Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques
Universitat de Barcelona

Utility functions and the St. Petersburg Paradox

Cristian Lorenzo Martínez

Director: José M. Corcuera Valverde
Realitzat a: Departament de Probabilitat, Lògica i Estadística. UB.
Barcelona, 20 de juny de 2013.

Contents

Part I. Abstract

The objective of this undergraduate thesis is to understand and review the fundamental aspects of the standard *Expected Utility Theory*. The expected utility theory is a model of the behaviour of the economic agent when choosing among uncertain or risky decisions. The roots of this theory can be found in Daniel Bernoulli's famous paper *Exposition of a New Theory on the measurement of Risk*. Inspired by The St. Petersburg Paradox, Bernoulli switched from the belief in an objective value of the money to the more subjective utility, which allowed to account personal differences in tastes, wealth and risk aversion in Economics, Finance and Actuarial Sciences.

The theory is built over a set of axioms that define what is a preference relation in a set. In the first chapter, we expose those axioms and discuss under which circumstances there exists a numerical representation of the preference relation.

In the second part, we define a special kind of numerical representations that are better suited to work with, specially when adopting the *monetary point of view*. That is, when we restrict the set of choices to lotteries with a monetary outcome. Those representations are called von Neumann-Morgenstern representations and require further axioms to guarantee its existence. We end this part studying the continuous case and its relationship with the weak topology.

The third part of the work defines the key concept of risk aversion and studies its relationship with concave functions. Also, we present the Arrow-Pratt Coeficient of Absolute Risk Aversion and use it to rank lotteries and obtain widely used utility functions.

The fourth part is devoted to see how the expected utility theory modifies the portfolio optimization problem. We construct the martingale and the dynamic programming methods and use them to compute the optimal terminal wealth of binomial markets. In particular, we use the binomial approximation to the Black-Scholes model to obtain the Merton's solution to the problem of maximizing the terminal utility of a portfolio.

Then we devote a whole part to study modern application of the expected utility theory. In particular,

- We analyse the mean-variance analysis under the prism of the expected utility theory.

The main result of the section is an implicit definition of certainty equivalent level curves that modify the Feasible Area and the Optimal Frontier.

- We introduce the Indifference Price Method of valuing derivatives and see that it is an extension of the risk neutral pricing in the sense that coincide with it in complete markets and allows us to set a range of buyer-seller prices in the incomplete ones.

- We end the part showing that, under the expected utility theory, the path dependent derivatives are suboptimal to risk averse agent with a fixed investment horizon.

The sixth part studies the main drawbacks of the expected utility theory and, when possible, tries to solve them refining the model. The guide of the exposition are the experimental Allais and Ellberg paradoxes, as well as the Markowitz hypothesis, and the proposed solutions are the Savage's Theory of Subjective Probabilites and the Machina's theorems on Fréchet differentiable numerical representations.

The last part of the text recovers the historical motivation of the expected utility theory, The St. Petersburg Paradox, and applies the expected utility theory to solve it, as Daniel Bernoulli did. Also, we show the insufficiency of the historical solution, via the construction of a Menger's Super-Petersburg Paradox, when not using bounded utility functions. We end discussing the implications of the boundedness hypothesis and how we obtain new paradoxes. In particular, we study signs of boundedness, without an explicit determination of the utility function, like Rabin's Calibration Theorem.

Methodologically, the sources of this work are both primary and secondary. The secondary sources are the main contributors to the exposition of the fundamentals of the theory while the primary are intensively used in the application and drawbacks of the model. Most of the results came directly from the bibliographic sources and were adapted to a common notation to keep the inner coherence of the text. The original results are limited to small propositions, expansions of known proofs, footnotes and the most part of the remarks. As a final note on the methodology, it is worth to note that the results of this work use the techniques of a rich variety of mathematical fields like Topology, Real Analysis and Measure Theory, Set Theory, Functional Analysis, Financial Engineering and Probability Theory. Hence, the understanding of the whole thesis requires some mathematical baggage.

Part II. Preference relations and expected utility

In this part, we aim to give a model that allows us to predict how an economic agent will act when facing uncertain scenarios in the future. The main idea of the model is to treat the possible scenarios as a set of random variables with different outcomes. This is a pretty general framework for work with it, as it applies to many of our day to day decisions. For example:

- It is Sunday evening and we have no food in the freezer. Should we go, for the first time, to the fancy new restaurant in the mall or should we buy pizza in the convenience shop?
- I am young and healthy. Is it good for me to contract a life insurance [1] ?
- I only have 24 EUR. What will make me happier, buy a new computer game or go to the cinema?

We do not know what will exactly happen if we choose to go to the cinema instead of buying the game, but even in the uncertainty, we have preferences and we are able to choose. And the reason is that we have expectations about the results of our actions.

To simplify the situation we will adopt the *monetary point of view*. That is, to identify the set of possible choices with a set of lotteries. The lotteries are random variables that reward or cost some money, the outcomes are real numbers, depending on the future states of nature. In the monetary point of view, the actors of the choices are economic agents that act according to a preference order defined in the set of lotteries.

This is a fruitful point of view as it helps us to attack the question about the choices and, at the same time, it gives us a strategy to solve related and important questions like how to price a product with random outcomes in incomplete markets. However, as we will see, it is not a path

[1] The problem of pricing an insurance is an historical motivator of this theory. In 1738, Daniel Bernoulli gave the first version of this model an wrote the following problem: *Suppose Caius, a Petersburg merchant, has purchased commodities in Amsterdam which he could sell for then thousand rubles if he had them in Petersburg. He therefore orders them to be shipped there by sea, but in doubt whether or not insure them.* The problem can be found in [3, p. 29] and it solved as a corollary of Proposition 5.1.

absent of difficulty; there are serious philosophical, empirical and mathematical drawbacks in the simplest form of the model.

To give an example of the kind of problems that we will face in this undergraduate thesis, lets consider a natural way to price the lotteries: the mathematical expectation [2]. Remember that, by the Law of large numbers, the mean of the outcomes of a lottery converges to its expectation. Then, it is reasonable, to consider the expectation as a fair price:

- The seller will not price the lottery far below its expectation because, otherwise, he will lose money in the long run,
- and the buyer will not accept a price far above its expectation for analogous reasons.

Unfortunately, this approach is flawed: does not reflect the reality as it does not account the diversity of behaviours we see every day. We do not need complex constructs to show this empirical failure. A simple example is the following one: lets flip a coin and consider a set Ω of two possible states of the nature:

$$\Omega = \{\omega_1 : \text{ the coin flips tails}, \omega_2 : \text{ the coin flips heads}\}.$$

Under the uncertainty of the future, we offer two possible lotteries:

(a) Do nothing, a 0 sure rewarding lottery, or

(b) Win 1 EUR if the coin flips tail and lose 1 otherwise.

The expectation of both lotteries is the same: 0 EUR. However, many will choose to play and many will not. This shows that we need to refine the process of pricing and that the expectation alone is not enough to achieve good approximations to reality; not everybody gives the same value.

Also, we have a related problem: the famous paradox of Saint Petersburg. In 1738 Daniel Bernoulli presented before the Imperial Academy of Sciences in Saint Petersburg a paper in

[2] The point of this introduction is to give a small insight of the contents of this work. For that reason, we will give only a naïve approach to illustrate the problems and we will ignore other considerations like differences of wealth between sellers and buyers, etc., that have a direct impact in the valuation of a game.

which he denies the common idea of a universal valuation of games or lotteries given by the expected value. He developed a groundbreaking theory in which the measure of money is exchanged with the measure of its utility. At the end of the paper, Bernoulli explains that the the motivation was a problem proposed by Nicholas Bernoulli:

"This is further confirmed by the following example which inspired these thoughts, and whose history is as follows: My most honorable cousin the celebrated Nicolas Bernoulli, Professor utrisque iusirs at the University of Basle, once submitted five problems to the highly distinguished mathematician Montmort. These problems are reproduced in the work L'analyse sur les jeux de hazard de M. de Montmort, p.402. The last of these problems runs as follows: Peter tosses a coin and continues to do so until it should lands 'heads' when it comes to the ground. He agrees to give Paul one ducat if he gets 'heads' on the very first throw, two ducats if he gets it on the second, four if on the third, eight if on the fourth, and so on, so that with each additional thrown the number of ducats he must pay is doubled. Suppose we seek to determine the value of Paul's expectation."

In the language of lotteries, the previous translates as follows. Let X be the lottery given by:

- Flip a fair coin until a head appears.
- If the head appears in the k-th flip, the lottery rewards 2^{k-1} EUR.

Lets compute the expected payoff of this lottery:

- We model the flip of a fair coin with a Bernoulli random variable with parameter $\frac{1}{2}$.
- Each flip is independent. Therefore, the chance of getting $k-1$ tails in a row and then a head is
$$\frac{1}{2^k}.$$
- As a result, the expectation, and the fair price, is infinite:

$$E[X] = \sum_{k=1}^{\infty} 2^{k-1} P[\text{Get } k-1 \text{ tails in a row and a head in the } k \text{ tail}] =$$

$$= \sum_{k=1}^{\infty} 2^{k-1} \frac{1}{2^k} = \sum_{k=1}^{\infty} \frac{1}{2} = \infty.$$

The paradox, is that such a bad game has an infinite fair price when, at the same time, people will hardly pay more than 3 EUR.

The good news are that both problems have a common possible solution: the utility functions . Grosso modo, utility functions are functions that modify the expected value of a lottery to reflect important factors, like the differences in the initial wealth of an economic agent or his aversion to the risk, that have a direct impact in how much an agent is willing to pay for a lottery.

The next chapters will be devoted to present the set of axioms and to discuss the intuition behind them. Also, we will study under what circumstances a preference between random variables is enough well behaved to have an associated utility functions and what are the main drawbacks of the utility function paradigm.

1 Preference relations, an axiomatic approach.

As stated before, we will give a set of axioms[3] that define what is a preference relation. The mere axioms does not lead us to the probability theory and lotteries. They refer only to a logical structure on sets and define a very general framework. The economic and stochastic point of view will be added later to refine our model to solve related problems like pricing portfolios in incomplete markets, answer some paradoxes, etc.

Let $\mathcal{X}$ be non-empty set. The elements of $\mathcal{X}$ are the possible choices of an economic agent. When facing two possible choices, the agent will choose one according to his preferences:

Definition 1.1. A **preference order** on $\mathcal{X}$ is a binary relation $\succ$ with the following properties:

- *Asymmetry*: If $x \succ y$, then $y \not\succ x$.
- *Negative transitivity*: If $x \succ y$ and $z \in \mathcal{X}$, then either $x \succ z$ or $z \succ y$ or both must hold.

[3] We owe these axioms to von Neumann and Morgenstern.

The first axiom is what defines an order. The second one, also known as *linearity*, states that if we have a preference between two choices and a third is added, then one of the former choices is still the most preferable (x if $x \succ z$) or the least preferable (y if $z \succ y$).
After setting a preference order, we automatically get a weak preference [4] and an indifference relation:

Definition 1.2. A preference order $\succ$ on $\mathcal{X}$ induces a **weak preference order** $\succeq$ defined by

$$x \succeq y \Longleftrightarrow y \not\succ x,$$

and an **indifference relation** $\sim$ given by

$$x \sim y \Longleftrightarrow x \succeq y \text{ and } y \succeq x.$$

The properties of $\succ$ are equivalent to the following properties of $\succeq$:

- *Completeness*: For all $x, y \in \mathcal{X}$, either $x \succeq y$ or $y \succeq x$.
- *Transitivity*: If $x \succeq y$ and $y \succeq z$, then either $x \succeq z$.

Remark 1.3. The induced $\succeq$ is not a total order on the set $\mathcal{X}$. Recall that a total order $\mathcal{R}$ is binary relation with the following properties:

(a) Completeness.

(b) Transitivity.

(c) Antisymmetry. I.e., if

$$x\mathcal{R}y \text{ and } y\mathcal{R}x,$$

then

$$x = y.$$

[4] Both the $\succ$ and $\succeq$ can be taken as the primitive relation from which the other can be derived. For example follmer chooses $\succ$ and chi-fu chooses $\succeq$.

The existence of the induced $\sim$ relation is what impedes $\succeq$ to be a total order, because

$$x \succ y \Leftrightarrow y \not\succ x, \qquad y \succ x \Leftrightarrow x \not\succ y,$$

implies, by definition of $\sim$,

$$x \sim y,$$

and (c) fails to be satisfied.

1.1 Existence of numerical representations

Many practical and theoretical questions not only require a preference order in a set. They also require a numerical representation of the order.

Definition 1.4. A **numerical representation** of a preference order $\succ$ is a function

$$U : \mathcal{X} \to \mathbb{R}$$

such that

$$x \succ y \Longleftrightarrow U(x) > U(y) \qquad \text{or equivalently} \qquad x \succeq y \Longleftrightarrow U(x) \geq U(y).$$

Remark 1.5. As any strictly increasing function keeps inequalities, we cannot have uniqueness when dealing with numerical representations. For example, if f is a strictly increasing real function and U is a numerical representation of a preference order $\succ$,

$$\tilde{U}(x) := f(U(x)),$$

is an equivalent numerical representation of $\succ$. It can be shown that $\succeq$ is a weak order if, and only if, $\succ$ has multiple numerical representations. That is, there exists a set $\mathcal{U}$ with more than one element such that

$$x \succ y \Leftrightarrow U(x) > U(y) \text{ for all } U \in \mathcal{U}.$$

In one of the following chapters, and given the scope of this work, we will see that taking the representations modulo positive affine transformation solves the uniqueness problem.

We now study a natural question: the existence problem. It is clear that the axioms of the previous section define a fairly general object. But, as usual, the generality comes with a price; we cannot guarantee, in general, the existence of a numerical representation.
The existence theorems require additional hypothesis. The hypothesis can be relative to

- the set of choices $\mathcal{X}$, like being finite or countable,
- or to the set of the outcomes of the lotteries involved, like being bounded.

Before giving a sufficient and necessary condition for the existence of a numerical representation, first note that the converse problem, when a real valued function on $\mathcal{X}$ represents a preference order, has an affirmative answer:

Proposition 1.6. *For any set $\mathcal{X}$ and any real valued function*

$$U : \mathcal{X} \to \mathbb{R},$$

the binary relation $\succ_U$ defined as

$$x \succ_U y \Leftrightarrow U(x) > U(y)$$

defines a preference order on $\mathcal{X}$.

Proof.

- *Asymmetry.* If $x \succ_U y$, then $U(x) > U(y)$, which implies that $U(y) > U(x)$ does not hold and $y \not\succ_u x$.
- *Negative transitivity.* By contradiction. Let $x \succ y$ and let $z \in \mathcal{X}$. If negative transitivity does not hold, then both

$$x \not\succ_U z \Rightarrow \neg(U(x) > U(z)) \Rightarrow U(z) \geq U(x)),$$

and

$$z \not\succ_U y \Rightarrow \neg(U(z) > U(y)) \Rightarrow U(y) \geq U(z)).$$

Therefore,

$$U(y) \geq U(x),$$

which contradicts $x \succ_U y$.

$\square$

The next theorem settles the question about the existence of a numerical representation of preference order on set $\mathcal{X}$.

Definition 1.7. Let $\mathcal{X}$ be an ordered set and let $\mathcal{Z} \subset \mathcal{X}$. We say that $\mathcal{Z}$ is an order dense subset of $\mathcal{X}$ if for all $x, y \in \mathcal{X} \setminus \mathcal{Z}$ such that

$$x \succ y,$$

there exists a $z \in \mathcal{Z}$ such that

$$x \succ z \succ y.$$

Theorem 1.8. *For the existence of a numerical representation of a preference relation $\succ$ it is necessary and sufficient that $\mathcal{X}$ contains a countable order dense subset $\mathcal{Z}$. In particular, any preference order admits a numerical representation if $\mathcal{X}$ is finite or countable.*

Proof. Suppose first that we are given a countable order dense subset $\mathcal{Z}$ of $\mathcal{X}$. For $x \in \mathcal{X}$, let

$$\overline{\mathcal{Z}}(x) := \{z \in \mathcal{Z} \mid z \succ x\} \text{ and } \underline{\mathcal{Z}}(x) := \{z \in \mathcal{Z} \mid x \succ z\}.$$

The relation $x \succeq y$ implies that

$$\overline{\mathcal{Z}}(x) \supseteq \overline{\mathcal{Z}}(y), \qquad \text{and} \qquad \underline{\mathcal{Z}}(y) \supseteq \underline{\mathcal{Z}}(x).$$

If the strict relation $x \succ y$ holds, then at least one of these inclusions is also strict. To see this, pick $z \in \mathcal{Z}$ with $x \succeq z \succeq y$, so that either either $x \succ z \succeq y$ or $x \succeq x \succ y$. In the first case,

$$z \in \underline{\mathcal{Z}}(x) \setminus \underline{\mathcal{Z}}(y)$$

while in the second

$$z \in \overline{\mathcal{Z}}(y) \setminus \overline{\mathcal{Z}}(x).$$

Now, take any strictly positive probability distribution μ on $\mathcal{Z}$ and let

$$U(x) := \sum_{z \in \underline{\mathcal{Z}}(x)} \mu(z) - \sum_{z \in \overline{\mathcal{Z}}(x)} \mu(z).$$

By the statement about the strict inclusions given $x \succ y$,

$$U(x) > U(y) \Leftrightarrow x \succ y,$$

so that U is the desired numerical representation.

For the proof of the converse, take a numerical representation U and let $\mathcal{J}$ denote the countable set

$$\mathcal{J} := \{[a, b] \mid a, b \in \mathbb{Q},\ a < b,\ U^{-1}([a, b]) \neq \emptyset\}.$$

For every interval $I \in \mathcal{J}$ we can choose some $z_I \in \mathcal{X}$ with $U(z_I) \in I$ and thus define the countable set

$$A := \{z_I \mid I \in \mathcal{J}\}.$$

At first glance it may seem that A is a good candidate for an order dense set. However, it may happen that there are $x, y \in \mathcal{X}$ such that $U(x) < U(y)$ and for which there is no $z \in \mathcal{X}$ with $U(x) < U(z) < U(y)$. In this case, an order dense set must contain al least one z with $U(z) = U(x)$ or $U(z) = U(y)$, a condition which cannot be guaranteed by A.

Let us define the set C of all pairs (x, y) which do not admit any $z \in A$ with $y \succ z \succ x$:

$$C := \{(x, y) \mid x, y \in \mathcal{X} \setminus A,\ y \succ x \text{ and } \nexists z \in A \text{ with } y \succ z \succ x\}.$$

Then $(x,y) \in C$ implies the apparently stronger fact that we cannot find any $z \in \mathcal{X}$ such that $y \succ z \succ x$: Otherwise we could find $a, b \in \mathbb{Q}$ such that

$$U(x) < a < U(z) < b < U(y),$$

so $I := [a,b]$ would belong to $\mathcal{J}$ and the corresponding z_I would be an element of A with $y \succ z_I \succ x$, contradicting the assumption that $(x,y) \in C$. It follows that all intervals $(U(x), U(y))$ with $(x,y) \in C$ are disjoint and non-empty. Hence, there can be only countably many of them. For each such interval J we pick now exactly one pair $(x^J, y^J) \in C$ such that $U(x^J)$ and $U(y^J)$ are the endpoints of J and we denote by B the countable set containing all x^J and y^J.
Finally we claim that $\mathcal{Z} := A \cup B$ is an order dense subset of $\mathcal{X}$. Indeed, if $x, y \in \mathcal{X} \setminus \mathcal{Z}$, then either there is some $z \in A$ such that $y \succ z \succ x$, or $(x,y) \in C$. In the latter case, there will be some $z \in B$ with $U(y) = U(z) > U(x)$ and, consequently, $y \succeq z \succ x$.

□

The previous solved the existence problem but has little or none application in the effective determination of either the numerical representation or the order-dense subset:

(a) Given a denumerable set, it is hard to compute the values of a numerical representation constructed as done in the previous theorem. For example, consider the standard order on the rational numbers:

$$x \succ y \Leftrightarrow x > y, \qquad \forall x, y \in \mathbb{Q}.$$

It is clear that we can construct a numerical representation with any summable series as follows. First, as $\mathbb{Q}$ is countable, index the elements

$$\mathbb{Q} = \{x_1, x_2, \ldots, x_n, x_{n+1}, \ldots\}. \tag{1}$$

Now, set the positive distribution

$$\mu(x_i) := \frac{1}{2^i}, \qquad i \text{ given by the indexing (1)}$$

and define

$$U(x) := \sum_{x_i \in \overline{\mathcal{Z}}(x))} \frac{1}{2^i} - \sum_{x_j \in \underline{\mathcal{Z}}(x))} \frac{1}{2^j}.$$

This is a numerical representation of the standard order but it is not feasible, nor practical, compute $U(x)$ for a given $x \in \mathbb{Q}$.

(b) On the other side, it is not always possible to determinate the order-dense subset of a given preference order. Recall that in the proof of Theorem 1.8, we define the countable collection of intervals

$$\mathcal{J} := \{[a,b] \mid a,b \in \mathbb{Q},\ a < b,\ U^{-1}([a,b]) \neq \emptyset\},$$

and, for every interval $I \in \mathcal{J}$ we choose some $z_I \in \mathcal{X}$ with $U(z_I) \in I$ to define a countable set A. In most cases, it is not easy, or even possible, to define the functional

$$\begin{array}{rccc} \psi : & \mathcal{I} \subset \mathcal{J} & \to & \mathcal{J} \\ & I & \mapsto & z_I . \end{array}$$

Therefore, we have to rely in some kind of Axiom of Choice, like the Axiom of denumerable choice [5] , to guarantee the existence of the dense order subset.

Anyway, it is still a capital result. It uses to be easier to prove theorems when dealing with countable sets and extend, via density, the result to more general sets. Hence, in most cases we will define a real functional on $\mathcal{X}$ according to empirical results, toy examples, and similar ones, consider the induced preference order of Theorem 1.6 and, by the previous Theorem, apply all the properties that derive from having an order-dense subset.

Remark 1.9. We dont need Theorem 1.8 to prove that if $\succ$ is a preference order in a finite or countable set $\mathcal{X}$, then exists a numerical representation U of $\succ$. Under the finite or countable hypothesis, we can directly construct U by the following recurrence:

[5] The Axiom of denumerable or countable choice, noted as AC_ω is a weaker version of the Axiom of choice. One formulation of AC_ω is what follows: Given a countable collection $\{A_n\}_{n\in\omega}$, there is a function ψ with domain ω and $\psi(n) \in A_n$ for each $n \in \omega$. This equivalent to state that there is some infinite subset $I \subset \omega$ and $\psi : I \to \cup_{n\in I} A_n$ with $\psi(n) \in A_n$ for every $n \in I$, which was how we chose the z_Is. See Herrlich, Horst, *Axiom of Choice*, Springer-Verlag. 2006., chapter 2.1 and 2.2.

1. Pick one element $x_1 \in \mathcal{X}$ and a real number b. Define

$$U(x_1) = b.$$

2. Now suppose that we have assigned the value of U for a set of $I := \{x_1, \ldots, x_n\}$ elements. For x_{n+1} proceed until one of the following cases happens:

   (a) If exists $x_k \in I$ such that $x_{n+1} \sim x_k$,

   $$U(x_{n+1}) = U(x_k).$$

   (b) If $x_{n+1} \succ x_k$, for all $k = 1, 2, \ldots, n$,

   $$U(x_{n+1}) = \sup_{x_i \in I} U(x_i) + 1.$$

   (c) If $x_k \succ x_{n+1}$, for all $k = 1, 2, \ldots, n$,

   $$U(x_{n+1}) = \inf_{x_i \in I} U(x_i) - 1.$$

   (d) If there are x_k, x_j such that $x_k \succ x_{n+1} \succ x_j$,

   $$U(x_{n+1}) = \frac{\inf_{x_k \succ x_{n+1}} U(x_k) + \sup_{x_{n+1} \succ x_j} U(x_j)}{2}.$$

It is clear that the function U defined above represents the preference relation $\succ$.

The only restriction we have considered in the existence theorem is on the set $\mathcal{X}$. In the next chapter we will see a well behaved kind of representations that require additional restrictions, this time in the set of outcomes.

2 Von Neumann-Morgenstern representations

In this section we exploit the idea of treating the set of choices as lotteries. The only lotteries that we consider are the ones that can be identified with simple probability distributions. The simple lotteries are enough to illustrate the key aspects of the theory. Those lotteries can be written as

$$\mu = \sum \alpha_i \delta_{x_i}, \qquad x_i \in \mathbb{R}, \qquad \sum \alpha_i = 1,$$

where α_i is the probability of getting an x_i outcome. For example,

$$\delta_{1000}$$

represents a lottery that always awards 1000 EUR. In this context, the set $\mathcal{X}$ of lotteries is a **convex set** of Borel probability measures defined in an interval $S \subset \mathbb{R}$ that contains all point masses δ_{x_i}. We will note the expected payoff, or *fair price*, of μ as

$$m(\mu) = \int_S z d\mu(z).$$

When dealing with lotteries, we will use a special class of numerical representations: *Von Neumann-Morgenstern representations.*

Definition 2.1 (von Neumann-Morgenstern representation)**.** A numerical representation of preference order $\succ$ is a von Neumann-Morgenstern representation if it can be written as:

$$U(\mu) = \int_\Omega u(z) d\mu(z), \qquad \text{for all } \mu \in \mathcal{X} \tag{2}$$

where u is a real valued function on Ω.

In probability language, a von Neumann-Morgenstern representation U is equivalent to

$$U(\mu) = E_\mu[u(z)],$$

and for that reason, they lead to the **expected utility representation** when paired with a utility function u. Before starting with the definition and the properties of a utility function, we need to ensure that $\succ$ admits a von Neumann-Morgenstern representation.

2.1 Existence and uniqueness theorem for von Neumann-Morgenstern representations

In what follows, we suppose that $\mathcal{X}$ is convex. A key property of the von Neumann-Morgenstern representations is that they are affine in $\mathcal{X}$:

Definition 2.2. A numerical representation U is affine in $\mathcal{X}$ if

$$U(\alpha\mu + (1-\alpha)\nu) = \alpha U(\mu) + (1-\alpha)U(\nu), \qquad \forall \mu, \nu \in \mathcal{X}, \quad \forall \alpha \in [0,1].$$

The main point of restricting the model to simple distributions is that we have the following equivalences:

- There exists an affine representation of $\succ$.
- There exists a von Neumann-Morgenstern representation of $\succ$.
- $\succ$ satisfies the Archimedean and the Independence Axiom.

To prove the equivalences, first we need to understand the statements. That is, to define the Archimedean and the Independence Axioms:

Definition 2.3. A preference relation $\succ$ on $\mathcal{X}$ satisfies the **independence** or **substitution** [6] axiom if for all $\mu, \nu \in \mathcal{X}$, $\mu \succ \nu$ implies

$$\alpha\mu + (1-\alpha)\tau \succ \alpha\nu + (1-\alpha)\tau,$$

for all $\tau \in \mathcal{X}$ and all $\alpha \in (0,1]$.

[6] This axioms fails to be satisfied in many empirical tests. As it is a must for having an expected utility representation, that failure is a major drawback. That topic is treated in Section 12

To understand the intuition behind the independence axiom, let us think in terms of lotteries. Suppose that μ, ν, τ are three lotteries with $\mu \succ \nu$ and $\alpha\mu + (1-\alpha)\tau$ is a compound lottery. Now we sample μ lottery with probability α and τ with probability $1-\alpha$. This is equivalent to playing directly the compound lottery. With probability $1-\alpha$, the distribution τ is drawn and there are no differences between $\alpha\mu + (1-\alpha)\tau$ and $\alpha\nu + (1-\alpha)\tau$. Otherwise, the μ is drawn and, from $\mu \succ \nu$, it seems reasonable to prefer the compound lottery with μ instead of the ν one. That shows that the satisfaction of the result of a choice in a given event does not depend on what the result would be if another event had happened: it is independent of the random value α.
The second axiom is Archimedean axiom.

Definition 2.4. A preference relation $\succ$ on $\mathcal{X}$ satisfies the *Archimedean axiom* if for any triple $\mu \succ \nu \succ \tau$, there are $\alpha, \beta \in (0,1)$ such that

$$\alpha\mu + (1-\alpha)\tau \succ \nu \succ \beta\mu + (1-\beta)\tau.$$

This axiom, that derives its name from the Archimedean principle of the Real Analysis, is also called the *continuity axiom* as it acts as a substitute of the continuity of $\succ$ in a suitable topology on $\mathcal{X}$. To be precise, if the topology [7] in $\mathcal{X}$ makes the convex combinations continuous, in other words, if

$$\alpha\mu + (1-\alpha)\tau$$

converges to τ as α decreases to zero and to μ if α increases to 1, then the continuity implies the Archimedean axiom.
As it is clear that affine representations always satisfies the independence and the Archimedean axioms, we just need to prove the converse to have the equivalences that we announced at the beginning of the chapter:

Theorem 2.5. *Suppose that $\succ$ is a preference relation in $\mathcal{X}$ satisfying both the Archimedean and the independence axiom. Then there exists an affine numerical representation U of $\succ$.*

[7] The exact topology is the weak topology. See the subsection 2.2 for a brief summary.

Moreover, U is unique up to positive affine transformations [8].

To prove this theorem we need a technical lemma:

Lemma 2.6. *Under the assumptions of Theorem 2.5, the following assertions are true:*

(a) If $\mu \succ \nu$ and $0 \leq \alpha < \beta \leq 1$, then

$$\beta\mu + (1-\beta)\nu \succ \alpha\mu + (1-\alpha)\nu.$$

(b) If $\mu \succ \nu$ and $\mu \succ \tau \succ \nu$, then there exists a unique $\alpha \in [0,1]$ with

$$\tau \sim \alpha\mu + (1-\alpha)\nu.$$

(c) If $\mu \sim \nu$, then

$$\alpha\mu + (1-\alpha)\tau \sim \alpha\nu + (1-\alpha)\tau$$

for all $\alpha \in [0,1]$ and all $\tau \in \mathcal{X}$.

Proof.

(a) Let $\lambda := \beta\mu + (1-\beta)\nu$. The independence axiom implies that

$$\lambda \succ \beta\nu + (1-\beta)\nu = \nu.$$

Hence, for $\gamma := \frac{\alpha}{\beta}$,

$$\beta\mu + (1-\beta) = (1-\gamma)\lambda + \gamma\lambda \succ (1-\gamma)\nu + \gamma\lambda = \alpha\mu + (1-\alpha)\nu.$$

(b) By (a), if α exists then it is unique. To show existence, we need only to consider the case $\mu \succ \lambda \succ \nu$, for otherwise we can take either $\alpha = 0$ or $\alpha = 1$. The candidate is

$$\alpha := \sup\{\gamma \in [0,1] \mid \lambda \succeq \gamma\mu + (1-\gamma)\nu\}.$$

[8] An affine transformation is a map of the form $f(U) = AU + b$, with $A > 0$.

If $\lambda \sim \alpha\mu + (1-\alpha)\nu$ is not true, the on of the following two possibilities must occur:

$$\lambda \succ \alpha\mu + (1-\alpha)\nu, \qquad \alpha\mu + (1-\alpha)\nu \succ \lambda.$$

In the first case, by the Archimedean axiom, we obtain $\beta \in (0,1)$ such that

$$\lambda \succ \beta[\alpha\mu + (1-\alpha)\nu] + (1-\beta)\mu = \gamma\mu + (1-\gamma)\nu \tag{3}$$

for

$$\gamma = 1 - \beta(1-\alpha).$$

Since $\gamma > \alpha$, it follows from the definition of α that

$$\gamma\mu + (1-\gamma)\nu \succ \lambda,$$

which contradicts (3). In the second case, the Archimedean axioms yields some $\beta \in (0,1)$ such that

$$\beta(\alpha\mu + (1-\alpha)\nu) + (1-\beta)\nu = \beta\alpha\mu + (1-\beta\alpha)\nu \succ \lambda. \tag{4}$$

As $\beta < 1$, $\beta\alpha < \alpha$ and the definition of α yields some $\gamma \in (\beta\alpha, \alpha]$ with $\lambda \succeq \gamma\mu + (1-\gamma)\nu$. Part (a) and the fact that $\beta\alpha < \gamma$ imply that

$$\lambda \succeq \gamma\mu + (1-\gamma)\nu \succ \beta\alpha\mu + (1-\beta\alpha)\nu,$$

which contradicts (4).

(c) We must exclude both of the following two possiblities

$$\alpha\mu + (1-\alpha)\lambda \succ \alpha\nu + (1-\alpha)\lambda \tag{5}$$

and

$$\alpha\nu + (1-\alpha)\lambda \succ \alpha\mu + (1-\alpha)\lambda. \tag{6}$$

To this end, we may assume that there exists some $\rho \in \mathcal{X}$ with

$$\rho \not\sim \mu \equiv \nu;$$

otherwise the result is trivial. Let us assume that $\rho \succ \mu \sim \nu$; the case in which $\mu \sim \nu \succ \rho$ is similar. Suppose that (5) would occur. The independence axiom yields

$$\beta\rho + (1-\beta)\nu \succ \beta\nu + (1-\beta)\nu = \nu \sim \mu,$$

for all $\beta \in (0,1)$. Therefore,

$$\alpha[\beta\rho + (1-\beta)\nu] + (1-\alpha)\lambda \succ \alpha\mu + (1-\alpha)\lambda, \qquad \forall\, \beta \in (0,1). \tag{7}$$

Using the assumption of (5), we obtain from part (b) a unique $\gamma \in (0,1)$ such that, for any fixed β,

$$\alpha\mu + (1-\alpha)\lambda \sim \gamma(\alpha[\beta\rho + (1-\beta)\nu] + (1-\alpha)\lambda) + (1-\gamma)[\alpha\nu + (1-\alpha)\lambda] =$$

$$\alpha[\beta\gamma\rho + (1-\beta\gamma)\nu] + (1-\alpha)\lambda \succ \alpha\mu + (1-\alpha)\lambda,$$

where we have used (7) for β replaced by $\beta\gamma$ in the last step. This is a contradiction. The possibility (6) is excluded by analogous argument.

□

Now we can proof the theorem 2.5:

Proof. We give a constructive proof. Fix two lotteries λ and ρ with $\lambda \succ \rho$ and define

$$\mathcal{X}(\lambda,\rho) := \{\mu \in \mathcal{X} \mid \lambda \succeq \mu \succeq \rho\}.$$

The assertion is trivial if no such pair $\lambda \succ \rho$ exists- just take a constant U. Otherwise, if

$\mu \in \mathcal{X}(\lambda, \rho)$, part (b) of Lemma 2.6 yields a unique $\alpha \in [0,1]$ such that

$$\mu \sim \alpha\lambda + (1-\alpha)\rho.$$

Then we set $U(\mu) := \alpha$. To prove that U is a numerical representation of $\succ$ on $\mathcal{X}(\lambda, \rho)$, we must show that for $\nu, \mu \in \mathcal{X}(\lambda, \rho)$, we have

$$U(\mu) > U(\nu) \Longleftrightarrow \mu \succ \nu.$$

To prove the sufficiency, we apply part (a) of Lemma 2.6 to conclude that

$$\mu \sim U(\mu)\lambda + (1-U(\mu))\rho \succ U(\nu)\lambda + (1-U(\nu))\rho \sim \nu.$$

Hence $\mu \succ \nu$.

Conversely, if $\mu \succ \nu$ the the preceding arguments already imply that we cannot have $U(\nu) > U(\mu)$. Thus, it suffices to rule out the case $U(\mu) = U(\nu)$. But if $U(\mu) = U(\nu)$, then by the definition of U, we have $\mu \sim \nu$, which contradicts $\mu \succ \nu$. We have proved that U is a numerical representation of $\succ$ restricted to $\mathcal{X}(\lambda, \rho)$.

Let us show now that $\mathcal{X}(\lambda, \rho)$ is a convex set. Take $\mu, \nu \in \mathcal{X}(\lambda, \rho)$ and $\alpha \in [0,1]$. Then

$$\lambda \succeq \alpha\lambda + (1-\alpha)\nu \succeq \alpha\mu + (1-\alpha)\nu,$$

using the independence axiom to handle the cases $\lambda \succ \nu$ and $\lambda \succ \mu$, and part (c) of Lemma 2.6 for $\lambda \succ \nu$ and for $\lambda \succ \mu$.By the same argument it follows that $\alpha\mu + (1-\alpha)\nu \succeq \rho$, which implies the convexity of the set $\mathcal{X}(\lambda, \rho)$. Therefore, $U(\alpha\mu + (1-\alpha)\nu)$ is well defined.

Now we have to show that U is affine. i.e.

$$U(\alpha\mu + (1-\alpha)\nu) = \alpha U(\mu) + (1-\alpha)U(\nu).$$

To this end, we apply part (c) of Lemma 2.6 twice:

$$\alpha\mu + (1-\alpha)\nu \sim \alpha(U(\mu)\lambda + (1-U(\mu))\rho) + (1-\alpha)(U(\nu)\lambda + (1-U(\nu))\rho) =$$

$$= [\alpha U(\mu) + (1-\alpha)U(\nu)]\lambda + [1 - \alpha U(\mu) - (1-\alpha)U(\nu)]\rho.$$

The definition of U and the uniqueness in part (b) of Lemma 2.6 imply that

$$U(\alpha\mu + (1-\alpha)\nu) = \alpha U(\mu) + (1-\alpha)U(\nu),$$

and U is an affine numerical representation of $\succ$ on $\mathcal{X}(\lambda,\rho)$.

Now we prove the uniqueness up to positive affine transformations of U on $\mathcal{X}(\lambda,\rho)$. Let $\tilde{U}$ be another affine numerical representation of $\succ$ on $\mathcal{X}(\lambda,\rho)$, and define

$$\hat{U} := \frac{\tilde{U}(\mu) - \tilde{U}(\rho)}{\tilde{U}(\lambda) - \tilde{U}(\rho)}, \qquad \mu \in \mathcal{X}(\lambda,\rho).$$

Then $\hat{U}$ is a positive affine transformation of $\tilde{U}$, and $\hat{U}(\rho) = 0 = U(\rho)$ as well as $\tilde{U}(\lambda) = 1 = U(\lambda)$. Hence, affinity of $\hat{U}$ and the definition of U imply

$$\hat{U}(\mu) = \hat{U}(U(\mu)\lambda + (1-U(\mu))\rho) = U(\mu)\hat{U}(\lambda) + (1-U(\lambda))\hat{U}(\rho) = U(\mu)$$

for all $\mu \in \mathcal{X}(\lambda,\rho)$. Thus $\hat{U} = U$.

Finally we extend U to the full space $\mathcal{X}$. To this end, we first take $\tilde{\lambda},\tilde{\rho} \in \mathcal{X}$ such that $\mathcal{X}(\tilde{\lambda},\tilde{\rho}) \supset \mathcal{X}(\lambda,\rho)$. By the arguments in the first part of this proof, there exists an affine numerical representation $\tilde{U}$ of $\succ$ on $\mathcal{X}(\tilde{\lambda},\tilde{\rho})$, and we may assume that $\tilde{U}(\lambda) = 1$ and $\tilde{U}(\mu) = 0$; otherwise we apply a positive affine transformation to $\tilde{U}$. By the previous step of the proof, $\tilde{U}$ coincides with U on $\mathcal{X}(\lambda,\rho)$, and so $\tilde{U}$ is the unique consistent extension of U. Since each lottery belongs to some set $\mathcal{X}(\tilde{\lambda},\tilde{\rho})$, the affine numerical representation U can be uniquely extended to all of $\mathcal{X}$.

□

As announced, we have solved the existence, and uniqueness up to positive affine transformations, of the von Neumann-Morgenstern representations when the lotteries of the set $\mathcal{X}$ are simple probability distributions:

Corollary 2.7. *Suppose that $\mathcal{X}$ is the set of all simple probability distributions on S and that $\succ$ is a preference order on $\mathcal{X}$ that satisfies both the Archimedean and the independence axiom. Then there exists a von Neumann-Morgenstern representation U. Moreover, both U and u are unique up to positive affine transformations.*

Proof. Let U be an affine numerical representation of $\succ$. That representation exists and is unique by theorem 2.5. We define

$$u(x) = U(\delta_x).$$

If μ is simple, X is of the form

$$\mu = \sum_{i=1}^{N} \alpha_i \delta_{x_i}, \qquad x_i \in S, \quad \sum_{i=1}^{N} \alpha_i = 1.$$

Then, the affine character of U implies

$$U(\mu) = U(\sum_{i=1}^{N} \alpha_i \delta_{x_i}) = \sum_{i=1}^{N} \alpha_i U(\delta_{x_i}) = \int_S u(z) d\mu(z).$$

□

Corollary 2.8. *Suppose that $\mathcal{X}$ is the set of all probability distributions on a finite set S and that $\succ$ is a preference order in $\mathcal{X}$ that satisfies both the Archimedean and the independence axiom. Then there exists a von Neumann-Morgenstern representation U. Moreover, both U an u are unique up to positive affine transformations.*

Proof. On finite sets, the probability distributions are simple. □

In the Theorem 1.8 we have shown that having a numerical representation of a $\succ$ is equivalent to the existence of an order-dense subset. Then, a natural question is if the order-dense subset of the simple lotteries is independent of the preference order $\succ$. The answer is negative as shown by the following example.

Example 2.9. Let $\mathcal{X}$ be the set of simple lotteries with outcomes in the closed interval $S = [0,1]$. Consider the following two numerical representations:

$$U(\mu) = \int_S x \mathbb{1}_{\mathbb{R}\setminus\mathbb{Q}}(x) d\mu(x), \qquad V(\mu) = \int_S x d\mu(x).$$

Those representation induce, respectively, the preference orders $\succ$ and $\tilde{\succ}$ in $\mathcal{X}$ that satisfies both the independence and Archimedean axiom, they have von Neumann-Morgenstern representations, but the order-dense subsets are not the same: it is clear that the degenerated lotteries

$$\delta_{x_i}, \qquad x_i \in \mathbb{Q},$$

are dense in $\tilde{\succ}$, but they cannot be dense in $\succ$ as all are equivalent to δ_0 with respect to $\succ$.

We have set heavy restrictions on $\mathcal{X}$ and S in order to have von Neumann-Morgenstern representations of $\succ$. The following examples show what kind of problems we can face in more general scenarios:

Example 2.10. Let $\mathcal{X}$ be the set on probability measures on $S := \mathbb{N}$ for which

$$U(\mu) := \lim_{k\uparrow\infty} k\mu(K)$$

exists and is finite. U is affine and induces a preference order on $\mathcal{X}$ which satisfies both the Archimedean and the independe axiom. However, U does not admit a von Neumann-Morgenstern representation.

Example 2.11. Let $\mathcal{X}$ be the set of all Borel probability measures on $S = [0,1]$, and denote by λ the Lebesgue measure on S. According to Lebesgue decomposition theorem, for every $\mu \in \mathcal{X}$, μ can be decomposed as

$$\mu = \mu_s + \mu_a,$$

where μ_s is singular with respect to λ and μ_a is absolutely continuous. We define

$$U(\mu) := \int_S x d\mu_a(x).$$

U is affine on $\mathcal{X}$ and induces a preference order $\succ$ on $\mathcal{X}$ which satisfies both the Archimedean and independence axioms. But $\succ$ cannot have a von Neumann-Morgenster representation. Recall that λ is singular with respect to all the δ_{x_i}, $x_i \in [0,1]$. That implies that $U(\delta_{xi}) = 0$ and u must be equal to zero. Then, the preference relation is the trivial in the sense

$$\lambda \equiv \mu, \qquad \forall \mu \in \mathcal{X},$$

but that contradicts, for example,

$$U(\lambda) = \frac{1}{2}, \qquad U(\delta_{\frac{1}{2}}) = 0.$$

2.2 A brief comment about the continuity of the preference orders

In the previous section we have proved that all affine representations of simple lotteries are von Neumann-Morgenstern representations too. But also we saw, with two examples, that we cannot expect to have, in general, von Neumann-Morgenstern representations without imposing additional conditions. One way to obtain representations is asking for continuity. As continuity is a topological property, we need to endow our set $\mathcal{X}$ with a good enough topology. As we will see, the weak topology is one possible choice. First let us define what is a continuous preference order.

Definition 2.12. Let $\mathcal{X}$ be a topological space. A preference relation $\succ$ is called **continuous** if for all $x \in \mathcal{X}$

$$\overline{\mathcal{B}}(x) := \{y \in \mathcal{X} \mid y \succ x\} \qquad \text{and} \qquad \underline{\mathcal{B}}(x) := \{y \in \mathcal{X} \mid x \succ y\}. \tag{8}$$

are open subset of $\mathcal{X}$.

If we recall the proof of the Theorem 1.8, we used similar pair of set to construct a numerical representation with a summable series or a positive distribution. We can use the same proof to construct a numerical representation if we suppose $\mathcal{X}$ connected.

Theorem 2.13. *Let $\mathcal{X}$ be a connected topological space with a continuous preference order $\succ$.*

Then every dense subset $\mathbb{Z}$ of $\mathcal{X}$ is also order dense in $\mathcal{X}$. In particular there exists a numerical representation of $\succ$ if $\mathcal{X}$ is separable.

Proof. Take $x, y \in \mathcal{X}$ with $y \succ x$, and consider the sets

$$\overline{\mathcal{B}}(x) := \{u \in \mathcal{X} \mid u \succ x\}, \qquad \underline{\mathcal{B}}(y) := \{v \in \mathcal{X} \mid v \succ y\},$$

given by the definition of continuous order. Since $y \in \overline{\mathcal{B}}(x)$ and $x \in \underline{\mathcal{B}}(y)$, neither are empty sets. Moreover, negative transitivity implies that

$$\mathcal{X} = \overline{\mathcal{B}}(x) \cup x \in \underline{\mathcal{B}}(y).$$

Hence, the open sets $\overline{\mathcal{B}}(x)$ and $\underline{\mathcal{B}}(y)$ cannot be disjoint as $\mathcal{X}$ is connected. Thus, the open set $\overline{\mathcal{B}}(x) \cap x \in \underline{\mathcal{B}}(y)$ must contain some element z of the dense set $\mathcal{Z}$, which then satisfies $y \succ z \succ x$. Therefore $\mathcal{Z}$ is an order dense subset of $\mathcal{X}$. If $\mathcal{X}$ is separable, then there exists a countable dense subset $\mathcal{Z}$ of $\mathcal{X}$, which is also order dense. Hence, we conclude from Theorem 1.8 that there exists a numerical representation of $\succ$. $\square$

With additional assumptions, we can also guarantee that the representation is also continuous:

Theorem 2.14. *Let $\mathcal{X}$ be a topological space which satisfies at least one of the following two properties:*

- *$\mathcal{X}$ has a countable base of open sets (Second countable space).*
- *$\mathcal{X}$ is separable and connected.*

Then every continuos preference order on $\mathcal{X}$ admits a continuous numerical representation.

Proof. A proof can be found on Debreu, G., *Continuity properties of paretian utility.* International Econ. Rev. 5 (1964), 285-293. $\square$

This continuous representation can be a von Neumann-Morgenstern representation if we choose the right topology: the *weak topology.* The weak topology is a kind of initial topology with respect to its dual topological space. That is, with respect to the set of continuous linear forms. The exact definition is what follows:

Definition 2.15. The **weak topology** on the set $\mathcal{M}(S)$ of non-negative finite measures defined in a metric space S is the coarsest topology for which all mappings

$$\mu \mapsto \int_S f \, d\mu, \qquad \mu \in \mathcal{M}(S), \quad f \text{ bounded and continuous on } S.$$

are continuous.

The weak convergence is equivalent to other kind of convergence familiar to anybody that has taken a course in probabilities: the convergence in law or in distribution. Recall that the convergence in law is defined as follows:

Definition 2.16. Let be $(X_n)_n$ a sequence of random variables and let be $(F_n)_n$ the corresponding sequence of distribution functions. We say that X_n converges to X in law if

$$\lim_n F_n(x) = F(x)$$

in every point $x \in \mathbb{R}$ at which F is continuous.

The exact relation between the convergence in law, the definition 2.15 and other characterizations is given by this capital theorem:

Theorem 2.17 (Portmanteau Lemma)**.** *Let S be a metric space with its Borel σ-algebra Σ. We say that a sequence of probability measures μ_n on (S, Σ) converges weakly to the probability measure μ if any of the following equivalent conditions is true:*

(a) $E_{\mu_n}[f] \to E_\mu[f]$ *for all bounded continuous functions f.*

(b) $E_{\mu_n}[f] \to E_\mu[f]$ *for all bounded and Lipschitz functions f.*

(c) $\limsup E_{\mu_n}[f] \leq E_\mu[f]$ *for every upper semi-continuous function f bounded from above.*

(d) $\liminf E_{\mu_n}[f] \geq E_\mu[f]$ *for every upper semi-continuous function f bounded from below.*

(e) $\limsup \mu_n(C) \leq \mu(C)$ *for all closed sets* C *of the space* S.

(f) $\liminf \mu_n(U) \leq \mu(U)$ *for all open sets* U *of the space* S.

(g) $\lim \mu_n(A) = \mu(A)$ *for all continuity sets* A *of the measure* μ.

Hence, when $S = \mathbb{R}$ with the usual topology, the convergence in law and the weak convergence are the same.

In our framework, the weak topology has all the hypothesis of the previous theorems:

Theorem 2.18. *The space* $\mathcal{M}(S)$ *is separable and metrizable for the weak topology. If* S *is complete, then so is* $\mathcal{M}(S)$. *Moreover, if* S_0 *is a dense subset of* S, *then the set*

$$\left\{ \sum_{i=1}^{n} \alpha_i \delta_{x_i} \mid \alpha_i \in \mathbb{Q}^+, \ x_i \in S_0, \ n \in \mathbb{N} \right\}$$

of simple measures on S_0 *with rational weights is dense in* $\mathcal{M}(S)$ *for the weak topology.*

And that allows us to announce the following theorem:

Theorem 2.19. *Let* $\mathcal{M} := \mathcal{M}_1(S)$ *be the space of all probability measures on* S *endowed with the weak topology, and let* $\succ$ *be a continuous preference order on* $\mathcal{M}$ *satisfying the independence axiom. Then there exists a von Neumann-Morgenstern representation*

$$U(\mu) = \int u(x) d\mu(x)$$

for which the function $u : S \to \mathbb{R}$ *is bounded and continuous. Moreover,* U *and* u *are unique up to positive affine transformations.*

This theorem involve only bounded functions u and assumptions about the continuity of $\succ$. But, nevertheless, it is general enough to construct, as a corollary, the following result:

Corollary 2.20. *Let* $\succ$ *be a preference order on the space* $\mathcal{M}_b(S)$ *of boundedly supported measures*

$$M_b(S) := \cup_{r>0} \mathcal{M}_1(\overline{B}(x_0, r)) = \{\mu \in \mathcal{M}_1(S) \mid \mu(\overline{B}(x_0, r)) = 1 \text{ for some } r \geq 0, \ x_0 \in S\}$$

whose restriction to each $\mathcal{M}_1(\overline{B}(x_0, r))$ is continuous with respect to the weak topology. If $\succ$ satisfies the independence axiom, then there exists a von Neumann-Morgenstern representation

$$U(\mu) = \int_S u(x) d\mu(x)$$

with a continuous function $u : S \to \mathbb{R}$. Moreover, U and u are unique up to positive affine transformations.

As a final word about the continuous case, we can try an axiomatic approach instead of a topological one. With a third axiom, in addition to the independence and Archimedean axioms, we can construct von Neumann-Morgenstern representations [9] : the *sure thing principle.*

Definition 2.21. For any $\mu, \nu \in \mathcal{M}$ and any measurable set A such that $\mu(A) = 1$, we have

- If $\delta_x \succ \nu$ for all $x \in A$, then $\mu \succ \nu$.
- If $\nu \succ \delta_x$ for all $x \in A$, then $\nu \succ \mu$.

Note that the sure-thing principle is violaten in both Examples 2.10 and 2.11.

Part III. Risk Aversion

3 Introduction and definitions

In this section we will introduce the key concepts of expected utility and risk aversion. To give an insight of the behaviours that we want to model with those concepts, let us consider an example. Like we did in the introduction, consider an economic agent with two possible choices:

(a) A sure 4 EUR lottery,

$$\mu = \delta_4,$$

[9] See Fishburn, Peter P., *Utility theory for decision making.* Pub. Operations Res. 18, Johm Wiley, New York 1970.

(b) and a lottery tied to the result of flipping a coin:

$$\nu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{10}.$$

The expected payoff of the lotteries is

$$m(\mu) = 4, \qquad m(\nu) = \frac{0+10}{2} = 5.$$

The second lottery has a higher expected payoff, but nevertheless, a large amount of agents will choose the first one.

When we observe this behaviour, we say that the agent shows *risk aversion* and it is equivalent to say that he will not accept to pay the fair price of a lottery. The risk aversion has direct consequences in many financial problems, for example:

- Given a risk aversion agent, how big has to be the expected return of a risky asset over the return of the riskless asset to buy it?
- How can we measure the risk aversion? Does it affect the form of von Neumann-Morgenstern representation?
- Is correct to optimize portfolios under the Mean Variance paradigm?
- ...

Our model can give an answer those questions. Recall that, under some assumptions, the preference order of an agent has a von Neumann-Morgenstern representation:

$$U(\mu) = \int u(z)d\mu(z), \qquad \forall \mu \in \mathcal{X}, \qquad u \text{ real valued.}$$

Because u is the only factor that varies across the possible preference orders of a given set $\mathcal{X}$, it is clear that the risk aversion has to be equivalent to some properties of the function u. In fact, it is equivalent to concavity.

Before proving that statement, we will define a property that, as we will see later, we expect to be true in most preference orders:

Definition 3.1. A preference relation $\succ$ on $\mathcal{X}$ is called **monotone** if

$$x > y \text{ implies } \delta_x \succ \delta_y.$$

The preference relation is called (strictly) **risk averse** if $\forall \mu \in \mathcal{X}$,

$$\delta_{m(\mu)} \succ \mu \quad \text{unless } \mu = \delta_{m(\mu)}.$$

Now we can prove the following equivalences:

Theorem 3.2. *Suppose that a preference relation $\succ$ has a von Neumann-Morgenstern representation*

$$U(\mu) = \int_S u d\mu.$$

Then:

(a) $\succ$ is monotone if and only if u is strictly increasing.

(b) $\succ$ is risk averse if and only if u is strictly concave.

Proof.

(a) Monotonicity means

$$u(x) = U(\delta_x) > U(\delta_y) = u(y) \qquad \text{for } x > y.$$

(b) If $\succ$ is risk averse,

$$\delta_{\alpha x+(1-\alpha)y)} \succ \alpha\delta_x + (1-\alpha)\delta_y$$

holds for all distinct $x, y \in S$ and $\alpha \in (0,1)$. Therefore,

$$u(\alpha x + (1-\alpha)y) > \alpha u(x) + (1-\alpha)u(y),$$

and u is strictly concave. Conversely, if u is strictly concave, Jensen's inequality implies risk aversion:

$$U(\delta_{m(X)}) = u\left(\int_S z d\mu(z)\right) \geq \int_S u(z) d\mu(z) = U(X),$$

with equality if and only if $\mu = \delta_{m(\mu)}$.

$\square$

Monotone and concave functions play an important role in finance and, for that reason, they have their own name:

Definition 3.3. A function $u : S \to \mathbb{R}$ is called a *utility function* if it is strictly concave, strictly increasing and continuous on S.
An *expected utility representation* is a von Neumann-Morgenstern representation

$$U(\mu) = \int_S u(z) d\mu(z)$$

where u is a utility function.

Remark 3.4. The utility functions are also called *Bernoulli utility functions* because Daniel Bernoulli, and independently Gabriel Cramer, introduced them to answer the St. Petersburg Paradox. We will review their proposed solutions in Section 15.

4 Measuring and understanding the Risk Aversion

By the definition of risk aversion, a risk averse agent prefers the *fair price* of a lottery, its expected payoff, over the lottery. But, due to the concavity of the utility function, he may still prefer an even lower sure amount. The amount of money that he is willing to sacrifice for a sure reward is called the *risk premium*:

Definition 4.1. Let u be a utility function and μ a lottery. The certainty equivalent $c(\mu, u)$ of μ relative to u is the real number that satisfies the following equality

$$u(c(\mu, u)) = \int_S u(z) \mu(dz).$$

The risk premium of X is the number

$$\rho(\mu, u) = m(\mu) - c(\mu, u).$$

Remark 4.2. From the definition 4.1, we get

$$\mu \sim c(\mu, u).$$

The next figure illustrates relates the concepts of certainty equivalent, utility and expected value:

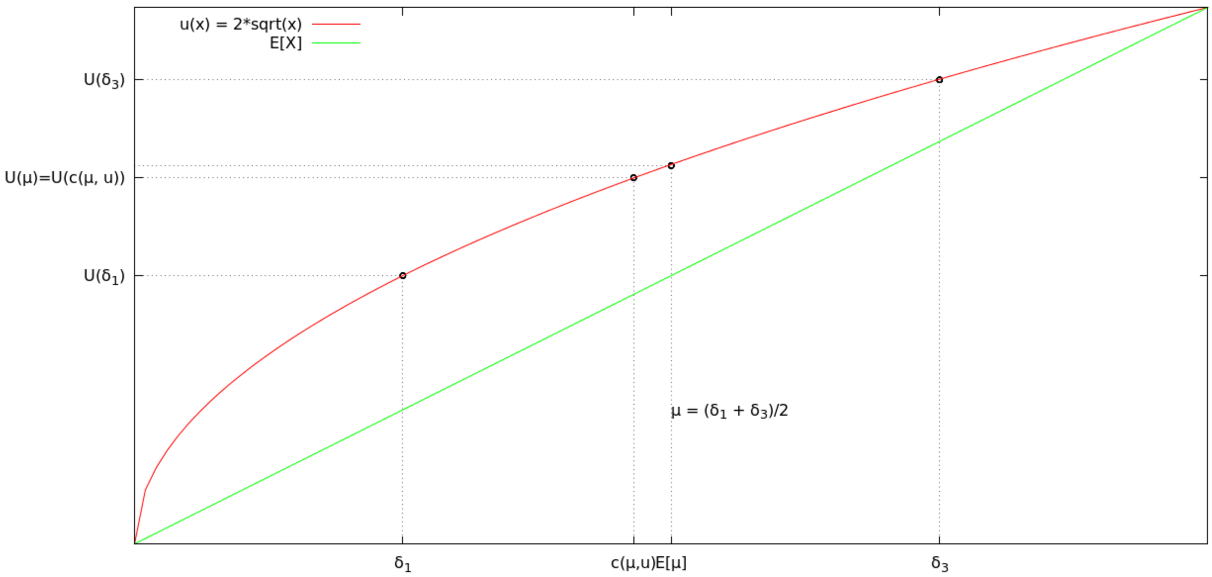

Fig. 1: Certainty equivalent of a Bernoulli lottery.

Observe that the gap between the certainty equivalent and the expected value corresponds to the risk premium of the lottery and that the expected utility is in the middle of the utility of the degenerated lotteries.

4.1 The Arrow-Pratt Measures

From the definition of the certainty equivalent, the risk premium can be viewed as the amount that an agent would pay to replace a lottery by its expected or fair value. Then, it is reasonable to expect that the risk premium can be used to compare the risk aversion between economic agents. To formalize this idea, lets consider the Taylor expansion of a sufficiently smooth utility function $u(x)$ around the fair price m of a lottery μ with finite variance,

$$u(x) = u(m) + u'(m)(x - m) + \dots .$$

Now evaluate at the certainty equivalent $c(\mu, u)$ and truncate the Taylor expansion to obtain

$$u(c(\mu, u)) \approx u(m) + u'(m)(c(\mu, u) - m) = u(m) - u'(m)\rho(\mu, u).$$

From the definition of the certainty equivalent and the Taylor expansion of u, we also get the following expression,

$$u(c(\mu, u)) \int u(x) d\mu(x) = \int \left[u(m) + u'(m)(x - m) + \frac{1}{2}u''(m)(x - m)^2 + r(x) \right],$$

where $r(x)$ is the remainder of the Taylor expansion of u. If we truncate again and evaluate the integral, we obtain

$$u(c(\mu, u)) \approx u(m) + \frac{1}{2}u''(m)\,\mathrm{var}(\mu).$$

Joining the two previous approximations, results

$$\rho(\mu, u) \sim -\frac{1}{2}\frac{u''(m)}{u'(m)}\,\mathrm{var}(\mu) = \frac{1}{2}\alpha(m)\,\mathrm{var}(\mu). \tag{9}$$

In many contexts, like the mean-variance paradigm, the risk of a random return is identified with its variance. Therefore, from (9), we see that α is the factor by which an economic agent weights the associated risk of a lottery and its premium risk.

Definition 4.3. Let u be a twice continuous differentiable utility function defined on S. Then,

$$\alpha(x) := -\frac{u''(x)}{u'(x)}$$

is called the **Arrow-Pratt coefficient of absolute risk aversion** of u at level x.

Remark 4.4. Given a twice differentiable function, the quotient of its second and first derivatives cancels any product by strictly positive scalars. That shows that the Arrow-Pratt coefficient of absolute risk aversion is invariant to strictly positive affine transformations and that the second derivative is not enough to characterize the risk aversion.

The Arrow-Pratt coefficient of absolute risk aversion is good tool to compare the risk aversion of two individuals:

Proposition 4.5. *Suppose that u and $\tilde{u}$ are two utility functions on S, which are twice continuous differentiable, and that α and $\tilde{\alpha}$ are the corresponding Arrow-Pratt coefficients of absolute risk aversion. Then the following conditions are equivalent*

(a) $\alpha(x) \geq \tilde{\alpha}(x)$, for all $x \in S$.

(b) $u = F \circ \tilde{u}$ for a strictly increasing concave function F.

(c) The respective risk premiums satify

$$\rho(\mu, u) \geq \rho(\mu, \tilde{u})$$

for all $\mu \in \mathcal{X}$.

Proof.

- (a) $\Rightarrow$ (b): Since $\tilde{u}$ is strictly increasing, it is a utility function, we may define its inverse function w. Then

$$F(t) := u(w(t))$$

is clearly increasing, twice differentiable, and satisfies

$$u = F \circ \tilde{u}.$$

Now we have to check that F is concave and strictly increasing. The first two derivatives of w are

$$w' = \frac{1}{\tilde{u}'(w)}, \qquad w'' = \tilde{\alpha}(w)\frac{1}{\tilde{u}'(w)^2}.$$

Therefore, the two first derivatives of F are, by the chain rule,

$$F' = u'(w) \cdot w' = \frac{u'(w)}{\tilde{u}'(w)} > 0,$$

$$F'' = \frac{\tilde{u}'(w)u''(w) - u'(w)\tilde{u}''(w)}{\tilde{u}'(w)^3} = \frac{u'(w)}{\tilde{u}'(w)^2}\left[\tilde{\alpha} - \alpha\right].$$

By hypothesis, $\alpha(x) \geq \tilde{\alpha}(x)$, for all $x \in S$ and u' is non negative and it follows

$$F'' \leq 0.$$

This proves that F is concave and strictly increasing.

- (b) $\Rightarrow$ (c): By the Jensen's inequality,

$$u(c(\mu,u)) = \int u d\mu = \int F \circ \tilde{u} d\mu \leq F\left(\int \tilde{u} d\mu\right) = F(\tilde{u}(c(\mu,\tilde{u})) = u(c(\mu,\tilde{u}). \tag{10}$$

Hence,

$$\rho(\mu,u) = m(\mu) - c(\mu,u) \geq m(\mu) - c(\mu,\tilde{u}) = \rho(\mu,\tilde{u}).$$

- (c) $\Rightarrow$ (a): Proceed by contradiction. If condition (a) is false, there exists an open interval $O \subset S$ such that

$$\tilde{\alpha}(x) > \alpha x, \qquad \text{for all } x \in O.$$

Let

$$\tilde{O} := \tilde{u}(O),$$

and denote again by w the inverse of $\tilde{u}$. Then the function

$$F(t) = u(w(t))$$

will be strictly convex in the open interval $\tilde{O}$, as seen in the first part of the proof. Thus, if μ is a measure with support in O, the inequality (10) is reversed and is even strict-unless μ is concentrated in a single point. It follows that

$$\rho(\mu, u) < \rho(\mu, \tilde{u}),$$

which contradict the hypothesis (c).

$\square$

4.2 Study of common families utility functions

The Arrow-Pratt coefficient of absolute risk aversion, α, is not only a way to compare utility functions. It also allows us to construct them. Instead of computing α from u, we can give a function

$$\alpha(x)$$

defined on the usual interval S and solve the differential equation

$$\alpha(x) = -\frac{u''(x)}{u'(x)}$$

to obtain the utility function u. The following are two standard examples of utility functions constructed this way:

Example 4.6 (Constant absolute risk aversion (CARA))**.** If we set a constant $\alpha(x) > 0$ for all $x \in S$, we obtain

$$\alpha = -\frac{u''(x)}{u'(x)} = -\log(u')'(x).$$

The functions

$$u(x) = a - be^{-\alpha x},$$

solve the ODE. If we apply an affine transformation, we can obtain

$$u(x) = -e^{-\alpha x}.$$

This is the **negative exponential utility function** and it is bounded from above.

Remark 4.7. The negative exponential utility function induces a preference on the lotteries that does not depend on the initial wealth w. We can interpret the initial wealth as translation factor in the outcomes of set of lotteries. Then, for all $\mu \in \mathcal{X}$,

$$\int u(x+w)d\mu(x) = \int -e^{-\alpha(x+w)}d\mu(x) = -e^{-\alpha w}\int -e^{-\alpha x}d\mu(x),$$

and,

$$\int u(x+w)d\mu(x) > \int u(x+w)d\nu(x) \Leftrightarrow \int -e^{-\alpha x}d\mu(x) < \int -e^{-\alpha x}d\nu(x).$$

which means that w does not impact at all in the preference order.

Example 4.8 (Hyperbolic absolute risk aversion(HARA))**.** Set

$$\alpha(x) = \frac{1-\gamma}{x}$$

on $S = (0, \infty)$ for some $\gamma \in [0, 1)$. Then, up to affine transformations, we have

$$u(x) = \log x \qquad \text{for } \gamma = 0, \qquad u(x) = \frac{1}{\gamma}x^{\gamma} \qquad \text{for } \gamma \in (0, 1).$$

Also, as

$$\frac{d}{dx}\alpha(x) = -\frac{1-\gamma}{x^2} < 0,$$

the agent shows a decreasing risk aversion.

A third widely used function is the quadratic utility function:

$$u(x) = x - \frac{b}{2}x^2, \qquad b > 0.$$

This function is not strictly increasing, as we setted in definition 3.3, but many authors do not impose that restriction on utility functions. In order to have non-negative utility, x must be less than $\frac{1}{b}$. If we study its α, we obtain

$$\alpha(x) = \frac{b}{1-bz}, \qquad \frac{d}{dx}\alpha(x) = \frac{b^2}{(1-bx)^2} > 0,$$

which means that the risk aversion is strictly increasing and random outcomes are very penalized with respect to sure or riskless lotteries.

Observe that the three previous examples show three different behaviours with respect the variation of x: increasing, decreasing and constant. This leads to a classification of the utility functions that is important in the two assets problem. That is, how to optimize a portfolio in a market were only two assets exists: a risky asset and a free risk asset. We will treat this problem in Section 5.

Definition 4.9. Let $u(x)$ be a utility function and $\alpha(x)$ its Arrow-Pratt coefficient of absolute risk aversion. Then u is

(a) A decreasing absolute risk aversion (**DARA**) function if

$$\frac{d}{dx}\alpha(x) < 0, \qquad \forall x \in S.$$

(b) A constant absolute risk aversion, (**CARA**) function if

$$\frac{d}{dx}\alpha(x) = 0, \qquad \forall x \in S.$$

(c) An increasing absolute risk aversion (**IARA**) function if

$$\frac{d}{dx}\alpha(x) > 0, \qquad \forall x \in S.$$

Part IV. Portfolio optimization

A common problem in financial engineering is how to optimize a portfolio. By optimization, we use to understand how to construct a portfolio that maximizes the terminal wealth, the value of the portfolio at a final trading period T, given an initial wealth.

The expected utility theory adds a new point of view to the portfolio optimization problem. In some sense, a portfolio can be interpreted as a compound lottery. This interpretation considers that each asset in the market is a lottery

$$X_1, X_2, \ldots, X_n,$$

that pays the value of the asset at time T. Hence, every possible portfolio is compound lottery

$$\sum \lambda_i X_i,$$

that can be ranked attending to the utility function of the agent. That is, to the terminal expected utility of the portfolio.

This part of the work analyses the simplest problem, the two asset-one period market, fixes the notation for the general case and explains strategies to optimize portfolio in more complex markets.

Remark 4.10. In all the the examples and techniques of this part, we assume that the utility function of the agent does not change with the oscillation of wealth that occur during the intermediate trading periods. Actually, this is in general an assumption that does not hold. The hypothesis that utility functions vary attending to changes in wealth is known as the *Markowitz hypothesis*. We analyse this hypothesis in Section 13.2.

5 The one period, two assets market

The simplest example in when we have a market with two assets: money in the bank that becomes a sure amount at he end of one period and a risky stock that pays a random outcome. We want to know when an agent will invest in the risky asset and we formalize this as follows. Consider a non-degenerated bounded from below lottery [10] X defined on some probability space $(\Omega, \mathcal{F}, P)$ and be c a certain amount. We define the compound lottery

$$X_\lambda = (1-\lambda)X + \lambda c, \qquad \lambda \in [0,1].$$

[10] Sometimes, we will use term lottery to refer either a random variable of its corresponding law.

and set μ_λ as the distribution function of X_λ. Clearly, λ represent the percentage of wealth invested in the risk-less asset.

Now we ask which is the λ that maximizes the expected utility? Or, equivalently, which is the maximum of the function

$$f(\lambda) := U(\mu_\lambda) = \int u d\mu_\lambda$$

defined on $[0,1]$ with u strictly concave? f is strictly concave and attains its maximum in a unique point $\lambda^* \in [0,1]$. The answer, in the next proposition:

Proposition 5.1.

(a) We have $\lambda^* = 1$ if $E[X] \leq c$ and $\lambda^* > 0$ if $c \geq c(u, X)$.

(b) If $u \in \mathcal{C}^1(\mathbb{R})$, then

$$\lambda^* = 1 \Longleftrightarrow E[X] \leq c$$

and

$$\lambda^* = 0 \Longleftrightarrow c \leq \frac{E[Xu'(X)]}{E[u'(X)]}.$$

Proof.

(a) By the Jensen's inequality and the linearity of the expectation,

$$f(\lambda) \leq u(E[X_\lambda]) = u((1-\lambda)E[X] + \lambda c),$$

with equality if and only if $\lambda = 1$. It follows that $\lambda^* = 1$ if the right-hand side is increasing in λ, i.e., if $E[X] \leq c$.

Strict concavity of u implies

$$f(\lambda) \geq E[(1-\lambda)u(X) + \lambda u(c)] = (1-\lambda)u(c(\mu, u)) + \lambda u(c),$$

with equality if and only if $\lambda \in \{0, 1\}$. The right-hand side is increasing in λ if $c \geq c(\mu, u)$, and this implies $\lambda^* > 0$.

(b) Clearly, as u is concave, we have $\lambda^* = 0$ if and only if the right-hand derivative f'_+ of f satisfies $f'_+(0) \leq 0$. Denote by $u'_\pm$ the left and right-hand derivatives of u. I.e.,

$$u'_-(x)- = \lim_{y\uparrow x} \frac{f(x)-f(y)}{x-y}, \qquad u'_+(x)- = \lim_{z\downarrow x} \frac{f(z)-f(x)}{z-x}.$$

First note that the difference quotients

$$\frac{u(X_\lambda)-u(\lambda)}{\lambda} = \frac{u(X_\lambda)-u(X)}{X_\lambda - X}(c-X),$$

converge to

$$u'_+(X)(c-X)^+ - u'_-(X)(c-X)^-.$$

To check this convergence, recall that

$$(c-X) = (c-X)^+ - (c-X)^- \qquad (c-X)^\pm \geq 0.$$

Then, from the definition of X_λ and the triangle inequality,

$$\lim_{\lambda\downarrow 0}\left|\frac{u(X_\lambda)-u(X)}{X_\lambda - X}(c-X) - (u'_+(X)(c-X)^+ - u'_-(X)(c-X)^-)\right| \leq$$

$$\leq \lim_{\lambda\downarrow 0}(c-X)^+\left|\frac{u(X+\lambda(c-X)))-u(X)}{X+\lambda(c-X)-X} - u'_+(X)\right| +$$

$$+\lim_{\lambda\downarrow 0}(c-X)^-\left|\frac{u(X+\lambda(c-X))-u(X)}{X+\lambda(c-X)-X} - u'_-(X)\right|$$

Now, if $(c-X)^+ \geq 0$, we have

$$c - X \geq 0 \Rightarrow \lambda(c-X) \geq 0,$$

and

$$X + \lambda(c-X) \downarrow X \qquad \text{as} \qquad \lambda \downarrow 0,$$

which implies that for any $\varepsilon > 0$, there exists δ_1 such that for all $\lambda \in (0, \delta_1)$,

$$(c - X)^+ \left| \frac{u(X + \lambda(c - X))) - u(X)}{X + \lambda(c - X) - X} - u'_+(X) \right| \leq (c - X)^+ \varepsilon.$$

Analogously, as $(c - X)^- \geq 0$ implies that $\lambda(c - X) \leq 0$, we have

$$X + \lambda(c - X) \uparrow X \qquad \text{as} \qquad \lambda \downarrow 0,$$

and again for any given $\varepsilon > 0$, there exists δ_2 such that for all $\lambda \in (0, \delta_2)$,

$$(c - X)^- \left| \frac{u(X + \lambda(c - X)) - u(X)}{X + \lambda(c - X) - X} - u'_-(X) \right| \leq (c - X)^- \varepsilon.$$

Therefore, taking $\delta = \min(\delta_1, \delta_2)$, we have

$$(c - X)^+ \left| \frac{u(X + \lambda(c - X))) - u(X)}{X + \lambda(c - X) - X} - u'_+(X) \right| +$$

$$+(c - X)^- \left| \frac{u(X + \lambda(c - X)) - u(X)}{X + \lambda(c - X) - X} - u'_-(X) \right| \leq 2|c - X|\varepsilon.$$

which proves the convergence. Also we have that the quotients are P-a.s. bounded by

$$u'_+(a)|c - X| \in \mathcal{L}^1(P) \qquad \text{if } a \leq c \wedge X.$$

To see this, just apply the mean value theorem to the quotient

$$\frac{u(X_\lambda) - u(X)}{X_\lambda - X}$$

and use the concavity of u to bound the derivative with its value at $a \leq c \wedge X$.

Therefore, by Lebesgue's dominated convergence theorem, this implies

$$f'_+(0) = \lim_{\lambda \to 0} \frac{f(\lambda) - f(0)}{\lambda} = \lim_{\lambda \to 0} \int \frac{u(X_\lambda) - u(X)}{X_\lambda - X}(c - X) d\mu =$$

$$= \int \lim_{\lambda \to 0} \frac{u(X_\lambda) - u(X)}{X_\lambda - X}(c - X)dP = E[u'_+(X)(c - X)^+] - E[u'_-(X)(c - X)^-].$$

If $u \in \mathcal{C}^1(\mathbb{R})$, or if the countable set $\{x \mid u'_+(x) \neq u'_-(x)\}$ has μ-measure 0, then we can conclude

$$f'_+(0) = E[u'(X)(c - X)],$$

that is, $f'(0)_+ \leq 0$ if and only if

$$c \leq \frac{E[Xu'(X)]}{E[u'(X)]}.$$

In the same way, we obtain

$$f'_-(1) = u'_(c)E[(X - c)^-] - u'_+(c)E[(X - c)^+].$$

If u is differentiable at c, then

$$f'_-(1) = u'(c)(c - E[X]).$$

This implies $f'_-(1) < 0$, and hence $\lambda^* < 1$, as soon as $E[X] > 0$.

□

Remark 5.2. Now we can answer an historical problem, analysed on the Bernouilli's paper, of under what circumstances an agent with utility function $u \in \mathcal{C}$ would buy an insurance to cover a possible loss Y of his initial wealth w.

Suppose that $0 \leq Y \leq w$ and $P[Y \neq E[Y]] > 0$. If insurance of λY is available at the insurance premium $\lambda\pi$, the final payoff is given by

$$X_\lambda := w - Y + \lambda(Y - \pi) = (1 - \lambda)(w - Y) + \lambda(w - \pi).$$

Full insurance is, by Proposition 5.1, optimal if and only if $\pi \leq E[Y]$. In the market, the insurance premium π will exceed the fair premium $E[Y]$. In this case, it will be optimal to insure only a fraction $\lambda^* Y$ of the loss with $\lambda^* \in (0,1)$. λ^* will be strictly positive as long as

$$\pi < \frac{E[Yu'(w-Y)]}{E[u'(w-Y)]}.$$

It is interesting to note that risk aversion can create demand for insurance even if the insurance premium π lies above the fair price $E[Y]$.

6 Setting the framework

The previous proposition dealt with the simplest example and allowed us to fix the upper and lower bounds where there are diversified portfolios. Now we want to treat more general examples and explicitly found, when possible, the exact maximal expected utility. To achieve that, first we need to set the notation and the main concepts. The framework is a finite probability space $(\Omega, \mathcal{F}, P)$. We consider only $\omega \in \Omega$ with strict positive probability:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}, \qquad P(\{\omega_i\}) > 0 \text{ for all } \omega_i \in \Omega.$$

The σ-field $\mathcal{F}$ is given by the parts of Ω:

$$\mathcal{F} = \mathcal{P}(\Omega).$$

We consider N trading periods and associate the filtration,

$$\mathcal{F}_0 = \{\Omega, \emptyset\} \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_{N-1} \subset \mathcal{F}_N = \mathcal{F}.$$

In the financial context, the filtration represents that the economic agent knows at any trading period what happened in the past. The market consists on $(d+1)$ assets whose price at time n is given by the non-negative random variables

$$S_n^0, S_n^1, \ldots, S_n^d$$

measurable with respect to $\mathcal{F}_n$. The zero super-index asset, S_i^0, corresponds to a riskless asset, a bank account, that pays r_i interest for sure at the end of the trading period i.

The next key concept is the strategy of investment.

Definition 6.1. Let be $(\Omega, \mathcal{F}, P)$ a probability space and let be $\{\mathcal{F}_n\}$ a filtration. A stochastic process $X = (X_0, \dots, X_N)$ is predictable if

$$\begin{cases} X_0 & \text{is } \mathcal{F}_0 - \text{measurable}, \\ X_n & \text{is } \mathcal{F}_{n-1} - \text{measurable for all } 1 \leq n \leq N. \end{cases}$$

Definition 6.2. A **strategy of investment** is a predictable stochastic process

$$\phi = ((\phi_n^0, \phi_n^1, \dots, \phi_n^d))_{0 \leq n \leq N} \in \mathbb{R}^{d+1},$$

where ϕ_n^i indicates the number of stocks, or units of money if $i = 0$, in the asset i at the instant n.

The predictability of the strategy means that the position in the portfolio at n is decided at $n-1$. The value of the portfolio at time n is given by the scalar product

$$V_n(\phi) = \phi_n \cdot S_n = \sum_{i=0}^{d} \phi_n^i S_n^i.$$

We denote $\tilde{V}_n(\phi)$ its discounted value. If r is fix in all the trade periods, this is equivalent to

$$\tilde{V}_n(\phi) = \frac{V_n(\phi)}{(1+r)^n}.$$

Analogously, $\tilde{S}_n$ denotes the vector of discounted stock prices.

We also ask for a self-financing condition in the conditioning of the problem. That is, that reinvestments at each period does not alter the value of the portfolio:

Definition 6.3. An investment strategy is said to be self-financing if

$$\phi_n \cdot S_n = \phi_{n+1} \cdot S_n, \qquad 0 \leq n \leq N-1.$$

With the new notation and concepts, we can announce the portfolio optimization problem as follows:
Given a utility function u, what is the self-financing strategy that optimizes the terminal utility. That is, find

$$\max_{\phi \in \mathcal{A}} \{E(u(V_N(\phi)) \mid V_0(\phi) = x\},$$

with x being the initial wealth of the agent and $\mathcal{A}$ the set of self-financing portfolios.

Remark 6.4. In classical portfolio optimization problems, it is usual to take the convention $S_0^0 = 1$ to normalize the computations. However, this is not correct when optimizing the terminal expected utility because risk aversion may vary with the initial wealth and, therefore, we can expect different maximal strategies given different values of S_0^0.

7 The martingale method

7.1 Arbitrage and martingales

When we try to formulate a mathematical model of the reality, we often start with some axioms or principles derived from our intuition, laws of the nature, simplifications,... In Financial Engineering, or Quantitative Finance, the main principle from which we obtain the formulas in our models is the principle of no-arbitrage.

Definition 7.1. A strategy ϕ is admissible if it is self-financing and $V_n(\phi) \geq 0$ for all $0 \leq n \leq N$.

Definition 7.2. An arbitrage (opportunity) is an admissible strategy ϕ with zero initial value and with final value different from zero. That is,

1. $V_0(\phi) = 0$,

2. $V_n(\phi) \geq 0$, for all $1 \leq n \leq N$,

3. $P(V_N(\phi) > 0) > 0$.

A market is said viable if it is free of arbitrage opportunities.

The no-arbitrage principle is a strong assumption that uniquely determines many, either stochastic or deterministic, financial models and formulas. What we will show now is that no-arbitrage in the stock market is equivalent to the existence of a probability such that the evolution of discounted stock prices are martingales [11] with respect to it.

Definition 7.3. A sequence of random variables $X = (X_n)_{0 \le n \le N}$ is adapted to a filtration $\{\mathcal{F}_n\}_{0 \le n \le N}$ if X_n is $\mathcal{F}_n$-measurable, $0 \le n \le N$.

Definition 7.4. An adapted stochastic process $(M_n)_{0 \le n \le N}$ in a probability space $(\Omega, \mathcal{F}, P)$ is a martingale if

$$E_P[|M_n|] < \infty, \qquad 0 \le n \le N,$$

and

$$E_P[M_{n+1} \mid \mathcal{F}_n] = M_n.$$

The martingale property is equivalent to have zero conditional expected variation. That is,

$$E[M_{n+1} - M_n \mid \mathcal{F}_n] = 0,$$

Also, it is easy to see that martingales are closed by linearity and that

$$E[M_{n+j} \mid \mathcal{F}] = M_n, \qquad \forall j \ge 0.$$

The next two theorems show why martingales play a capital role in Financial Engineering.

Definition 7.5. Let be P and P^* probability measures on a measurable space $(\Omega, \mathcal{F})$. P and P^* are equivalent if

$$P \ll P^* \text{ and } P^* \ll P.$$

That is, the null sets with respect to P and P^* coincide.

Definition 7.6. A market is complete if any derivative, a random payoff h tied to the evolution of the stocks, is replicable. That is, if there exists an admissible strategy ϕ such that replicates

[11] The martingales get its name from a family of betting strategies that were popular in the 18th century France. The simplest one, consisted in doubling the bet in each flip of a coin.

h:

$$V_N(\phi) = h.$$

Theorem 7.7. *A financial market is viable, free of arbitrage, if and only if there exists P^* equivalent to P such that discounted prices of the stocks $(\tilde{S}_n^j)$, $j = 1, \ldots, d$ are P^*-martingales. P^* is named the risk neutral probability.*

Theorem 7.8. *A viable market is complete if and only if there is a unique probability P^* equivalent to P under which the discounted prices are martingales.*

The theorems also allows us to equal the price of a derivatives to its expected discounted prices with respect to the risk neutral probability:

$$E_{P^*}[\tilde{X} \mid \mathcal{F}_n] = C_n.$$

7.2 Derivation of the martingale method

The martingale method of portfolio optimization requires that the market is complete. By Theorems 7.7 and 7.8, the completeness of the market is equivalent to:

(a) The uniqueness of the risk neutral probability P^* equivalent to P,

(b) The replicability of random payoffs.

Both facts allows us to define a maximization strategy. Let be V_x the set of random variables that can be replied with initial wealth x. We proceed in two steps:

1 Find the maximal terminal utility $\hat{Y}$ defined as

$$E[u(\hat{Y})] = \max_{Y \in V_x} E[u(Y)].$$

2 Compute the self-financing portfolio that replies $\hat{Y}$.

By Theorem 7.8, the completeness of the market is equivalent to

$$E_{P^*}\left[\frac{Y}{(1+r)^N}\right] = x, \qquad P^*\text{risk neutral probability, } Y \in V_x.$$

Hence, we have can transform the portfolio optimization problem in the equivalent constrained optimization problem

$$\text{maximise} \quad E[u(Y)],$$

subject to

$$E_{P^*}\left[\frac{Y}{(1+r)^N}\right] = x.$$

A standard way to solve the problem, assuming that u is enough differentiable, is to set the Lagrangian

$$\mathcal{L}(Y,\lambda) := E[u(Y)] - \lambda\left(E_{P^*}\left[\frac{Y}{(1+r)^N}\right] - x\right).$$

This is equivalent, by the Radon-Nikodym Theorem [12] , to

$$\mathcal{L}(Y,\lambda) := E\left[u(Y)] - \lambda L\frac{Y}{(1+r)^N} - x\right],$$

where $L = \frac{dP^*}{dP}$ is the Radon-Nikodym derivative of P^* with respect to P. Now we maximize $\mathcal{L}(Y,\lambda)$ and get the extremal conditions

$$u'(Y) = \frac{\lambda L}{(1+r)^N}, \tag{11}$$

$$x = E\left[\frac{Y}{(1+r)^N}L\right]. \tag{12}$$

If we denote $I := (u')^{-1}(s)$, the equations turn

$$Y = I\left(\frac{\lambda L}{(1+r)^N}\right), \tag{13}$$

[12] The Randon-Nikodym Theorem states that two σ-finite measures satisfy $\mu \ll \nu$ if and only if there exists a measurable function in $f \in L^1(\nu)$ such that

$$\mu(A) = \int_A f d\nu.$$

We say that f is the Radon-Nikodym derivative of μ with respect to ν and note $f = \frac{d\mu}{d\nu}$.

and

$$x = E\left[\frac{I\left(\frac{\lambda L}{(1+r)^N}\right)}{(1+r)^N} L\right]. \tag{14}$$

From (14), we can isolate λ and replace it in (13) to find Y. The Y found is the possible maximum and now, we construct the strategy that replies it, which exists by the completeness of the market.

Remark 7.9. It is not always easy to calculate the Radon-Nikodym derivative of two arbitrary measures. Fortunately, this is not our case. If we have two equivalent finite and discrete probability measures μ and ν, we can compute directly $\frac{d\mu}{d\nu}$. Let be $h = \frac{d\mu}{d\nu}$. It satisfies

$$\mu(\omega) = h(\omega)\nu(\omega), \qquad \forall \omega \in \Omega.$$

Hence,

$$h(\omega) = \frac{\mu(\omega)}{\nu(\omega)},$$

and we just need to make the quotient of the probabilities to obtain $\frac{d\mu}{d\nu}$.

7.3 Examples

Now we calculate the optimal wealth with respect to the three basic examples of utility functions.

Example 7.10 (Exponential utility)**.** Consider $u(x) = -e^{-ax}$. Then,

$$u'(x) = ae^{-ax}, \qquad I(y) = \frac{1}{a}\log\left(\frac{1}{y}\right).$$

The maximal terminal utility is

$$Y = \frac{1}{a}\log\left(\frac{(1+r)^N}{\lambda L}\right).$$

To find λ, we have

$$x = E\left[\frac{L}{a(1+r)^N}\log\left(\frac{(1+r)^N}{\lambda L}\right)\right].$$

Hence,

$$ax(1+r)^N = \log\left(\frac{(1+r)^N}{\lambda}\right) E\left[L\right] - E[L\log L],$$

and

$$\log\left(\frac{(1+r)^N}{\lambda}\right) = \frac{ax(1+r)^N + E[L\log L]}{E[L]} = ax(1+r)^N + E[L\log L].^{13}$$

Replacing in the first equation gives us

$$Y = x(1+r)^N + \frac{E[L\log L] - \log L}{a}.$$

Example 7.11 (Quadratic utility)**.** Given $u(x) = x - \frac{b}{2}x^2$, we have

$$u'(x) = 1 - bx, \qquad I(y) = \frac{1-y}{b}.$$

The maximal utility is

$$Y = \frac{1}{b} - \frac{\lambda L}{b(1+r)^N}.$$

We calculate λ,

$$x(1+r)^N = E\left[\frac{L}{b} - \frac{\lambda L^2}{b(1+r)^N}\right],$$

$$\lambda = \frac{1}{E[L^2]}\left((1+r)^N - xb(1+r)^{2N}\right).$$

Hence,

$$Y = \frac{1}{b} - \frac{L}{E[L^2]}\left(b - x(1+r)^N\right).$$

We break the hyperbolic utility in two cases.

---

[13] The expected value of a Randon-Nikodym derivative L of a probability measure with respect to the dominating measure is always 1. Just recall that if $\nu \ll \mu$,

$$\nu(A) = \int_A L d\mu.$$

Therefore,

$$E_\mu[L] = \int_\Omega L d\mu = \nu(\Omega) = 1.$$

Example 7.12 (Logarithmic utility)**.** We have $u(x) = \log x$,

$$u'(x) = \frac{1}{x}, \qquad I(y) = \frac{1}{y}.$$

Then,

$$x = E\left[\frac{(1+r)^N}{\lambda L}\frac{1}{(1+r)^N}L\right] = \frac{1}{\lambda},$$

and

$$Y = \frac{x(1+r)^N}{L}.$$

Example 7.13 (Monomial utility)**.** Let be $u(x) = \frac{x^\gamma}{\gamma}$,

$$u'(x) = x^{\gamma-1}, \qquad I(y) = y^{\frac{1}{\gamma-1}}.$$

Then,

$$Y = \left(\frac{\lambda}{(1+r)^N}L\right)^{\frac{1}{\gamma-1}}.$$

We find λ,

$$x = E\left[\frac{\left(\frac{\lambda}{(1+r)^N}L\right)^{\frac{1}{\gamma-1}}}{(1+r)^N}L\right],$$

that is

$$\lambda^{\frac{1}{\gamma-1}} = x(E[(L(1+r)^{-N}]^{\frac{\gamma}{\gamma-1}}))^{-1},$$

and

$$Y = \frac{x(L(1+r)^{-N})^{\frac{1}{\gamma-1}}}{E[(L(1+r)^{-N}]^{\frac{\gamma}{\gamma-1}}}.$$

8 The Cox-Ross-Rubinstein Binomial Model

The binomial model is a simple, yet powerful, tool to work with discrete time stochastic processes. It is used to:

- Price derivatives on stocks, securities and commodities, like swaptions and options.

- Discretize and stress the hedging of continuous time models like the Black-Scholes model [14].
- Via the option theory, evaluate non-financial investment and decision problems like pricing a lease on a gold mine [15] stressing risk factors or toll a power plant.

To summarize the binomial model, let be

$$X = (X_t)_{0 \leq t \leq N}$$

a real valued discrete time stochastic process. At each time t, we have two possible future states:

- A *up state* with probability p.
- A *down state* with probability $1 - p$.

This schema generates a binomial lattice

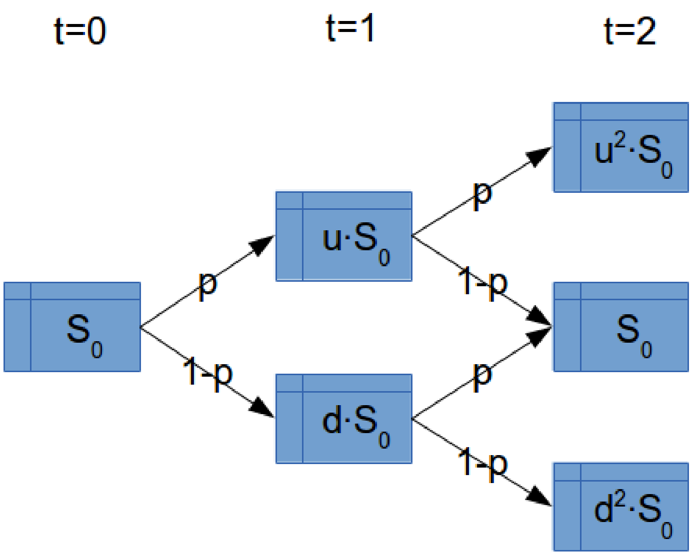

Fig. 2: A Binomial Lattice of 3 periods.

[14] The Black Scholes model has big problems. One of them is that continuous-time trade is not a realistic assumption and we need to discretize the trade periods. This impedes perfectly hedge the option and, depending on the realized path of the discretized process, we can incur in big loses or wins.

[15] To see an example of option theory applied to decision problems, check the popular example *Simplico gold mine case* in Luenberger, *Investment Science*, OUP 1998.

with

$$1 + 2 + \ldots + N + (N+1) = \frac{(N+1)(N+2)}{2}$$

nodes. To further simplify the model, we will suppose that:

(a) The probability q do not depend on the time t.

(b) Given a present state $X_t = x$, we obtain the next up state multiplying by a factor u and the down state multiplying by d. In most cases, d will equal $\frac{1}{u}$.

To summarize, the marginal law of X_t can be written as follows:

$$X_t = \begin{cases} uX_{t-1} & \text{with probability} \quad p, \\ dX_{t-1} & \text{with probability} \quad 1-p. \end{cases}$$

To simplify, we will assume that u and d probabilities do not depend on the time t, nor p does. In our context, we will suppose that the binomial lattice shows the evolution of the price of a risky asset, like an stock. Furthermore, we will also suppose that there exists a risk-less asset, a bank account, that pays an interest r at the end of each trading period.
Now, we check if a market modelled this way, is complete. This is equivalent, by Theorem 7.8 to the existence of a unique risk neutral probability. And, by the assumptions of having p, u and u independent of time, is enough to solve the problem for this 1 period market:

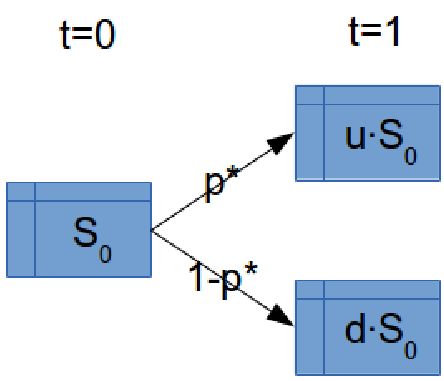

Fig. 3: Single period market with risk neutral probability.

That is, to find p^* such that

$$E_P^*[X_1] = p^* \frac{u}{1+r} + (1-p^*) \frac{d}{1+r} = 1 = X_0.$$

This has the unique solution

$$p^* = \frac{(1+r)-d}{u-d}, \qquad 1+r \in (u,d).$$

Now that we have found the risk neutral probability and ensured that the market is complete, the martingale method gives us the maximal terminal utility.

Example 8.1. Let us take logarithmic utility function. Checking the last column of the binomial lattice, or by direct computation, it is clear that each path ω is completely determined by the number U_N of ups that occurred until time N and, therefore, the respective probabilities can be computed with the joint distribution of N Bernoulli random variables with parameter p:

$$P\{\text{number of } ups = U_N\} = \binom{N}{U_N} p^{U_N} \cdot (1-p)^{N-U_N}.$$

The same argument applies when the binomial lattice follows the risk neutral probability, from which follows that the Radon-Nikodym derivative L, the quotient of the laws, is

$$L = \left(\frac{p^*}{p}\right)^{U_N} \cdot \left(\frac{1-p^*}{1-p}\right)^{N-U_N}.$$

By Example 7.12, the maximal terminal utility Y is

$$Y = \frac{x(1+r)^N}{L} = x(1+r)^N \left(\frac{p}{p^*}\right)^{U_N} \cdot \left(\frac{1-p}{1-p^*}\right)^{N-U_N}.$$

Because the market is complete, we can find a self-financing portfolio that replicates the terminal utility Y. As usual, the wealth invested in the risky asset is

$$\phi_N^1 = \frac{V_N^u - V_N^d}{S_{N-1}(u-d)}.$$

Then

$$\phi_N^1 S_{N-1} = \frac{V_N^u - V_N^d}{(u-d)} = \frac{x(1+r)^N}{(u-d)} \left(\frac{p}{p^*}\right)^{U_{N-1}} \cdot \left(\frac{1-p}{1-p^*}\right)^{(N-1)-U_{N-1}} \left(\frac{p-p^*}{p^*(1-p^*)}\right).$$

The value of the portfolio that replies the payoff Y in $N-1$ is given by the risk neutral probability:

$$V_{N-1} = E_{P^*}\left[\frac{Y}{1+r}\middle| \mathcal{F}_{N-1}\right] =$$

$$= \frac{x(1+r)^N}{(u-d)} \left(\frac{p}{p^*}\right)^{U_{N-1}} \cdot \left(\frac{1-p}{1-p^*}\right)^{(N-1)-U_{N-1}} \left(\frac{p}{p^*}p^* + \frac{1-p}{1-p^*}(1-p^*)\right) =$$

$$= \frac{x(1+r)^N}{(u-d)} \left(\frac{p}{p^*}\right)^{U_{N-1}} \cdot \left(\frac{1-p}{1-p^*}\right)^{(N-1)-U_{N-1}}.$$

Putting all together,

$$\frac{\phi_N^1 S_{N-1}}{V_{N-1}} = \frac{(1+r)(p-p^*)}{p^*(1-p^*)(b-a)}.$$

Observe that this quantity does not depend on time N, from where we obtain that the ratio of investment in the risky asset is constant:

$$\frac{\phi_N^1 S_{N-1}}{V_{N-1}} = \frac{\phi_{N-1}^1 S_{N-2}}{V_{N-2}} = \ldots = \frac{\phi_1^1 S_0}{x}.$$

.

8.1 The Binomial Model and Dynamic Programming

In this section we prepare the tools to derive the well known Merton's solution to the problem of maximizing the terminal utility of a portfolio. Merton's original problem is set in continuous time, but we will be able to reach the same result in discrete time applying methods of dynamic programming to the binomial model.

We refer as dynamic programming a set of techniques extensively used in Mathematics and Applied Sciences. The idea that bonds all those techniques is to break the problem in simpler sub-problems. For example, the Dijkstra algorithm switch the problem

- *Find the shortest path from vertex a_0 to vertex a_n in a connected graph.*

with this set of simpler problems

- *Set $S = \{a_0\}$. Now find the nearest vertex a_k to the set S, record its shortest path and add the vertex to the set S. Continue until a_n has been added to the set.*

Consider now a market with N trading periods, d risky assets and 1 risk-less asset. The Dynamic Programming strategy that we apply to the portfolio optimization problem is to break the original problem in tail optimization sub-problems. That is, define

$$U_n(y) = \max\{E[u(V_N(\phi))|\mathcal{F}_n], \phi \in \mathcal{A}_n^y\},$$

where $n = 0, 1, \ldots, N-1$ is the trading period and $\mathcal{A}_n^y$ is the set of self-financing portfolios made from period n with value y at time n. Due to the fact that the value of every self-financing portfolio depends on its initial value and the predictable sequence,

$$(\phi_n^1, \phi_n^2, \ldots, \phi_n^d)_{0 \le n \le N},$$

we have

$$V_N(\phi) = (1+r)^N \left(\frac{y}{(1+r)^n} + \sum_{j=n+1}^{N} \phi_j \cdot \Delta \tilde{S}_j \right).$$

Therefore,

$$U_n(y) = \max\{E[u(V_N(\phi))|\mathcal{F}_n], \quad (\phi_j^1, \phi_j^2, \ldots, \phi_j^d)_{n+1 \le j \le N} \text{ predictable}, V_n = y\}.$$

From this expression we can deduce the functional equation of the problem and define a well-posed the optimization method:

Proposition 8.2.

$$U_n(y) = \max_{\phi \mathcal{F}_n - measurable} \{E[U_{n+1}(V_{n+1}(\phi) \mid \mathcal{F}_n], \ V_n = y\} =$$

$$= \max_{\phi \mathcal{F}_n-measurable} \left\{ E\left[U_{n+1}((1+r)^N \left(\frac{y}{(1+r)^n} + \sum_{j=n+1}^{N} \phi_j \cdot \Delta\tilde{S}_j \right) \mid \mathcal{F}_n \right] \right\}.$$

Proof. By definition, $U_N(y) = u(y)$, then

$$U_{N-1} = \max \left\{ E\left[u(V_N(\phi)) \mid \mathcal{F}_{N-1}\right],\ (\phi_j^1, \phi_j^2, \ldots, \phi_j^d)_{j=N} \text{ predictable},\ V_{N-1} = y \right\} =$$

$$= \max \left\{ E\left[U_N(V_N(\phi)) \mid \mathcal{F}_{N-1}\right],\ (\phi_j^1, \phi_j^2, \ldots, \phi_j^d)_{j=N} \text{ predictable},\ V_{N-1} = y \right\}.$$

Also,

$$U_n(y) = \max_{\substack{\phi\ \mathcal{F}_n-\text{measurable} \\ (\varphi_j)-\text{measurable}}} E\left[u\left((1+r)^N \left(\frac{y}{(1+r)^n} + \phi \cdot \Delta\tilde{S}_{n+1} + \sum_{j=n+2}^{N} \varphi_j \cdot \Delta\tilde{S}_j \right) \right) \middle| \mathcal{F}_n \right] =$$

$$= \max_{\phi \mathcal{F}_n-\text{measurable}} \left\{ \max_{(\varphi_j)-\text{measurable}} E\left[u\left((1+r)^N \left(\frac{y}{(1+r)^n} + \phi \cdot \Delta\tilde{S}_{n+1} + \right.\right.\right.\right.$$
$$\left.\left.\left.\left. + \sum_{j=n+2}^{N} \varphi_j \cdot \Delta\tilde{S}_j \right) \right) \middle| \mathcal{F}_n \right] \right\} =$$

$$= \max_{\phi\ \mathcal{F}_n-\text{measurable}} \left\{ \max_{(\varphi_j)-\text{measurable}} E\left[E\left[u\left((1+r)^N \left(\frac{y}{(1+r)^n} + \phi \cdot \Delta\tilde{S}_{n+1} + \right.\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\left. + \sum_{j=n+2}^{N} \varphi_j \cdot \Delta\tilde{S}_j \right) \right) \middle| \mathcal{F}_{n+1} \right] \middle| \mathcal{F}_n \right] \right\} =$$

$$= \max_{\phi\ \mathcal{F}_n-\text{measurable}} \left\{ E\left[\max_{(\varphi_j)-\text{measurable}} E\left[u\left((1+r)^N \left(\frac{y}{(1+r)^n} + \phi \cdot \Delta\tilde{S}_{n+1} + \right.\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\left. + \sum_{j=n+2}^{N} \varphi_j \cdot \Delta\tilde{S}_j \right) \right) \middle| \mathcal{F}_{n+1} \right] \middle| \mathcal{F}_n \right] \right\} =$$

$$= \max_{\phi\ \mathcal{F}_n-\text{measurable}} E\left[U_{n+1}\left((1+r)^{n+1} \left(\frac{y}{(1+r)^n} + \phi \cdot \Delta\tilde{S}_{n+1} \right) \right) \middle| \mathcal{F}_n \right].$$

In the last equality we used the fact that, if Z is a random variable which depends on (ϕ_j),

$$\max_{(\varphi_j)\text{ previsible}} E[Z((\varphi_j))|\mathcal{F}_n] = E\left[\max_{(\varphi_j)\text{ previsible}} Z((\varphi_j))|\mathcal{F}_n \right].$$

To prove this statement, for all (φ_J) predictable, its is clear that

$$E\left[\max_{(\varphi_j)\text{ previsible}} Z((\varphi_j))|\mathcal{F}_n\right] \geq E\left[Z((\varphi_j))|\mathcal{F}_n\right]$$

and

$$E\left[\max_{(\varphi_j)\text{ previsible}} Z((\varphi_j))|\mathcal{F}_n\right] \geq \max_{(\varphi_j)\text{ previsible}} E[Z((\varphi_j))|\mathcal{F}_n].$$

For the converse inequality, let be (ψ_j) predictable that satisfies

$$E\left[\max_{(\varphi_j)\text{ previsible}} Z((\varphi_j))|\mathcal{F}_n\right] = E\left[Z((\psi_j))|\mathcal{F}_n\right].$$

Obviously,

$$E\left[\max_{(\varphi_j)\text{ previsible}} Z((\varphi_j))|\mathcal{F}_n\right] = E\left[Z((\psi_j))|\mathcal{F}_n\right] \leq \max_{(\varphi_j)\text{ previsible}} E\left[Z((\varphi_j))|\mathcal{F}_n\right].$$

□

As we have announced, by the previous proposition, solve the original problem is equivalent to solve the sequence of tail sub-problems. That is, we only just need to compute the backward recurrence

$$U_{N-1} \Rightarrow U_{N-2} \Rightarrow \ldots \Rightarrow U_0,$$

to solve the whole problem.

In the general framework of Dynamic Programming, this property is named *The Principle of Optimality* and states that the optimal policy of the tail subproblem coincides with the corresponding portion of the solution of the original problem. The intuition behind this principle is simple: if a new policy could outperform the original policy on the tail sub-problem, the original problem could be improved by replacing the corresponding portion with the new policy.

Remark 8.3. Observe that, under the assumption we have made at beginning of this section, the stochastic processes modelled by the Binomial Lattice are Markov Processes. That is,

$$P[X_t = x_t \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots, X_1 = x_1] = P[X_t = x_t \mid X_{t-1} = x_{t-1}].$$

This property allows us to guarantee that the dynamical programming strategy is well posed and that backward recurrence can solve the optimization problem.

Example 8.4. We recover the Example 8.1 to check that the martingale and the dynamic programming methods coincide. We have

$$U_n(y) = \max \left\{ E\left[\log(V_N(\phi)) \mid \mathcal{F}_n\right], \phi \in \mathcal{A}_n^y \right\}.$$

Then,

$$U_{N-1}(V_{N-1}) = \max_{\phi_N^1} E\left[\log\left((1+r)^N(\tilde{V}_{N-1} + \phi_N^1 \Delta \tilde{S}_N^1)\right) | \mathcal{F}_{N-1}\right] =$$

$$= \max_{\phi_N^1} \left(\log[(1+r)V_{N-1} + \phi_N^1(d-1-r)S_{N-1}^1)p] + \log[(1+r)V_{N-1} + \phi_N^1(u-1-r)S_{N-1}^1](1-p)\right).$$

If we calculate the first order conditions,

$$\frac{p(u-1-r)S_{N-1}^1}{(1+r)V_{N-1} + \phi_N^1(u-1-r)S_{N-1}^1} + \frac{(1-p)(d-1-r)S_{N-1}}{(1+r)V_{N-1} + \phi_N^1(d-1-r)S_{N-1}^1} = 0.$$

We obtain from the last equality

$$\frac{\phi_N^1 S_{N-1}}{V_{N-1}} = (1+r)\left(\frac{p}{1+r-d} - \frac{1-p}{u-1-r}\right),$$

and substituting ϕ_N^1,

$$U_{N-1}(V_{N-1}) = p \log\left(\frac{V_{N-1}(1+r)p(u-d)}{(1+r-d)}\right) + (1-p)\left(\frac{V_{N-1}(1+r)(1-p)(u-d)}{u-1-r}\right) =$$

$$= \log(V_{N-1}) + C, \qquad C \text{ constant}.$$

Applying an induction argument,

$$\frac{\phi_N^1 S_{N-1}^1}{V_{N-1}} = \frac{\phi_{N-1}^1 S_{N-2}^1}{V_{N-2}} = \ldots = \frac{\phi_1^1 S_0^1}{x},$$

which is the same formula that we have derived from the martingale method.

8.2 The binomial approximation to the Black-Scholes model

The Black-Scholes equations are a well known, and used despite its limitations, model for pricing derivatives. Applies to a two assets market:

(1) A risk-less asset, like a bank account, that pays a fixed interest r,

(2) A risky asset, like a stock, which its price is a random variable.

Also, it assumes:

(a) The evolution of the risky asset follows a Geometric Brownian Motion with constant drift and volatility.

(b) There are no trade fees and the stock can be traded in continuous time.

(c) The stock does not pay dividends [16].

In our language, the risk-less asset evolves as

$$S_t^0 = e^{rt},$$

and the risky asset as

$$S_t^1 = S_0^1 \exp\left(\mu + \sigma B_t\right), \qquad t \geq 0,$$

where B_t is a Brownian Motion.

Definition 8.5. A Brownian Motion is a stochastic process on some probability space with these properties:

(a) The process starts at 0: $P[B_0 = 0] = 1$.

(b) The increments are independent. If

$$t_0 \leq t_1 \leq \ldots \leq t_k,$$

[16] The original Black-Scholes model did not include dividends, but it can be modified to include them. The inclusion of dividends has a big impact in the valuation of derivatives like American options, but we do not need to consider that case in this work.

then

$$P[B_{t_i} - B_{t_{i-1}} \in H_i,\ i \leq k] = \prod_{i \leq k} P[B_{t_i} - B_{t_{i-1}} \in H_i].$$

(c) For $0 \leq s < t$ the increment $B_t - B_s$ is normally distributed with mean 0 and variance $t - s$:

$$P[B_t - B_s \in H] = \frac{1}{\sqrt{2\pi(t-s)}} \int_H e^{\frac{-x^2}{2(t-s)}} dx.$$

Definition 8.6. A Geometric Brownian Motion (GBM) with drift μ is a stochastic process

$$Y = S_0 e^{X(t)}$$

with $S_0 > 0$, $X(t) = \mu t + B_t$ and B_t being a Brownian Motion.

Hence, we want to set a binomial lattice that approximates the paths of a Geometric Brownian Motion to apply the previous optimization results.

A binomial lattice, with the restrictions set on Section 8, is completely determined by the parameters u, d and p. Our duty is now calculate them to approximate the GBM S on a time interval $[0, T]$. Take an arbitrary $n \in \mathbb{N}$ and break the time interval $(0, T]$ into n equal ΔT sized intervals

$$\left(0, \frac{T}{n}\right], \left(\frac{T}{n}, \frac{2T}{n}\right], \ldots, \left(\frac{(n-1)}{n}T, T\right].$$

It is clear that the risk-less asset increases by a factor of $\exp(r\Delta T)$ at the end of each subinterval. To simplify the notation, we set $t_i := n\frac{T}{n}$. We have

$$L_i := \frac{S_{t_i}}{S_{t_{i-1}}} = \frac{\exp(\mu t_i + \sigma B_t)}{\exp(\mu t_{i-1} + \sigma B_{t-1})} = \exp\left(\mu \Delta t + \sigma (B_{t_i} - B_{t_{i-1}})\right)$$

Taking logarithms and applying property (b) of Definition 8.5,

$$\log(L_i) = \mu \Delta t + \sigma(B_{t_i} - B_{t_{i-1}}) = \mu \Delta t + \mathcal{N}(0, \sigma^2 \Delta T) =$$

$$= \mathcal{N}\left(\mu \Delta t, \sigma^2 \Delta T\right).$$

This means that L_i is a $\mathcal{L}\wr\}\mathcal{N}(\mu\Delta T, \sigma^2\Delta T)$ and, due to the equality

$$S_t^1 = S_0 L_1 L_2 \cdot L_t,$$

we deduce that the Geometric Brownian motion can be discretized as a product of i.i.d. lognormal random distributions. Now recall that a lognormal distribution, like the normal distribution from where it derives, is completely determined by its mean and variance or, equivalently, by setting its first and second order moment. With some calculations, for example using moment-generating functions, we can obtain those both values

$$E[L] = \exp\left(\mu\Delta T + \frac{\sigma}{2}\Delta T\right), \qquad E[L^2] = \exp\left(2\mu\Delta T + 2\sigma\Delta T\right).$$

We choose now the parameters u, d and p that match those moments at each Bernoulli node of the Binomial tree solving the system of equations

$$pu + (1-p)d = \exp\left(\mu\Delta T + \frac{\sigma}{2}\Delta T\right), \qquad pu^2 + (1-p)d^2 = \exp\left(2\mu\Delta T + 2\sigma\Delta T\right).$$

In general, this system has infinite solutions as we have two equations and three variables. But in our case, we assumed that the binomial models we use satisfy

$$u = \frac{1}{d}.$$

With this additional condition, the system has the unique solution for p

$$p = \frac{\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta T\right) - d}{u - d}.$$

Plugging this on the equations yields,

$$u = \frac{1}{2}\left(\exp\left(-\left(\mu - \frac{\sigma^2}{2}\right)\Delta T\right)\right) + \exp\left(\left(\mu + \frac{\sigma^2}{2}\right)\Delta T\right) +$$

$$+\frac{1}{2}\sqrt{\left(\exp\left(-\left(\mu-\frac{\sigma^2}{2}\right)\Delta T\right)\right)+\exp\left(\left(\mu+\frac{\sigma^2}{2}\right)\Delta T\right)^2-4}.$$

This value of u, and hence d, can be approximated, for enough small ΔT, by

$$u = e^{\sigma\sqrt{\Delta T}}, \qquad d = e^{-\sigma\sqrt{\Delta T}}.$$

Remark 8.7. Those values converge to the paths of GBM with volatility σ and drift μ. We will not prove this result and limit us to give a small outline of the prove. When n is large enough, by the Central Limit Theorem,

$$\log(B_1B_2\cdots B_n) = n\log(B_1) \approx N(\mu T, \sigma^2 T).$$

Taking exponentials on both sides,

$$S_n^1 = S_0^1\prod_{i=1}^n B_i \approx e^{X(T)}.$$

Now, it can be shown that

$$S_n^1 \overset{n\uparrow\infty}{\longrightarrow} S_T^1 \text{ in distribution},$$

using again the Central Limit Theorem and the fact that

$$E\left[\log\left(\prod_{i=1}^n B_i\right)\right] = nE[\log(B_1)] \overset{n\uparrow\infty}{\longrightarrow} \mu T \text{ and } \operatorname{Var}\left[\log\left(\prod_{i=1}^n B_i\right)\right].$$

Example 8.4 tells us what is the optimal investment in a risky asset following a GBM under a logarithmic utility function. We have proved that the optimal fraction of the initial wealth x invested in the risky asset is

$$\frac{\phi_1^1 S_0^1}{x} = (1+r)\left(\frac{p}{1+r-d} - \frac{1-p}{u-1-r}\right).$$

Using the parameters that converge to the binomial model, the formula becomes,

$$\frac{\phi_1^1 S_0^1}{x} = e^{r\Delta T}\left(\frac{p}{e^{r\Delta T} - e^{-\sigma\sqrt{\Delta T}}} - \frac{1-p}{e^{\sigma\sqrt{\Delta T}} - e^{r\Delta T}}\right).$$

with

$$p = \frac{\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta T\right) - e^{-\sigma\sqrt{\Delta T}}}{e^{\sigma\sqrt{\Delta T}} - e^{-\sigma\sqrt{\Delta T}}}.$$

If we take the limit $\Delta T \to 0$, the optimal wealth converges to

$$\frac{\mu - r}{\sigma^2} \tag{15}$$

The same Dynamic Programming procedure results, when paired with a Constant Relative Risk Aversion (CARRA) [17] utility function,

$$u(x) = \frac{1}{1-\gamma}x^{1-\gamma},$$

in the following optimal portfolio investment

$$\frac{\phi_1^1 S_0^1}{x} = \frac{\mu - r}{\gamma\sigma^2}.$$

This is the famous Merton's solution to the optimization problem that he found in a continuous time setting and that we were able to derive from the discretization of the GBM. Observe that (15) is a particular case that can be obtained as a limit case $\gamma \to 1$.

[17] The Arrow-Pratt Coefficient of Relative Risk Aversion is a measure of risk aversion related that is obtained multiplying the Arrow-Pratt Coefficient of Absolute Risk Aversion by the current wealth of the agent. Hence, a CARRA utility function is a solution of the ODE

$$-x\frac{u''(x)}{u'(x)} = \alpha.$$

Part V. Other applications of the theory

9 Indifference price

One common problem in Financial Engineering is how to price derivatives or contingent claims. Derivatives are a kind of lotteries bonded to the evolution of a, real or not, underlying assets. For example, the European Put Option allows to exercise the right to sell an asset with price S_t in a future time T at a fixed price, called the strike price, K. In our language, is a lottery that pays

$$X = \max(K - S_T, 0).$$

Derivatives are very important in modern days and offer undoubted advantages. At the same time, they can be pretty complex and hard to manage: non-linear payoffs, multiple exercise times, synthetic underlying assets, reliance on complex numerical methods and so on. The market models that we have seen in this work, like the Binomial Lattice or the Black-Scholes, are complete [18] and exists a unique arbitrage free price. Then, under the completeness hypothesis, the expected utility theory does not add anything worth to note on the pricing problem. I am not saying that the expected utility theory is irrelevant in complete markets; the theory still ranks the derivatives and can give birth to impressive results that discard whole families of derivatives. But the price is unique and independent of the utilities functions of the agents.

The incomplete market are a complete different story. In the incomplete markets we cannot replicate all the derivatives and, in consequence, some financial products have a range of arbitrage free prices. Within this range, the supply and the demand fix the final price of the derivatives and the mathematical models are more complex and pricing require more assumptions than the mere absence of arbitrage.

If we assume that agents are maximizers of utility functions, then the expected utility theory gives a new way to fix the feasible range of prices of derivative: *the utility indifference price*. Given a contingent claim, C_T, the utility indifference buy price is the price at which the agent is indifferent (in the sense that his expected utility) between paying nothing and not buying

[18] See Subsection 7.1 for the main definitions and results.

the claim C_T and paying p^b now to buy the claim C_T at time T. Analogously, the utility indifference sell price is price at with the agent is indifferent between receiving nothing and not selling the claim C_T and receive p^s now to sell the claim C_T at time T.

Definition 9.1. Let $k > 0$ be the units of the claim C_T and let x be the initial wealth of an economic agent with utility function u. Define

$$V(x,k) = \sup_{X_T \in \mathcal{A}(x)} E[u(X_T + kC_T)],$$

where $\mathcal{A}(x)$ are the self-financing portfolios at time T with initial value x.
The utility indifference buy price $p^b(k)$ is the solution of

$$V(x - p^b(k), k) = V(x, 0). \tag{16}$$

The utility indifference sell price $p^s(k)$ is the solution of

$$V(x + p^s(k), -k) = V(x, 0). \tag{17}$$

Observe that, by definition, $p^b(k) = -p^s(-k)$.

Example 9.2. Consider an agent with exponential utility function $u = -\exp(-x)$ and initial wealth w. Suppose that we have a two assets- on period market: one risk-less asset with $r = 0$ and a risky asset S_t such that

$$S_0 = 100, \qquad S_1 = \begin{cases} 90, & \text{with probability } \frac{1}{2}, \\ 110, & \text{with probability } \frac{1}{2} \end{cases}.$$

We want to compute the utility indifference buy price of a single call option C_T with strike price $K = 105$. Then

$$V(w,0) = \max_{\alpha + \beta S_0 = w} E[-\exp(-\alpha - \beta S_1)] =$$

$$= \max_{\substack{\beta \\ \alpha = w - 100\beta}} \left[-\exp(100\beta - w - 110\beta)\frac{1}{2} - \exp(100\beta - w - 90\beta)\frac{1}{2} \right] =$$

$$= \max_{\substack{\beta \\ \alpha = w - 100\beta}} -\frac{\exp(-w)}{2} \left[\exp(-10\beta) + \exp(10\beta)\right].$$

$-\exp(-w)$ is a non positive number. Then, the maximum is attained in the minimum of $\left[\exp(-10\beta)\frac{1}{2} + \exp(10\beta)\frac{1}{2}\right]$, that is, clearly, $\beta = 0$:

$$V(w, 0) = -\exp(-w).$$

Now the second part,

$$V(w - p^b(1), 1) = \max_{\alpha + \beta S_0 = w - p^b(1)} E[-\exp(-\alpha - \beta S_1 - C_T)] =$$

$$= \max_{\substack{\beta \\ \alpha = w - 100\beta - p^b(1)}} \left[-\exp(-w - 10\beta + p^b(1) - 5)\frac{1}{2} - \exp(-w + 10\beta + p^b(1))\frac{1}{2}\right] =$$

$$= \max_{\substack{\beta \\ \alpha = w - 100\beta - p^b(1)}} -\frac{\exp(-w)\exp(p^b(1))}{2}[\exp(-10\beta - 5) + \exp(10\beta)].$$

Again, the maximum is attained at the minimum of the inner part with $\beta = -\frac{1}{4}$:

$$V(w - p^b(1), 1) = -\exp(-w)\exp(p^b(1))\exp(-\frac{5}{2}).$$

Equating both sides we get $p^b(1) = \frac{5}{2}$.

Observe that this price coincides with the risk neutral price. This is a property of the utility indifference price: when the market is complete, the seller and buyer prices match the martingale pricing. Therefore, the utility indifference price is an extension of the standard theory.

The utility indifference price is an active research topic. We now summarize its main properties:

(a) *Non-linear pricing.* The indifference price is not linear on k due to the concavity of the utility function. That is, the buyer will not pay the twice to double his position on the derivative and the seller requires more than twice to double his exposition to the derivative. This is an expected behaviour of risk averse agents that other models for pricing contingent claims in incomplete markets do not account for.

(b) *Recovery of the complete market price.* As in the example, when the market is complete or the claim is replicable, the indifference and the complete market price coincide. Let R_T denote the value at time T of one unit of currency invested at time 0. If $X_T \in \mathcal{A}(w)$, we can write $X_T = xR_T + \tilde{X}_T$ for some $\tilde{X}_T \in \mathcal{A}(0)$ is the set of claim which can be replicated with zero initial wealth. Since C_T is replicable,

$$C_T = pR_T + \tilde{X}_T^C,$$

with p being the complete market price and $\tilde{X}_T^C \in \mathcal{A}(0)$. Then,

$$X_T + kC_T = (w + kp)R_T + \tilde{X}_T + k\tilde{X}_T^C = (x + kp)R_T + \tilde{X}'_T,$$

where $\tilde{X}'_T \in \mathcal{A}(0)$. Then $X_T + kC_T \in \mathcal{A}(x + kp)$ and

$$V(w,k) = \sup_{X_T \in \mathcal{A}(w)} E[u(X_T + kC_T)] = \sup_{X_T \in \mathcal{A}(w+kp)} E[u(X_T)] = V(w + kp, 0).$$

Hence, $p(k) = kp$.

(c) *Monotonicity.* If $C_T^1 \leq C_T^2$, then $p_1^b(k) \leq p_2^b(k)$.

(d) *Concavity.* If $p_\lambda^b(k)$ is the price for the claim $\lambda C_T^1 + (1-\lambda)C_T$ where $\lambda \in [0,1]$, then

$$p_\lambda^b(k) \leq \lambda p_1^b(k) + (1-\lambda)p_2^b(k).$$

This is a consequence of the continuity of u and the Jensen's inequality. If we consider seller prices, then we have convexity.

To further explore the concept of utility indifference price, we need market models more complex than The Cox-Ross-Rubinstein Binomial Model. Those models are out of the scope of this work and we refer the reader to [8] and its extensive bibliography.

10 Comparison with the mean-variance paradigm

The mean-variance paradigm, or modern portfolio theory, is an investment strategy that use pack of analytical and statistical tools to rank the possible portfolios under the motto of making the trade-off between risk and return as favourable as possible to the investor.

To summarize the theory, assume that we have a market with n risky assets, $S_1, \ldots, S_n$ and 1 monetary asset S_0 that pays an interest rate μ_0. Let be $\overline{\mu}$ the vector of mean returns,

$$\overline{\mu} := (\mu_0, E[S_1], E[S_2], \ldots, E[S_n]) = (\mu_0, \mu_1, \mu_2, \ldots, \mu_n)$$

and let be V the matrix of variances and covariances,

$$V = \begin{pmatrix} \text{var}(S_1) & \text{covar}(S_1, S_2) & \ldots & \text{covar}(S_1, S_n) \\ \text{covar}(S_2, S_1) & \text{var}(S_1) & \ldots & \text{covar}(S_2, S_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{covar}(S_n, S_1) & \text{covar}(S_2, S_n) & \ldots & \text{var}(S_n) \end{pmatrix}$$

To simply the notation, we will denote the components of V as

$$V = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{pmatrix}$$

A portfolio in this model equals to a vector ϕ

$$\phi = (\phi_0, \phi_1, \phi_2, \ldots, \phi_n)$$

such that $\sum_i \phi_i = 1$. We identify the return μ_ϕ of the portfolio with its the expected return and the risk with its variance σ_ϕ:

$$\mu_\phi = \overline{\mu} \cdot \phi = \sum_{i=0}^{n} \mu_i \phi_i, \qquad \sigma_\phi^2 = \phi \cdot V\phi = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \phi_i \phi_j.$$

With this setting, the problem of investment can be stated as any of the following optimization problems:

(a) Minimize risk ensuring a target return r:

$$\min_\phi \sigma_\phi^2 = \min_\phi \phi \cdot V\phi = \min_\phi \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \phi_i \phi_j.$$

subject to

$$\mu_\phi = \overline{\mu} \cdot \phi = \sum_{i=0}^{n} \mu_i \phi_i \qquad \text{and} \qquad \sum_{i=0}^{n} \phi_i = 1.$$

(b) Maximize return ensuring a bound to the risk $\tilde{\sigma}^2$:

$$\max_\phi \mu_x = \overline{\mu} \cdot \phi = \sum_{i=0}^{n} \mu_i \phi_i$$

subject to

$$\sigma_\phi^2 = \phi \cdot V\phi = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \phi_i \phi_j \leq \tilde{\sigma}^2 \qquad \text{and} \qquad \sum_{i=0}^{n} \phi_i = 1.$$

(c) Maximize a risk-aversion adjusted return τ

$$\max_\phi \{\mu_\phi - \tau \sigma_\phi^2\} = \max_\phi \sum_{i=0}^{n} -\tau \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} \phi_i \phi_j \right)$$

subject to

$$\sum_{i=0}^{n} \phi_i = 0.$$

All those problems can be easily solved with Lagrangian multipliers. In particular, solve problem (a) with Lagrangians equals to solve a linear system of $n + 2$ equations.

We do not enter in details, but if we graph the return of the portfolio as a function of the variance,

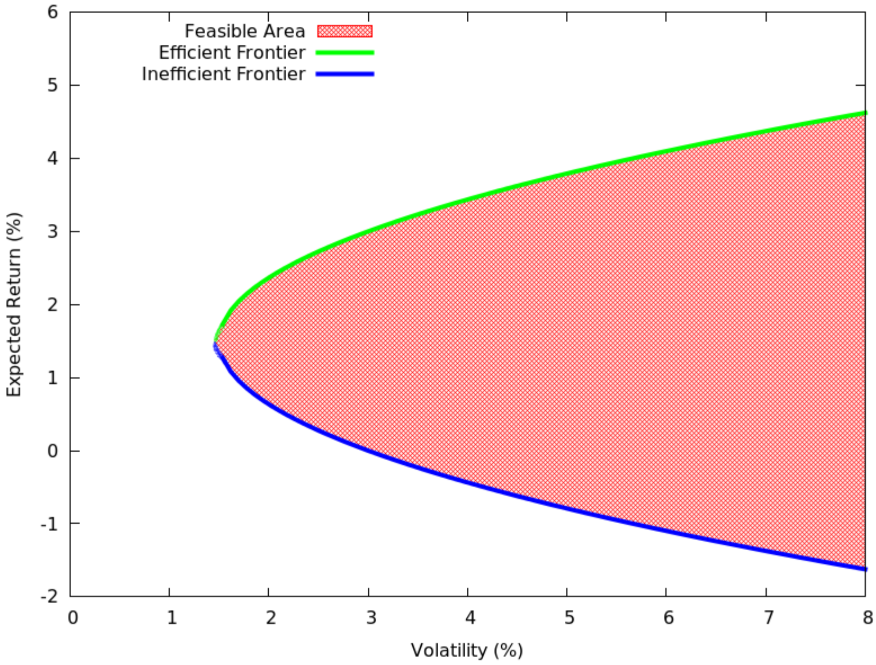

Fig. 4: Return of a portfolio as a function of its variance.

we can observe:

1. There is a feasible area. That is, the only constructable portfolio are those which lay inside the filled area.

2. The frontier of the feasible can be divided in two pieces: the Efficient Frontier, which contains the performance of the feasible portfolios that solve the optimization problem, and the Inneficient frontier, which contains the worst feasible portfolios.

Despite some calibration statistic problems, the optimizations methods multiply the estimation errors and can result in very bad estimated efficient frontiers, the mean-variance is quite approximate to the reality when the stocks follow an elliptical distribution [19] , in particular, when the n assets are distributed as a normal multivariate random variable. In this case, the relationship between the expected utility and the mean-variance analysis becomes clear. The certainty equivalent of a portfolio ϕ is defined by the equality

$$u(c(\phi, u)) = E_\phi[u(x)].$$

[19] An elliptical distribution is a distribution with ellipsoid level curves.

When μ is a normal distribution, we can write

$$E_\phi[u(x)] = E[u(\mu_\phi) + \sigma_\phi \mathcal{Z})],$$

where $\mathcal{Z}$ is a standard normal distribution. This allows us to parametrize the expected utility through the expected return and the standard deviation. If we differentiate the parametrization to study the behaviour with respect μ_ϕ and σ_ϕ, results

$$\frac{\partial}{\partial \mu_\phi} E_\phi[u(x)] = E[u'(\phi)] \geq 0, \qquad \frac{\partial}{\partial \sigma_\phi} E_\phi[u(x)] = E[u'(\phi)\mathcal{Z}] \leq 0.$$

In the first equality we used the monotonicity of u and in the second the monotonicity and the concavity u and the symmetry of $\mathcal{Z}$. This is the exact behaviour that we should expect when we used Taylor series to find a second order approximation to the certainty equivalent:

$$u(c(\phi, u)) \approx u(m) + \frac{1}{2} u''(m) \operatorname{var}(\phi).$$

Under the normality assumption, and the related parametrization, the slope of the certainty equivalent curves, curves of portfolios with the same certainty equivalent, can be computed from

$$E[u'(\phi)] d\mu_\phi + E[u'(\phi)\mathcal{Z}] d\sigma_\phi = 0.$$

This results in

$$\frac{d}{d\sigma} \mu_\phi = -\frac{E u'(\phi)\mathcal{Z}}{E[u'(\phi)]} > 0,$$

from where we deduce this interesting property: the certainty equivalent curves are ordered in an increasing sense, have a positive slope and are convex. Again, this is what we expect from the utility theory.

The certainty equivalent curves, also known as *indifference curves*, impact directly in the feasible area when we consider the interest rate of the risk-free asset. It is clear that any portfolio with minor expected return than the interest rate is unacceptable. This means, that the indifference curve of the interest rate is a lower bound and only portfolios who are over indifference curves of greater level, but still in the feasible area, are admissible. In particular, if the interest rate indifference curve and the efficient frontier intersect in a single point, then there exists a unique optimal portfolio.

Remark 10.1. The previous analysis also yields in a more general case: when the portfolio distributions belong to the same family of the asset distributions. This is the case of normal variables, which are closed for linearity.

11 Suboptimality of path-dependent pay-offs

The financial industry is an innovative economic sector that generates new and complex derivatives. Despite some of the products have a legitimate demand and a financial justification, others are just a big messy nearer to casino games than to helpful products. In some cases, the complexity of the new derivatives is so high that only track the evolution of the product requires thousands of lines of source code [20] . Not to mention that we do not have good models to price them known to the general public and, as a consequence, they put in danger systemic banking and investment companies.
In this section, we aim to analyse a whole family of derivatives and discard them as suboptimal with respect to the expected utility theory. In particular, we will see that under some market hypothesis, utility maximizers will always choose derivatives whose pay-offs depends on the state of the stock at the investment horizon over path dependent derivatives [21].

11.1 The framework of the problem: Lévy Processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space and consider a market with a single risky asset and a risk free interest $r \geq 0$. Suppose that there are no transactions costs, dividends and restrictions to short sales and borrowing. As usual, the random price of the risky asset at time t is defined as

$$S_t = S_0 e^{X(t)},$$

[20] Examples of such products are *Squared Collateralized Debt Obligation*- CDO^2- and CDO^n.
[21] One well known example of this kind of products are the *Contingent Convertible bonds*, also known as CoCos.

where $X(t)$ is a Lévy process. A Lévy process is a class of stochastic process that includes the Brownian Motion:

Definition 11.1. A Lévy process $X(t)$ is a stochastic process on some probability space with these properties:

(a) The process starts at 0: $P[X(0) = 0] = 1$.

(b) The increments are independent. If

$$t_0 \leq t_1 \leq \ldots \leq t_k,$$

then

$$P[X(t_i) - X(t_{i-1}) \in H_i,\ i \leq k] = \prod_{i \leq k} P[X(t_i) - X(t_{i-1}) \in H_i].$$

(c) For $0 \leq s < t$ the increment $X(t) - X(s)$ is stationary:

$$X(t) - X(s) = X(t - s),$$

11.2 Statement of the problem

In this market, consider an agent which is a maximizer of a utility function u. The agent has a fixed investment horizon T and faces the decision at time $t = 0$ of investing in a path dependent pay-off P_g defined as

$$Z_g = g(S_{t_i} \mid 0 \leq t_i \leq T,\ i = 1, 2, \ldots, n),$$

for some function g. We claim that the agent will always prefer the alternative pay-off

$$E_p[Z_g \mid S_T],$$

which depends only on the price of the asset S at the final time T. To see this, we prove that both derivatives have the same risk neutral price. Then, by the properties of the conditional expectation and Jensen's Inequality, results

$$E_P[u(P_Z)] = E_P[E_P[u(Z_g) \mid S_T]] \leq E_P[u(E_P[Z_g \mid S_T])], \tag{18}$$

for all concave increasing function u. Therefore, at equal price, the agent will choose $E[Z_g \mid S_T]$ over Z_g and the claim follows.

We now prove for the continuous and the discrete time settings that the price of both derivatives is the same. To simplify the notation, for any vector of times $t = (t_1, t_2, \ldots, t_n)$, and for any vector of values $x = (x_1, x_2, \ldots, x_n)$ we define its distribution function as

$$F_t(x) = P[X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_n) \leq x_n].$$

When $(X(t_1), X(t_2), \ldots, X(t_n))$ is a continuous random vector, we define $f_t(x)$ as its density. If the vector is discrete, we define $f_t(x)$ as

$$f_t(x) = P[X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_n) = x_n].$$

We suppose that all needed density and moment generators functions exists.

The Esscher transform will play a fundamental role on the proof:

Definition 11.2. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space. The Esscher transform Q of P is a probability measure that satisfies

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = \frac{e^{hX(t)}}{E_P[e^{hX(t)}]} = \frac{e^{hX(t)}}{e^{\psi(h)t}}.$$

The Esscher transform is well defined for Lévy processes due to the independent increments property.

We now use the Esscher transform to construct a risk neutral probability as follows. First we claim that the Radon-Nikodym process, i.e. the sequence of Radon-Nikodym derivatives with respect to the filtration, is a P-martingale. Let

$$Y_n = \left(\frac{dQ}{dP}\right)_{|\mathcal{F}_n},$$

be the sequence of Radon-Nikodym derivatives. We have

$$Q(A) = \int_\Omega Y_n \mathbb{1}_A dP = \int_\Omega Y_{n-1}\mathbb{1}_A dP, \qquad \forall A \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$$

Then, if $Z = E_P[\Delta Y_n \mid \mathcal{F}_{n-1}]$, by definition of conditional expectation

$$E_P[Z\mathbb{1}_A] = E_P[\Delta Y_n \mathbb{1}_A] = \int_\Omega \Delta Y_n \mathbb{1}_A dP = \int_A (Y_n - Y_{n-1})dP = 0, \qquad \forall A \in \mathcal{F}_{n-1}$$

and the claim follows. Apply the previous result to obtain

$$\left.\frac{dQ}{dP}\right|_{\mathcal{F}_t} = E_P\left[\left.\frac{dQ}{dP}\right| \mathcal{F}_t\right].$$

Then, because $X(t)$ is a Lévy process,

$$\frac{E_P[e^{hX(T)} \mid F_T]}{E_P[e^{hX(T)}]} = \frac{E_P[e^{h(X(T)-X(t))}]e^{hX(t)}}{E_P[e^{hX(T)}]} = \frac{e^{hX(t)}}{E_P(e^{hX(t)})}.$$

We want Q to be risk neutral probability. i.e.,

$$E_Q[e^{-rT}S_T \mid \mathcal{F}_t] = e^{-rt}S_t.$$

By definition of the process (S_t),

$$E_Q[e^{-rT+X(T)}S_0 \mid \mathcal{F}_t] = e^{-rt+X(t)}S_0.$$

We have, by the Bayes rule, that

$$E_Q[X \mid \mathcal{F}_t] = \frac{E_P[Z_t X \mid \mathcal{F}_t]}{Z_T}, \qquad Z_t := \left.\frac{dQ}{dP}\right|_{\mathcal{F}_t}.$$

Then,

$$\frac{E_P[e^{-rT+X(T)}e^{hX(T)} \mid \mathcal{F}_t]}{e^{\psi(h)T}} \frac{e^{\psi(h)t}}{e^{hX(t)}} = e^{-rt+X(t)}.$$

Hence,

$$e^{\psi(h)(t-T)}e^{-rT}E_P[e^{(1+h)X(T)} \mid \mathcal{F}_t] = e^{-rt+(1+h)X(t)},$$

$$e^{\psi(h)(t-T)}E_P[e^{(1+h)(X(T)-X(t))}]e^{(1+h)X(t)} = e^{-r(t-T)+(1+h)X(t)},$$

$$e^{\psi(h)(t-T)+\psi(1+h)(T-t)} = e^{-r(t-T)}.$$

Taking logarithms at both sides of the equality, we can always find a h for Lévy Processes that solves the problem

$$\psi(1+h) - \psi(h) = r.$$

We denote $h*$ the solution of this equation and P^* the related risk neutral probability.

Remark 11.3. When the density of $X(t)$ exist, the density of the Esscher transform equals to

$$f_t^{(h)}(x) := \frac{e^{hx}f_t(x)}{m_t(h)} \tag{19}$$

where

$$m_t(h) := m_{X_t}(h) = \int_\Omega e^{hx}f_t(x)dx,$$

is the moment generator function of $X(t)$. The multidimensional density is equal setting the variable $x = x_t$.

With the help of the Esscher transform, we prove now the main theorem of this section:

Theorem 11.4. *Let $C(\cdot)$ denote the risk neutral price of a derivative. Then,*

$$C(E_P[Z_g \mid S_T]) = C(Z_g).$$

Proof. Is sufficient to prove $E_P[Z_g \mid S_T] \equiv E_{P^*}[Z_g \mid S_T]$ and apply the equalities

$$E_{P^*}[Z_g] = E_{P^*}[E_{P^*}[Z_g \mid S_T]] = E_{P^*}[E_P[Z_g \mid S_T]].$$

Consider a Esscher transform $f_t^{(h^*)}$ of parameter h^* of $X(t)$ such that the evolution of the discounted stock prices are martingales. To prove the equality, we will check that for all $x = (x_1, \ldots, x_n)$ and $t = (t_1, \ldots, t_n)$, with $0 < t_1 < \ldots < t_n < T$, and $y \in \mathbb{R}$ we have

$$f_t(x \mid X_T = y) = f_t^{(h^*)}(x \mid X_T = y).$$

Setting the convention $t_0 = 0$ and $x_0 = 0$, and using the properties of Lévy processes,

$$f_t^{(h^*)}(x \mid X_T = y) = \frac{1}{f_T^{(h^*)}(y)} \left(\prod_{i=1}^{n} f_{t_i - t_{i-1}}^{(h^*)}(x_i - x_{i-1}) \times f_{T-t_n}^{(h^*)}(y - x_n) \right) =$$

$$= \prod_{i=1}^{n} \frac{f_{t_i - t_{i-1}}(x_i - x_{i-1}) e^{h^*(x_i - x_{i-1})}}{m_{t_i - t_{i-1}}(h^*)} \times \frac{f_{T-t_n}(y - x_n) e^{h^*(y - x_n)}}{m_{T-t_n}(h^*)} \times \frac{m_T(h^*)}{f_T(y) e^{h^* y}} =$$

$$= \frac{1}{f_T(y)} \left(\prod_{i=1}^{n} f_{t_i - t_{i-1}}(x_i - x_{i-1}) \times f_{T-t_n}(y - x_n) \right) = f_t(x \mid X_T = y).$$

□

Corollary 11.5. *Under the hypothesis of this section, risk averse agent who prefer more to less will always choose path independent pay-offs over the related path dependent ones.*

Remark 11.6. We can give a direct proof of the result if the Lévy process is a binomial lattice without relying in auxiliary techniques like the Esscher transform or the content of the next subsection. If $X(S_0, \ldots, S_n)$ is a random payoff, then

$$E_P[X(S_1, \ldots, S_n) \mid S_n = s_n] = \sum X(S_1, \ldots, S_n) P(S_1 = s_1, \ldots, S_n = s_n \mid S_n = s_n) =$$

$$= \frac{1}{P(S_n = s_n)} \sum X(S_1, \ldots, S_n) P(S_1 = s_1, \ldots, S_n = s_n),$$

and

$$E_{P^*}[X(S_1, \ldots, S_n) \mid S_n = s_n] = \ldots = \frac{1}{P^*(S_n = s_n)} \sum X(S_1, \ldots, S_n) P^*(S_1 = s_1, \ldots, S_n = s_n).$$

Hence, it is sufficient to check that the equality holds term by term. i.e.

$$\frac{P(S_1 = s_1, \dots, S_n = s_n \mid S_n = s_n)}{P^*(S_1 = s_1, \dots, S_n = s_n \mid S_n = s_n)} = \frac{P(S_n = s_n)}{P^*(S_n = s_n)}.$$

to have equal risk neutral prices. To see this, recall to facts about the binomial trees: the phase space of the last period is fully determined by the number of ups in the trajectory and Ω is in a bijection with the sets of vectors $\{0,1\}^n = (i_1, i_2, \dots, i_n)$. If we rewrite the equality in terms of zeros and ones,

$$\frac{P(i_1, i_2, \dots, i_n)}{P^*(i_1, i_2, \dots, i_n)} = \frac{p^{i_1}(1-p)^{1-i_i} \cdots p^{i_n}(1-p)^{1-i_n}}{p^{*i_1}(1-p^*)^{1-i_i} \cdots p^{*i_n}(1-p^*)^{1-i_n}} =$$

$$= \frac{\binom{N}{i_1+\ldots+i_n} p^{i_1+\ldots+i_n}(1-p)^{N-(i_1+\ldots+i_n)}}{\binom{N}{i_1+\ldots+i_n} p^{*i_1+\ldots+i_n}(1-p^*)^{N-(i_1+\ldots+i_n)}} = \frac{P(S_n = i_1 + \ldots + i_n)}{P^*(S_n = i_1 + \ldots + i_n)}.$$

11.3 A direct result on processes without density

A main drawback of the previous theorem is that we have made assumption about the existence of the all the densities and moments that we needed in the construction of the Esscher transform and the proof. Another strategy, which is hidden is to check a condition on the Radon-Nikodym derivative.

Proposition 11.7. *Let $X \geq 0$ be a payoff. Consider a model where the risk neutral probability Q satisfies*

$$\frac{\mathrm{d}Q}{\mathrm{d}P} \in \sigma(S_T),$$

then

$$\mathbb{E}_Q(X|S_T) = \mathbb{E}_P(X|S_T).$$

Proof. First

$$Z = \mathbb{E}_Q(X|S_T) \Longleftrightarrow \mathbb{E}_Q(YZ) = \mathbb{E}_Q(YX) \text{ for all } Y \geq 0,\, Y \in \sigma(S_T),$$

then

$$\mathbb{E}_Q(YZ) = \int_\Omega YZ\mathrm{d}Q = \int_\Omega Y\frac{\mathrm{d}Q}{\mathrm{d}P}Z\mathrm{d}P = \int_\Omega \bar{Y}Z\mathrm{d}P = \int_\Omega \bar{Y}X\mathrm{d}P,$$

with $\bar{Y} \geq 0$ and $\bar{Y} \in \sigma(S_T)$ arbitrary, so $Z = \mathbb{E}_P(X|S_T)$. □

Corollary 11.8. *If the risk neutral probability satisfies* $\frac{\mathrm{d}Q}{\mathrm{d}P} \in \sigma(S_T)$, *path-dependent payoffs are* dominated, *in the sense that is there is another payoff with the same initial price and more terminal utility.*

Proof. Given a payoff X we can take $\bar{X} := \mathbb{E}_Q(X|S_T)$. Then, the price is the same, assuming $r = 0$,

$$\mathbb{E}_Q(X_T) = \mathbb{E}_Q(\mathbb{E}_Q(X|S_T)).$$

Now, by the previous proposition

$$\bar{X} = \mathbb{E}_Q(X|S_T) = \mathbb{E}_P(X|S_T),$$

and given a utility function u

$$\mathbb{E}_P(u(\bar{X})) = \mathbb{E}_P(u(\mathbb{E}_P(X|S_T))) \geq \mathbb{E}_P(\mathbb{E}_P(u\left(X\right)|S_T)) = \mathbb{E}_P(u\left(X\right)),$$

where the inequality follows from the Jensen inequality since u is concave. □

Part VI. Drawbacks of the expected utility theory

12 Empirical drawbacks of expected utility: The Allais and the Ellsberg Paradoxes.

The first set of drawbacks of the model are the Allais and the Ellsberg Paradoxes. We call them paradoxes because they do not attack the mathematical foundations of the model. Instead, they present a series of experiments that show that our model fails to model how economic agents act when facing the risk or the uncertainty.

In the original papers of Allais and Ellberg, the theory of the utility functions discussed is a refinement made by Savage and Friedman. Nevertheless, the paradoxes still apply to the model we saw in this text and, simple as they are, suppose a big problem.

12.1 The Allais Paradox

The Allais paradox was set by Maurice Allais in 1953 and challenges the assumption that rational economic agents follows the independence axiom. Remember that we have shown that a preference order has a von Neumann-Morgenstern representation if and only if it satisfies both the Archimedean and independence axioms. Hence, the whole integral representation must be discarded, or modified, if independence axiom does not hold. Lets us examine the paradox.

Example 12.1 (The Allais Paradox)**.** The Allais Paradox is an experiment in which the interviewers offered a set of lotteries and accounted the preferences. First, if we offer the following two lotteries [22]:

$$\nu_1 := 0.33\delta_{2500} + 0.66\delta_{2400} + 0.01\delta_0, \quad \mu_1 := \delta_{2400},$$

most people prefer μ_1 over ν_1 even if the expected payoff favours ν_1. At the same time, if the same people now has to choose between the following 2 lotteries,

$$\mu_2 := 0.34\delta_{2400} + 0.66\delta_0, \quad \nu_2 := 0.33\delta_{2500} + 0.67\delta_0,$$

they tend to prefer ν_2 over μ_2, choosing more expectation over risk. In particular, 65% of the people stated [23] both

$$\mu_1 \succ \nu_1 \qquad \text{and} \qquad \nu_2 \succ \mu_2.$$

This observed behaviour contradicts the model:

Proposition 12.2. *The Allais paradox is an experimental violation of the axiom of independence.*

[22] We use the rescaled example of [7, p.59]. Allais used other values in the original paper.
[23] The numbers are due to Allais and independently confirmed by D. Kahnemann and A. Tversky.

Proof. If the axiom of independence holds, then necessarily

$$\alpha\mu_1 + (1-\alpha)\nu_2 \succ \alpha\nu_1 + (1-\alpha)\nu_2 \succ \alpha\nu_1 + (1-\alpha)\mu_2$$

for all $\alpha \in (0,1)$. By taking $\alpha = \frac{1}{2}$, we get

$$\frac{1}{2}(\mu_1 + \nu_2) \succ \frac{1}{2}(\nu_1 + \mu_2)$$

which is a contradiction to the fact that

$$\frac{1}{2}(\mu_1 + \nu_2) = \frac{1}{2}(\nu_1 + \mu_2).$$

□

12.2 The Savage refinement: preferences on asset profiles

This paradox introduces a new ingredient in the model that we did not take in account. So far, when we have wrote about lotteries, we always considered the case that either the probability distribution, or the random variable if we fixed a probability space, was absolutely known by the economic agent; the agent only faced a risk that he was able to calibrate. To explain the Allais Paradox, we can add uncertainty in the sense that the agent does not know exactly the underlying probability distributions.

L.J. Savage introduced a refinement of the theory that we review in what follows. Instead of taking a set of probability distributions as $\mathcal{X}$, we now fix a measurable space $(\Omega, \mathcal{F})$ and denote $\mathcal{X}$ as a set of bounded measurable functions X. The Xs are what we call assets: measurable functions which associate real valued payoffs to possible scenarios.

As usual, we assume a preference relation on $\mathcal{X}$ which we assume monotone in the following sense:

$$Y \succeq X \text{ if } Y(\omega) \geq X(\omega) \text{ for all } \omega \in \Omega.$$

Without entering in all the details, if we assume some axioms and continuity conditions, we can guarantee a numerical representation of the following kind:

Definition 12.3. Let be $(\Omega, \mathcal{F})$ be a measurable space and let be $\mathcal{X}$ a set of bounded measurable functions defined on $(\Omega, \mathcal{F})$. We say that a preference order $\succ$ has a **Savage representation** U if it can be written as

$$U(X) = E_{\mathcal{Q}}[u(X)] = \int u(X(\omega))\mathcal{Q}(d\omega), \qquad \forall X \in \mathcal{X},$$

where $\mathcal{Q}$ is a probability measure on $(\Omega, \mathcal{F})$ and u a real valued function.

The probability $\mathcal{Q}$ must be thought as the subjective view of the probabilities of the events and allows us to interpret the Allais Paradox as follows. Assume that the agent accepts the view that the scenarios $\omega \in \Omega$ are generated in accordance to an objective probability P. In this case, P is the Lebesgue measure on $\Omega = [0,1]$ and $\mathcal{X}$ is the space of bounded right continuous increasing function on $[0,1]$. Let $\mu_{P,X}$ denote the distribution of X under P. Consider the next Lemma:

Lemma 12.4. *Suppose X is a real valued random variable on a probability space $(\Omega, \mathcal{F}, P)$ with distribution function*

$$F_X(x) = P[X \leq x],$$

and let q_X denote the right-continuous inverse of F_X

$$q_X(s) := \inf\{x \in \mathbb{R} \mid F_X(x) > s\}.$$

Let U be a random variable on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ with a uniform distribution on $(0,1)$. Then

$$\tilde{X}(\tilde{\omega}) := q_X(U(\tilde{\omega}))$$

has the same distribution as X.

Proof. See section 2.4 of [7]. □

By this lema, every probability measure on $\mathbb{R}$ with bounded support is of the form $\mu_{P,X}$ for some $X \in \mathcal{X}$:

$$\mathcal{M}_b(\mathbb{R}) = \{\mu_{P,X} \mid X \in \mathcal{X}\}.$$

If we pair $X \in \mathcal{X}$ with the lottery $\mu_{P,X}$, the preference relation on $\mathcal{X}$ induces a preference relation on $\mathcal{M}_b(\mathbb{R})$ with numerical representation

$$U^*(\mu_{P,X}) := U(X).$$

To explain the paradox, assume now that the agent distort P with a pessimistic bias. This is a legitimate assumption because many people think that they are deceived when facing experiments like the paradox one. We formalize this bias replacing P with the subjective measure

$$\mathcal{Q} := \alpha\delta_0 + (1-\alpha)P$$

for some $\alpha \in (0,1)$. The corresponding Save representation is

$$U^*(\mu_{P,X}) = E_{\mathcal{Q}}[u(X)] = \int u d\mu_{\mathcal{Q},X} = \alpha u(X(0)) + (1-\alpha)\int u d\mu_{P,X}.$$

Note that $X(0) = \ell(\mu_{P,X})$ with

$$\ell(\mu) := \inf(\operatorname{supp}\mu) = \sup\{a \in \mathbb{R} \mid \mu((-\infty, a)) = 0\}.$$

Therefore, replacing P by $\mathcal{Q}$, we obtain the non-linear distortion that maps the lottery $\mu_{P,X}$ to the lottery $\mu^* = \mu_{\mathcal{Q},X}$ given by

$$\mu^* = \alpha\delta_{\ell(\mu)} + (1-\alpha)\mu.$$

The preference relation has the numerical representation

$$U^*(\mu) = \int u(x)\mu^*(dx), \qquad \mu \in \mathcal{M}_b(\mathbb{R}).$$

Under this new apparatus, recall the lotteries of the Allais Paradox

$$\nu_1 := 0.33\delta_{2500} + 0.66\delta_{2400} + 0.01\delta_0, \quad \mu_1 := \delta_{2400},$$

$$\mu_2 := 0.34\delta_{2400} + 0.66\delta_0, \quad \nu_2 := 0.33\delta_{2500} + 0.67\delta_0.$$

and distort them as follows:

$$\mu_1^* = \mu_1, \qquad \nu_1^* = \alpha\delta_0 + (1-\alpha)\nu_1,$$

$$\mu_2^* = \alpha\delta_0 + (1-\alpha)\nu_1, \qquad \nu_2^* = \alpha\delta_0 + (1-\alpha)\nu_2.$$

If we set $u(x) = x$, we have

$$U^*(\nu^2) > U^*(\mu_2)$$

and for $\alpha > \frac{9}{2409} \approx 0.0037$ we obtain

$$U^*(\mu_1) > U^*(\nu_1)$$

as in the Allais Paradox.

12.3 The Ellberg Paradox

The Ellberg Paradox is, as the Allais Paradox, the result of an experiment.

Example 12.5 (The Ellberg Paradox)**.** Suppose we have an urn containing 30 red balls and 60 other balls that are either black or yellow. The number of black and yellow balls is unknown. All 90 balls have the same chance to be drawn. Now we offer two lotteries:

- μ_1: Win 100 EUR if a red ball is drawn.
- ν_1: Win 100 EUR if a black ball is drawn

Later, we offer two more lotteries:

- μ_2: Win 100 EUR if a red or yellow ball is drawn.

- ν_2: Win 100 EUR if a black or yellow ball is drawn.

The paradox is what follows: when surveyed, most people strictly prefer μ_1 over ν_1 and ν_2 over μ_2.

Proposition 12.6. *The Ellberg paradox violates the paradigm of the expected utility. That is, economic agents act as maximizers of a utility function.*

Proof. Suppose that an economic agent is a maximizer of a utility function u, that is, a strictly increasing continuous concave function. Let B, R, Y be the probabilities of drawing, respectively, a black, a red or a yellow ball from the urn. If the economic agents shows the behaviour of the paradox, we have

$$\mu_1 \succ \nu_1 \Leftrightarrow Ru(100) + (1-R)u(0) > Bu(100) + (1-B)u(0).$$

Because u is strictly increasing, we obtain

$$R(u(100) - u(0)) > B(u(100) - u(0))$$

and

$$R > B.$$

At the same time, as the agent prefers ν_2 to μ_2,

$$\nu_2 \succ \mu_2 \Leftrightarrow Bu(100) + Yu(100) + Ru(0) > Ru(100) + Yu(100) + Bu(0).$$

Which again simplifies to

$$B(u(100) - u(0)) > R(u(100) - u(0)).$$

That implies $R < B$, which is a contradiction and the economic agent cannot follow the expected utility paradigm if the Ellberg paradox holds. $\square$

As in the Allais Paradox, if the economic agent does not know all the probabilities for sure, then our model cannot apply and we need to consider how the beliefs in probabilities, like being the experiment a trick or have a lucky day, impact the theory. But, the possible solution of Savage is not enough to solve this paradox; the Ellberg paradox is much more stronger than the Allais one, as it does not depend on any utility function of u of the agent, nor on the risk aversion.

One possible solution is to go further than the Savage representation and adopt *worst case approach* when evaluating expected utilities. The worst case approach requires a set of subjective probabilities $\mathcal{M}$ on $(\Omega, \mathcal{F})$ and set

$$U(x) := \inf_{\mathcal{Q} \in \mathcal{M}} E_{\mathcal{Q}}[u(X)].$$

We are not going to see the details in this work, but with an intelligent, and reasonable, choice of $\mathcal{M}$ it is possible to give an answer to the Ellsberg paradox[24].

13 Other violations of the Independece Axiom and a proposed solution

We saw in the previous section two systematic violations of the Independence Axiom: the Allais and the Ellberg Paradoxes. However, those are not the only violations and we devote this section to review the most popular: those that can be found on the article [11] of Mark Machina.

13.1 Oversensitivity to changes in small probabilities

The Allais and Ellberg Paradoxes are, in words of Machina, examples of oversensitivity to changes to the probabilities of low-probability events.

The so-called subjective expected utility models are a second source of systematic violations. Like in the Savage refinement, such models assume that the agents transform the set of objective probabilities into subjective probabilities. The key aspect with respect to the study of the

[24] The details can be found in Section 2.5 of [7].

Independence axioms is that the axiom requires the transformations of objective to subjective probabilities to be linear. Hence, any observed non linear transform is against the model. And the non linearity is what empirical tests show: agents overemphasize small probabilities and underemphasize the large ones.

Some authors argued that this is compatible with standard expected utility theory if those agents maximize convex functions, but experiments designed to explicitly distinguish between behaviour due to curvature or to overemphasizing small probabilities resulted in more evidence against the curvature hypothesis[25].

Once the transform of subjective probabilites are allowed to be nonlinear to fit the experiments, the whole frameworks is damaged as it also loses the monotonicity property in the sense of stochastic dominance. That is, that a lottery μ dominates a lottery ν if its is preferred with respect to all utility functions.

13.2 The role of the past and the Markowitz hypothesis

The axioms of the expected utility theory implies that the ordering in lotteries correspond to the expectation of a fixed utility function defined over the terminal levels of wealth. Friedman and Savage observed that people of all income levels buy insurances and lotteries and defined a von Neumann-Morgenstern representation which explained that behaviour:

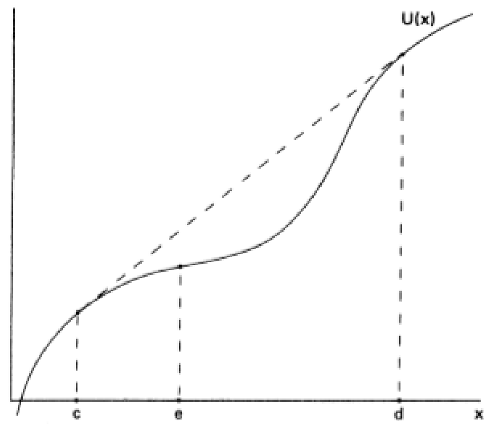

Fig. 5: Observed behaviour. Source: [11, p. 283.]

The function u associated with this von Neumann-Morgenstern representation is not, strictly speaking, a utility function; its is locally concave, linear or convex attending to the initial wealth

[25] We direct the reader to [11, p. 291] for a deeper explanation of this phenomenon and extensive bibliography.

levels of the agent. Still, Markowitz observed that the assumption that this utility function is defined over ultimate wealth levels is not consistent with the fact that individuals of all wealth levels purchase lottery tickets and buy insurance.

Agents with wealth levels less than c, poor agents, or greater that d, rich agents, would never accept fair bets. Agents with wealth just d would be willing to take insurances. Insurances involve an expected loss to cover against big losses. Also, agents with wealth near $\frac{c+d}{2}$ would prefer all fair bets up to at least $\frac{d-c}{2}$, which contradicts the fact that most people do not accept fair bets.

Markowitz noted that individuals of all wealth levels behaved as if their initial wealth was near the inflection point e and hypothesized that changes in wealth caused the utility function to shift horizontally to keep the inflection point near the current wealth level. Further experimental evidence suggest that individual gambling behaviour at different initial wealth levels is more indicative of a shifting utility function than of movements along a fixed utility function[26]. This implies that changes in initial wealth cause the agent to change his preference order on the set of probability measures.

14 Smooth preferences

In the last chapter, we saw that one of the key ingredients of the theory, the Independence axiom, is also its experimental Achilles' Ell. However, we can recover some fundamental results, like the Arrow-Pratt Coefficient of Risk Aversion without relying in the axiom. In order to achieve it, we need to impose additional smooth conditions on the numerical representations and move away from the von Neumann-Morgenstern representation paradigm.

We set now $\mathcal{X}$ as the set of all probability distribution functions $F(\cdot)$ over the bounded interval $[0, M]$. We assume that we have a preference order on $\mathcal{X}$ representable by a real functional $V(\cdot)$ on $D[0, M]$. We set on $D[0, M]$ the topology of weak convergence [27] , i.e.,

$$(F_n(\cdot))_n \to F(\dot{)} \text{ weakly } \Leftrightarrow F_n(\cdot) \to F(\cdot) \text{ point-wise on continuity points of } F(\cdot).$$

[26] [11, p. 285]

[27] Check the Portmanteau Lemma, Theorem 2.17, for equivalences.

The condition of differentiability on the numerical representation $V(\cdot)$ also requires the existence of a norm in the space

$$\Delta D[0,M] = \{\lambda(F^* - F) \mid F, F^* \in D[0,M],\ \lambda \in \mathbb{R}\}.$$

With additional Functional Analysis, we can prove that both this norm, and the weak convergence, is induced by the L^1 metric

$$d(F,F^*) = \int |F(x) - F(x)|dx,$$

and has the form

$$||\lambda(F^* - F)|| = |\lambda| d(F,F^*).$$

We have now all the elements to define the smooth condition: be Fréchet differentiable on the space $D[0,M]$.

Definition 14.1. The functional $V(\cdot)$ is said to be Fréchet differentiable on the space $D[0,M]$ if there exists a functional $\psi(\cdot;F)$ defined on $\Delta[0,M]$ such that

$$\lim_{||F^*-F||\to 0} \frac{|V(F^*) - V(F) - \psi(F^* - F; F)|}{||F^* - F||} = 0 \tag{20}$$

The Fréchet derivative is the natural extension of the $\mathbb{R}^n$-differentiation to Banach Spaces, because it just adapts the standard definition to the correspondent norms. From its mere definition, we can recover a local version of the idea of agents as maximizers of utility functions. We can rewrite (20) as

$$V(F^*) - V(F) = \psi(F^* - F; F) + o(||F^* - F||) \tag{21}$$

$\Delta[0,M]$ is a linear subspace of $L^1[0,M]$. Thus, by the Riesz representation theorem on $L^1[0,M]$, we have that for any $F^* \in D[0, M]$,

$$\psi(F^* - F; F) = \int (F^*(x) - F(x))h(x; F)dx, \qquad h(x; F) \in L^\infty[0, M] \tag{22}$$

which is equivalent, by the Radon-Nikodym theorem, to

$$\psi(F^* - F; F) = -\int (F^*(x) - F(x))dU(x; F). \tag{23}$$

We have

$$U(x, F) := -\int_0^x h(s; F)ds \tag{24}$$

from which it follows that $U(\cdot; F)$ is absolutely continuous and differentiable almost everywhere on $[0, M]$. Substituting (22) into (21) and integrating by parts, we get

$$V(F^*) - V(F) = \int U(x; F)(dF^*(x) - dF(x)) + o(||F^* - F||). \tag{25}$$

The last equality shows that the differential change from the distribution $F(\cdot)$ to a distribution $F^*(\cdot)$ changes the value of the numerical representation by

$$\int U(x; F)(dF^*(x) - dF(x)).$$

That is, by the difference in the expected value of $U(x; F)$ with respect to the distribution $F^*(\cdot)$ and $F(\cdot)$. Or equivalently, the agent ranks near distributions as would an expected utility maximizer with local utility function $U(x; F)$. This is not a surprise because the differential of a functional is its best local linear approximation. Therefore, by the arguments of the section 13.2, linearity is equivalent to expected utility maximization.

One price that we have paid after shifting to this new theory is that the utility function $U(\cdot, F)$ is only local. To derive global consequences from local properties, we can act as in standard Differential Analysis. For example, we can prove that a function is monotone increasing in an interval if its derivative is positive in each point, even when the derivative is everywhere different.

The general method in our case is to use path integrals in the space $D[0,M]$. If the path

$$\{F(\cdot;\alpha) \mid \alpha \in [0,1]\}$$

is smooth enough so the term $||F(\cdot;\alpha) - F(\cdot;\alpha^*)||$ is differentiable in α at $\alpha = \alpha^*$, we get from (25)

$$\frac{d}{d\alpha}(V(F(\cdot;\alpha)))\bigg|_{\alpha^*} = \frac{d}{d\alpha}\left(\int U(x;F(\cdot;a^*)dF(x;\alpha)\right)\bigg|_{\alpha^*} + (o(||F(\cdot;\alpha - F(\cdot;\alpha^*)||))|_{\alpha^*} =$$

$$= \frac{d}{d\alpha}\left(\int U(x;F(\cdot;a^*)dF(x;\alpha)\right)\bigg|_{\alpha^*}.$$

Applying the Fundamental Theorem of Calculus,

$$v(F(\cdot;1)) - V(F(\cdot;0)) = \int_0^1 \left[(U(x;F(\cdot;\alpha^*))dF(x;\alpha))|_{\alpha^*}\right] d\alpha^*,$$

which illustrates how the agent's reaction to the sfhift from $F(\cdot;1)$ to $F(\cdot;0)$ depends on the properties of the local utility function at each point along the path $\{F(\cdot;\alpha) \mid \alpha \in [0,1]\}$.
With this approach, we can prove and recover the two main properties of utility functions: monotony and risk aversion:

Theorem 14.2 (Monotonicity)**.** *Let $V(\cdot)$ be a Fréchet differentiable preference function on $D[0,M]$. Then $V(F^*) \geq V(F)$ whenever $F^*(\cdot)$ stochastically dominates $F(\cdot)$ if and only if $U(x;F)$ is nondecreasing in x for all $F(\cdot) \in D[0,M]$.*

Proof. See [11, Appendix]. □

Theorem 14.3 (Risk Aversion)**.** *Let $V(\cdot)$ be a Fréchet differentiable preference function on $D[0,M]$. Then $V(F^*) \geq V(F)$ whenever $F^*(\cdot)$ differs from $F(\cdot)$ by a mean preserving increase in risk if and only if $U(x;F)$ is a concave function of x for all $F(\cdot) \in D[0,M]$.*

Proof. See [11, Appendix]. □

To keep this work in a reasonable size, we end this section with an theorem that states under what circumstances the Arrow-Pratt Theorem, Proposition 4.5, holds:

Definition 14.4. If $F(\cdot)$ and $F^*(\cdot)$ are two cumulative distribution functions over a wealth interval $[0, M]$, then F^* is said to differ from F by a **simple compensated spread** if the individual is indifferent between F and F^* and if $[0, M]$ may be partitioned into disjoint intervals I_L and I_R (with I_L to the lefy of I_R) such that $F^*(x) \geq F(x)$ for all $x \in I_L$ and $F^*(x) \leq F(x)$ for all $x \in I_R$.

Theorem 14.5. *The following conditions on a pair of Fréchet differentiable preference functionals $V(\cdot)$ and $V^*(\cdot)$ on $D[0, M]$ with respective local utility functions $U(x; F)$ and $U^*(x; F)$ are equivalent:*

(a) For arbitrary distributions $F(\cdot), F^{**}(\cdot) \in D[0, M]$ and positive probability p, if c and c^* solve

$$V((1-p)F^{**}+pF) = V((1-p)F^{**}+p\mathbb{1}_{x\geq c} \text{ and } V^*((1-p)F^{**}+pF) = V^*((1-p)F^{**}+p\mathbb{1}_{x\geq c},$$

then $c \leq c^*$ (the conditional certainty equivalents for $V(\cdot)$ are never greater than the corresponding ones for $V^*(\cdot)$).

(b) For all $F(\cdot) \in D[0, M]$, $U(x; F)$ is at least as concave a function of x as $U^*(x; F)$. That is, for all F, $U(x; F)$ is a concave transform of $U^*(x; F)$ so that if these functions are twice differentiable in x, then the analogous to the Arrow-Pratt Coefficient of Risk Aversion satisfy

$$-\frac{\frac{\partial^2}{\partial x^2}U(x; F)}{\frac{\partial}{\partial x}U(x; F)} \geq -\frac{\frac{\partial^2}{\partial x^2}U^*(x; F)}{\frac{\partial}{\partial x}U^*(x; F)}$$

(c) If the distribution $F^*(\cdot)$ differs from $F(\cdot)$ by a simple compensated spread from the point of view of $V^*(\cdot)$ so that $V^*(F^*) = V^*(F)$, then $V(F^*) \leq V(F)$.

Part VII. The Saint Petersburg Paradox and the Expected Utility Theory

As we saw in the introduction, Daniel Bernoulli presented in 1738 before the Imperial Academy of Sciences in Saint Petersburg a paper inspired in the St. Petersburg Paradox. The St. Petersburg Paradox was proposed by Nicholas Bernoulli and is one of the most fruitful problems in Mathematics and Economics. Recall the original problem: *Peter tosses a coin and continues to do so until it should lands "heads" when it comes to the ground. He agrees to give Paul one ducat if he gets "heads"' on the very first throw, two ducats if he gets it on the second, four if on the third, eight if on the fourth, and so on, so that with each additional thrown the number of ducats he must pay is doubled. Suppose we seek to determine the value of Paul's expectation.* And remember that we have shown that it has infinite expectation:

$$E[X] = \sum_{k=1}^{\infty} 2^{k-1} P[\text{Get } k-1 \text{ tails in a row and a head in the } k \text{ tail}] =$$

$$= \sum_{k=1}^{\infty} 2^{k-1} \frac{1}{2^k} = \sum_{k=1}^{\infty} \frac{1}{2} = \infty.$$

In this part of the work, we list some proposed solutions and discuss, when possible, their correctness.

15 Expected Utility

The whole Daniel Bernoulli's paper is devoted to develop the first version of the expected utility theory. He reasons that the wealth has not to be considered in its absolute value. Instead, has to be accounted in the utility it has for Paul. In words of Gabriel Cramer, who independently answered the paradox,

"the mathematicians estimate money in proportion to its quantity, and men of good sense in proportion to the usage that they may make of it."

In the paper, Bernoulli considered two possible utility functions. The square root utility

$$u(x) = \sqrt{x},$$

and the logarithmic utility

$$u(x) = \log x.$$

If Paul has no wealth before playing the lottery, for example if he has only the right to play, or sell, the lottery, the expected utility becomes, under the square root utility,

$$E[\sqrt{X}] = \sum_{k=1}^{\infty} 2^{\frac{k-1}{2}} \frac{1}{2^k} = \sum_{k=1}^{\infty} 2^{\frac{-k-1}{2}} = \sum_{k=1}^{\infty} \left(\frac{1}{\sqrt{2}}\right)^{k+1}.$$

This is a summable series:

$$E[\sqrt{x}] = \frac{1}{\sqrt{2}} \sum_{k=1}^{\infty} \left(\frac{1}{\sqrt{2}}\right)^k = \frac{1}{\sqrt{2}} \left(\frac{1}{1-\frac{1}{2}} - 1\right) = \frac{1}{2-\sqrt{2}}.$$

Bernoulli goes one step further and introduces the concept of certainty equivalent:

However this magnitude is not the equivalent we seek, for this equivalent need not be equal to my moral expectation but should rather be of such a magnitude that the pain caused by its loss is equal to the moral expectation of the pleasure I hope to derive from my gain. Therefore, the equivalent must on our hypothesis, amount to

$$\left(\frac{1}{2-\sqrt{2}}\right)^2 \approx 2.9.$$

With the same initial wealth hypothesis, the certainty equivalent of the lottery under the logarithmic utility equals to 2.

16 The necessity of bounded utility functions

The expected utility solution to the St. Petersburg Paradox did not satisfy Nicholas Bernoulli, but, nevertheless, is considered the historical solution and is cited in most of the books of expected utility theory.

However, this solution is not satisfactory at all. It supposes that economic agents has bounded utility functions. If the utility function u is not bounded, we can always offer a new St. Petersburg lottery with infinite expected utility. Just consider the lottery

$$\mu_u = \sum_{k=1}^{\infty} u^{-1}(\delta_{2^k})\frac{1}{2^{k+1}}.$$

Then, the expected utility is

$$U[\mu_u] = \sum_{k=1}^{\infty} u(u^{-1}(2^k))\frac{1}{2^{k+1}} = \sum_{k=1}^{\infty} 2^{k-1}\frac{1}{2^k} = \sum_{k=1}^{\infty} \frac{1}{2} = \infty,$$

and the St. Peterburg Paradox reappears rescaled. This kind of lottery is also known as Menger's Super-Petersburg Paradox.

Of course, we can impose the boundedness to the set of utility functions for the sake of removing Super-Petersburg Paradoxes from our model. However this can lead to unwanted side effects. To have unbounded utility functions is helpful in some statistical choice model. Also, imposing additional conditions on what a rational behaviour is, restricting further the idea that to behave rationally equals to maximize a utility function, can model away from reality and worsen the predictions.

Some authors, instead of imposing boundedness, assume that there exists some favourable bet, in the sense of having strictly positive fair price, that is declined at every level of wealth. This hypothesis is interesting because:

(a) It implies that the utility function is bounded from above.

(b) We can easily test if it holds, due to a Theorem proved by M. Rabin, with a simple tail or heads lottery.

(c) As a consequence, the model predicts the following behaviour: at high levels of wealth, the agent would reject a bet with huge potential gain even though the potential loss is a tiny part of his wealth. This, which can be proved only imposing that u is bounded from above, contradicts the fact that very rich people buy insurances, travel to Las Vegas or invest in the stock exchange.

First we prove (a) and (b) in the following theorem:

Theorem 16.1. *If a favorable bet μ bounded from below is rejected at any level of wealth, then the utility function u is bounded from above, and there exists $A > 0$ such that the bet*

$$v := \frac{1}{2}\left(\delta_{-A} + \delta_{\infty}\right)$$

is rejected at any level of wealth.

Proof. As μ is bounded from below, it mass is concentrated on $[a, \infty)$ for som $a < 0$, where a is in the interior of an unbounded from above interval S. Moreover, we can choose $b > 0$ such that

$$\tilde{\mu}(B) := \mu(B \cap [a, b]) + \delta_b(B) \cdot \mu((b, \infty))$$

is still favourable. Since u is increasing, we have

$$\int u(w + x)\tilde{\mu}(dx) \leq \int u(w + x)\mu(dx) < u(w),$$

where $w \geq 0$ is the initial wealth. That is, $\tilde{\mu}$ is still rejected at any level of wealth. It follows that

$$\int_{[0,b]} [u(w + x) - u(w)]\tilde{\mu}(dx) \leq \int_{[a,0]} [u(w) - u(w + x)]\tilde{\mu}(dx).$$

Let us assume for simplicity that u is differentiable, the general case requires minor modifications. Then, by the previous inequality,

$$u'(w + b)m^{+}(\tilde{\mu}) < u'(w + a)m^{-}(\tilde{\mu}),$$

where

$$m^+(\tilde{\mu}) := \int_{[0,b]} x\tilde{\mu}(dx) > \int_{[a,0]} (-x)(\tilde{\mu}) =: m^-(\tilde{\mu}),$$

due to the fact that $\tilde{\mu}$ is favourable. Thus,

$$\frac{u'(w+b)}{u'(w-|a|)} < \frac{m^-(\tilde{\mu})}{m^+(\tilde{\mu})} =: \gamma < 1,$$

for any w, hence

$$u'(x + n(|a| + b)) < \gamma^n u'(x)$$

for any x in the interior of S. This exponential decay of the derivative implies

$$u(\infty) := \lim_{x\uparrow\infty} u(x) < \infty.$$

More precisely, if

$$A := n(|a| + b)$$

for some n, then

$$u(\infty) - u(x) = \sum_{k=0}^{\infty} \int_{x+kA}^{x+(k+1)A} u'(y)dy = \sum_{k=0}^{\infty} \int_{x-A}^{x} u'(z + (k+1)A)dz <$$

$$< \sum_{k=0}^{\infty} \gamma^{(k+1)n} \int_{x-A}^{x} u'(z)dz = \frac{\gamma^n}{1-\gamma^n}(u(x) - u(x-A)).$$

Take n such that $\gamma^n \leq \frac{1}{2}$. Then we obtain

$$u(\infty) - u(x) < u(x) - u(x-a),$$

therefore

$$\frac{1}{2}\left(u(\infty) + u(x-A)\right) < u(x)$$

for all x such that $x - A \in S$. $\square$

For an exponential utility $u(x) = 1 - e^{-\alpha x}$, the vet

$$v := \frac{1}{2}\left(\delta_{-A} + \delta_{\infty}\right)$$

is rejected at any wealth level as soon as

$$A > \frac{1}{\alpha}\log 2.$$

This theorem shows that trying to solve a paradox can produce, as a side effect, more paradoxes and put in serious troubles the descriptive aspects of the expected utility theory. In the next section, we revisit the relation between rejection of favourable lotteries and boundedness from more general point of view. To be precise, we prove that we can check if there exists Super-Petersburg paradoxes without knowing the exact utility function and relying only in the rejection of a set of test lotteries.

17 A calibration Theorem

In the previous section, we have announced that simple tail or head lottery is enough to check if the agent displays a bounded utility function. To give an example, if at any given wealth an agent refuses to play the following lottery,

$$\mu := \frac{1}{2}(\delta_{-100} + \delta_{110}),$$

then he will refuse any lottery of the form

$$\nu := \frac{1}{2}(\delta_{-1000} + \delta_{\infty}).$$

The big point of this result is that we do not need to know the utility function at all; the concavity, the risk aversion, is enough to prove the result. Also, it is worth to note that this is a big drawback of the model because it shows that the expected utility theory is over sensitive to small stakes. The difference between accepting or rejecting a small bet is that massive, that can be reasonable to do not apply the theory at all.

Now we announce and prove the theorem.

Theorem 17.1. *Suppose that for all w, $U(w)$ is strictly increasing and weakly concave. Suppose that there exist $\overline{w} > \underline{w}$, $g > l > 0$, such that for all $w \in [\underline{w}, \overline{w}]$,*

$$\frac{1}{2}\left(U(w-l) + U(w+g)\right) < U(w).$$

Then for all $w \in [\underline{w}, \overline{w}]$, for all $x > 0$,

(i) if $g \leq 2l$, then

$$U(w) - U(w-x) \geq 2 \sum_{i=1}^{k^*(x)} \left(\frac{g}{l}\right)^{i-1} r(w),$$

if $w - \underline{w} + 2l \geq x \geq 2l$, and

$$U(w) - U(w-x) \geq 2\left[\sum_{i=1}^{k^*(w-\underline{w}+2l)} \left(\frac{g}{l}\right)^{i-1} r(w)\right] + [x - (w - \underline{w} + l)] \left(\frac{g}{l}\right)^{k^*(w-\underline{w}+2l} r(w)$$

if $x \geq w - \underline{w} + 2l$.

(ii) Otherwise,

$$U(w) - U(w-x) \leq \sum_{i=1}^{k^{**}(x)} \left(\frac{l}{g}\right)^{i} r(w),$$

if $x \leq \overline{w} - w$, and

$$U(w) - U(w-x) \leq \sum_{i=1}^{k^{**}(\overline{w})} \left(\frac{l}{g}\right)^{i} r(w) + [x - \overline{w}] \left(\frac{l}{g}\right)^{k^{**}(\overline{w})} r(w),$$

if $x \geq \overline{w} - w$.

where int(y) denote the smallest integer less than or equal to y, $k^*(x) := int\left(\frac{x}{2l}\right)$, $k^{**} := int\left(\frac{x}{g} + 1\right)$, and $r(w) := U(w) - U(w-l)$.

Proof.

(i) For notational ease and without loss of generality, let $r(w) = U(w) - U(w-l) \equiv 1$. Then clearly $U(w-l) - U(w-2l) \geq 1$ by the concavity of U. Also, since $2l > g > l$, we know that $w - 2l \in (w-l, sw)$, and by the concavity of U,

$$U(w-2l+g) - U(w-l) \geq \frac{g-l}{l} = \frac{g}{l} - 1.$$

Hence,

$$U(w-2l+g) - U(w-2l) \geq \frac{g}{l} - 1 + 1 = \frac{g}{l}.$$

Therefore, if $w - 2l \geq \underline{w}$, we know that

$$U(w-2l) - U(w-3l) \geq \frac{g}{l}$$

since by assumption,

$$U(w-2l-l) + U(w-2l+g) \leq 2u(w-2l).$$

By concavity, we also know that

$$U(w-3l) - U(w-4l) \geq \frac{g}{l}.$$

More generally, if $w - 2kl \geq \underline{w}$, then

$$U(w-(2k-1)l) - U(w-2kl) \geq U(w-2(k-1)l) - U(w-(2k-1)l) \Rightarrow$$

$$\Rightarrow U(w-2kl+g) - U(w-2kl) \geq \frac{g}{l}\left[U(w-2(k-1)l) - U(w-(2k-1)l)\right] \Rightarrow$$

$$U(w-2kl) - U(w-(2k+1)l) \geq \frac{g}{l}\left[U(w-2(k-1)l) - U(w-(2k-1)l)\right].$$

These lower bounds on margin utilities yield the lower bound on total utilities $U(w) - U(w-x)$ in part (i) of the Theorem.

(ii) Again let $r(w) = U(w) - U(w-l) \equiv 1$. Then

$$U(w+g) - U(w) \leq 1.$$

By the concavity of U,

$$U(w+g) - U(w+g-l) \leq \frac{l}{g}.$$

But if $w + g \leq \overline{w}$, this implies by assumption that

$$U(w+2g) - U(w+g) \leq \frac{l}{g},$$

since

$$U(w+g-l) + U(w+2g) \leq 2U(w+g).$$

More generally, we know that if $w + mg \leq \underline{w}$, then

$$U(w+mg+g) - U(w+mg) \leq \frac{l}{g}[U(w+mg) - U(w+mg-g)].$$

These upper bounds on marginal utilities yield upper bounds on utilities $U(w+x) - U(w)$ in part (ii) of the Theorem.

□

Corollary 17.2. *Suppose that for all w, $U(w)$ is strictly increasing and weakly concave. Suppose there exists $g > l > 0$ such that for all w,*

$$\frac{1}{2}\left(U(w-l) + U(w+g)\right) < U(w).$$

Then for all positive integers k, for all $m < m(k)$,

$$\frac{1}{2}\left(U(w-2kl) + U(w+mg)\right) < U(w),$$

where

$$m(k) := \begin{cases} \frac{\log\left[1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i}\right]}{\log\frac{l}{g}} - 1 & \textit{if } 1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i} > 0 \\ \infty & \textit{if } 1-\left(1-\frac{l}{g}\right)2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i} \leq 0. \end{cases}$$

Proof. From the proof of Theorem 17.1, we know

$$U(w) - U(w-2kl) \geq 2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i-1} r(w)$$

and

$$U(w+mg) - U(w) \leq \sum^{m+1} i = 0\left(\frac{l}{g}\right)^{i} r(w).$$

Therefore, if $U(w) - U(w-2kl) < U(w+mg) - U(w)$, then

$$2\sum_{i=1}^{k}\left(\frac{g}{l}\right)^{i-1} \leq \sum^{m+1} i = 0\left(\frac{l}{g}\right)^{i}.$$

Solving for m yields the formula. Note that if $g > 2l$, we only need

$$U(w) - U(w-2kl) \geq 2k(U(w) - U(w-1))$$

to get the result. □

Theorem 17.1, and its corollary, gives upper and lower bounds on the utility of increases of wealth. The source paper [13] comes with tables that help us to understand the exact meaning of the results and how it poses in trouble the descriptive aspect of the theory when dealing with small and big bets at the same time.

Table 1: If averse to 50-50 lose 100/Gain G bets for all wealth levels,
will turn down 50-50 lose L/gain G bets; G's entered in table

L/G	101	105	110	125
400	400	420	550	1250
600	600	730	990	∞
800	800	1050	2090	∞
1000	1010	1570	∞	∞
2000	2320	∞	∞	∞
4000	5750	∞	∞	∞
6000	11810	∞	∞	∞
8000	34940	∞	∞	∞
10000	∞	∞	∞	∞
20000	∞	∞	∞	∞

Part VIII. Final words

In this work we saw the mainstream expected utility theory and discussed what, probably, are the major applications and drawbacks of the theory. One notable aspect of the theory is that it originated as a side product of other problems, like solving the St. Petersburg Paradox, on Bernoulli and Cramer's works, or defining the fundamentals of the Theory of Games on John von Neumann and Oskar Morgenstern *Theory of Games and Economic Behavior*.

Being a side product, it is surprising to see how the theory uses a wide set of mathematical fields and its big impact in social sciences. But even more surprising is that, being a reasonable and normative appealing idea, when applied results in even bigger paradoxes that the ones that he was supposed to solve. This implies the necessity of further refinements or a whole new theory of what a rational behaviour is. This is a fruitful tension and, probably, is one of the reasons that still, nowadays, the expected utility theory is studied and well alive.

The lecture of this undergraduate thesis will give a nice scope of the theory. However, time, knowledge and space constrains kept out of the work some interesting topics like,

- The role of the stochastic dominance and the convex ordering.
- Further explore the axioms of the subjective probability and their impact in the theory.
- The bijection between risk measures and the utility of acceptance sets.

- The Yaari's Dual Theory of Choice under Risk, which reverses the roles of the payment and the probabilities.

- The Generalised Expected Utility as a response to the Allais and Ellberg paradoxes.

- Economics and Behavioural critiques to the idea that rational behaviour equals to maximize utility functions.

- Finite St. Petersburg Lotteries and under what circumstances two risk averse player will buy and sell the gamble.

To conclude, I want to express my gratitude to my Director and the regular tutorials that he offered during the elaboration of this work.

Part IX. Bibliography

References

[1] Allais, Maurice, *Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine*. Econometrica, Vol. 21, No. 4 (Oct., 1953), pp. 503-546

[2] Barucci, Emilio, *Financial Markets Theory. Equilibrium, Efficiency and Information*. Springer-Verlag, 2003.

[3] Bernoulli, Daniel, *Exposition of a New Theory on the measurement of Risk*. Econometrica, Vol. 22, No. 1 (Sep., 1954), pp. 23-36.

[4] Billingsley, Patrick, *Probability and Measure. Third Edition*. Wiley-Interscience, 1995.

[5] Corcuera, José Manuel, *Lecture Notes on Introduction to Quantitative Finance*. Personal Webpage, http://www.mat.ub.edu/ corcuera/

[6] Ellsberg, Daniel, *Risk, Ambiguity, and the Savage Axioms*. Quarterly Journal of Economics, Vol. 75, No. 4 (Nov. 1961), 643–669.

[7] Föllmer, Hans, and Schied, Alexander, *Stochastic Finance. An introduction in discrete time*. Walter de Gruyter, 2002.

[8] Henderson, Vicky, and Hobson, David *Utility Indifference Pricing - An Overview*. to appear in Volume on Indifference Pricing, (ed. R. Carmona), Princeton University Press, 2004

[9] Haugh, Martin, and Iyengar, Garud, *Lecture notes and slides on Financial Engineering and Risk Management*. Coursera, 2013.

[10] Huang, Chi-fu, and Litzenberger, Robert H., *Fundations for Financial Economics*. Prantice Hall, 1988.

[11] Machina, Mark J., *Expected Utility" Analysis without the Independence Axiom*. Econometrica, Vol. 50, No. 2. (Mar., 1982), pp. 277-323.

[12] Maj, Mateus, and Schoutens, Wim, *A note on the suboptimality of path-dependent pay-offs in Lévy Markets*. Nov., 2008.

[13] Rabin, Matthew, *Risk Aversion and Expected-Utility Theory: A Calibration Theorem.* Econometrica, Vol. 68, No. 5 (Sep., 2000), pp. 1281-1292.

[14] Samuelson, Paul A., *St. Petersburg Paradoxes: Defanged, Dissected, and Historically Described.* Journal of Economic Literature, Vol. 15, No. 1, Mar., 1977

[15] Sigman, Karl, *Lecture Notes on Financial Engineering.* Personal Webpage, http://www.columbia.edu/ ks20/FE-Notes/FE-Notes-Sigman.html

[16] Sigman, Karl, *Lecture Notes on Financial Engineering.* Personal Webpage, http://www.columbia.edu/ ks20/FE-Notes/FE-Notes-Sigman.html

[17] Tsanakas, A., and Desli, E., *Risk Measures and Theories of Choice.* British Actuarial Journal, Vol. 9, Issue 04, Oct., 2003, pp 959-991.