1]$ (write $(1-q)$ for $1-q$). We show that $ql + (1-q)l' \in \mathbf{L}$.

Finite support. We claim $\operatorname{supp}[ql + (1-q)l'] = \operatorname{supp} l \cup \operatorname{supp} l'$. Indeed, c belongs to the left-hand support iff $ql(c) + (1-q)l'(c) > 0$, i.e. $l(c) > 0$ or $l'(c) > 0$.

Non-negativity. For every c, $(ql + (1-q)l')(c) \geq 0$.

Unit mass. With $\mathcal{S} := \operatorname{supp}[ql + (1-q)l']$,

$$\sum_{c \in \mathcal{S}} \big(ql + (1-q)l'\big)(c) = q \sum_{c \in \mathcal{S}} l(c) \; + \; (1-q) \sum_{c \in \mathcal{S}} l'(c) = q \cdot 1 + (1-q) \cdot 1 = 1,$$

because $\mathcal{S}$ contains $\operatorname{supp} l$ and $\operatorname{supp} l'$.

□

Remark 7.6 (Compound lotteries). Convex combinations $\tilde{l} = ql + (1-q)l'$ of lotteries can be represented by *compound lotteries*, having lotteries as prizes. If $l = \{c_1, p;\, c_2, 1-p\}$ and $l' = \{c_3, p';\, c_4, 1-p'\}$, then

$$\tilde{l} = \{l, q;\, l', 1-q\} = \{c_1, pq;\, c_2, (1-p)q;\, c_3, p'(1-q);\, c_4, (1-p')(1-q)\}.$$

Lotteries can thus be depicted with decision trees, and compound lotteries are decision trees whose terminal nodes are themselves lotteries.

Definition 7.7 (Trivial lotteries). The risk-free lottery $\{c, 1\}$ is denoted δ_c. Clearly,

$$l(c) = \sum_{d \in \operatorname{supp} l} l(d)\, \delta_d(c) = \sum_{d \in \mathbf{C}} l(d)\, \delta_d(c).$$

Remark 7.8. Identifying the risk-free lottery $\{c, 1\}$ with the prize c itself, the prize space coincides with the set of risk-free lotteries: $\mathbf{C} \equiv \{\delta_c : c \in \mathbf{C}\} \subseteq \mathbf{L}$.

7.2 Decisions

We start from the quartet $(\mathbf{A}, \mathcal{A}, \mathbf{L}, \rho)$ representing an action space $\mathbf{A}$, a consequence function $\rho : \mathbf{A} \to \mathbf{L}$ that assigns to each action a its random consequence $\rho(a) \in \mathbf{L}$, and a collection $\mathcal{A}$ of action sets $A \subseteq \mathbf{A}$. Denote by $\mathbf{A}_c \subseteq \mathbf{A}$ the set of actions whose consequences are certain, i.e. $\rho(a) \in \mathbf{C}$.
A decision maker (DM) has three universal preferences:

- $\succsim$ over actions in $\mathbf{A}$,
- $\dot{\succsim}$ over consequences in $\mathbf{C}$,
- $\ddot{\succsim}$ over lotteries in $\mathbf{L}$.

The third drives the other two:

i) Since $\mathbf{C} \subseteq \mathbf{L}$, for any $c, c' \in \mathbf{C}$ we have $c \dot{\succsim} c' \iff \delta_c \ddot{\succsim} \delta_{c'}$.

ii) We assume **random consequentialism**, ensuring that actions are ranked according to their random consequences: **actions that induce the same distribution of consequences are indifferent**. For all actions a and b, $\rho(a) = \rho(b) \implies a \sim b$.[23] Under this assumption, we have $a \succsim b \iff \rho(a) \ddot{\succsim} \rho(b)$, so that $\ddot{\succsim}$ uniquely determines $\succsim$.

Once the preference relation $\ddot{\succsim}$ is defined, $(\mathbf{A}, \mathcal{A}, \mathbf{L}, \rho)$ is extended to $(\mathbf{A}, \mathcal{A}, \mathbf{L}, \rho, \succsim, \ddot{\succsim})$. If we then introduce a utility function $\ddot{u} : \mathbf{L} \to \mathbb{R}$, we can define $u : \mathbf{A} \to \mathbb{R}$, given by $u = \ddot{u} \circ \rho$, and $\dot{u} : \mathbf{C} \to \mathbb{R}$, which is the restriction of $\ddot{u}$ to $\mathbf{C}$.
From these definitions, we derive:

i) the decision environment under risk is $(\mathbf{A}, \mathcal{A}, \ddot{u} \circ \rho)$ and the decision problem under risk is, for each $A \in \mathcal{A}$, $(A, \ddot{u} \circ \rho)$.

ii) $\ddot{u} \circ \rho$ over <u>actions</u> serves as the decision criterion, and the DM has to solve the optimization problem $\max_a \ddot{u}(\rho(a)) \quad \text{sub } a \in A$.

iii) Let $\mathcal{D}$ be the collection of choice sets where the optimization problem has a solution. Then, the functions $\sigma : \mathcal{D} \rightrightarrows \mathbf{L}$ and $v : \mathcal{D} \to \mathbb{R}$ are defined as:

$$\sigma(A) = \arg\max_{a \in A} \ddot{u}(\rho(a)) \quad \text{and} \quad v(A) = \max_{a \in A} \ddot{u}(\rho(a)).$$

We now shift to the reduced-form, lottery-based analysis by expressing $\ddot{\succsim}$ in terms of $\succsim$. In this framework, any action $a \in \rho^{-1}(l)$ is identified with the corresponding lottery l. Preferences are thus considered directly over lotteries.

i) Define $\mathcal{L} = \{\rho(A) : A \in \mathcal{A}\}$, the collection of lottery sets corresponding to action sets.

ii) The decision environment under risk in reduced form is $(\mathbf{L}, \mathcal{L}, \ddot{u})$ and the decision problem under risk is, for each $L \in \mathcal{L}$, $(L, \ddot{u})$.

iii) $\ddot{u}$, now defined over *lotteries*, serves as the decision criterion. The DM solves the optimization problem: $\max_l \ddot{u}(l) \quad \text{sub } l \in L$. Once a solution $\hat{l}$ is found, any action $\hat{a} \in \rho^{-1}(\hat{l}) \cap A$ solves the original optimization problem, and vice versa.

8 Expected Utility Theory

8.1 The St. Petersburg Paradox

In principle, to value a lottery with monetary prices, it should be enough to use the expected value $\mathbb{E}(l) = c_1 l(c_1) + \cdots + c_n l(c_n)$, since $n\,\mathbb{E}(l)$ is approximately the value we would get if playing n times, with n sufficiently large. However, this reasoning fails to account for personal risk aversion, diminishing marginal utility, or other individual preference criteria. Moreover, the expected value can paradoxically become infinite in certain cases.

Example 8.1. Example (The St. Petersburg Paradox): Consider a game where a coin is tossed until the first heads appears. The player receives 2^n ducats if heads occurs on the n-th toss. The utility, i.e. the expected value of this game, is infinite, meaning that the DM would be willing to bet all his fortune: $\mathbb{E} = \displaystyle\sum_{n=1}^{\infty} \frac{2^n}{2^n} = \infty$.
To resolve this, one could use a logarithmic scale as a measure of the DM's benefit from owning c units of money. The logarithmic utility accounts for diminishing marginal utility. The expected utility in this case becomes:

$$\mathbb{E}[\log(c)] = \log(2) \cdot 2 = 2\log(2) \approx 1.4.$$

Based on this reasoning, the decision maker would be willing to pay at most $c \leq e^{2\log(2)} = 4$ ducats to participate in the game.
A better measure for evaluating lotteries is the **expected utility criterion** $\ddot{u} : \mathbf{L} \to \mathbb{R}$ defined as

$$\ddot{u}(l) = \sum_{c \in \operatorname{supp} l} \dot{u}(c) l(c) = \mathbb{E}_l\, \dot{u}.$$

[23] Note that $\mathbf{A}_c \subseteq \mathbf{A}$, so that random consequentialism implies outcome consequentialism.

From this point forward, u will denote preferences over prizes, replacing $\dot{u}$, and $\succsim$ will denote preferences over lotteries, replacing $\ddot{\succsim}$.
To proceed in our analysis, we need to introduce some axioms for $\succsim$. These will lead, thanks to the fundamental representation theorem of **von Neumann** and **Morgenstern**, to the existence of a utility function $\ddot{u}$ over lotteries, which we call **Bernoulli** and denote as $\ddot{u}$, that admits an expected utility representation through the function u on prizes, which we call *von Neumann-Morgenstern (vN-M) utility function*.
The proof of the vN-M theorem extends to more abstract settings, enhancing our understanding of the structure of expected utility. This generalization allows applications of the theorem far beyond the scope of risk theory, providing a foundation for modern decision theory.

8.2 Axioms

(B.1) Weak Order: $\succsim$ is complete and transitive.

(B.2) Independence: for all $l, l', l'' \in \mathbf{L}$ and $0 < p < 1$,

$$l \succ l' \implies pl + (1-p)l'' \succ pl' + (1-p)l''$$

(Comment): This can be interpreted using decision trees and compound lotteries. According to this axiom, the ranking is not affected by the common part, and what matters is only the other part, where they differ, i.e. on the ranking between l and l'.

(B.3) Archimedean for all $l, l', l'' \in \mathbf{L}$ with $l \succ l' \succ l''$, there exist $p, q \in (0,1)$ such that

$$pl + (1-p)l'' \succ l' \succ ql + (1-q)l''.$$

Again, the intuition is that there are no infinitely preferred or infinitely despised lotteries.

(B.2 bis) Strong Independence For all $l, l', l'' \in \mathbf{L}$ and $p \in (0,1)$,

$$l \succsim l' \iff pl + (1-p)l'' \succsim pl' + (1-p)l''.$$

In particular,

$$l \sim l' \iff pl + (1-p)l'' \sim pl' + (1-p)l''$$

(Comment:) B.2 and B.3 combined imply B.2 bis. The proof, however, relies on the abstract vN-M theorem. This axiom is particularly appealing because it trivially shows that $\succsim$ is affine (the proof is left as an exercise). As a consequence, $\ddot{u}$ is affine.

Certainty Effect the **Allais Paradox** is an example of violation of the independence axiom. It's due to a "certainty effect" affecting the ranking between l and l' that features a risk-free alternative. Empirical evidence shows that DMs will tend to violate the independence axiom when one lottery is risk-free, but this pattern disappears when both alternatives are risky.

8.3 Von Neumann-Morgenstern Representation Theorem

Theorem 8.2 (Von Neumann & Morgenstern). *Let $\succsim$ be a preference on the set $\mathbf{L}$ of all lotteries defined on a prize space $\mathbf{C}$. The following conditions are equivalent:*

i) $\succsim$ satisfies axioms B.1, B.2, B.3;

ii) There exists a function $u : \mathbf{C} \to \mathbb{R}$ such that the function $\ddot{u} : \mathbf{L} \to \mathbb{R}$ defined by

$$\ddot{u}(l) = \sum_{c \in supp\ l} u(c)l(c) = \mathbb{E}_l\, u \tag{vN-M}$$

represents $\succsim$. Moreover, the function u (and so $\ddot{u}$ [24]*) is cardinal, i.e. unique up to a positive affine transformation.*[25]

Corollary 8.3. *under axioms B.1, B.2, B.3, we have*

$$l \succsim l' \iff \mathbb{E}_l\, u \geq \mathbb{E}_{l'}\, u \iff \ddot{u}(l) \geq \ddot{u}(l') \iff \sum_{c \in \mathbf{C}} u(c)l(c) \geq \sum_{c \in \mathbf{C}} u(c)l'(c).$$

Corollary 8.4. *under axioms B.1, B.2, B.3, $\ddot{u}(l) = \mathbb{E}_l\, u$ if and only if $\ddot{u}(l)$ is affine.*

[24] By linearity of the expectation, if $u' = au + b$, then $\ddot{u}'(l) = \mathbb{E}_l(au + b) = a\ddot{u} + b$.

[25] Note that vN-M utility function is the restriction on prizes of the Bernoulli utility function. However, under the 3 axioms, such restriction uniquely determines the Bernoulli utility function, so that there's a perfect correspondence between the two utilities.

Proof.
If $\ddot{u}$ is affine, then, for every $l \in \mathbf{L}$:

$$\ddot{u}(l) = \ddot{u}\left(\sum_{c \in \text{supp } l} l(c)\delta_c\right) = \sum_{c \in \mathbf{C}} l(c)\ddot{u}(\delta_c) = \sum_{c \in \mathbf{C}} u(c)l(c) = \mathbb{E}_l\, u,$$

where the first equality is l written in terms of δ_c, the second is the affinity of $\ddot{u}$, and the third is setting $u := \ddot{u} \circ \delta$. Conversely, if $\ddot{u}(l) = \sum_{c \in \mathbf{C}} u(c)l(c)$, then

$$\ddot{u}(ql+(1-q)l') = \sum_{c \in \mathbf{C}} u(c)[ql(c)+(1-q)l'(c)] = q\left(\sum_{c \in \mathbf{C}} u(c)l(c)\right)+(1-q)\left(\sum_{c \in \mathbf{C}} u(c)l'(c)\right) = q\ddot{u}(l)+(1-q)\ddot{u}(l'),$$

implying that $\ddot{u}$ is affine. $\square$

Example 8.5. under axioms B.1-B.3, take

$$l = \left\{1, \frac{1}{3}; 3, \frac{2}{3}\right\} \quad \text{and} \quad l' = \left\{2, \frac{1}{4}; -1, \frac{1}{2}; 0, \frac{1}{4}\right\}$$

By the von Neumann-Morgenstern Representation Theorem, there exists a vN-M utility function u such that $l \succsim l'$ if and only if

$$\ddot{u}(l) = \frac{1}{3}u(1) + \frac{2}{3}u(3) \geq \frac{1}{4}u(2) + \frac{1}{2}u(-1) + \frac{1}{4}u(0) = \ddot{u}(l').$$

If the vN-M utility function u is $u(c) = c^2$, we obtain $\ddot{u}(l) = \frac{1}{3} + \frac{2 \cdot 9}{3} = \frac{19}{3}$ and $\ddot{u}(l') = 1 + \frac{1}{2} + 0 = \frac{3}{2}$. Therefore, $\ddot{u}(l) > \ddot{u}(l')$, and we conclude that $l \succ l'$.

8.4 Importance of the Von Neumann-Morgenstern Representation Theorem

The vN-M theorem established the basis for modern choice theory under risk. It states that, under suitable axioms, the DM ranks lotteries according to their expected utilities. Prior to this theorem, it was not clear, except for the analytical convenience and intuitive appeal, why not to consider besides the mean also the variance, or other moments, of the distribution of the prizes' utilities. For centuries, the expected utility criterion lacked a solid axiomatic foundation.
The vNM theorem directly addresses this issue, providing a rigorous axiomatic basis. It demonstrates that a decision-maker who satisfies four core axioms should rank lotteries solely by their expected utilities. This result shows that the expected utility criterion possesses a transparent and compelling preference-based foundation. In contrast, other criteria, such as those involving variance, violate these axioms.
Additionally, this foundation gives expected utility a clear experimental content, since its peculiar independence axiom can be tested experimentally. This is evident in Allais-type paradoxes or other certainty effects, which illustrate preferences that deviate from the vNM model.

8.5 On cardinality

As previously mentioned, a vN-M utility function is cardinally unique, i.e. unique up to a positive affine transformation. To prove that u is cardinal, one should prove two things:

i) $\mathbb{E}(au + b)$ is a utility representation of $\succsim$ for every $a > 0$ and $b \in \mathbb{R}$;

ii) if $\mathbb{E}(v)$ is a utility representation of $\succsim$, then there exist $a > 0$ and $b \in \mathbb{R}$ such that $v = au + b$.

If f is strictly increasing but not affine, then $f \circ u$ and $f \circ \ddot{u}$ are still ordinal (rather than cardinal) utility functions on prizes and lotteries, but they are no longer connected through the expected utility formula. Actually, any strictly increasing but nonlinear transformation won't work.

9 Abstract Von Neumann Morgenstern Theorem

We now present a version of the von Neumann-Morgenstern (vN-M) Representation Theorem in a general setting X, which is a convex subset of a vector space; x, y, z are elements of X, while α, β, γ are elements of $[0, 1]$; For convenience, we use the notation $\overline{\alpha} := 1 - \alpha$ for any $\alpha \in [0, 1]$.
$\succsim$ is a binary relation on X which satisfies axioms $A.1$, $A.2$, $A.4$ and $A.8$, where $A.8$ is the independence axiom, which can be stated in this form

(A.8) Independence: For all $x, y, z \in X$ and $\alpha \in (0,1)$, $x \succ y \implies \alpha x + (1-\alpha)z \succ \alpha y + (1-\alpha)z$

Abstract vN-M Theorem: $\succsim$ satisfies axioms $A.1$, $A.2$, $A.4$ and $A.8$ if and only if $\succsim$ admits an affine utility function $u : X \to \mathbb{R}$. Moreover, u is cardinal.

If $\succsim$ satisfies axioms $A.1$, $A.2$, $A.4$, $A.8$, the following lemmas are true:	
Lemma 0: If $x \succ y$, then $x \succ \alpha x + \overline{\alpha} y \succ y$.	We directly apply $A.8$, inserting $z = x$ and $z = y$. If $z = x$ we get $x \succ \alpha y + \overline{\alpha} x = \beta x + \overline{\beta} y$, with $\beta = \overline{\alpha}$. If $z = y$ we get $\alpha x + \overline{\alpha} y \succ y$.
Lemma 1: If $x \succ y$, then $\alpha > \beta \iff \alpha x + \overline{\alpha} y \succ \beta x + \overline{\beta} y$	If $\alpha = 1$, then $x \succ \beta x + \overline{\beta} y$, true by **lemma 0**. If $\alpha = 0$, then $\alpha x + \overline{\alpha} y \succ y$, true by **lemma 0**. If $0 < \beta < \alpha < 1$, then $0 < \beta/\alpha < 1$, and, by **lemma 0**, $\alpha x + \overline{\alpha} y \succ y$. Now apply $A.8$ setting $x \mapsto \alpha x + \overline{\alpha} y$, $y \mapsto y$, $z \mapsto \alpha x + \overline{\alpha} y$, and $\alpha \mapsto 1 - \beta/\alpha$. If you believe enough, you'll reach the conclusion... Conversely, suppose that $\alpha x + \overline{\alpha} y \succ \beta x + \overline{\beta} y$. if $\alpha = \beta$ then $\alpha x + \overline{\alpha} y \sim \beta x + \overline{\beta} y$, contradiction. If $\alpha < \beta$, then $\beta x + \overline{\beta} y \succ \alpha x + \overline{\alpha} y$, contradiction.
Lemma 2, continuity: If $x \succsim z \succsim y$, with $x \succ y$, then there exists a unique $\alpha \in [0,1]$ such that $z \sim \alpha x + \overline{\alpha} y$.	Uniqueness directly follows from **lemma 1**, since $z \sim \alpha x + \overline{\alpha} y \sim \beta x + \overline{\beta} y$ with $\alpha \neq \beta$ is clearly a contradiction of the previous lemma. As for the existence, suppose $x \succ z \succ y$, otherwise just take $\alpha = 0$ or $\alpha = 1$. Define $A = \{\alpha \mid \alpha x + \overline{\alpha} y \succ z\}$ and $B = \{\beta \mid z \succ \beta x + \overline{\beta} y\}$. Note that A and B are disjoint and $1 \in A$, $0 \in B$. We now prove that they are both intervals. For any $\alpha \in A$, take $\xi \in (\alpha, 1]$. Since $\xi > \alpha$, then, by **lemma 1**, $\xi x + \overline{\xi} y \succ \alpha x + \overline{\alpha} y \succ z$, so that $\xi \in A$. $[\alpha, 1] \subseteq A$, hence A is an interval, and similarly for B. If we prove that A and B are open, then two open disjoint intervals cannot cover $[0,1]$ leading the existence of $\lambda \in [0,1]$ such that $\lambda \notin A \cup B$, meaning that $z \sim \lambda x + (1-\lambda) y$. (Since they are open, then $B = [0, b)$ and $A = (a, 1]$ for some $a, b \in [0,1]$. If $a = b$, then $\lambda = a = b$ works. If $b < a$, then $\lambda = b$ works.) To prove that A and B are open, we take $\alpha_0 \in A$ and we find $\alpha_1 < \alpha_0$ such that $\alpha_1 \in A$. This can be achieved by the **Archimedean** axiom, which guarantees the existence of $\xi \in (0,1)$ such that $$\xi(\alpha_0 x + \overline{\alpha}_0 y) + \overline{\xi} y \succ z \implies \alpha_0 \xi x + \overline{\alpha_0 \xi} y \succ z,$$ meaning that $\alpha_0 \xi \in A$. But $\alpha_0 > \alpha_0 \xi$ since $\xi < 1$. Setting $\alpha_1 := \alpha_0 \xi$, we conclude that A is open. Similarly you conclude that B is open.
Lemma 3: If $x \sim y$, then $\alpha x + (1-\alpha) y \sim y$.	If $\alpha = 0$ or $\alpha = 1$, it's obvious. If not, suppose by contradiction that $\alpha x + \overline{\alpha} y \succ x \sim y$ or $x \sim y \succ \alpha x + \overline{\alpha} y$. By applying twice $A.8$ we get $$\alpha x + \overline{\alpha} y = \alpha(\alpha x + \overline{\alpha} y) + \overline{\alpha}(\alpha x + \overline{\alpha} y) \succ \alpha x + \overline{\alpha}(\alpha x + \overline{\alpha} y) \succ \alpha x + \overline{\alpha} y.$$
Lemma 4: If $x \sim y$, then $\alpha x + \overline{\alpha} z \sim \alpha y + \overline{\alpha} z$.	If $x \sim z$, the result would follow from **lemma 3**. Suppose $z \prec x$ (a similar argument holds if $z \succ x$). Suppose, by contradiction, that $\alpha x + \overline{\alpha} z \prec \alpha y + \overline{\alpha} z$ (a similar argument shows that $\alpha x + \overline{\alpha} z \succ \alpha y + \overline{\alpha} z$ leads to a contradiction). By **lemma 0**, $z \prec x$ implies $z \prec \alpha x + \overline{\alpha} z$. Since $$z \prec \alpha x + \overline{\alpha} z \prec \alpha y + \overline{\alpha} z,$$ by **lemma 2** there exists a unique $\beta \in (0,1)$ such that $$\alpha x + \overline{\alpha} z \sim \beta(\alpha y + \overline{\alpha} z) + \overline{\beta} z = \beta \alpha y + \overline{\alpha \beta} z.$$ by **lemma 0**, $z \prec x \sim y$ implies $\beta y + \overline{\beta} z \prec y \sim x$. By $A.8$ this implies $$\alpha(\beta y + \overline{\beta} z) + \overline{\alpha} z \prec \alpha x + \overline{\alpha} z.$$ But LHS simplifies to $\alpha \beta y + \overline{\alpha \beta} z$, so $\alpha x + \overline{\alpha} z \sim \alpha \beta y + \overline{\alpha \beta} z \prec \alpha x + \overline{\alpha} z$, contradiction.

If $\succsim$ satisfies the axioms, then it admits an affine utility representation.	
If $x \sim y$ for every $x, y \in X$, then u represents $\succsim$ if and only if u is a constant	Obvious.
If not, there exist $b \succ w$ in X. We consider the **preference interval** $[w, b]$. Such interval is convex.	$[w, b] := \{x \vert w \precsim x \precsim b\}$. If $x \prec y \in [w, b]$, then $w \precsim x \prec y \precsim b$. By **lemma 1**, $w \precsim x \prec \alpha x + \overline{\alpha} y \prec y \precsim b$, so it's a convex interval.
We define $u_{wb}(x) = \alpha_x$, where α_x is such that $x \sim \alpha_x b + \overline{\alpha}_x w$.	Since $[w, b]$ is convex, we apply **lemma 2**: there exists a unique α_x satisfying that property
$u_{wb}(x) = \alpha_x$ is a valid **utility** function. Thus it represents $\succsim$ on $[w, b]$.	For every $x, y \in [w, b]$, we have $x \succ y \iff \alpha_x b + \overline{\alpha}_x w \succ \alpha_y b + \overline{\alpha}_y w \iff \alpha_x > \alpha_y \iff u_{wb}(x) > u_{wb}(y)$, by **lemma 1**.
$u_{wb}(x)$ is also **affine** on $[w, b]$. u on $[w, b]$ is exactly what we want. We now need to extend this function to any other interval.	For every x, y and $\lambda \in [0, 1]$, $\lambda x + (1-\lambda) y \sim \lambda(\alpha_x b + \overline{\alpha}_x w) + (1-\lambda)(\alpha_y b + \overline{\alpha}_y w) = b(\lambda \alpha_x + (1-\lambda)\alpha_y) + w(\overline{\lambda \alpha_x + (1-\lambda)\alpha_y})$. This means that $u_{wb}(\lambda x + (1-\lambda) y) = \lambda u_{wb}(x) + (1-\lambda) u_{wb}(y)$.
Fix $x_0 \prec x_1 \in X$ and define on them the **family of intervals** $\mathcal{I} = \{[w, b] : w \prec b \text{ and } x_0, x_1 \in [w, b]\}$. Now a general result: if Y is a convex subset of X and u, v are two affine representations of $\succsim$ on Y, then if they coincide in two points, they coincide everywhere. To use this result, we need to **renormalize** u. Up to now, for every w, b, we have $u_{wb}(w) = 0$ and $u_{wb}(b) = 1$.	Suppose $x_0 \prec x_1 \in Y$ and $u(x_0) = v(x_0)$ and $u(x_1) = v(x_1)$. We need to prove that $u(x) = v(x)$ for every $x \in Y$. There are three cases: • If $x_0 \precsim x \precsim x_1$, then by **lemma 2** there exists a unique α such that $x \sim \alpha x_0 + \overline{\alpha} x_1$. Thus $u(x) = \alpha u(x_0) + \overline{\alpha} u(x_1) = \alpha v(x_0) + \overline{\alpha} v(x_1) = v(x)$, by affinity. • If $x \prec x_0 \prec x_1$, then by **lemma 2** there exists a unique α such that $x_0 \sim \alpha x_1 + \overline{\alpha} x$. Thus $u(x_0) = \alpha u(x_1) + \overline{\alpha} u(x)$ and $v(x_0) = \alpha v(x_1) + \overline{\alpha} v(x)$. Taking the difference and noting that $\overline{\alpha} \neq 0$, we get the conclusion. • If $x_0 \prec x_1 \prec x$, you argue in the same way.
For every $[w, b] \in \mathcal{I}$, we define v_{wb} as $$v_{wb}(x) = \frac{u_{wb}(x) - u_{wb}(x_0)}{u_{wb}(x_1) - u_{wb}(x_0)}.$$	Since u_{wb} is affine, then v_{wb} is affine. v_{wb} is a utility representation on $[w, b]$. Finally, $v_{wb}(x_0) = 0$ and $v_{wb}(x_1) = 1$, exactly what we want.
For every $x \in [w, b] \in \mathcal{I}$, define $u(x) = v_{wb}(x)$, losing the dependence of the representation on w and b.	u is a function rather than a correspondence. Suppose $x \in [a, b] \in \mathcal{I}$ and $x \in [c, d] \in \mathcal{I}$. Then v_{ab} and v_{cd} are affine utilities and take value 0 on x_0 and 1 on x_1. Therefore, they coincide. This implies that $u(x) = v_{ab} = v_{cd}$.
This function is well defined for every $x \in X$.	We use completeness and transitivity to be able to order x, y, x_0, x_1. We define $w^* = \min_{\succsim}\{x, y, x_0, x_1\}$ and $b^* = \max_{\succsim}\{x, y, x_0, x_1\}$. Then $x, y, x_0, y_0 \in [w^*, b^*] \in \mathcal{I}$. Thus $u(x) = v_{w^*b^*}(x)$.
This function is a utility representation of $\succsim$ and it's affine.	Since $x, y \in [w^*, b^*]$, then $x \succsim y \iff v_{w^*b^*}(x) \geq v_{w^*v^*}(y) \iff u(x) \geq u(y)$. For every $x, y \in X$ and $\alpha \in [0, 1]$, take $w^*, b^* \in X$ such that $x, y \in [w^*, b^*]$. Being this a convex set, $u(\alpha x + \overline{\alpha} y) = v_{w^*b^*}(\alpha x + \overline{\alpha} y) \overset{!}{=} \alpha v_{w^*b^*}(x) + \overline{\alpha} v_{w^*b^*}(y) = \alpha u(x) + \overline{\alpha} u(y)$. Therefore, u is affine (at ! we used the affinity of $v_{w^*b^*}$).

If $\succsim$ admits an affine utility function, then it satisfies axioms $A.1$, $A.2$, $A.4$ and $A.8$	
$\succsim$ is a weak order.	This is a necessary condition for the existence of a utility representation that we have already discussed.
$\succsim$ is Archimedean.	Take $x \succ y \succ z$ elements of X (if they don't exist, then there's nothing to prove). Then $u(x) > u(y) > u(z)$. Since $\mathbb{R}$ is Archimedean[26], there exist $\alpha, \beta \in (0,1)$ such that $$\alpha u(x) + \overline{\alpha} u(z) > u(y) > \beta u(x) + \overline{\beta} u(z).$$ By affinity, $u(\alpha x + \overline{\alpha} z) > u(y) > u(\beta x + \overline{\beta} z) \iff \alpha x + \overline{\alpha} z \succ y \succ \beta x + \overline{\beta} z$.
$\succsim$ satisfies the independence axiom.	Take $x, y, z \in X$ and $\alpha \in (0,1)$. If $x \succ y$, then $u(x) > u(y)$, so $\alpha u(x) + \overline{\alpha} u(z) > \alpha u(y) + \overline{\alpha} u(z)$. By affinity, $u(\alpha x + \overline{\alpha} z) > u(\alpha y + \overline{\alpha} z) \implies \alpha x + \overline{\alpha} z \succ \alpha y + \overline{\alpha} z$.

If there exists an affine representation u, then u is cardinally unique.	
If u is an affine representation of $\succsim$, then $au + b$ is an affine representation of $\succsim$ for every $a > 0$ and $b \in \mathbb{R}$	Since u is affine, then $au + b$ is affine. Moreover, $x \succsim y \iff u(x) \geq u(y) \iff au(x) + b \geq au(y) + b$, thus $au + b$ is a utility representation of $\succsim$ on X.
If $u, v : X \to \mathbb{R}$ are affine representations of $\succsim$, then they coincide up to an affine transformation, i.e. there exist $a > 0$ and $b \in \mathbb{R}$ such that $u = mv + q$.	If one of them is constant, then the other also is, and thus they are equal up to a multiplicative constant. If not, let $x_0 \prec x_1 \in X$, so that $u(x_0) \prec u(x_1)$ and $v(x_0) \prec v(x_1)$. Now normalize both in the same way we did before, so that $\hat{u}(x_0) = 0$, $\hat{u}(x_1) = 1$, $\hat{v}(x_0) = 0$ and $\hat{v}(x_1) = 1$.
$\hat{u}$ and $\hat{v}$ are affine representations of $\succsim$ on X. By the previous lemma, $\hat{u}$ and $\hat{v}$ coincide on X.	The normalizing multiplicative constant is positive, so that $\hat{u} = au + b$ for some $a > 0$ and $b \in \mathbb{R}$, and $\hat{v} = cv + d$ for some $c > 0$ and $d \in \mathbb{R}$. For what we said before, $\hat{u}$ and $\hat{v}$ are affine representations of $\succsim$.
As $\hat{u}(x) = \hat{v}(x)$ for every $x \in X$, then u and v coincide up to an affine transformation.	$$\hat{u} = au + b = cv + d = \hat{v} \implies u = \frac{c}{a} v + \frac{d - b}{a}, \quad c/a > 0.$$

9.1 Probabilistic version of the vN-M theorem

In this section, we explore the elegant connection between the abstract von Neumann-Morgenstern Representation Theorem and its application to risk theory. For a comprehensive discussion on this topic, refer to the appendix, where we extend the vN-M theorem further by introducing the Choquet Expected Utility framework.

Instead of considering a generic convex subset X of a vector space, we now focus on $\Delta_0(\mathbf{C}) = \Delta_0$, the set of probability measures $p : 2^{\mathbf{C}} \to [0,1]$.[27] Clearly, Δ_0 coincides with the previously defined lottery space $\mathbf{L}$, through the identification:

$$l_p(c) = \begin{cases} p(c) & \text{if } c \in \operatorname{supp} p, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, the prize space $\mathbf{C}$ is embedded within Δ_0 by identifying each prize $c \in \mathbf{C}$ with the corresponding Dirac measure δ_c, as we did before. For a comparison of the different but equivalent representations of lotteries, refer to Table 4.

Now, if $\succsim$ satisfies the axioms $B.1$, $B.2$, and $B.3$, the abstract vN-M theorem applied to $(\Delta_0, \succsim)$ guarantees the existence of a cardinally unique affine utility function $\ddot{u} : \Delta_0 \to \mathbb{R}$ such that $p \succsim q \iff \ddot{u}(p) \geq \ddot{u}(q)$, for every $p, q \in \Delta_0$. For every $p \in \Delta_0$, we have:

$$\ddot{u}(p) = \ddot{u}\left(\sum_{c \in \operatorname{supp} p} p(c) \delta_c \right) = \sum_{c \in \operatorname{supp} p} p(c) \ddot{u}(\delta_c) \stackrel{!}{=} \sum_{c \in \operatorname{supp} p} p(c) u(c) = \mathbb{E}_p[u],$$

where ! follows by setting $u(c) := \ddot{u}(\delta_c)$. Cardinal uniqueness of u follows from that of $\ddot{u}$, which comes from the linearity of the expectation. The converse is analogous.[28]

[26] If $x > y > z$, then take $\alpha := 1 - \frac{1}{2}\frac{x-y}{x-z} = \frac{x+y-2z}{2(x-z)}$ and $\beta := \frac{1}{2}\frac{y-z}{x-z}$. Then $1 - \alpha = \frac{1}{2}\frac{x-y}{x-z}$ and $1 - \beta = \frac{2x-y-z}{2(x-z)}$. Starting from $\alpha x + \overline{\alpha} z > y > \beta x + \overline{\beta} z$, you'll get $(y-z)(x-z) < 0$ and $(x-z)(x-y) > 0$, proving the statement.

[27] For the definition of a probability measure, refer to Section 13

[28] Compare this proof with "corollary 2" of vN-M theorem. You will observe that it follows the same logical steps, differing only in the notation used.

Table 4: Different ways to represent lotteries

Lotteries	**Format**	**Expected utility** $\ddot{u}(l)$
Risky prospects	$l = \{c_1, p_1; \dots; c_n, p_n\}$	$\sum_{i=1}^n u(c_i)p_i$
Density functions	$l : \mathbf{C} \to [0,1]$	$\mathbb{E}_l\, u = \sum_{c \in \text{supp } p} u(c)l(c)$
Simple prob. measures	$p : 2^{\mathbf{C}} \to [0,1]$	$\mathbb{E}_p\, u = \sum_{c \in \text{supp } p} u(c)p(c)$

10 Risk Aversion

10.1 Utility of Money

In this section, we restrict our model to lotteries involving monetary gains and losses. A financial asset or investment can be modeled as a monetary lottery $l = \{w_1, p_1; \dots; w_n, p_n\}$. For each prize w_i, there will be an optimal consumption bundle $\hat{c}_i \in B(\pi, w_i)$. therefore, we can directly consider the consumption lottery

$$\hat{l} = \{\hat{c}_1, p_1; \dots, \hat{c}_n, p_n\}.$$

Income lotteries are ranked based on the utility of the optimal consumption bundles they determine. Specifically: $l \succeq_\pi l' \iff \hat{l} \succsim \hat{l}'$, where $\succeq_\pi$ reflects preferences over monetary lotteries, dependent on the price vector π. Introducing a vN-M utility function over consumption lotteries, we see that

$$l \succeq_\pi l' \iff \hat{l} \succsim \hat{l}' \iff \ddot{u}(\hat{l}) \geq \ddot{u}(\hat{l}') \iff \sum_{i=1}^n u(\hat{c}_i)p_i \geq \sum_{i=1}^m u(\hat{c}_i')p_i' \iff \sum_{i=1}^n v_\pi(w_i)p_i \geq \sum_{i=1}^n v_\pi(w_i')p_i',$$

so that v_π is the vN-M utility function of $\succeq_\pi$.

10.2 Risk Attitudes

In this section, we aim to formalize the behavioral aversion of decision makers (DMs) to taking risks, both in absolute and relative terms. To achieve this, we introduce two new axioms: Monotonicity (B.4) and Certainty Equivalence (B.5).

Remark 10.1. Pay attention to the assumptions: if explicitly said "under B.4 and B.5", we are considering a monotone preorder $\succsim$ satisfying the two axioms (but not necessarily B.1); if instead we mention a utility function, then necessarily we are considering a weak order.

(B.4) Strong Monotonicity: For all $c, c' \in \mathbf{C}$, we have $c \gg c' \implies c \succ c'$ and $c \geq c' \implies c \succsim c'$. $\succsim$ on $\mathbf{L}$ is said monotone if it satisfies B.4; it is said monotone vN-M if it satisfies B.1-B.4.

(B.5) Certainty equivalent: For every $l \in \mathbf{L}$ there exists $c_l \in \mathbf{C}$ such that $l \sim c_l$. Under B.4, it's also unique (prove it as an exercise).

Comment: as a consequence, $\ddot{u}(c_l) = u(c_l) = \ddot{u}(l)$, and, when u is strictly increasing, $c_l = u^{-1}(\ddot{u}(l))$. Thus, this axiom is equivalent to assuming the continuity of the vN-M function u, which is equivalent to the topological continuity of $\succsim$ on $\mathbf{C}$. When u is convex, concave, or affine, it's a mild assumption.

Definition 10.2 (Risk premium). Under B.5, we define the risk premium $\pi_l := \mathbb{E}(l) - c_l$. Under B.4 and B.5, we define a certainty equivalent function $\mathbf{c} : \mathbf{L} \to \mathbb{R}$ and a corresponding risk premium function $\boldsymbol{\pi} : \mathbf{L} \to \mathbb{R}$ such that $\mathbf{c}(l) = c_l$ and $\boldsymbol{\pi}(l) = \pi_l$.

Remark 10.3. c_l is the sure amount of money that makes the DM indifferent, or the lowest price at which the DM is willing to sell l. It coincides, by definition, with $\mathbb{E}(l) - \pi_l$. Therefore, π_l is the highest discount the DM is willing to grant.

Example 10.4. If $l = \{c_1, 1/2; c_2, 1/2\}$, and $u(c) = \log c$, then the certainty equivalent of l is

$$c_l = u^{-1}(\ddot{u}(l)) = e^{\frac{1}{2}\log c_1 + \frac{1}{2}\log c_2} = e^{\log\sqrt{c_1 c_2}} = \sqrt{c_1 c_2}.$$

Thus, more generally,

$$\mathbf{c}(l) = \prod_{i=1}^n c_i^{p_i} \quad \text{and} \quad \boldsymbol{\pi}(l) = \sum_{i=1}^n c_i p_i - \prod_{i=1}^n c_i^{p_i}$$

Definition 10.5 (Risk Aversion). Let $\mathbb{E}(l)$ be the expectation of the lottery l, and let $\mathbb{E}(l)$ also indicate the lottery $l = \{\mathbb{E}(l), 1\}$ (the difference should be clear by the context). A preference $\succsim$ on $\mathbf{L}$ is called:

i) Risk averse if $l \precsim \mathbb{E}(l)$, for every $l \in \mathbf{L}$;

ii) Risk loving if $l \succsim \mathbb{E}(l)$, for every $l \in \mathbf{L}$;

iii) Risk neutral if $l \sim \mathbb{E}(l)$, for every $l \in \mathbf{L}$;

iv) Strictly risk averse if $l \prec \mathbb{E}(l)$, for every $l \in \mathbf{L}$;

v) Strictly risk loving if $l \succ \mathbb{E}(l)$, for every $l \in \mathbf{L}$.

Remark 10.6. there may be DMs who do not fit in any of the three categories.

Definition 10.7 (Comparative risk aversion). A preference $\succsim_1$ is more risk averse than a preference $\succsim_2$ if, for all $c \in \mathbf{C}$ and $l \in \mathbf{L}$,

$$l \succsim_1 c \implies l \succsim_2 c \qquad \text{and} \qquad l \succ_1 c \implies l \succ_2 c$$

However, under B.4 and B.5, the first comparison is sufficient and implies the second.

Theorem 10.8. *The following conditions are equivalent (dual versions hold for risk loving preference):*

i) $\succsim$ *is risk averse;*

ii) $\mathbb{E}(l) \succsim l$ *for every* $l \in \mathbf{L}$*;*

iii) $u(\mathbb{E}(l)) \geq \ddot{u}(l)$*, under B.1 - B.3;*

iv) u *is concave, under B.1 - B.3;*[29]

v) $\pi(l)$ *is nonnegative, i.e.* $\pi_l \geq 0$ *for every* $l \in \mathbf{L}$*, under B.1 - B.5;*

vi) $\succsim$ *is more risk averse than a risk neutral preference;*

Proof.
We prove the equivalence step by step.
$(i) \iff (ii)$: it holds by definition of risk aversion.
$(ii) \iff (iii)$: by definition of Bernoulli utility function, $\mathbb{E}(l) \succsim l$ if and only if $\ddot{u}(\mathbb{E}(l)) = u(\mathbb{E}(l)) \geq \ddot{u}(l)$.
$(ii) \iff (v)$: if $\mathbb{E}(l) \succsim l$ for every $l \in \mathbf{L}$, by B.5 and transitivity we have $\mathbb{E}(l) \succsim c_l$. By B.4, we conclude $\mathbb{E}(l) \geq c_l$. The same argument reversed proves the converse.
$(ii) \iff (vi)$: not required for the exam. One argues by contradiction.
$(iv) \implies (iii)$: If u is concave, for every $l \in \mathbf{L}$, by Jensen's inequality

$$u(\mathbb{E}(l)) = u\left(\sum_{i=1}^{k} cl(c)\right) \geq \sum_{i=1}^{k} u(c)l(c) = \ddot{u}(l) \implies u(\mathbb{E}(l)) \geq \ddot{u}(l).$$

$(i) \implies (iv)$: If $\succsim$ is risk averse, consider the lottery $l = \{c, \lambda; c', 1-\lambda\}$, with $c, c' \in \mathbf{C}$ and $\lambda \in (0,1)$. By risk aversion, $\lambda c + \overline{\lambda} c' = \mathbb{E}(l) \succsim l$. By the vN-M Representation Theorem,

$$u(\lambda c + \overline{\lambda} c') = u(\mathbb{E}(l)) \geq \ddot{u}(l) = \lambda u(c) + \overline{\lambda} u(c') \implies u \text{ concave by Jensen.}$$

□

Theorem 10.9. *A monotone vN-M preference is risk neutral if and only if it has a utility* $u(x) = x$*, meaning that it ranks lotteries according to their expected values.*[30]

Proof. If $u(x) = x$, then u is both concave and convex. In particular, $\succsim$ is both risk averse and risk loving, thus risk neutral. Conversely, if $\succsim$ is risk neutral, then u is affine, so that $u(c) = ac + b$. Monotonicity implies $a > 0$, and cardinality allows us to normalize u setting $a = 1$ and $b = 0$. □

[29] assuming B.4, this interpretation is equivalent to say $\{c, 1\} \succsim \{c + e, 1/2; c - e, 1/2\}$, meaning that a risk averse person is averse to any kind of additional risk, intended as a shift from the mean.

[30] Note that for a risk neutral DM, the expected value $\mathbb{E}(l)$ of a lottery l is its certainty equivalent c_l. So, expected values can be regarded as risk neutral certainty equivalents.

Theorem 10.10. *The following conditions are equivalent:*

i) $\succsim_1$ *is more risk averse than* $\succsim_2$*;*

ii) $\succsim_2$ *is more risk loving than* $\succsim_1$*, under B.4;*

iii) $\mathbf{c}_1 \leq \mathbf{c}_2$ *and* $\boldsymbol{\pi}_1 \geq \boldsymbol{\pi}_2$*, under B.4 and B.5;*

iv) There exists a strictly increasing and concave transformation $f : Im\ u_2 \to \mathbb{R}$ *such that* $u_1 = f \circ u_2$*, under B.1 - B.3.*

Proof.
We prove the equivalence step by step.
$(iv) \implies (i)$: Suppose that $u_1 = f \circ u_2$, where f is strictly increasing and concave. We want to show that $l \succsim_1 c$ implies $l \succsim_2 c$ for every $l \in \mathbf{L}$ and $c \in \mathbf{C}$. Showing that $l \succ_1 c$ implies $l \succ_2 c$ requires analogous steps. Passing to the utility functions, we aim to show that

$$\sum_{c \in \mathbf{C}} f(u_2(c))l(c) \geq f(u_2(c)) \implies \sum_{c \in \mathbf{C}} u_2(c)l(c) \geq u_2(c).$$

To do so, we employ this chain of inequalities:

$$f\left(\sum_{c \in \mathbf{C}} u_2(c)l(c)\right) \overset{1}{\geq} \sum_{c \in \mathbf{C}} f(u_2(c))l(c) \overset{2}{\geq} f(u_2(c)) \overset{3}{\implies} \sum_{c \in \mathbf{C}} u_2(c)l(c) \geq u_2(c),$$

where 1 is the Jensen inequality (concavity of f), 2 is the assumption, and 3 is the strict monotonicity of f which allows us to simplify f on both sides.
$(i) \implies (iv)$: Not requested.
$(i) \iff (iii)$: not requested.
$(i) \iff (ii)$: not requested. One should use the equivalence between (i) and (iv) and notice that the inverse of a strictly increasing and convex (concave) function is strictly increasing and concave (convex). From this, the conclusion is trivial. □

Remark 10.11. If the utility functions are twice differentiable on an open interval $\mathbf{C}$ and positive, we don't need to check the concavity of the vN-M utility functions, but rather a special index in terms of the Bernoullian utility functions, known as **Arrow-Pratt index** $\lambda : \mathbf{C} \to \mathbb{R}$, defined by

$$\lambda(c) = -\frac{u''(c)}{u'(c)}.$$

If $\lambda_1 \geq \lambda_2$, then there exists a differentiable and strictly increasing concave function g such that $u_1 = g \circ u_2$, and vice versa. Thus, a higher Arrow-Pratt index implies a higher risk aversion.[31] In particular, $\succsim$ is risk averse if and only if $\lambda \geq 0$.

11 Portfolio Decision Problem

11.1 Two-asset portfolio problem

Imagine a DM who may invest in two possible assets. His investor action is represented by a **portfolio** $a = (a_1, a_2) \in \mathbb{R}^n_+$, where a_i is the amount he invests in the i-th asset. The first asset is a lottery $l_1 = \{r_1, p_1; \dots; r_k, p_k\}$, while the second is a risk-free asset $l_2 = \{r_f, 1\}$. Any possible combination of gross returns, weighted with its probabilities, characterizes the monetary lottery which describes the **gross return** of the whole portfolio:

$$l_a = \{a_1 r_1 + a_2 r_f, p_1; \dots; a_1 r_k + \dots + a_2 r_f, p_k\}.$$

The consequence function $\rho : \mathbf{A} \to \mathbf{L}$ is defined by $\rho(a) = l_a$, where $\mathbf{A}$ is the set of portfolios and $\mathbf{L}$ is the set of lotteries with monetary prize space $\mathbf{C} = \mathbb{R}$.
Suppose he has a wealth $w \geq 0$. The action set depends on it, so that $A(w) = \{a \in \mathbb{R}^n_+ : a_1 + a_2 = w\}$, and $\mathcal{A} = \{A(w) : w \geq 0\}$.

[31] To make this index independent from the monetary unit of account, we should multiply by c, obtaining the relative Arrow-Pratt index. Notice also that it's invariant under positive affine transformations.

The **investor portfolio decision problem** is $(A(w), u)$, and he has to maximize, for a given w, $\max_a u(a) \quad \text{sub } a \in A(w)$, where $u : \mathbf{A} \to \mathbb{R}$ is given by

$$u(a) = (\ddot{u} \circ \rho)(a) = \sum_{i=1}^{k} \dot{u}(a_1 r_i + a_2 r_f) p_i,$$

where $\dot{u} : \mathbb{R} \to \mathbb{R}$ is a vN-M utility function strictly increasing and strictly concave. Thus, we assume our investors **strictly risk averse**. We also assume them being **analytical**, meaning that $\dot{u}$ is twice differentiable, thus $\dot{u}' > 0$ and $\dot{u}'' < 0$. An optimal portfolio $\hat{a} \in A(w)$ is such that $u(\hat{a}) \geq u(a)$ for all $a \in A(w)$.

11.2 Portfolio solution

Let $\alpha \in [0, w]$ be the risky investment. Now the portfolio is $(\alpha, w - \alpha)$. Our optimization problem becomes

$$\max_{\alpha} \sum_{i=1}^{k} \dot{u}(\alpha r_i + (w - \alpha) r_f) p_i = \max_{\alpha} u_w(\alpha) \quad \text{sub } \alpha \in [0, w].$$

This problem has a unique solution. The solution is unique since u is **strictly concave**. Moreover, strict concavity implies continuity, i.e. u_w is continuous. Since $[0, w]$ is a compact interval, the existence of a solution is guaranteed by Weierstrass' Theorem. Thus, we can reasonably define a demand function

$$d(w) = (d_1(w), d_2(w)) = (\hat{a}_w, w - \hat{a}_w) \in \mathbb{R}^2_+,$$

associating the optimal portfolio to each wealth level $w \in \mathbb{R}_+$.

11.3 Portfolio analysis

The demand function is directly related with the risk aversion of the DM. If $\succsim_1$ and $\succsim_2$ are **analytical** investors and $\succsim_1$ is more risk averse than $\succsim_2$, then $d_{11}(w) \leq d_{21}(w)$ for all $w \geq 0$, meaning that the less risk averse investor should invest more in the risky asset at all wealth levels.

Theorem 11.1 (Expected Excess Return). $\overline{\mathbb{E}}(l) = \mathbb{E}(l) - r_f \leq 0$ *if and only if* $d_1(w) = 0$ *for all* $w \geq 0$.

Proof.
Suppose $\hat{\alpha} = 0$ for every $w \geq 0$. This implies that the right derivative of u at $\alpha = 0$ is ≤ 0:

$$\begin{aligned} 0 \geq u'_+(0) &= \left(\sum_{i=1}^{k} \dot{u}(\alpha(r_i - r_f) + w r_f) p_i \right)'_+ \Bigg|_{\alpha=0} = \sum_{i=1}^{k} \dot{u}'_+(\alpha(r_i - r_f) + w r_f) p_i (r_i - r_f) \Bigg|_{\alpha=0} \\ &= \dot{u}'_+(0) \cdot \left(\sum_{i=1}^{k} p_i r_i - \sum_{i=1}^{k} p_i r_f \right) = \dot{u}'(w r_f) \cdot (\mathbb{E}(l) - r_f) = \dot{u}'(w r_f) \overline{\mathbb{E}}(l). \end{aligned}$$

The DM is analytic, thus $\dot{u}'_+ > 0$, implying $\overline{\mathbb{E}}(l) \leq 0$.
Conversely, suppose $\overline{\mathbb{E}}(l) \leq 0$. If $w = 0$, the result is obvious. Fix $w > 0$. We obtain

$$\begin{aligned} u(\alpha) &= \sum_{i=1}^{k} \dot{u}(\alpha r_i + (w - \alpha) r_f) p_i \\ &\overset{1}{<} \dot{u} \left(\sum_{i=1}^{k} (\alpha r_i + (w - \alpha) r_f) p_i \right) = \dot{u} \left(w r_f + \alpha \sum_{i=1}^{k} (r_i p_i - r_f p_i) \right) = \dot{u}(\alpha \overline{\mathbb{E}}(l) + r_f w) \\ &\overset{2}{\leq} \dot{u}(r_f w) = u(0), \end{aligned}$$

where the first inequality if the Jensen Inequality, using the strict concavity of $\dot{u}$, and the second inequality, valid when $\alpha > 0$, uses the strict monotonicity of $\dot{u}$ and the fact that $\overline{\mathbb{E}}(l) \leq 0$. We deduce that $\hat{\alpha} = 0$ is the unique solution of the portfolio optimization problem. □

Remark 11.2. Morally, this means that a strictly risk averse investor invests in the risky asset only when he expects from it a higher return than that of the risk-free asset.

12 Wealth Effects

Here, WA stands for wealth-adjusted. DARA, CARA and IARA stand for decreasing / constant / increasing absolute risk averse preferences.

Definition 12.1 (WA preference). Given a wealth level $w \geq 0$, the wealth-adjusted preference $\succsim_w$ on $\mathbf{L}$ is defined by

$$l \succsim_w l' \iff l^w \succsim l'^w,$$

where $l^w = \{c_1 + w, p_1; \ldots; c_n + w, p_n\}$ is the **wealth-adjusted** lottery. Intuitively, the DM ranks lotteries through the impact that their prizes have on the DM wealth w. In particular, $\succsim_0 = \succsim$.

Definition 12.2 (WA certainty equivalent). We assume the existence of c_l^w such that $c_l^w \sim_w l$. Indicating with c_{l^w} the certainty equivalent of l^w (B.5), we get, by definition of $\succsim_w$ and using B.1 and B.4, $c_l^w + w = c_{l^w}$. To sum up,

$$c_l^w \sim_w l \iff c_l^w := c_{l^w} - w \iff c_l^w + w = c_{l^w}$$

Definition 12.3 (WA risk premium). $\pi_l^w := \mathbb{E}(l) - c_l^w$. We can define a corresponding function $\boldsymbol{\pi}(w, l) = \pi_l^w$.

Theorem 12.4 (WA risk aversion). *$\succsim_w$ is risk averse if and only if $\pi_l^w \geq 0$, under B.1, B.4 and B.5*

Proof. Fix $w \geq 0$. $\succsim_w$ is risk averse when

$$\mathbb{E}(l) \succsim_w l \iff \mathbb{E}(l^w) \succsim l^w \iff \mathbb{E}(l^w) \geq c_{l^w} \iff 0 \leq \mathbb{E}(l^w) - c_{l^w} = \mathbb{E}(l) - c_{l^w} + w = \mathbb{E}(l) - c_l^w = \pi_l^w,$$

where in the second passage we used transitivity and axioms B.4 and B.5. □

Definition 12.5 (Decreasing Risk Aversion). The following conditions are equivalent (and dual conditions hold for increasing absolute risk aversion):

i) If $w \leq w'$, then $\succsim_w$ is more risk averse than $\succsim_{w'}$.

ii) the WA risk premium function $\boldsymbol{\pi}(w, l) = \pi_l^w$ is decreasing in w, i.e. the same lottery requires a lower risk premium as the DM becomes richer and richer, under B.1, B.4 and B.5.

iii) The **Arrow-Pratt** index $\lambda : \mathbf{C} \to \mathbb{R}$ is decreasing, under B.1 - B.5 and assuming u twice continuously differentiable and strictly increasing.

Remark 12.6. a risk lover may be decreasing risk averse, meaning that his risk loving attitude will become stronger as wealth increases.

Remark 12.7. CARA functions have constant Arrow-Pratt index, meaning that they satisfy the differential equation $\dot{u}'' = -\lambda \dot{u}'$. The solution of this differential equation is $\dot{u}(c) = -\frac{k_1}{\lambda} e^{-\lambda c} + k_2$, with $k_1 \geq 0$ and $k_2 \in \mathbb{R}$. Up to affine transformation, they can be written in this unique form:

$$\dot{u}(c) = \begin{cases} -e^{-\lambda c} & \text{if } \lambda > 0 \\ c & \text{if } \lambda = 0 \\ e^{-\lambda c} & \text{if } \lambda < 0 \end{cases}.$$

13 Measure Theory

If not explicitly stated, S is a given set, $\mathbf{\Sigma} \subseteq 2^S$ is an algebra on S, $f : S \to \mathbb{R}$ is a function from S to $\mathbb{R}$.

Definition 13.1 (Sigma Algebra). A sigma algebra $\mathbf{\Sigma}$ is a collection of subsets of S satisfying these three properties:

i) *non-empty*: $S \in \mathbf{\Sigma}$;

ii) *complement closeness*: If $A \in \mathbf{\Sigma}$, then $A^c \in \mathbf{\Sigma}$;

iii) *countable union closeness*: if $A_1, A_2, \dots \in \mathbf{\Sigma}$, then $\displaystyle\bigcup_{i=1}^{\infty} A_i \in \mathbf{\Sigma}$.

Remark 13.2. If we relax the requirement of closure under countable unions to only closure under finite unions ($A, B \in \mathbf{\Sigma}$ imply $A \cup B \in \mathbf{\Sigma}$), we obtain an algebra, which is a more general structure but not suitable for applications in probability theory. In decision theory, considering algebras is sufficient. Note that properties ii and iii combined imply closeness under countable intersections. The elements of $\mathbf{\Sigma}$ are said $\mathbf{\Sigma}$-measurable.

Definition 13.3 (Measurable Function). $f : S \to \mathbb{R}$ is $\mathbf{\Sigma}$-measurable if, for every interval $I \subseteq \mathbb{R}$, then $f^{-1}(I) \in \mathbf{\Sigma}$. In probability theory, $\mathbf{\Sigma}$-measurable functions are called random variables.

Definition 13.4 (Simple Function). $f : S \to \mathbb{R}$ is simple if it is $\mathbf{\Sigma}$-measurable and $\mathrm{Im}(f) = f(S)$ is finite, i.e. it takes finitely many values. Equivalently, f is simple if $\mathrm{Im}(f) = f(S)$ is finite and $f^{-1}(x) \in \mathbf{\Sigma}$ for every $x \in \mathbb{R}$. In decision theory, acts are simple functions.

For our purposes, we consider the following spaces, all of which (except the last one) are normed vector spaces, with the supremum norm defined as $\|f\| := \|f\|_\infty = \sup_{x\in S}|f(x)|$:

$$\boldsymbol{B}_0(\mathbf{\Sigma}) \quad \subseteq \quad \boldsymbol{B}(\mathbf{\Sigma}) \quad \subseteq \quad \boldsymbol{B}(S) \quad \subset \mathbb{R}^S$$

- The space $\boldsymbol{B}(S) \subset \mathbb{R}^S$ is the space of bounded functions $f : S \to \mathbb{R}$
  - the pair $(\boldsymbol{B}(S), \|\cdot\|)$ is a Banach vector space, i.e. a complete normed vector space.
- The space $\boldsymbol{B}_0(\mathbf{\Sigma}) \subset \boldsymbol{B}(S)$ is the space of simple functions.
  - There exists a canonical representation of $f \in \boldsymbol{B}_0(\mathbf{\Sigma})$, as $f = \sum_{i=1}^n a_i \mathbf{1}_{E_i}$, where $a_1 < \cdots < a_n \in \mathbb{R}$ and $\{E_1, \ldots, E_n\}$ is a partition of measurable subsets of S.[32]
- The space $\boldsymbol{B}(\mathbf{\Sigma}) := \overline{\boldsymbol{B}_0(\mathbf{\Sigma})}$ is the closure of $\boldsymbol{B}_0(\mathbf{\Sigma})$ in the Banach space of bounded functions $\boldsymbol{B}(S)$, i.e. the space of bounded functions which can be written as the limit of a sequence of simple functions. More explicitly,

$$\boldsymbol{B}(\mathbf{\Sigma}) := \left\{ \lim_{n\to+\infty} f_n : \ \forall n, f_n \text{ is simple} \right\} = \left\{ f \in \boldsymbol{B}(S) : \exists \{f_n\}_{n\in\mathbb{N}}, f_n \in \boldsymbol{B}_0(\mathbf{\Sigma}) | \lim_{n\to+\infty} \sup_{x\in S} |f_n(x) - f(x)| = 0 \right\}$$

Remark 13.5. If $\mathbf{\Sigma}$ is an algebra, then every bounded and measurable function is the limit of a sequence of simple functions. Moreover, if $\mathbf{\Sigma}$ is a σ-algebra, then every function which is the limit of a sequence of simple functions is bounded and measurable. In other words, $\{f \in \boldsymbol{S}, f \text{ measurable}\} \subseteq \boldsymbol{B}(\mathbf{\Sigma})$, and $\{f \in \boldsymbol{S}, f \text{ measurable}\} = \boldsymbol{B}(\mathbf{\Sigma})$ if $\mathbf{\Sigma}$ is a σ-algebra.

Definition 13.6 (Charge). A charge is a finitely additive function $\mu : \mathbf{\Sigma} \to \mathbb{R}$, i.e. if $A \cap B = \emptyset$ and $A, B \in \mathbf{\Sigma}$, then $\mu(A \cup B) = \mu(A) + \mu(B)$. We define $\mathrm{ba}(\mathbf{\Sigma})$ as the space of bounded charges, i.e.

$$\mathrm{ba}(\mathbf{\Sigma}) := \left\{ f \in \mathbb{R}^{\mathbf{\Sigma}} : \|f\| = \sup_{A\in\mathbf{\Sigma}} |f(A)| < \infty \quad \text{and} \quad f(A \cup B) = f(A) + f(B) \ \forall A, B \in \mathbf{\Sigma}, A \cap B = \emptyset \right\}.$$

Definition 13.7 (Integral of a Simple Function). We define the *integral* of a function $\varphi \in \boldsymbol{B}_0(\mathbf{\Sigma})$ with respect to a charge $\mu : \mathbf{\Sigma} \to \mathbb{R}$ as the operator $T_\mu : \boldsymbol{B}_0(\mathbf{\Sigma}) \to \mathbb{R}$ defined as

$$T_\mu(\varphi) = \int_S \varphi \mathrm{d}\mu = \sum_{i=1}^n \alpha_i \mu(A_i),$$

where $\varphi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ is a representation of φ.

Theorem 13.8. *Let $T_\mu : \boldsymbol{B}_0(\mathbf{\Sigma}) \to \mathbb{R}$ be defined as before. Then:*

i) T_μ is well defined, that is, it's independent of the representation of φ chosen.

ii) T_μ is linear.

iii) If $\mu \in ba(\mathbf{\Sigma})$, then T_μ is continuous (in fact, Lipschitz Continuous).

iv) If $\mu \geq 0$, then T_μ is monotone.

Proof.
(i) Take two different representations of φ, namely $\varphi = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} = \sum_{j=1}^k \beta_j \mathbf{1}_{B_j}$. for simplicity, here

[32] In fact, any simple function admits infinitely many representations. For instance, $f(x) = 2\mathbf{1}_{\mathbb{Q}}(x) + \mathbf{1}_{\mathbb{R}\setminus\mathbb{Q}}(x) = \mathbf{1}_{\mathbb{R}}(x) + \mathbf{1}_{\mathbb{Q}}(x)$.

we will assume $\{A_1, \dots, A_n\}$ and $\{B_1, \dots, B_j\}$ be measurable partitions of S. If not, the proof is the same but more involved with notations. In this case,

$$\begin{aligned}\sum_{i=1}^{n} \alpha_i \mu(A_i) &= \sum_{i=1}^{n} \alpha_i \mu\left(\bigcup_{j=1}^{k} A_i \cap B_j\right) \overset{1}{=} \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{k} \mu(A_i \cap B_j)\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{k} \alpha_i \mu(A_i \cap B_j) \overset{2}{=} \sum_{i=1}^{n}\sum_{j=1}^{k} \beta_j \mu(A_i \cap B_j) \\ &= \sum_{j=1}^{k} \beta_j \left(\sum_{i=1}^{n} \mu(A_i \cap B_j)\right) \overset{1}{=} \sum_{j=1}^{k} \beta_j \mu\left(\bigcup_{i=1}^{n} A_i \cap B_j\right) = \sum_{j=1}^{k} \beta_j \mu(B_j),\end{aligned}$$

where 1 uses the fact that B_j's are disjoint and the additivity of μ and 2 uses the fact that, for every s, $\sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i} = \varphi(s) = \sum_{j=1}^{k} \beta_j \mathbf{1}_{B_j}$, and, consequently, $\alpha_i = \beta_j$ on $A_i \cap B_j$.
(ii) Take a representation $\varphi = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$ and a representation $\psi = \sum_{j=1}^{k} \beta_j \mathbf{1}_{B_j}$. Then $\alpha\left(\sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}\right) + \beta\left(\sum_{j=1}^{k} \beta_j \mathbf{1}_{B_j}\right)$ is a representation of $\alpha\varphi + \beta\psi$. Since T is independent from the representation chosen, then

$$T_\mu(\alpha\varphi + \beta\psi) = T_\mu\left(\sum_{i=1}^{n}\sum_{j=1}^{k} (\alpha\alpha_i + \beta\beta_i)\mathbf{1}_{A_i \cap B_j}\right) = \alpha \sum_{i=1}^{n} \alpha_i \mu(A_i) + \beta \sum_{j=1}^{k} \beta_j \mu(B_j) = \alpha T_\mu(\varphi) + \beta T_\mu(\psi).$$

(iii) We show that T_μ is Lipschitz continuous. Applying linearity of T_μ,

$$|T_\mu(\varphi) - T_\mu(\psi)| = |T_\mu(\varphi - \psi)| \overset{?}{\leq} C\|\varphi - \psi\|.$$

Set $\zeta := \varphi - \psi$, and let $\zeta = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$ be the canonical representation of ζ. Then

$$|T_\mu(\zeta)| = \left|\sum_{i=1}^{n} \alpha_i \mu(A_i)\right| \overset{1}{\leq} \sum_{i=1}^{n} |\alpha_i||\mu(A_i)| \overset{2}{\leq} \sum_{i=1}^{n} \|\zeta\||\mu(A_i)| = \|\zeta\| \sum_{i=1}^{n} |\mu(A_i)| \overset{3}{\leq} \|\zeta\| \underbrace{2\|\mu\|}_{C},$$

where 1 is the triangular inequality, 2 is the definition of $\|\zeta\| = \|\zeta\|_\infty$, and 3 is a nontrivial lemma.[a]
(iv) Suppose $\varphi \leq \psi$. We want to show that $0 \leq T_\mu(\varphi) - T_\mu(\psi) = T_\mu(\varphi - \psi)$, using linearity. Take the canonical representation of $T_\mu(\varphi - \psi) = \sum_{i=1}^{n} \alpha_i \mu(A_i)$. $\mu(A_i) \geq 0$ by hypothesis, and $\alpha_i \geq 0$ since $\varphi \geq \psi$. □

[a] Split $\sum \mu(A_i)$ in $P + N$, where P is the sum of the positive terms and N the sum of the negative terms. Note that $P = \mu(A)$ and $N = \mu(B)$ for some $A, B \in \mathbf{\Sigma}$. Then $\sum |\mu(A_i)| = P - N = \mu(A) - \mu(B) \leq \|\mu\| + \|\mu\|$.

Remark 13.9. Up to now we have shown that the integral operator is a linear and continuous function for every $\mu \in \mathrm{ba}(\mathbf{\Sigma})$. In linear algebra, the dual V^* of a vector space V is the space of linear functions $f : V \to \mathbb{R}$. The topological dual V' of a vector space V is the space of linear and continuous functions $f : V \to \mathbb{R}$. So far we have shown that $\{T_\mu : \mu \in \mathrm{ba}(\mathbf{\Sigma})\} \subseteq \boldsymbol{B}_0(\mathbf{\Sigma})'$. Now we want to define an extension of T_μ on $\boldsymbol{B}(\mathbf{\Sigma})$, that is, we want to define the integral not only for simple functions, but also for functions that are the limit of simple functions. We call this extension Stieltjes Integral, denoted $\tilde{T}_\mu$. We will show that such extension is well defined, linear and continuous. This would suffice to prove that $\{\tilde{T}_\mu : \mu \in \mathrm{ba}(\mathbf{\Sigma})\} \subseteq \boldsymbol{B}(\mathbf{\Sigma})'$. The last big question is whether the opposite holds, that is, whether any linear and continuous function $f : \boldsymbol{B}(\mathbf{\Sigma})$ is $\tilde{T}_\mu$ for some μ (i.e. is the integral operator with respect to some charge). The answer is affirmative.

For the next definition, we first recall the notion of convergence on the space of simple functions: let $\{\varphi_n\}_{n\in\mathbb{N}}$ be a sequence of simple functions. Then

$$f = \lim_{n\to+\infty} \varphi_n \iff \forall\varepsilon > 0, \exists N_\varepsilon \in \mathbb{N} : \forall n \geq N_\varepsilon : \sup_{x\in S} |\varphi_n(x) - \varphi(x)| < \varepsilon.$$

Definition 13.10 (Stieltjes Integral). Let $\mu \in \mathrm{ba}(\mathbf{\Sigma})$ and f be the limit of a sequence of simple functions $\{\varphi_n\}_{n\in\mathbb{N}}$. The Stieltjes integral of f is the operator $\tilde{T}_\mu : \boldsymbol{B}(\mathbf{\Sigma}) \to \mathbb{R}$ defined as

$$\tilde{T}_\mu(f) = \lim_{n\to+\infty} T_\mu(\varphi_n) = \lim_{n\to+\infty} \int_S \varphi_n \mathrm{d}\mu.$$

What needs to be proven is that this operator is well defined, that is, it doesn't depend on the sequence $\{\varphi_n\}_{n\in\mathbb{N}}$, linear and continuous.

Proof.
We will provide only a sketch of the proof.
$(\star)$ $\varphi_n \to f$ can be equivalently expressed by stating the existence of a decreasing sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ such that $\lim\limits_{n\to\infty} \varepsilon_n = 0$ and $|f - \varphi_n| \leq \varepsilon_n$.
$(\star\star)$ We define the lower and upper integrals as

$$I_*(f) := \sup\left\{\int_S \varphi \mathrm{d}\mu : \varphi \leq f, \varphi \in \boldsymbol{B}_0(\boldsymbol{\Sigma})\right\}, \quad \text{and} \quad I^*(f) := \inf\left\{\int_S \varphi \mathrm{d}\mu : \varphi \geq f, \varphi \in \boldsymbol{B}_0(\boldsymbol{\Sigma})\right\},$$

and we prove that $I_*(f) = I^*(f) \in \mathbb{R}$, using $(\star)$:

$$\int (\varphi_n - \varepsilon_n)\mathrm{d}\mu \leq I_*(f) \overset{?}{\leq} I^*(f) \leq \int (\varphi_n + \varepsilon_n)\mathrm{d}\mu \quad \text{and}$$

$$\left|\int (\varphi_n - \varepsilon_n)\mathrm{d}\mu - \int (\varphi_n + \varepsilon_n)\mathrm{d}\mu\right| = \left|\int -2\varepsilon_n \mathrm{d}\mu\right| = |2\varepsilon_n \mu(S)| \to 0.$$

Since they coincide, they are finite and $I_*(f) \leq \tilde{T}_\mu(f) \leq I^*(f)$, then $I_*(f) = I = I^*(f) = \tilde{T}_\mu(f)$, that is, $\tilde{T}_\mu(f)$ does not depend on $\{\varphi_n\}_{n\in\mathbb{N}}$, and it's a real number ($\neq \pm\infty$).
$(\star\star\star)$ The linearity of $\tilde{T}_\mu(f)$ follows from the linearity of the limit operator. As for the continuity, we need to use $\mathrm{ba}(\boldsymbol{\Sigma})$. Since $\tilde{T}_\mu$ is linear, it suffices to prove $|\tilde{T}_\mu(f)| \leq 2\|\mu\|\|f\|$, as we did for T_μ.

$$\begin{aligned}|\tilde{T}_\mu(f)| &= |\int_S f \mathrm{d}\mu| = |\int : S \lim \varphi_n \mathrm{d}\mu| = |\lim \int_S \varphi \mathrm{d}\mu| = \lim |\int_S \varphi_n \mathrm{d}\mu| \\ &\leq \lim 2\|\mu\|\|\varphi_n\| = 2\|\mu\| \lim \|\varphi_n\| = 2\|\mu\|\| \lim \varphi_n\| = 2\|\mu\|\|f\|\end{aligned}$$

From the continuity of $|\cdot|$ and $\|\cdot\|$. □

Theorem 13.11 (Riesz Representation Theorem). *$T \in \boldsymbol{B}(\boldsymbol{\Sigma})'$ if and only if there exists $\mu \in ba(\boldsymbol{\Sigma})$ such that $T = \tilde{T}_\mu$. In such case, μ is unique.*

Proof.
We already know that the Stieltjes integral is linear and continuous for every $\mu \in \mathrm{ba}(\boldsymbol{\Sigma})$. We need to prove that for every linear and continuous operator $T : \boldsymbol{B}(\boldsymbol{\Sigma}) \to \mathbb{R}$ there exists $\mu = \mu_T$ such that $T = \tilde{T}_{\mu_T}$, and that such μ is unique.
Let $\mu = \mu_T : \boldsymbol{\Sigma} \to \mathbb{R}$ such that $A \mapsto T(\mathbf{1}_A)$. μ is a bounded charge: it's additive since, if $A \cap B = \emptyset$, then $T(\mathbf{1}_{A\cup B}) = T(\mathbf{1}_A + \mathbf{1}_B) = T(\mathbf{1}_A) + T(\mathbf{1}_B)$, from the linearity of T; it's bounded since

$$\|\mu\| = \sup_{A\in\boldsymbol{\Sigma}} |\mu(A)| = \sup_{A\in\boldsymbol{\Sigma}} |T(\mathbf{1}_A)| \overset{!}{\leq} \sup_{A\in\boldsymbol{\Sigma}} C\|\mathbf{1}_A| \leq C \sup_{A\in\boldsymbol{\Sigma}} \|\mathbf{1}_A| \leq C,$$

where C is the Lipschitz constant and ! is the Lipschitz property of T (a linear function is continuous if and only if it's Lipschitz continuous).
We prove that $T = \tilde{T}_\mu$ starting from indicator functions $\mathbf{1}_A$, then simple functions $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ and finally limits of simple functions $f = \lim\limits_{n\to\infty} \varphi_n$.

$$T(\mathbf{1}_A) = \mu(A) = \mu_T(A) = \int_S \mathbf{1}_A \mathrm{d}\mu_T = T_{\mu_T}(\mathbf{1}_A) = \tilde{T}_{\mu_T}(\mathbf{1}_A)$$

$$\begin{aligned}T(f) = T\left(\sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}\right) &= \sum_{i=1}^n \alpha_i T(\mathbf{1}_{A_i}) = \sum_{i=1}^n \alpha_i \mu_T(A_i) = \sum_{i=1}^n \alpha_i \tilde{T}_{\mu_T}(\mathbf{1}_{A_i}) \\ &= \sum_{i=1}^n \alpha_i \left(\int_S \mathbf{1}_{A_i} \mathrm{d}\mu_T\right) = \sum_{i=1}^n \alpha_i \mu(A_i) = T_\mu(f) = \tilde{T}_{\mu_T}(f).\end{aligned}$$

$$T(f) = T\left(\lim_{n\to+\infty} \varphi_n\right) \overset{!}{=} \lim_{n\to+\infty} T(\varphi_n) = \lim_{n\to+\infty} \tilde{T}_{\mu_T}(\varphi_n) \overset{!}{=} \tilde{T}_{\mu_T}\left(\lim_{n\to+\infty} \varphi_n\right) = \tilde{T}_{\mu_T}(f),$$

where ! is the definition of continuity. Now assume by contradiction the existence of μ_1 and μ_2 such that $T = \tilde{T}_{\mu_1} = \tilde{T}_{\mu_2}$. Then, for every $A \in \boldsymbol{\Sigma}$, $\mu_1(A) = \int_S \mathbf{1}_A \mathrm{d}\mu_1 = \tilde{T}_{\mu_1}(\mathbf{1}_A) = \tilde{T}_{\mu_2}(\mathbf{1}_A) = \int_S \mathbf{1}_A \mathrm{d}\mu_2 = \mu_2(A)$, proving the uniqueness of the charge. $\square$

Remark 13.12. for every linear and continuous operator $T : \boldsymbol{B}(\boldsymbol{\Sigma}) \to \mathbb{R}$ there exists a unique $\mu \in \mathrm{ba}(\boldsymbol{\Sigma})$ such that $T = \tilde{T}_\mu$. On the other hand, for every $\mu \in \mathrm{ba}(\boldsymbol{\Sigma})$, the Stieltjes integral $\tilde{T}_\mu : \boldsymbol{B}(\boldsymbol{\Sigma}) \to \mathbb{R}$ is linear and continuous.

Remark 13.13. This theorem guarantees the existence and uniqueness of a (bounded additive) measure μ such that a linear and continuous operator T can be written as the integral of its argument with respect to μ. To prove that $\{\tilde{T}_\mu : \mu \in \mathrm{ba}(\boldsymbol{\Sigma})\} = \boldsymbol{B}(\boldsymbol{\Sigma})'$, it was sufficient to show the existence part of this theorem. The uniqueness part guarantees that $\boldsymbol{B}(\boldsymbol{\Sigma})' \simeq \mathrm{ba}(\boldsymbol{\Sigma})$, that is, there is a linear bijection that associates to every bounded charge a linear and continuous operator $\tilde{T}_\mu : \boldsymbol{B}(\boldsymbol{\Sigma}) \to \mathbb{R}$. It turns out that this bijection is also continuous and preserves the norm, that is, $\|\mu\| = \|\tilde{T}_\mu\|$.

Corollary 13.14. *With the same assumptions of the previous theorem, T is also monotone if and only if μ is also nonnegative.*

Proof.

If T is monotone, fixing $A \in \boldsymbol{\Sigma}$, then $\mu(A) = \int_S \mathbf{1}_A \mathrm{d}\mu = \tilde{T}_\mu(\mathbf{1}_A) = T(\mathbf{1}_A) \geq T(0) = 0$, from the linearity of T. If $\mu \geq 0$, then T_μ is monotone in $\boldsymbol{B}_0(\boldsymbol{\Sigma})$. We must extend this property to $\boldsymbol{B}(\boldsymbol{\Sigma})$. Recall that $\int f \mathrm{d}\mu = I_*(f) = I^*(f)$. Therefore, if $f_1 \leq f_2$,

$$\tilde{T}_\mu(f_1) = \int_S f_1 \mathrm{d}\mu = I_*(f_1) = \sup\{T_\mu(\varphi) : \varphi \leq f_1\} \overset{!}{\leq} \sup\{T_\mu(\varphi) : \varphi \leq f_2\} = I_*(f_2) = \tilde{T}_\mu(f_2),$$

where $\varphi \in \boldsymbol{B}_0(\boldsymbol{\Sigma})$, and ! holds since the set on the right is bigger, hence its sup is bigger. $\square$

Corollary 13.15. *$T : \mathbb{R}^n \to \mathbb{R}$ then T is linear if and only if there exists $a \in \mathbb{R}^n$ such that $T(x) = a \cdot x$. In addition, a is unique.*

Proof.

We rely on the following nontrivial fact: every linear function from $\mathbb{R}^n$ to $\mathbb{R}$ is continuous, meaning that $(\mathbb{R}^n)^* = (\mathbb{R}^n)'$. Take $S = \{1, 2, \ldots, n\}$ and $\boldsymbol{\Sigma}$ being the power set 2^S. Now

$$\boldsymbol{B}_0(\boldsymbol{\Sigma}) = \{f : S \to \mathbb{R} \text{ simple}\} = \boldsymbol{B}(\boldsymbol{\Sigma}) = \mathbb{R}^S \simeq \mathbb{R}^n \implies (\mathbb{R}^n)^* = (\mathbb{R}^n)' \simeq \boldsymbol{B}(\boldsymbol{\Sigma})' = \{\tilde{T}_\mu : \mu \in \mathrm{ba}(\boldsymbol{\Sigma})\},$$

where the last passage is the Riesz Representation Theorem. Therefore, $T : \mathbb{R}^n \to \mathbb{R}$ is linear if and only it is isomorphically equivalent to an operator $\tilde{T}_\mu$ such that

$$\tilde{T}_\mu(f) = \int_S f \mathrm{d}\mu = \sum_{i=1}^{n} f(i)\mu(\{i\}) = \underbrace{(f(1), f(2), \ldots, f(n))}_{x \in \mathbb{R}^n} \cdot \underbrace{(\mu(\{1\}), \ldots, \mu(\{n\}))}_{a \in \mathbb{R}^n} = a \cdot x$$

$\square$

Definition 13.16. Given a set S and a function $P : \boldsymbol{\Sigma} \to [0, 1]$, such that $P(\emptyset) = 1 - P(S) = 0$, consider the following three properties:

i) **Finite Additivity:** for every disjoint $A, B \in \boldsymbol{\Sigma}$, $P(A \cup B) = P(A) + P(B)$.[33]

ii) **Countable Additivity** (or σ-additivity): for any disjoint sequence of events $A_1, \ldots,$ such that $\bigcup_{i=1}^{\infty} A_i \in \boldsymbol{\Sigma}$, we have $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

iii) **Finite Carrier** (or finite support): There exists a (smallest) finite set $\mathrm{supp}\, P \in \boldsymbol{\Sigma}$ such that $P(\mathrm{supp}\, P) = 1$.[34]

[33] As a trivial consequence, $A \subseteq B$ implies $P(A) \leq P(B)$.

[34] As a trivial consequence, the probability of any set in $\boldsymbol{\Sigma}$ can be determined by simply adding the probabilities of the singletons that form the support.

Then, the following taxonomy holds:

- P is a **probability** if it satisfies property i, and the set of such functions is $\Delta(S)$. Some authors refer to those as *finitely additive probability measures*, *normalized positive charges* or simply *probability charges*. These are typically used when defined on an algebra $\mathbf{\Sigma}$ that is not σ-additive, such as in the Anscombe-Aumann setting.
- P is a **probability measure** if it satisfies property ii, and the set of such functions is $\Delta^\sigma(S)$. Some authors refer to those as *σ-additive probability measures*.
- P is a **simple probability measure** if it satisfies properties ii and iii, and the set of such functions is $\Delta_0(S)$. Some authors refer to those as *simple probabilities*.

Remark 13.17. It always holds $\Delta_0(S) \subseteq \Delta^\sigma(S)$, and it holds $\Delta^\sigma(S) \subseteq \Delta(S)$ when $\mathbf{\Sigma}$ contains every singleton of S, typically a mild assumption. Moreover, by definition, $\Delta(S) \subseteq [0,1]^{\mathbf{\Sigma}} \subseteq \mathrm{ba}(\mathbf{\Sigma})$. In particular, the integral with respect to a probability of a simple function is the particular case of T_μ when $\mu \in \Delta(S)$. For functions which are not simple, we should rather use $\tilde{T}_\mu$; however, in the specific case in which $p \in \Delta_0(S)$, then

$$\int_S f \mathrm{d}p = \mathbb{E}_p\, f = \sum_{s \in \operatorname{supp}\, p} f(s)p(s).$$

Remark 13.18. For our purposes, we assume that S is countable, $\mathbf{\Sigma}$ includes every singleton in 2^S, and lotteries are treated as simple probability measures on $\mathbf{C}$, forming the set $\Delta_0(\mathbf{C})$.

14 Decisions Under Uncertainty

The following chapters employ a mathematical formalism and definitions that are introduced in Section 13. If any notions appear unclear, please refer to that section for clarification.

14.1 State Uncertainty

In this section, we consider a DM who must choose among a set of alternative actions in an action set $\mathbf{A}$, whose consequences c, within a consequence space $\mathbf{C}$, depend on some state s beyond his control, where s belongs to a state space S endowed with an algebra $\mathbf{\Sigma}$ of subsets of S, called *events*. For instance, suppose the DM has to decide which means of transportation to use to go to work; the consequence, namely the time spent to reach the workplace, depends on the weather, which is not under his control. Among the possible states, one and only one is the *true state*, but the DM does not know which one it is. However, we assume that the DM is aware of the entire state space, abstracting from unforeseen contingencies. Finally, we partition the action space $\mathbf{A}$ into a collection $\mathcal{A}$ of subsets A of $\mathbf{A}$, within which the DM may have to choose.

Definition 14.1 (Repercussion Function). $\gamma : \mathbf{A} \times S \to \mathbf{C}$ associates to each action and state a consequence.

Definition 14.2 (Act). A (simple) act is a simple $\mathbf{\Sigma}$-measurable map $f : S \to \mathbf{C}$. We denote the set of all (simple) acts interchangeably as $\mathbf{F}_0$ or $\boldsymbol{B}_0(\mathbf{\Sigma})$.

To illustrate, consider partitioning the state space S into subsets $\{E_1, \dots, E_n\}$, corresponding to mutually disjoint events in $\mathcal{A}$. Let c_i be the consequence that the DM would receive if E_i occurs, or more formally, if the *true state* s belongs to E_i. In this case, the act f can be expressed as:

$$f(s) = \sum_{i=1}^n c_i \mathbf{1}_{E_i}(s),$$

where $\mathbf{1}_{E_i}$ is the indicator function of event E_i. As in risk theory, we identify a consequence $c \in \mathbf{C}$ with the constant act that delivers c in every state. Accordingly, we embed $\mathbf{C}$ into $\mathbf{F}_0$ by $\mathbf{C} \subseteq \mathbf{F}_0$.

Definition 14.3 (Consequence Function). If we fix an action a, the repercussion function becomes an act $\gamma_a : S \to \mathbf{C}$ which describes the consequence of that action in various states $s \in S$. Therefore, the consequence function $\rho : \mathbf{A} \to \mathbf{F}_0$ of a is the act obtained by fixing a, representing an uncertain outcome that depends on s:

$$\rho(a) = \gamma_a.$$

To get an outcome, one should specify which state we are considering:

$$\rho(a)(s) = \gamma_a(s) = c_s.$$

We assume ρ to be surjective, meaning *inter alia* that the consequences of any action are finitely many.

Preferences: We have preferences over actions, consequences and acts.

i) Once endowed $\mathbf{C}$ in $\mathbf{F}_0$, the preference over consequences is simply a restriction of the preference over acts.

ii) We assume **state consequentialism**: $\rho(a) = \rho(b) \implies a \sim b$. By doing so, ranking actions corresponds to ranking the acts that they determine:

$$a \succsim b \iff \rho(a) \,\dot{\succsim}\, \rho(b).$$

iii) The decision framework and decision environment are, respectively,

$$(\mathbf{A}, \mathcal{A}, S, \mathbf{C}, \gamma) \qquad \text{and} \qquad (\mathbf{A}, \mathcal{A}, S, \mathbf{C}, \gamma, \succsim, \dot{\succsim}).$$

iv) Assuming there exists a utility representation $\dot{u}$ for preferences over acts, then $u = \dot{u} \circ \rho$ is the corresponding utility representation for preferences over actions. The extended form of the decision environment and decision problem under uncertainty in utility form are, respectively,

$$(\mathbf{A}, \mathcal{A}, \dot{u} \circ \rho) \qquad \text{and} \qquad (A, \dot{u} \circ \rho).$$

v) In reduced form, we identify each action $a \in \rho^{-1}(f)$ with the act f delivered. The dual reduced forms of the latter structures are

$$(\mathbf{F}_0, \mathcal{F}, \dot{u}) \qquad \text{and} \qquad (F, \dot{u}),$$

where $\mathcal{F} = \{\rho(A) : A \in \mathcal{A}\}$.

15 Anscombe - Aumann

In this section and the next, we will examine two fundamental results: the Anscombe-Aumann theorem, with its Subjective Expectation (SEU), and the Gilboa-Schmeidler theorem, with its Robust Expectation (MEU). While the second theorem introduces an additional layer of uncertainty, formally defined as ambiguity, both theorems share a similar structure, so that they should be studied together. Indeed, both theorems rely on a key lemma, here labelled as Chisini Lemma, central in both proofs, which is based on the concept of Chisini Mean. Although abstract, this result will play a pivotal role in our analysis.

15.1 Anscombe-Aumann Setting

Anscombe-Aumann (AA) uncertainty adds an additional layer of uncertainty by assuming that acts do not output consequences in $\mathbf{C}$ but lotteries in $\Delta_0(\mathbf{C}) = \Delta_0$. This means that consequences themselves are probabilistic. From a mathematical perspective, the AA model is a special case of the previous (detailed by Savage), where consequences are explicitly modeled as lotteries. By contrast, Savage's framework imposes no specific assumptions on the nature of the consequence space $\mathbf{C}$, leaving it general enough to potentially include lotteries. The repercussion function becomes *stochastic*, associating to each pair (a, s), the lottery $l(a, s)$, describing the probability distribution of getting various possible consequences:

$$\gamma : \mathbf{A} \times S \to \Delta_0,$$

where $l(a, s)(c) = \gamma(a, s)(c)$ is the probability of obtaining consequence c when action a is taken in state s. Alternatively, to emphasize that the state s is given and beyond the decision maker's control, the notation $l_a(c \mid s)$ is sometimes used.[35] An Anscombe-Aumann decision problem under uncertainty is therefore defined by the quartet $(\mathbf{A}, S, \mathbf{C}, \gamma)$.

Definition 15.1 (Anscombe-Aumann Act). It's a simple $\mathbf{\Sigma}$-measurable map $f : S \to \Delta_0$, where $\mathbf{\Sigma}$ is an algebra. This map associates a lottery to each state $s \in S$. We consider only acts that are simple and $\mathbf{\Sigma}$-measurable, meaning that for every act f, there exists a collection of disjoint events $\{E_1, \ldots, E_n\} \subseteq \mathcal{A}$ that partitions S, such that f can be expressed as:

$$f(s) = \sum_{i=1}^{n} l_i \mathbf{1}_{E_i}(s) \implies f(s)(c) = \sum_{i=1}^{n} l_i(c) \mathbf{1}_{E_i}(s), \tag{15.1}$$

where the first expression denotes a lottery which depends on s, and the second expression specifies the probability of obtaining a particular consequence c in the act f, given the state s.

[35] Note that, if we knew the probability of being in state s for each $s \in S$, then we could compute the probability of getting c with the action a by simply summing the probabilities in each s: $\sum_{s \in S} \gamma(a, s)(c)$. However, such distribution is not known.

Definition 15.2 (Act Consequentialism). If two actions induce the same distribution of consequences in every state, then they are indifferent. Therefore

$$\gamma(a,s) = \gamma(b,s) \quad \forall s \in S \implies a \sim b.$$

Remark 15.3. Thanks to act consequentialism, we no longer need to take decisions based on actions $a \in \mathbf{A}$, but rather based on the acts $\gamma_a = f \in S$ that they determine through ρ. Moreover, since $l \in \Delta_0$ can be identified with the constant act $f(s) = l \ \forall s \in S$, the primitive preference is on acts $f \succsim g$.

Remark 15.4 (Abstract Setting). To generalize our framework, we will consider a general convex subset X of a vector space V, rather than restricting to the prototypical example $X = \Delta_0(\mathbf{C})$ of lotteries (if it bothers you, you may substitute X with $\Delta_0(\mathbf{C})$). Here, We denote the set of all (simple) acts from S to X (or Δ_0) interchangeably as

$$\mathbf{F}_0 = \boldsymbol{B}_0(\boldsymbol{\Sigma}) = \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, \Delta_0) = \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, X) = \boldsymbol{B}_0(X).$$

We will employ the notation $\boldsymbol{B}_0(X)$ to differentiate it from $\boldsymbol{B}_0(u(X))$, which denotes the set of acts from S to $u(X) = \text{Im } u$, where $u : X \to \mathbb{R}$.
By identifying the consequence $x \in X$ with the constant act $x\mathbf{1}_S(s)$, it follows that $X \subseteq \mathbf{F}_0$. Consequently, the primitive preference $\succsim$ on acts subsumes the risk preference $\ddot{\succsim}$ on lotteries.

Proposition 15.5. *If X is a convex subset of a vector space V, then $\mathbf{F}_0$ is a convex subset of V^S.*

Remark 15.6. As a consequence, acts such as $\frac{1}{2}f + \frac{1}{2}g$ are simple acts. This particular act can be interpreted as the DM choosing between the acts f and g by tossing a fair coin.

Proof. Taking $f, g \in \mathbf{F}_0$ and $\alpha \in [0,1]$, for every $s \in S$:

$$(\alpha f + \overline{\alpha} g)(s) = \alpha \underbrace{f(s)}_{\in X} + \overline{\alpha} \underbrace{g(s)}_{\in X} \in X \implies \alpha f + \overline{\alpha} g \in X.$$

The proof that $\alpha f + \overline{\alpha} g$ is simple and measurable, i.e. that there exists a finite partition $\{E_1, \ldots, E_n\}$ of events for which $\alpha f + \overline{\alpha} g = \sum l_i \mathbf{1}_{E_i}$, is left as an exercise. $\square$

15.2 Abstract Anscombe-Aumann

Here we consider the same axioms for the Abstract von Neumann Morgenstern Theorem, specifically A.1, A.2, A.4, A.8, and we introduce an additional one:

(A.16) Monotonicity For any $f, g \in \mathbf{F}_0$, if $f(s) \ddot{\succsim} g(s)$ for every $s \in S$, then $f \dot{\succsim} g$.

Comment: Monotonicity states that an act yielding better lotteries in every state is always preferred. For example, a medical treatment that offers higher chances of survival, regardless of the specific pathology, is better.

Theorem 15.7 (Abstract Anscombe-Aumann). *The binary relation $\succsim$ on $\boldsymbol{B}_0(S, \boldsymbol{\Sigma}, X) = \mathbf{F}_0$ satisfies axioms A.1, A.2, A.4, A.8, A.16 if and only if there exists an affine function $u : X \to \mathbb{R}$ and a probability $P : \boldsymbol{\Sigma} \to [0,1]$ such that*

$$U(f) = \int_S u(f(s)) \mathrm{d}P(s)$$

represents $\succsim$. In this case, u, and so U, is cardinally unique, and P is unique unless $f \sim g$ for every $f, g \in \mathbf{F}_0$.[36]

Theorem 15.8 (Anscombe-Aumann). *The binary relation $\succsim$ on $\boldsymbol{B}_0(S, \boldsymbol{\Sigma}, \Delta_0)$ satisfies axioms A.1, A.2, A.4, A.8, A.16 if and only if there exists a function $\dot{u} : \mathbf{C} \to \mathbb{R}$ and a probability measure $P : \boldsymbol{\Sigma} \to [0,1]$ such that the function $u : \mathbf{F}_0 \to \mathbb{R}$ given by*

$$u(f) = \int_S \left(\sum_{c \in supp \ p} \dot{u}(c) p(c) \right) \mathrm{d}P(s)$$

represents $\succsim$. Moreover, the probability P is unique and the function $\dot{u}$, and so u, is cardinal, unless $f \sim g$ for every $f, g \in \mathbf{F}_0$.

[36] If $f \sim g$ for every $f, g \in \mathbf{F}_0$, then $c \dot{\sim} d$ for every $c, d \in X$, u is constant and any P does the job. If not, then clearly u cannot be constant. Sometimes, we discard this degenerate case by assuming that there exists $c, d \in \mathbf{C}$ such that $c \dot{\succ} d$. Such axiom is known as nontriviality.

Remark 15.9 (SEU criterion). The Anscombe-Aumann theorem establishes a subjective expected utility (SEU) criterion to rank acts, and consequently actions, as each act corresponds to an action. Here, P represents the decision-maker's subjective probability over states, used to average, state by state, the utility u assigned to lotteries. This probability reflects the decision-maker's beliefs about the likelihood of events, based on an implicit, unmodeled interpretation of available information. The utility u on lotteries is an affine Bernoulli utility function, which itself can be expressed as a von Neumann-Morgenstern (vN-M) utility on consequences:

$$u(f(s)) = \sum_{c \in \text{supp } f(s)} \dot{u}(c) f(s)(c).$$

Remark 15.10. Note that:

- Writing $f(s) = \sum_{i=1}^{n} x_i \mathbf{1}_{E_i}(s)$, we obtain

$$U(f) = \int_S u(f(s)) \mathrm{d}P(s) = \int_S u\left(\sum_{i=1}^{n} x_i \mathbf{1}_{E_i}(s)\right) \mathrm{d}P(s) = \sum_{i=1}^{n} u(x_i) P(E_i) = \mathbb{E}_P(u(f)).$$

- Applying the theorem on constant lotteries, we see that, for $x, y \in X$, then $x \succsim y \iff x\mathbf{1}_S \succsim y\mathbf{1}_S \iff u(x)P(S) \geq u(y)P(Y) \iff u(x) \geq u(y)$. This implies that u is a von Neumann-Morgenstern affine utility function.

Definition 15.11 (Chisini Mean). Let $K \subset \mathbb{R}$ be an interval such that 0 is an interior point of K. A chisini mean is a function $I : \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, K) \to \mathbb{R}$ that is normalized and monotone, where $\boldsymbol{B}_0(S, \boldsymbol{\Sigma}, K)$ is the space of simple $\boldsymbol{\Sigma}$-measurable acts $f : S \to K \subseteq \mathbb{R}$:

i) $I(c\mathbf{1}_S) = c$ for every $c \in K$, i.e. trivial (constant) acts are "fixed points" for I.

ii) If $f, g \in \boldsymbol{B}_0(S; \boldsymbol{\Sigma}, K)$ and $f(s) \geq g(s)$ for every $s \in S$, then $I(f) \geq I(g)$.

Theorem 15.12 (Extending a Chisini Mean). *Let $K \subset \mathbb{R}$ be an interval such that 0 is an interior point of K. Let $I : \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, K) \to \mathbb{R}$ be such that $I(\alpha\varphi) = \alpha I(\varphi)$ for every $\alpha \in [0, 1]$ and $\varphi \in \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, K)$. Then there exists a unique positively homogeneous extension $\hat{I} : \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, \mathbb{R}) \to \mathbb{R}$. Moreover,*

i) $\hat{I}$ is a Chisini Mean if and only if I is a Chisini Mean;

ii) $\hat{I}$ is linear if and only if I is affine.

Proof.
1: define a valid extension. Let $k > 0$ be such that $[-k, k]$ is in the interior of K. For every $\varphi \in \boldsymbol{B}_0(\mathbb{R})$, we define the supremum norm as $\|\varphi\| = \sup_{s \in S} |\varphi(s)|$. Now define, for every $\varphi \in \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, \mathbb{R})$:

$$\hat{I}(\varphi) := I\left(\frac{k}{\|\varphi\| + 1} \cdot \varphi\right) \cdot \frac{\|\varphi\| + 1}{k}.$$

For every $s \in S$ and every $\varphi \in \boldsymbol{B}_0(\mathbb{R})$ we see that $-k \leq \frac{k\varphi(s)}{\|\varphi\|+1} \leq k$, so that the argument of I is in the interior of $\boldsymbol{B}_0(K)$.
2: show that it's an extension. To show that $\hat{I}$ is an extension of I, fix $\varphi \in \boldsymbol{B}_0(K)$ and use the homogeneity property of I

$$\hat{I}(\varphi) = I\left(\frac{k}{\|\varphi\| + 1} \cdot \varphi\right) \cdot \frac{\|\varphi\| + 1}{k} = \cancel{\frac{k}{\|\varphi\| + 1}} \cancel{\frac{\|\varphi\| + 1}{k}} I(\varphi) = I(\varphi)$$

3: show that it's unique. To show that $\hat{I}$ is unique, suppose there exists an extension J such that $J(\varphi) = I(\varphi)$ for $\varphi \in \boldsymbol{B}_0(K)$. Therefore, J is affine and positively homogeneous on $\boldsymbol{B}_0(K)$. Now, for any $\varphi \in \boldsymbol{B}_0(\mathbb{R})$ there exists $\alpha > 0$ sufficiently small such that $\alpha\varphi \in \boldsymbol{B}_0(K)$. Then

$$J(\varphi) = J\left(\frac{1}{\alpha}(\alpha\varphi)\right) \overset{1}{=} \frac{1}{\alpha} J(\alpha\varphi) \overset{2}{=} \frac{1}{\alpha} I(\alpha\varphi) \overset{3}{=} \frac{1}{\alpha} \hat{I}(\alpha\varphi) \overset{4}{=} \hat{I}(\varphi),$$

where 1 is the positive homogeneity of J on $\boldsymbol{B}_0(K)$, 2 is the fact that $J \equiv I$ on $\boldsymbol{B}_0(K)$, 3 is the fact that $\hat{I} \equiv I$ on $\boldsymbol{B}_0(K)$, and 4 is the positive homogeneity of $\hat{I}$ on $\boldsymbol{B}_0(K)$.

4: show that it's normalized.

$$\hat{I}(c) = I\left(\frac{k}{|c|+1}\cdot c\right)\cdot\frac{|c|+1}{k} = c\frac{\cancel{k}}{\cancel{|c|+1}}\cdot\frac{\cancel{|c|+1}}{\cancel{k}} = c,\ \forall c\in\mathbb{R}$$

5: show that it's monotone. For any $\varphi \geq \psi$ in $\boldsymbol{B}_0(\mathbb{R})$, we can find $\alpha > 0$ sufficiently small such that $\alpha\varphi, \alpha\psi \in \boldsymbol{B}_0(K)$, and then we use the monotonicity of I:

$$\hat{I}(\varphi) = \frac{1}{\alpha}I(\alpha\varphi) \geq \frac{1}{\alpha}I(\alpha\psi) = \hat{I}(\psi).$$

6: show that it's additive. To show that $\hat{I}$ is additive, we take $\varphi, \psi \in \boldsymbol{B}_0(\mathbb{R})$ and $\alpha > 0$ sufficiently small such that $\alpha\varphi + \alpha\psi \in \boldsymbol{B}_0(K)$ (for instance, $\alpha = (\|\varphi\| + \|\psi\| + 1)^{-1}$). Then we rely on affinity of I on $\boldsymbol{B}_0(K)$:

$$\hat{I}(\varphi+\psi) = 2\hat{I}\left(\frac{1}{2}(\varphi+\psi)\right) = \frac{2}{\alpha}I\left(\frac{1}{2}(\alpha\varphi)+\frac{1}{2}(\alpha\psi)\right) = \frac{\cancel{2}}{\alpha}\left(\frac{1}{\cancel{2}}I(\alpha\varphi)+\frac{1}{\cancel{2}}I(\alpha\psi)\right) = \hat{I}(\varphi)+\hat{I}(\psi).$$

7: show that it's homogeneous. To show that $\hat{I}$ is linear, we only need to show negative homogeneity ($\hat{I}(0\varphi) = 0$ is obvious), i.e. fixing $\varphi \in \boldsymbol{B}_0(\mathbb{R})$ and $\alpha > 0$, then $\hat{I}(-\alpha\varphi) = -\alpha\hat{I}(\varphi)$. This trivially comes from additivity:

$$\hat{I}(-\alpha\varphi) + \alpha\hat{I}(\varphi) = \hat{I}(-\alpha\varphi) + \hat{I}(\alpha\varphi) = \hat{I}(-\alpha\varphi+\alpha\varphi) = \hat{I}(0) = 0 \implies \hat{I}(-\alpha\varphi) = -\alpha\hat{I}(\varphi).$$

□

Lemma 15.13 (Chisini). *Let $I : \boldsymbol{B}_0(S,\boldsymbol{\Sigma},K) \to \mathbb{R}$ be a Chisini Mean. Then*

i) I is affine if and only if there exists a unique probability $P \in \Delta(\boldsymbol{\Sigma})$ such that, for every $f \in \boldsymbol{B}_0(S,\boldsymbol{\Sigma},K)$,

$$I(f) = \int_S f\mathrm{d}P = \sum_{i=1}^n a_i P(E_i), \qquad \text{(SEU criterion)}$$

where a_i and E_i are defined as in (15.1).

ii) I is concave and constant affine if and only if there exists a set $\mathcal{C} \subseteq \Delta(\boldsymbol{\Sigma})$ of probabilities such that, for every $f \in \boldsymbol{B}_0(S,\boldsymbol{\Sigma},K)$,

$$I(f) = \min_{P\in\mathcal{C}}\int_S f\mathrm{d}P = \min_{P\in\mathcal{C}}\sum_{i=1}^n a_i P(E_i) \qquad \text{(Robust Expectation)}$$

Proof.
We only prove the first claim. Note that affinity is much weaker than linearity. Note that P is a normalized nonnegative charge, that is, $P \in \Delta(\boldsymbol{\Sigma}) = \{\mu \in \mathrm{ba}(\boldsymbol{\Sigma}) : \mu \geq 0, \mu(S) = 1\}$. Suppose there exists $P \in \Delta(\boldsymbol{\Sigma})$ such that I admits a SEU representation. Then

$$I(\alpha\varphi+\overline{\alpha}\psi) = \int_S(\alpha\varphi+\overline{\alpha}\psi)\mathrm{d}P = T_P(\alpha\varphi+\overline{\alpha}\psi) = \alpha T_P(\varphi)+\overline{\alpha}T_P(\psi) = \alpha I(\varphi)+\overline{\alpha}I(\psi).$$

Conversely, suppose I is normalized, affine and monotone. We will show that I is linear and continuous.

- Setting $\alpha = 1/2$ and $\psi = -\varphi$ yields $0 = 2I(0) = I(\varphi) + I(-\varphi) \implies I(-\varphi) = -I(\varphi)$, by affinity and normalization.
- Setting $\psi = 0$, yields $I(\alpha\varphi) = \alpha I(\varphi)$, for $\alpha \in [0,1]$. If $\alpha > 1$, then $I(\varphi) = I(\frac{1}{\alpha}\alpha\varphi) = \frac{1}{\alpha}I(\alpha\varphi) \implies I(\alpha\varphi) = \alpha I(\varphi)$. If $\alpha < 0$, then $I(\alpha\varphi) = -I(-\alpha\varphi) = -(-\alpha)I(\varphi) = \alpha I(\varphi)$. This proves homogeneity.
- As for additivity,
$$I(f+g) = I\left(2\left(\frac{f}{2}+\frac{g}{2}\right)\right) = 2\left(\frac{I(f)}{2}+\frac{I(g)}{2}\right) = I(f)+I(g).$$
- Continuity is equivalent to Lipschitz Continuity, that is, $|I(f-g)| = |I(f)-I(g)| \leq C\|f-g\| \iff$

$|I(f)| \leq C\|f\|$. Thanks to monotonicity, $|I(f)| = \max\{I(f), -I(f)\} = \max\{I(f), I(-f)\} \leq I(\|f\|) \stackrel{!}{=} \|f\| = 1 \cdot \|f\|$, where ! is the normalization.

$I : \boldsymbol{B}_0(\boldsymbol{\Sigma}) \to \mathbb{R}$ is linear and continuous, thus, by the Riesz Representation Theorem, there exists a unique charge $\mu \in \mathrm{ba}(\boldsymbol{\Sigma})$ such that $I(\varphi) = \int_S \varphi \mathrm{d}\mu$. Since I is monotone, $\mu \geq 0$. Moreover, $\mu(S) = \int_S \mathbf{1}_S \mathrm{d}\mu = I(\mathbf{1}_S) = 1$, by normalization. Therefore, $\mu \in \Delta(\boldsymbol{\Sigma})$. $\square$

If $\succsim$ satisfies the axioms, then it admits a subjective utility representation.	
For every $f \in \boldsymbol{B}_0(X)$ there exists $x_f \in X$ such that $f \sim x_f$.	Fix $f \in \mathbf{F}_0 = \boldsymbol{B}_0(X)$. Write it as $f = \sum_{i=1}^n x_i \mathbf{1}_{E_i}$, with $x_1 \succsim \cdots \succsim x_n$. By A.16, $x_1 \succsim f \succsim x_n$. The preference $\succsim$ on acts satisfies A.1 - A.8, therefore, using **Lemma 2** of the proof of abstract vN-M, there exists a unique $\alpha \in [0,1]$ such that $f \sim \alpha x_1 + \overline{\alpha} x_n$. Denote $x_f := \alpha x_1 + \overline{\alpha} x_n$.
There exists an **affine** representation $u : X \to \mathbb{R}$ of $\succsim$ on X. Moreover, there exists a utility representation $V : \boldsymbol{B}_0(X) \to \mathbb{R}$ of $\succsim$ on $\boldsymbol{B}_0(X)$, defined as $V(f) = u(x_f)$.	The first claim directly follows from the abstract vN-M theorem applied on X. To show that V is well defined, if there exist $x_f \sim f$ and $x'_f \sim f$, then $x_f \sim x'_f$ and $u(x_f) = u(x'_f)$. This assures that V is well defined. Moreover, $$f \succsim g \iff x_f \succsim x_g \iff u(x_f) \geq u(x_g) \iff V(f) \geq V(g),$$ thus it's a valid utility representation.
Let $\boldsymbol{B}_0(u(X)) = \{u \circ f : f \in \mathbf{F}_0\}$ be the set of acts from S to $u(X)$, that is, for every $f = \sum a_i \mathbf{1}_{E_i} \in \mathbf{F}_0$, we consider the act $\varphi = \sum u(a_i)\mathbf{1}_{E_i}$. Then the function $I : \boldsymbol{B}_0(u(X)) \to \mathbb{R}$ defined as $I(\varphi) = V(f)$, where $\varphi = u \circ f$, is well defined and **monotone**.	To prove monotonicity, we need to prove that, if $u \circ f \geq u \circ g$, then $I(u \circ f) \geq I(u \circ g)$. To prove that I is well defined we need to prove that, if $u \circ f = u \circ g$, then $V(f) = V(g)$. As for the first one, note that $$I(u \circ f) \geq I(u \circ g) \iff V(f) \geq V(g) \iff u(x_f) \geq u(x_g) \iff x_f \succsim x_g$$ $$\iff f \succsim g \stackrel{A.16}{\iff} f(s) \succsim g(s) \iff u(f(s)) \geq u(g(s)) \iff u \circ f \geq u \circ g,$$ With the same steps one proves that $u \circ f = u \circ g \implies V(f) = V(g)$, assuring that I is well defined.
I is **Affine** and **Normalized**, i.e. for every trivial act $\varphi = u(x)$, then $I(\varphi) = \varphi$. In particular, $I(\alpha\varphi) = \alpha I(\varphi)$ for every $\alpha \in [0,1]$. Just use affinity with φ and 0.	To show that I is normalized, note that, for every $x \in X$, $V(x) = u(x)$. Therefore, for every trivial act $\varphi = u(x)$, we have $I(\varphi) = V(x) = u(x) = \varphi$. To show that I is affine, take $\alpha \in [0,1]$ and $\varphi = u \circ f$, $\psi = u \circ g$. We need two observations: 1) for every $s \in S$, $u((\alpha f + \overline{\alpha} g)(s)) = u(\alpha f(s) + \overline{\alpha} g(s)) = \alpha u(f(s)) \overline{\alpha} u(g(s)) = \alpha\varphi(s) + \overline{\alpha}\psi(s) = (\alpha\varphi + \overline{\alpha}\psi)(s)$. 2) there exist $x_f, x_g \in X$ such that $f \sim x_f$ and $g \sim x_g$. By **Lemma 4** of vN-M applied on $\boldsymbol{B}_0(X)$ we have $\alpha f + \overline{\alpha} g \sim \alpha x_f + \overline{\alpha} g \sim \alpha x_f + \overline{\alpha} x_g$. Therefore $$I(\alpha\varphi + \overline{\alpha}\psi) = V(\alpha f + \overline{\alpha} g) = u(\alpha x_f + \overline{\alpha} x_g) = \alpha u(x_f) + \overline{\alpha} u(x_g)$$ $$= \alpha V(f) + \overline{\alpha} V(g) = \alpha I(\varphi) + \overline{\alpha} I(\psi).$$
$u(X)$ is an interval containing 0 in its interior, unless $f \sim g$ for every $f, g \in \mathbf{F}_0$.	If $f \sim g$ for every $f, g \in \mathbf{F}_0$, then $c \sim d$ for every $c, d \in X$. Thus $u(c)$ is constant and Im u is a singleton. In this case, $U(f) = \int_S u(f(s))\mathrm{d}P(s) = u(c)$, for any $P : \boldsymbol{\Sigma} \to [0,1]$. If there exist $f \succ g \in \mathbf{F}_0$, then take $a \succsim f(s)$ for every $s \in S$ and $b \precsim g(s)$ for every $s \in S$. By monotonicity, $a \succsim f \succ g \succsim b$. Since $a \succ b$, then $u(a) > u(b)$. Since u is affine, we can normalize it by setting $u(a) = 1$ and $u(b) = -1$. Since u is affine, it's also continuous, thus $u(X) = \mathrm{Im}\ u \supseteq [-1,1]$.
There exists a unique $P \in \Delta(\boldsymbol{\Sigma})$ such that, for every $f \in \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, \mathbb{R})$, $\succsim$ is represented by the SEU criterion.	I is an affine Chisini Mean on $\boldsymbol{B}_0(u(X))$, $u(X)$ is an interval containing 0 in its interior. So there exists a linear Chisini Mean extension on $\boldsymbol{B}_0(\mathbb{R})$ and thus, by the Chisini Lemma, we conclude the proof, since $V(f) = I(u \circ f)$.

If exists a SEU representation for $\succsim$, then U and u are cardinally unique, and P is unique	
Assume there are two SEU representations U and V for $\succsim$, one with $u : X \to \mathbb{R}$ affine and $P \in \Delta(S, \mathbf{\Sigma})$, and the other with $v : X \to \mathbb{R}$ affine and $Q \in \Delta(S, \mathbf{\Sigma})$. Then u and v coincide up to an affine transformation.	This relies on vN-M theorem applied on X, once we notice that u and v are affine representations of $\succsim$ on X, thanks to the equivalence, for every $x, y \in X$, $x \succsim y \iff x\mathbf{1}_S \succsim y\mathbf{1}_S \iff u(x) \geq u(y) \iff v(x) \geq v(y)$ Now the cardinal uniqueness follows from vN-M theorem.
Assuming there exist $a \succ b \in X$, for every event $E \in \mathbf{\Sigma}$, $P(E) = Q(E)$.	Take $a \succ b \in X$. We then have $U(\mathbf{1}_E a + \mathbf{1}_{E^c} b) = u(a) \int_S \mathbf{1}_E \mathrm{d}P + u(b) \int_S \mathbf{1}_{E^c} \mathrm{d}P$ $= u(a)P(E) + u(b)P(E^c) = P(E)u(a) + \overline{P(E)}u(b)$ $\overset{1}{=} u(P(E)a + \overline{P(E)}b) = U(P(E)a + \overline{P(E)}b),$ where 1 is affinity of u. Similarly, $V(\mathbf{1}_E a + \mathbf{1}_{E^C} b) = v(P(E)a + \overline{P(E)}b) = V(P(E)a + \overline{P(E)}b)$. These two results imply that $\mathbf{1}_E a + \mathbf{1}_{E^c} b \sim P(E)a + \overline{P(E)}b \sim Q(E)a + \overline{Q(E)}b.$ in particular $u(P(E)a + \overline{P(E)}b) = u(Q(E)a + \overline{Q(E)}b),$ Now affinity and some trivial computations lead to $(u(a) - u(b))(P(E) - Q(E)) = 0$, leading to the conclusion since $a \succ b$.
U and V are cardinally unique.	If $v = au + b$, with $a > 0$ and $b \in \mathbb{R}$, then $V(f) = \int_S v(f(s)) \mathrm{d}Q(s) = \mathbb{E}_Q(v(f)) = \mathbb{E}_Q(au(f) + b) = a\,\mathbb{E}_Q(u(f)) + b = a\,\mathbb{E}_P(u(f)) + b = aU(f) + b$.

16 Ambiguity

16.1 Ellsberg two-urn experiment

There are two urns, I and II, each with 100 balls, either white or black. It's known that in urn I there are 50 white and 50 black balls. The DM draws one ball in each urn but before he chooses one among the following bets:

i) Bet $\mathbf{1}_{IB}$: it pays 1 euro if the ball drawn from urn I is black;

ii) Bet $\mathbf{1}_{IW}$: it pays 1 euro if the ball drawn from urn I is white;

iii) Bet $\mathbf{1}_{IIB}$: it pays 1 euro if the ball drawn from urn II is black;

iv) Bet $\mathbf{1}_{IIW}$: it pays 1 euro if the ball drawn from urn II is white;

The next table summarizes the (money) consequences of each action in each state:

	BB	BW	WB	WW
$\mathbf{1}_{IB}$	1	1	0	0
$\mathbf{1}_{IW}$	0	0	1	1
$\mathbf{1}_{IIB}$	1	0	1	0
$\mathbf{1}_{IIW}$	0	1	0	1

Here, state BW stands for drawing a black ball in urn I and a white ball in urn II, and similarly for the other states. The information provided is sufficient to conclude that:

$$P(BB \cup BW) = P(WW \cup WB) = \frac{1}{2}.$$

On the other hand, symmetry reasons lead the DM to argue that:

$$P(BB \cup WB) = P(BW \cup WW) = \frac{1}{2},$$

since there is no reason to consider drawing a white ball or a black ball from urn II as more likely. Both events must be equal to 1/2, as they are disjoint events whose union is the entire state space.
Here's the paradox: the urns are now perfectly symmetric in terms of probabilities. As a consequence, the SEU criterion will result in $u(\mathbf{1}_i) = 1/2$ for every event i. For example:

$$u(\mathbf{1}_{IB}) = \dot{u}(1)P(BB \cup BW) + \dot{u}(0)P(WW \cup WB) = \frac{1}{2},$$

and similarly for any other event. In this configuration, the DM is indifferent between betting on urn I, for which he has solid information, and betting on urn II, for which he has limited information, despite the significant difference in the quality of information available.
Experimental evidence confirms that most subjects prefer betting on urn I over urn II, a choice incompatible with the Savage's Sure Thing Principle.

Schmeidler Two-Coin Experiment: A similar example involves betting on coins. For a tested coin, the DM knows that Heads and Tails are equally likely. However, if the coin is untested, while the DM might still assume equal likelihood due to symmetry, the confidence in this assumption, and consequently the willingness to bet, is much lower. This contrasts sharply with a tested coin, even though, according to the SEU criterion, both environments should be perfectly symmetric.
The model is analogous to the previous, but now the actions are $\mathbf{1}_I$ and $\mathbf{1}_{II}$, that is, betting 1 that coin I will get a head, or that coin II will get a head.

16.2 Ellsberg Single-Urn Experiment

Consider a single urn containing 90 balls, which can be red, blue, or green. It is known that there are 30 red balls, but the distribution of blue and green balls is unknown. The decision-maker (DM) must draw one ball and choose among the following bets:

i) Bet $\mathbf{1}_R$: pays 1 euro if the ball drawn is red;

ii) Bet $\mathbf{1}_{B\cup G}$: pays 1 euro if the ball drawn is not red;

iii) Bet $\mathbf{1}_B$: pays 1 euro if the ball drawn is blue;

iv) Bet $\mathbf{1}_{R\cup G}$: pays 1 euro if the ball drawn is not blue.

The DM has significantly better information about the event R than about B. Behavioural experiments confirm that this asymmetry leads to preferences such as $\mathbf{1}_R \succ \mathbf{1}_B$, meaning the DM prefers betting on red over blue. Similarly, for the complements, $\mathbf{1}_{B\cup G} \succ \mathbf{1}_{R\cup G}$, as the DM consistently prefers bets associated with better-understood events.
However, observe the following:

$$\frac{1}{2}\mathbf{1}_B + \frac{1}{2}\mathbf{1}_G = \frac{1}{2}\mathbf{1}_{B\cup G} \succ \mathbf{1}_{R\cup G}. \tag{16.1}$$

According to the Independence Axiom, this implies:

$$\mathbf{1}_R \succ \mathbf{1}_B \implies \frac{1}{2}\mathbf{1}_R + \frac{1}{2}\mathbf{1}_G \succ \frac{1}{2}\mathbf{1}_B + \frac{1}{2}\mathbf{1}_G \iff \frac{1}{2}\mathbf{1}_{R\cup G} \succ \frac{1}{2}\mathbf{1}_{B\cup G}. \tag{16.2}$$

Combining (16.1) and (16.2), we conclude that this experiment violates the Independence Axiom of Anscombe-Aumann. The violation arises due to the DM's aversion to ignorance about the states, as they avoid bets tied to poorly understood probabilities.

17 Gilboa and Schmeidler Theorem

The SEU criterion does not take into account the quality of the DM's information and therefore his degree of confidence. However, the DM may prefer bets tied to better-understood events, as demonstrated by the single-urn experiment, where the Sure Thing Principle and Independence Axiom can be violated under ambiguous information. *Ambiguity* refers to situations such as those involving urn II or the untested coin, where beliefs lack complete confidence. There are two main approaches to address this issue:

i) We can relax the assumption that P is additive. For instance, suppose $P(R) = 1/3$, $P(B) = P(G) = 0$, and $P(B \cup G) - P(B) - P(G) = 2/3$. This approach leads to the *Choquet Expected Utility Criterion*. However, nonadditive probabilities are mathematically challenging to handle.

ii) Alternatively, we can relax the assumption that P is unique and instead represent beliefs using a set of probabilities, $\mathcal{C}$. The size of $\mathcal{C}$ reflects the degree of ambiguity in the decision problem. For example, $\mathcal{C}$ could represent all probabilities that assign $1/3$ to R. This leads to the Maximin Expected Utility (MEU) criterion.

In this section, we introduce the **Maximin Expected Utility (MEU)** criterion, also known as **Robust Expectation**, expressed as:

$$\min_{P \in \mathcal{C}} \int_S u(f(s))\,\mathrm{d}P(s),$$

and it is saying that the DM evaluates the act using the probability from $\mathcal{C}$ that yields the minimum SEU, reflecting a cautious attitude driven by ambiguity aversion. This approach shifts from Bayesian Statistics to Robust Statistics, acknowledging the DM's ignorance about the precise probabilities. To implement this framework, we modify the Anscombe-Aumann setup by removing the Independence Axiom. As a result, we no longer compute a traditional SEU criterion but instead a Robust Expectation. Axiom A.8 is weakened and replaced with two technical axioms to account for this modification:

(GS.6) Uncertainty Aversion If $f \sim g$, then $\alpha f + \overline{\alpha} g \succsim f$.

Comment: In the Anscombe-Aumann framework, by applying **Lemma 3** of the abstract von Neumann-Morgenstern theory, we derived that $f \sim g \implies \alpha f + \overline{\alpha} g \sim f$. This axiom, instead, asserts that ambiguity is modeled as a preference for hedging and randomization. The DM always prefers to mix acts rather than making a definitive choice in the presence of ambiguity. For example, if the DM is uncertain about which act to select between two equally plausible options, he would delegate the decision to a fair coin.

(GS.3) Certainty Independence For every $f, g \in \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, X)$ and $x \in X$,

$$f \succ g \iff \alpha f + \overline{\alpha} x \succ \alpha g + \overline{\alpha} x.$$

In particular,

$$f \sim g \iff \alpha f + \overline{\alpha} x \sim \alpha g + \overline{\alpha} x.$$

Comment: This axiom makes sense because constant acts are not subject to state uncertainty or informational ambiguity.

Theorem 17.1 (Gilboa-Schmeidler). *The binary relation $\succsim$ on $\boldsymbol{B}_0(S, \boldsymbol{\Sigma}, X) = \mathbf{F}_0$ satisfies axioms A.1, A.2, A.4, A.16, GS.3, GS.6 if and only if there exists an affine function $u : X \to \mathbb{R}$ and a set of probabilities $\mathcal{C} \subseteq \Delta(S, \boldsymbol{\Sigma})$ such that*

$$U(f) = \min_{P \in \mathcal{C}} \int_S u(f(s))\mathrm{d}P(s)$$

represents $\succsim$. In this case, u is cardinally unique, and $\mathcal{C}$ is unique up to convex closure, unless $f \sim g$ for every $f, g \in \mathbf{F}_0$ (In that case, u is constant and any $\mathcal{C}$ works. Compare abstract AA)

Remark 17.2 (Comparison with AA). In the Anscombe-Aumann framework, we assumed the stronger axiom A.8, which implies that the only $\mathcal{C}$ that works is a singleton $P \in \Delta(S, \mathcal{A})$. The set $\mathcal{C}$ is a singleton if and only if $\succsim$ satisfies A.8. In this case, the Maxmin Expected Utility (MEU) criterion simplifies to the Subjective Expected Utility (SEU) form. A decision-maker following MEU who has no interest in hedging ambiguity effectively behaves as an SEU decision-maker.

Remark 17.3. $\mathcal{C}$ is unique up to convex closure in the following sense: if $(u, \mathcal{C})$ and $(v, \mathcal{D})$ represent $\succsim$, then

$$\overline{\mathrm{co}}(\mathcal{C}) = \overline{\mathrm{co}}(\mathcal{D}),$$

where $\mathrm{co}(\mathcal{C})$ is the convex hull (or convex envelope) of $\mathcal{C}$, i.e., the set of all convex combinations of elements of $\mathcal{C}$. The notation $\overline{\mathrm{co}}(\mathcal{C})$ represents the closure of the convex hull of $\mathcal{C}$. Intuitively, we take the closure so that a minimum exists (for "Weierstrass").

Theorem 17.4 (Omnibus). *The binary relation $\succsim$ on $\boldsymbol{B}_0(S; \boldsymbol{\Sigma}, X) = \boldsymbol{F}_0$ satisfies axioms A.1, A.2, A.4, A.16 and GS.3 if and only if there exists $T \neq \emptyset$, a correspondence $\mathcal{C} : T \rightrightarrows \Delta(S, \boldsymbol{\Sigma})$ and $u : X \to \mathbb{R}$ affine such that*

$$U(f) = \max_{t \in T} \min_{P \in \mathcal{C}_t} \int_S u(f(s)\mathrm{d}P(s)$$

represents $\succsim$.[37]

[37]The interpretation is the following: T is a set of strategies. The DM can choose both the strategy $t \in T$ and the act $f \in \boldsymbol{F}_0$. Upon adopting $t \in T$, he will conclude that the probability on S must belong to $\mathcal{C}_t$, so that his action affects the probabilities, restricting them to some family. Once $\mathcal{C}_t$ is chosen, a precautional evaluation of f $\min \int_S u(f)\mathrm{d}P_t$ is performed.

- ***(vN-M)*** *If S is a singleton, then there exists $u : X \to \mathbb{R}$ affine representing $\succsim$;*
- ***(GS)*** *If $\succsim$ satisfies GS.6 then we can choose $\mathcal{C}(T)$ constant for every T, that is, $U(f) = \min_{P\in\mathcal{C}} \int_S u(f(s))\mathrm{d}P(s)$;*
- ***(AA)*** *If $\succsim$ satisfies uncertainty neutrality, i.e.* ***lemma 3*** *of vN-M, then $\mathcal{C}$ can be chosen to be a constant singleton P, that is, $U(f) = \int_S u(f(s))\mathrm{d}P(s)$.*

If $\succsim$ satisfies the axioms, then it admits a robust maximin utility representation.	
For every $f \in \boldsymbol{B}_0(X)$ there exists $x_f \in X$ such that $f \sim x_f$.	Fix $f \in \mathbf{F}_0 = \boldsymbol{B}_0(X)$. Write it as $f = \sum_{i=1}^n x_i \mathbf{1}_{E_i}$, with $x_1 \succsim \cdots \succsim x_n$. By A.16, $x_1 \succsim f \succsim x_n$. The preference $\succsim$ on X satisfies A.1 - A.8 (since GS.3 becomes A.8 on X), therefore, using **Lemma 2** of the proof of abstract vN-M, there exists a unique $\alpha \in [0,1]$ such that $f \sim \alpha x_1 + \overline{\alpha} x_n$. Denote $x_f := \alpha x_1 + \overline{\alpha} x_n$.
There exists an **affine** representation $u : X \to \mathbb{R}$ of $\succsim$ on X. Moreover, there exists a utility representation $V : \boldsymbol{B}_0(X) \to \mathbb{R}$ of $\succsim$ on $\boldsymbol{B}_0(X)$, defined as $V(f) = u(x_f)$.	The first claim directly follows from the abstract vN-M theorem applied on X. To show that V is well defined, if there exist $x_f \sim f$ and $x_f' \sim f$, then $x_f \sim x_f'$ and $u(x_f) = u(x_f')$. This assures that V is well defined. Moreover, $$f \succsim g \iff x_f \succsim x_g \iff u(x_f) \geq u(x_g) \iff V(f) \geq V(g),$$ thus it's a valid utility representation.
V is constant-affine, that is, affine when mixed with a constant.	Given $f \in \boldsymbol{B}_0(X)$, and $x \in X$, **Certainty independence** and $f \sim x_f$ imply $\alpha f + \overline{\alpha} x \sim \alpha x_f + \overline{\alpha} x$ for every $\alpha \in [0,1]$. Therefore $V(\alpha f + \overline{\alpha} x) = u(\alpha x_f + \overline{\alpha} x) = \alpha u(x_f) + \overline{\alpha} u(x) = \alpha V(f) + \overline{\alpha} V(x)$.
Let $\boldsymbol{B}_0(u(X)) = \{u \circ f : f \in \mathbf{F}_0\}$ be the set of acts from S to $u(X)$, that is, for every $f = \sum a_i \mathbf{1}_{E_i} \in \mathbf{F}_0$, we consider the act $\varphi = \sum u(a_i)\mathbf{1}_{E_i}$. Then the function $I : \boldsymbol{B}_0(u(X)) \to \mathbb{R}$ defined as $I(\varphi) = V(f)$, where $\varphi = u \circ f$, is well defined and **monotone**.	To prove monotonicity, we need to prove that, if $u \circ f \geq u \circ g$, then $I(u \circ f) \geq I(u \circ g)$. To prove that I is well defined we need to prove that, if $u \circ f = u \circ g$, then $V(f) = V(g)$. As for the first one, note that $$I(u \circ f) \geq I(u \circ g) \iff V(f) \geq V(g) \iff u(x_f) \geq u(x_g) \iff x_f \succsim x_g$$ $$\iff f \succsim g \overset{A.16}{\Longleftarrow} f(s) \succsim g(s) \iff u(f(s)) \geq u(g(s)) \iff u \circ f \geq u \circ g,$$ With the same steps one proves that $u \circ f = u \circ g \implies V(f) = V(g)$, assuring that I is well defined.
I is **normalized** and **constant affine**, meaning that, for all $\varphi \in \boldsymbol{B}_0(u(X))$, $\lambda \in u(X)$ and $\alpha \in [0,1]$, we have $I(\alpha\varphi + \overline{\alpha}\lambda) = \alpha I(\varphi) + \overline{\alpha}\lambda$.	To show that I is normalized, note that, for every $x \in X$, $V(x) = u(x)$. Therefore, for every trivial act $\varphi = u(x)$, we have $I(\varphi) = V(x) = u(x) = \varphi$. To prove constant affinity, we take $\varphi = u \circ f$ and $\lambda = u(x)$, with $f \in \boldsymbol{B}_0(X)$ and $\lambda \in u(X)$. Then, by affinity of u, $$u \circ (\alpha f + \overline{\alpha} x) = \alpha(u \circ f) + \overline{\alpha} u(x) = \alpha\varphi + \overline{\alpha}\lambda,$$ consequently, using constant affinity of V, $$I(\alpha\varphi + \overline{\alpha}\lambda) = V(\alpha f + \overline{\alpha} x) = \alpha V(f) + \overline{\alpha} V(x) = \alpha I(\varphi) + \overline{\alpha} I(\lambda)$$
I admits a unique extension $\hat{I} : \boldsymbol{B}_0(S, \boldsymbol{\Sigma}, \mathbb{R}) \to \mathbb{R}$ which, in addition, is a constant affine Chisini Mean.	Even if I is not affine, Constant affinity guarantees that, for every $\alpha \in [0,1]$, $I(\alpha\varphi) = I(\alpha\varphi + \overline{\alpha}0) = \alpha I(\varphi) + \overline{\alpha}0 = \alpha I(\varphi)$. This is enough to apply lemma "Extending a Chisini Mean".
If $V(f) = V(g)$, then $V(\alpha f + \overline{\alpha} g) \geq \alpha V(f) + \overline{\alpha} V(g)$.	If $V(f) = V(g)$, then $f \sim g$, so we can apply GS.6 to conclude that $\alpha f + \overline{\alpha} g \succsim f \iff V(\alpha f + \overline{\alpha} g) \geq V(f) = \alpha V(f) + \overline{\alpha} V(g)$.
If $I(\varphi) = I(\psi)$, then $I(\alpha\varphi + \overline{\alpha}\psi) \geq \alpha I(\varphi) + \overline{\alpha} I(\psi)$.	Take $\varphi, \psi \in \boldsymbol{B}_0(u(X))$. Then, by definition of $\boldsymbol{B}_0(u(X))$, there exist $f, g \in \boldsymbol{B}_0(X)$ such that $\varphi = u(f)$ and $\psi = u(g)$. By affinity, $u(\alpha f + \overline{\alpha} g) = \alpha u(f) + \overline{\alpha} u(g)$. Therefore $I(\alpha\varphi + \overline{\alpha}\psi) = V(\alpha f + \overline{\alpha} g) \geq \alpha V(f) + \overline{\alpha} V(g) = \alpha I(\varphi) + \overline{\alpha} I(\psi)$.
If $\varphi, \psi \in \boldsymbol{B}_0(\mathbb{R})$, then there exists λ sufficiently small such that both $\lambda\varphi, \lambda\psi \in \boldsymbol{B}_0(u(X))$. Using this, one shows that if $\hat{I}(\varphi) = \hat{I}(\psi)$, then $\hat{I}(\alpha\varphi + \overline{\alpha}\psi) \geq \alpha\hat{I}(\varphi) + \overline{\alpha} I(\psi)$.	The first claim is shown in point 6 of the proof of "Extending a Chisini Mean". Therefore $$\lambda\hat{I}(\alpha\varphi + \overline{\alpha}\psi) = \hat{I}(\alpha\lambda\psi + \overline{\alpha}\lambda\psi) = I(\alpha\lambda\varphi + \overline{\alpha}\lambda\psi) \geq$$ $$\geq \alpha I(\lambda\varphi) + \overline{\alpha} I(\lambda\psi) = \alpha\hat{I}(\lambda\varphi) + \overline{\alpha}\hat{I}(\lambda\psi) = \lambda\alpha I(\varphi) + \lambda\overline{\alpha} I(\psi).$$ Dividing by λ we get the result.

<table>
<tr><th colspan="2">If $\succsim$ satisfies the axioms, then it admits a robust maximin utility representation.</th></tr>
<tr><td>$\hat{I}$ is concave, that is, the previous point holds even if $\hat{I}(\varphi) \neq \hat{I}(\psi)$.
Since $\hat{I}$ is a concave and constant affine Chisini Mean on $\boldsymbol{B}_0(\mathbb{R})$, we conclude by the Chisini Lemma that there exists a set $\mathcal{C} \subseteq \Delta(\boldsymbol{\Sigma})$ of probabilities such that,for every $\varphi \in \boldsymbol{B}_0(\mathbb{R})$, $\hat{I}(\varphi) = \min_{P\in\mathcal{C}} \int_S \varphi \mathrm{d}P$. Recalling that $V(f) = \hat{I}(u(f))$, then $V(f) = \min_{P\in\mathcal{C}} \int_S u(f)\mathrm{d}P$.</td>
<td>Define $\chi = \psi + \hat{I}(\varphi) - \hat{I}(\psi)$. Now we compute both $\hat{I}(\chi)$ and $\hat{I}(\alpha\varphi + \overline{\alpha}\chi)$, thanks to constant affinity: $\hat{I}(\chi) = \hat{I}(\psi + \hat{I}(\varphi) - \hat{I}(\psi)) = \hat{I}(\varphi)$, and $\hat{I}(\alpha\varphi + \overline{\alpha}\chi) = \hat{I}(\alpha\varphi + \overline{\alpha}\psi) + \overline{\alpha}(\hat{I}(\varphi) - \hat{I}(\psi))$. From here, we have
$$\hat{I}(\alpha\varphi + \overline{\alpha}\psi) = \hat{I}(\alpha\varphi + \overline{\alpha}\chi) - \overline{\alpha}\hat{I}(\varphi) + \overline{\alpha}\hat{I}(\psi).$$
Since $\hat{I}(\chi) = \hat{I}(\varphi)$, we use previous point to conclude that
$$\hat{I}(\alpha\varphi + \overline{\alpha}\chi) \geq \alpha\hat{I}(\varphi) + \overline{\alpha}\hat{I}(\chi) = \alpha\hat{I}(\varphi) + \overline{\alpha}\hat{I}(\varphi) = \hat{I}(\varphi).$$
Combining this inequality with the previous equality we get
$$\hat{I}(\alpha\varphi + \overline{\alpha}\psi) \geq \hat{I}(\varphi) - \overline{\alpha}\hat{I}(\varphi) + \overline{\alpha}\hat{I}(\psi) = \alpha\hat{I}(\varphi) + \overline{\alpha}\hat{I}(\psi).$$</td></tr>
</table>

18 Savage and de Finetti subjective probability

In this section, we introduce Savage's framework for evaluating acts based on the subjective expected utility (SEU) criterion, which operates without imposing strong assumptions on the consequence space (in contrast, Anscombe-Aumann assumed the consequence space X to be a convex subset of a vector space). We first present de Finetti's treatment of subjective probability, which established this concept as a fundamental tool for studying decision-making under uncertainty. Although Savage's SEU criterion was developed before Anscombe-Aumann's SEU, its greater mathematical complexity makes it more suitable to introduce after the simpler framework of Anscombe-Aumann.

18.1 De Finetti's Axioms

Let $\succsim^*$ be a binary relation over an algebra $\boldsymbol{\Sigma}$ of events of the state space S, called *qualitative probability*, and representing a comparative belief of the DM about likelihood of events to occur. $E \succsim^* F$ means "event E is at least as likely as event F". We introduce some axioms on it, following de Finetti (1931).

(DF.1) Weak Order: $\succsim^*$ is complete and transitive.

(DF.2) Normalization: for all events $E \in \boldsymbol{\Sigma}$, $S \succsim^* E \succsim^* \emptyset$, with $S \succ^* \emptyset$.

Comment: it's needed to set $P(S) = 1 - P(\emptyset) = 1$.

(DF.3) Additivity: For all events E, F and H, with $E \cap H = F \cap H = \emptyset$,

$$E \succsim^* F \iff E \cup H \succsim^* F \cup H.$$

Comment: It is a basic independence assumption, which says that adding a common disjoint event shouldn't affect the ranking. Note that in Ellsberg axiom, we have $R \succ^* B \sim^* G$ and $R \cup G \prec^* B \cup G$, meaning this property does not hold.

(DF.4) Equidivisibility: For each $n \geq 1$, there is a partition into 2^n equally-likely events.

Comment: As a trivial consequence, the state space must be infinite. It's a highly nontrivial technical axiom.

(DF.4 bis) Divisibility: For all events $E \succ^* F$, there exists an event partition $\{H_i\}_{i=1}^n$ such that $E \succ^* F \cup H_i$ for every $i = 1, \ldots, n$.

Comment: Under DF.1 - DF.3, DF.4 bis is a stronger version of DF.4, proposed by Savage (1954) to complete de Finetti's partial analysis.

18.2 Existence of a probability measure representation

Usually, we employ utility functions $u : X \to \mathbb{R}$ to represent preferences over elements in X. Here, a probability measure takes the role of the utility function because we are considering preferences over subsets instead of individual elements. In this section, we outline the necessary and sufficient conditions for the existence of such a probability measure.

Proposition 18.1 (Monotonicity). *Under axioms DF.1, DF.2, and DF.3, it holds that* $E \subseteq F \implies E \precsim^* F$, *meaning larger events are more likely. Indeed:*

$$F = (F \setminus E) \cup E \succsim^* \emptyset \cup E = E.$$

A probability measure is a *representation* of $\succsim^*$ if, for all events E and F,

$$E \succsim^* F \iff P(E) \geq P(F).$$

Proposition 18.2. *A probability measure representation of a qualitative probability exists only if axioms DF.1 to DF.3 are satisfied.*

Theorem 18.3 (De Finetti). *Under DF.1 - DF.4, there exists a* <u>*unique*</u> *and* <u>*nonatomic*</u> *probability measure* $P : \mathbf{\Sigma} \to [0, 1]$ *such that, for all events* E *and* F,

$$E \succsim^* F \implies P(E) \geq P(F).$$

Remark 18.4. Although this result is only a partial representation, its importance lies in its historical role in axiomatizing subjective probability. De Finetti was one of the first to use an axiomatic method based on a binary relation. In this theorem, we make the technical assumptions that $\mathbf{\Sigma}$ is a σ-algebra and P is nonatomic.[38] 30 years later, a counterexample demonstrated that axioms DF.1 to DF.4 are not sufficient for a complete probability measure representation.

Theorem 18.5 (De Finetti-Savage). $\succsim^*$ *satisfies DF.1, DF.2, DF.3 and DF.4 bis if and only if there exists a* <u>*convex-ranged*</u> *probability measure* $P : \mathbf{\Sigma} \to [0, 1]$ *representing* $\succsim^*$. *Moreover,* P *is unique.*

Remark 18.6. This theorem completes de Finetti's analysis, providing a sufficient characterization for $\succsim^*$ being representable through a probability measure. Two technical restrictions are required: $\mathbf{\Sigma}$ to be a σ-algebra, and P to be convex-ranged.

Definition 18.7 (Convex-Ranged Probability Measure). $P : \mathbf{\Sigma} \to [0, 1]$ is convex-ranged if, for every collection of events $\{E_1, \ldots, E_n\}$ in $\mathbf{\Sigma}$, there exists $F \in \mathbf{\Sigma}$ such that $P(F) = \sum_{i=1}^{n} \alpha_i P(E_i)$, where $\alpha_i \geq 0$ and $\alpha_1 + \cdots + \alpha_n = 1$.

Example 18.8. For instance,

- For any two events E and F, there exists G such that $P(G) = \alpha P(E) + \overline{\alpha} P(F)$.
- Equivalently, for any event E and any $\alpha \in (0, 1)$, there exists another event F such that $P(F) = \alpha P(E)$.
- It can be shown that if a probability is **nonatomic**, then it is **convex-ranged**.
- Both the nonatomic and convex-ranged properties imply this operational characterization, which will be used in Savage's proof: for every lottery $l \in \Delta_0(\mathbf{C})$, there exists an act $f \in \boldsymbol{B}_0(\mathbf{\Sigma})$ such that, for every $c \in \mathbf{C}$, the probability of obtaining c with f, i.e., the probability that the true state is one of the states that results in c, is exactly $l(c)$. More explicitly,

  $$l(c) = P(\{f = c\}), \quad \text{for every } c \in \mathbf{C}. \tag{18.1}$$

  Operationally, this means that for such P and S, any lottery $l \in \Delta_0(\mathbf{C})$ can be represented by some act $f \in \boldsymbol{B}_0(\mathbf{\Sigma})$. In other words, any probability distribution can be obtained by mapping states to consequences via an act, using P to compute the probabilities.

18.3 Savage's Axioms

In 1954, Savage combined de Finetti's analysis of subjective probability with von Neumann and Morgenstern's expected utility, formulating the first axiomatic theory for the subjective expected utility (SEU) criterion:

$$u(f) = \int_S \dot{u}(f(s))\,\mathrm{d}P, \tag{18.2}$$

where $\dot{u}$ is a utility function defined on consequences, and P is a monotone subjective probability measure over states. Recall that Savage's decision environment $(\mathbf{F}_0, \mathcal{F}, \succsim)$ considers simple[39] $\mathbf{\Sigma}$-measurable acts $f : S \to \mathbf{C}$, without placing any specific assumptions on the set of consequences $\mathbf{C}$, but assuming that $\mathbf{\Sigma}$ is a σ-algebra.

[38] A probability measure P is **nonatomic**, or diffuse, if, for every $A \in \mathbf{\Sigma}$ such that $P(A) > 0$, there exists a subset $B \subseteq A$ such that $0 < P(B) < P(A)$. Intuitively: 1. The events in $\mathbf{\Sigma}$ are infinitely divisible (see DF.4). 2. The probability measure is not concentrated on discrete points (atoms) but is distributed smoothly across subsets.

[39] In his original derivation, he considers generic $\mathbf{\Sigma}$-measurable acts in $\mathbf{F}$, but we will focus on simple acts by simplicity.

First, Savage introduces a qualitative probability relation, defined as follows: for a pair of acts f and g and an event E, we write:

$$fEg = \begin{cases} f(s) & \text{if } s \in E, \\ g(s) & \text{if } s \notin E. \end{cases}$$

This representation has a clear interpretation when applied to consequences (i.e., constant acts): cEd represents a bet on event E, which pays c if E occurs and d otherwise. If $c \dot{\succ} d$, then $cFd \succsim cEd$ indicates that the decision-maker subjectively prefers betting on F rather than on E, since he perceives F as more likely than E. As a result, this comparison aligns with de Finetti's qualitative relation $F \succsim^* E$.
More formally, Savage defines the qualitative probability relation $\succsim^*$ on $\mathbf{\Sigma}$ as follows:

$$E \succsim^* F \iff cEd \succsim cFd$$

for $c \dot{\succ} d$. He defines an event as **null** if, for all $f, g \in \mathbf{F}_0$, $fEh \sim gEh$. This means that the DM is indifferent between betting on f or g when restricted to E, as they do not consider E to be possible.
Then, he introduces the following axioms:

(P.1) Weak Order $\succsim$ is complete and transitive

(P.2) Sure-Thing Principle Given any acts $f, g, h, h' \in \mathbf{F}_0$ and any event E,

$$fEh \succsim gEh \iff fEh' \succsim gEh'.$$

(Comment:) The sure-thing principle is an independence axiom asserting that a reversal in preference cannot occur when the common part of two acts is replaced by a different common part. It's the analogous of DF.3. It allows to properly define the following conditional preference relation:

Conditional Preference The preference relation on $\mathbf{F}_0$ conditioned to event E is given by

$$f \succsim_E g \iff \exists h \in \mathbf{F}_0 : fEh \succsim gEh.$$

This preference depends only on what happens when E obtains.

(P.3) State Independence if E is not null, $f \in \mathbf{F}_0$, for all consequences $c, d \in \mathbf{C}$,

$$cEf \succsim dEf \iff c \dot{\succsim} d$$

(Comment): abstracting from *state dependence issues*, the evaluation of consequences is not affected by the states under which they realize. In particular, $c \succsim_E d \iff c \succsim d \iff c \succsim_F d$.

(P.4) Stake Independence If $a \dot{\succ} b$ and $c \dot{\succ} d$, for any events E and F,

$$aEb \succsim aFb \iff cEd \succsim cFd.$$

(Comment:) this axiom ensures well-posedness of the interpretation of $aEb \succsim aFb$, with $a \dot{\succ} b$, as $E \succsim^* F$, by ensuring that what matters is the relative ranking of bets' consequences, not their relative magnitude.

(P.5) Nontriviality There exist consequences $c, d \in \mathbf{C}$ such that $c \dot{\succ} d$, ensuring nontriviality of P.4.

(P.6) Divisibility Let $f, g \in \mathbf{F}_0$ be acts such that $f \succ g$. Given any consequence $c \in \mathbf{C}$, there is an event partition $\{E_i\}_{i=1}^n$ of S such that both $cE_if \succ g$ and $f \succ cE_ig$ for each $i = 1, \ldots, n$.

(Comment:) The first five axioms are necessary conditions for a subjective expected utility (SEU) representation.[40] This technical axiom ensures sufficiency. Essentially, this axiom asserts that it is possible to divide the state space S into events so finely that modifying these slices does not disrupt the overall preference ranking. It allows the decision-maker's preferences to remain consistent even when the state space is broken into very small slices (events). For any act f that is preferred over g, you can create a partition of the state space such that adding a "small modification" (represented by cE_i) to either f or g does not change the fact that f is preferred over g.

[40] If there is a probability measure $P : \mathbf{\Sigma} \to [0,1]$ and a nonconstant function $\dot{u} : \mathbf{C} \to \mathbb{R}$ such that (18.2) holds, then P.1 - P.5 hold, and $P(E) = 0$ if and only if E is null.

Remark 18.9 (Comparison DF.4 bis and P.6). DF.4 bis states: for any events $E \succ^* F$, there exists a partition $\{H_i\}_{i=1}^n$ such that $E \succ^* F \cup H_i$ for every i. This ensures that preferences between events remain stable even with small perturbations (H_i) added to F.

Remark 18.10. P.6 similarly asserts that for any acts $f \succ g$, a partition $\{E_i\}_{i=1}^n$ exists such that modifying f or g with a consequence c (via cE_if or cE_ig) does not reverse the preference ranking.

18.4 Savage's Subjective Expected Utility

Savage's Representation theorem establishes the equivalence of P.1-P.6 with the existence of a utility function u on consequences and a convex-ranged subjective probability measure P on states such that acts f are ranked according to the SEU criterion.

Theorem 18.11 (Savage). *$\succsim$ satisfies P.1 - P.6 if and only if there exists a nonconstant function $\dot{u} : \mathbf{C} \to \mathbb{R}$ and a convex-ranged probability measure $P : \mathbf{\Sigma} \to [0,1]$ such that the function $u : \mathbf{F}_0 \to \mathbb{R}$ given by*

$$u(f) = \int_S \dot{u}(f(s))\mathrm{d}P(s) \tag{18.3}$$

represents $\succsim$. Moreover, the subjective probability P is unique and the function $\dot{u}$ (which is a utility function over consequences), and so u, is cardinally unique.

Remark 18.12 (Comparison with Anscombe-Aumann). Savage's Theorem is evidently more general. In particular:

i) It applies even when $\mathbf{C}$ is not convex;

ii) The utility function $\dot{u}$ is not required to be affine;

iii) P is required to be convex-ranged, $\mathbf{\Sigma}$ must be a σ-algebra, and the elements of $\mathbf{\Sigma}$ must satisfy P.6.

Remark 18.13 (Subjective probability). P represents a convex-ranged subjective probability that reflects beliefs about events. Specifically, by P.5, if $c \dot{\succ} d$, we normalize $\dot{u}$ such that $\dot{u}(c) = 1$ and $\dot{u}(d) = 0$. Using this normalization, and applying the SEU criterion, we deduce:

$$E \succsim^* F \iff cEd \succsim cFd \iff P(E) \geq P(F).$$

Proof.

We provide only a sketch of the proof.

($\star$) If $\succsim$ satisfies axioms P.1–P.6, then $\succsim^*$ satisfies DF.1–DF.3 and DF.4 bis. From this, the De Finetti-Savage theorem implies the existence of a unique *convex-ranged* subjective probability measure $P : \mathbf{\Sigma} \to [0,1]$ that represents $\succsim^*$.

Given acts $f(s) = \displaystyle\sum_{i=1}^n c_i \mathbf{1}_{E_i}(s)$ and $g(s) = \displaystyle\sum_{j=1}^m d_i \mathbf{1}_{F_i}(s)$, then P induces the lotteries P_f and P_g, represented by the uncertainty prospects $\{c_1, P(E_1); \ldots; c_n, P(E_n)\}$ and $\{d_1, P(F_1); \ldots; d_n, P(F_n)\}$. Moreover, every lottery $l \in \Delta_0(\mathbf{C})$ is induced by some act, thanks to the range convexity of P (see (18.1)).

($\star\star$) Suppose P has the property that, for every two acts f, g, $P_f = P_g \implies f \sim g$. This property ensures that preferences over acts are determined solely by their corresponding lotteries. This consequentialism induces a preference over lotteries:

$$P_f \ddot{\succsim} P_g \iff f \succsim g.$$

($\star\star\star$) Assume that axioms P.1–P.6 guarantee $\ddot{\succsim}$ satisfies axioms A.1, A.2, A.4, and A.8. By applying the abstract von Neumann-Morgenstern (vN-M) theorem to $\ddot{\succsim}$, there exists a utility function such that:

$$\begin{aligned} f \succsim g \iff P_f \ddot{\succsim} P_g &\overset{\text{vN-M}}{\iff} \sum_{c \in \text{supp } P_f} \dot{u}(c) P_f(c) \geq \sum_{c \in \text{supp } P_g} \dot{u}(c) P_g(c) \\ &\iff \sum_{i=1}^n \dot{u}(c_i) P(E_i) \geq \sum_{j=1}^m \dot{u}(d_j) P(F_j) \iff \int_S \dot{u}(f(s))\mathrm{d}P(s) \geq \int_S \dot{u}(g(s))\mathrm{d}P(s) \end{aligned}$$

□

Remark 18.14 (Importance). While the Anscombe-Aumann theorem is more versatile, Savage's theorem came first, laying the groundwork for modern decision theory. Its innovation lies in unifying two foundational pillars: the von Neumann-Morgenstern expected utility theorem and de Finetti's subjective probability. Savage's framework showed how to derive subjective probability from qualitative preferences, turning uncertainty into measurable risk without relying on numerical assumptions. This approach established the idea that rational choice requires expressing uncertainty in Bayesian terms.

19 Appendix

In this section we have a look to more advanced topics which are not essential but still included in the syllabus.

Definition 19.1 (Strict Transitivity). for all $x, y, z \in X$, $x \succ y$ and $y \succ z$ imply $x \succ z$.[41]

Remark 19.2. the necessity of strict transitivity arises when considering threshold effects in perception. For instance, imagine a good x_0 of 1000 grams of sugar and a good x_1 of 999 grams of sugar. A DM may be essentially indifferent between them. Now consider x_n as an amount of $1000 - n$ grams of sugar. With the same logic, we may conclude by transitivity that

$$x_0 \sim x_1 \sim x_2 \sim \ldots \sim x_n \sim \ldots \sim x_{1000},$$

which is a clear contradiction. To avoid such paradoxes, we discard the stronger version of transitivity of the indifference preference, relying only on the sharper version of transitivity.

Theorem 19.3 (Existence of a Utility Function). *A preference relation on $\mathbb{R}^n_+$ has a strongly monotone and continuous utility function $u : \mathbb{R}^n_+ \to \mathbb{R}$ if and only if it is transitive, complete, strongly monotone and Archimedean.*

Proof.

Here we prove the direct implication. If $u : \mathbb{R}^n_+$ is a strongly monotone and continuous utility function for $\succsim$, then $\succsim$ is complete, transitive (which is a necessary condition for u to exist) and strongly monotone (since $x \gg y \implies u(x) > u(y) \implies x \succ y$ and $x \geq y \implies u(x) \geq u(y) \implies x \succsim y$)
We only need to show that $\succsim$ is Archimedean. Let $x, y, z \in X$ with $x \succ y \succ z$. We have to show the existence of $\alpha, \beta \in (0, 1)$ such that

$$\alpha x + \overline{\alpha} z \succ y \succ \beta x + \overline{\beta} z.$$

Consider the set

$$\mathcal{S} = \{\lambda \in [0, 1] : \lambda x + \overline{\lambda} z \succsim y\}.$$

Lemma: $\mathcal{S}$ is closed. Take a sequence $\{\alpha_n\} \in \mathcal{S}$ converging to $\alpha \in [0, 1]$. $\mathcal{S}$ is closed if and only if $\alpha \in \mathcal{S}$, that is, $\alpha x + \overline{\alpha} z \succsim y$. We have

$$\alpha_n x + \overline{\alpha}_n z \succsim y \overset{1}{\iff} u(\alpha_n x + \overline{\alpha}_n z) \geq u(y) \overset{2}{\iff} u(\alpha x + \overline{\alpha} z) \geq u(y) \overset{3}{\iff} \alpha x + \overline{\alpha} z \succsim y,$$

where 1 is the definition of u, 2 uses the linearity of limit, the continuity of u and the permanence of inequalities via limits, and 3 is the strong monotonicity of u.
By completeness, we can easily characterize the complement $\mathcal{S}^c$ of $\mathcal{S}$ in $[0, 1]$:

$$\mathcal{A} = \mathcal{S}^c \cap [0, 1] = \{\lambda \in [0, 1] : \lambda x + \overline{\lambda} z \prec y\}.$$

Note that $0 \in \mathcal{A}$, since $y \succ z$. Since $\mathcal{S}^c$ is open, there exists $\varepsilon > 0$ such that $[0, \varepsilon) \subseteq \mathcal{A}$. Taking $\beta = \varepsilon/2$ we manage to find $\beta \in [0, 1]$ such that $\beta x + \overline{\beta} z \prec y$.
Similarly, defining $\mathcal{S}' = \{\lambda \in [0, 1] : \lambda x + \overline{\lambda z} \precsim y\}$, we can show that $\mathcal{S}'$ is closed, and therefore the set

$$\boldsymbol{B} = \mathcal{S}'^c \cap [0, 1] = \{\lambda \in [0, 1] : \lambda x + \overline{\lambda} z \succ y\}$$

is open. Since $1 \in \boldsymbol{B}$, we can find $\varepsilon > 0$ such that $(\varepsilon, 1] \in \boldsymbol{B}$, and thus $\alpha = 1 - \varepsilon/2 \in \boldsymbol{B}$ satisfies $\alpha x + \overline{\alpha} z \succ y$. □

Definition 19.4 (Additional Remarks on Rational Correspondence). We can define

$$C_{\succsim} = \{x \in X : \forall y \in X, x \succsim_X y \text{ or } y \succsim_X x\} \quad \text{and} \quad \sigma_\infty(X) = \sigma(X) \cap C_{\succsim} = \{\hat{x} \in X : \forall y \in X, \hat{x} \succsim_X y\},$$

[41] As an exercise, you can show that the transitivity of $\succsim$ implies the transitivity of $\succ$, and that the converse holds if we assume that $\succsim$ is also antisymmetric.

that is, $C_{\succsim}$ is the subset of alternatives which are universally comparable in X, and σ_∞ is the collection of optimal alternatives in it, that is, among the optimal alternatives, we pick only those that we are sure are preferred to anything else, discarding those that simply are not comparable.[42] Unfortunately, $\sigma_\infty(X)$ may be empty even if $\sigma(X)$ is not. However, with the additional assumption that $\succsim$ is complete, then $C_{\succsim} = X$ and $\sigma(X) = \sigma_\infty(X) = \{\hat{x} \in X : \forall x \in X, \hat{x} \succsim_X x\}$.

Definition 19.5 (Ascent algorithm). It's an algorithm to show that optimal alternatives exist when X is finite, say of n elements. Initially, we take a random element i_0 from X, and we set $j_0 := i_0$. Then we take a second random element i_1 from $X - \{i_0\}$ and we define j_1 as the optimal between j_0 and i_1. Now we take a third random element i_2 from $X - \{i_0, i_1\}$ and we define j_2 as the optimal between j_1 and i_2. And so on. We finally choose j_n.

Definition 19.6 (Convex menu correspondence). A menu correspondence $\varphi : \Theta \rightrightarrows \mathbf{X}$ is (sharply) quasi-convex if, for each $\alpha \in [0, 1]$,

$$\varphi(\alpha\theta + (1-\alpha)\theta') \subseteq \varphi(\theta) \cup \varphi(\theta'), \quad \theta, \theta' \in \Theta.$$

We have that, if the menu correspondence is quasi-convex, then the value function is quasi-convex.[43]

Theorem 19.7 (Maximum). *If $\mathbf{X}$ and Θ are metrizable, the parametric utility function $u : Gr\varphi \to \mathbb{R}$ and the menu correspondence $\varphi : \Theta \rightrightarrows \mathbf{X}$ are both continuous, then $D = \Theta$, the value function $v : \Theta \to \mathbb{R}$ is continuous and σ is compact-valued and upper hemicontinuous (so, continuous when a function).*

Definition 19.8 (Optimal bundles). An optimal bundle $\hat{c} \in B(p, w)$ is:

i) **internal** when $\hat{c} \gg \mathbf{0}$;

ii) **boundary** when $\hat{c}_i = 0$ for some $i = 1, \ldots, n$

iii) **corner** when it is a vertex of the budget set. The lexicographic optimal is corner.

Definition 19.9 (Nonlinear Pricing). If markets of goods are competitive, consumers can buy any amount of any good without affecting its price. In particular, the price of any bundle c is given by $p \cdot c$, where p is a price vector. We can actually associate a **market value function** $\mathrm{p} : \mathbb{R}^n_+ \to [0, +\infty)$ which associates to each consumption bundle its price, i.e. $p \cdot c$. Such function is linear if and only if the price of a mixture of two consumption bundles is proportionally determined by their prices, without any distortions. For instance, $\mathrm{p}(c/2) = \mathrm{p}(c)/2$.
Quantity discounts or other economic phenomena lead to nonlinear market value functions, which do not have a dot product representation. In particular, it's not sufficient to know the vector of unit prices $p = (\mathrm{p}(e_1), \ldots, \mathrm{p}(e_n)) \in \mathbb{R}^n_+$ to determine any other price. Still, the optimality condition does not change too much:

$$\frac{\partial u(\hat{c})}{\partial c_k} \Big/ \frac{\partial u(\hat{c})}{\partial c_j} = \frac{\partial \mathrm{p}(\hat{c})}{\partial c_k} \Big/ \frac{\partial \mathrm{p}(\hat{c})}{\partial c_j}$$

Definition 19.10 (Choquet Simplex). A convex subset X of a vector space V is a Choquet simplex if one of the following equivalent conditions hold:

i) Each of its elements can be uniquely written as a convex combination of extreme points of X;

ii) For every $x \in X$ there is a unique probability measure $p_x \in \Delta(\mathrm{ext}X)$ such that

$$x = \sum_{z \in \mathrm{supp}\ p_x} p_x(z) z.$$

The conditions are equivalent since z are the extreme points in X representing x and $p_x(z)$ are the weights you put on each z.

Theorem 19.11 (Abstract vN-M (2)). *if X is a Choquet simplex, the following conditions are equivalent:*

i) $\succsim$ satisfies axioms A.1, A.2, A.4, A.8;

ii) There exists a function $\zeta : ext\ X \to \mathbb{R}$ such that the function $u : X \to \mathbb{R}$ given by

$$u(x) = \sum_{z \in supp\ p_x} \zeta(z) p_x(z)$$

represents $\succsim$. Moreover, ζ and u are cardinal. ext X are the extreme points of X.

It's the most abstract form of the expected utility representation.

[42] For instance, an employer might well be unable to rank all job candidates but, still, be able to rank one candidate above all the other ones.

[43] The budget correspondence is an important example of a continuous and quasi-convex menu correspondence.