THE VON NEUMANN MORGENSTERN THEOREM

GORDON RITTER

Abstract. We review the celebrated von Neumann-Morgenstern theorem, which shows that (in a sense to be made precise), any rational investor behaves as if they have a utility function, and they are making decisions that maximize expected utility.

1. Convex sets

We begin with a little math review that we are going to need for what follows. The mathematics of convex sets and convex functions is extremely important in finance. Most of the optimization problems we'll encounter are convex, and convex optimization problems enjoy many favorable properties that are not shared by their non-convex counterparts. Each non-convex optimization problem has a dual that is convex. Moreover, model-free no-arbitrage relationships guarantee that many derivative pricing functions are convex. We do not assume previous knowledge of convex spaces or functions, so start with the basic definitions.

The line segment between x_1 and x_2 in a vector space V is all points $x = \theta x_1 + (1-\theta)x_2$ with $0 \le \theta \le 1$. A set C is *convex* if it contains the line segment between any two points in the set, i.e.

$$x_1, x_2 \in C \;\Rightarrow\; \theta x_1 + (1-\theta)x_2 \in C \;\forall\, \theta \in [0,1]$$

A *convex combination* of $x_1, \dots, x_k$ is any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \cdots + \theta_k = 1, \theta_i \ge 0$.

The *convex hull* of a set S, denoted $\mathrm{conv}(S)$, is the set of all convex combinations of points in S. The convex hull is also the smallest convex set that contains S, or the intersection of all convex sets containing S.

Date: February 5, 2019.

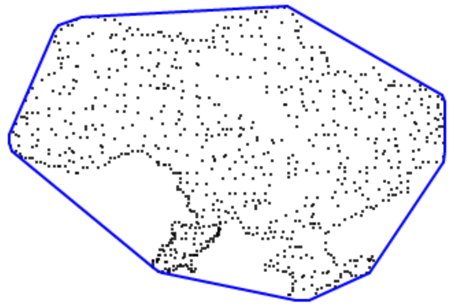

FIGURE 1.1. When S is a bounded subset of the plane, the convex hull may be visualized as the shape formed by a rubber band stretched around S.

2. Preference relations

Choosing between lotteries *can* be done by optimizing expected utility, but that isn't necessarily the only reasonable way of choosing between lotteries. In other words, one might wonder whether there could exist a "rational" way of choosing between lotteries that does not go by way of the expectation of some utility function. To determine whether this is possible, first we have to define what we mean by *rational*.

Suppose the reward space is finite; $\mathcal{X} = (x_1, \ldots, x_n)$. Recall that in this case, a lottery is simply a set of probabilities $\boldsymbol{p} = (p_1, \ldots, p_n)$, where we receive reward x_i with probability p_i. Let $\mathcal{P} = \Delta(\mathcal{X}) \subset \mathbb{R}^n$ denote the space of lotteries, ie. discrete probability distributions.

If lottery p' is preferred over lottery p, we write $p \prec p'$. If p' is either preferred over or viewed with indifference relative to p, we write $p \preceq p'$. If the agent is indifferent between p and p', we write $p \sim p'$. A collection of preferences specified by $\sim, \prec, \preceq$ is called a *preference relation*. This is obviously subjective: each of us has his or her own preference relation over lotteries.

Observe that given two lotteries p and p', any convex combination of them: $\alpha p + (1-\alpha)p'$ with $\alpha \in [0, 1]$ is also a lottery. This can be viewed simply as stating the mathematical fact that $\mathcal{P} = \Delta(\mathcal{X})$ is a convex set. Moreover, $\mathcal{P}$ is the convex hull of its extreme points – points of the form $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, etc which correspond to degenerate lotteries where one of the rewards is certain.

We can also view $\alpha p + (1-\alpha)p'$ more explicitly as a compound lottery, summarizing the overall probabilities from two successive events: first, a flip of a weighted (ie. unfair) coin with weight $\alpha, 1-\alpha$ that determines whether the lottery p or p' should be used to determine the ultimate consequences; second, either the lottery p or p'.

Definition 2.1. A preference relation is said to be *complete* if

- For any lotteries p, p', exactly one of the following holds:

$$p \prec p', \;\; p' \prec p, \quad \text{or } p \sim p'$$

A preference relation is said to be *transitive* if

- If $p \preceq p'$ and $p' \preceq p''$, then $p \preceq p''$.

Definition 2.2. A preference relation $\succeq$ is said to be *continuous* if

- If $p \succeq p' \succeq p''$, then there exists $\alpha \in [0, 1]$ such that

$$\alpha p + (1 - \alpha)p'' \;\sim\; p'$$

The intuition behind the continuity property is that if p is preferred to p', then a lottery "close" to p (a short distance away in the direction of p'' for instance) will still be preferred to p'. This sounds reasonable, though there are situations where it might be doubtful.

Consider the following example. Suppose p is a gamble where you get \$100 for sure, p' is a gamble where you nothing for sure, and p'' is a gamble where you get killed for sure. Naturally $p \succ p' \succ p''$. But this means that there is some α, with $0 < \alpha < 1$, such that you would be indifferent between (a) nothing happens, and (b) playing a dangerous game where you either make \$100 with probability α or die with probability $1 - \alpha$. Technically, I should have written $0 \leq \alpha \leq 1$, but neither $\alpha = 0$ nor $\alpha = 1$ satisfy the required indifference. If $\alpha = 5/6$, this amounts to charging an opponent \$100 to watch you play Russian roulette!

Now, if $1 - \alpha$ is sufficiently small, most of us would probably play the game. I think that, for me, $1 - \alpha = 10^{-8}$ would already make me want to play the game (ie. prefer the game to nothing). after all, we take those kind of chances (or worse) every time we fly. Presumably, then, there is some slightly higher value for $1 - \alpha$ where I'm indifferent between playing or not.

The main point here is, that the continuity axiom isn't trivial. If death is an element of the reward space, then an ultraconservative agent who will never accept any probability of death, no matter how small, would violate the axiom (and hence might not have a utility function). In finance, the reward space $\mathcal{X}$ will exclusively consist of various amounts of wealth gained or lost, which makes it impossible to construct wild Russian-roulette-type counterexamples, and which makes the continuity axiom easier to justify.

Given that preferences are complete, transitive and continuous, they can be represented by a *preference function* $U : \mathcal{P} \to \mathbb{R}$, where $p \succeq p'$ if and only if $U(p) \geq U(p')$; we shall prove this later in the course of proving Theorem 3.1.

Definition 2.3. A preference function $U : \mathcal{P} \to \mathbb{R}$ has an *expected utility form* (or is a *von Neumann–Morgenstern preference function*) if there is some function $u : \mathcal{X} \to \mathbb{R}$ such that

$$U(\boldsymbol{p}) = \mathbb{E}_{\boldsymbol{p}}[u(x)] = \sum_{i=1}^{n} p_i u(x_i) \quad \text{for all} \quad \boldsymbol{p} \in \mathcal{P} \tag{2.1}$$

A preference relation $\preceq$ is said to have an *expected utility form* if its comparisons among lotteries are described by a preference function U, and furthermore, that preference function has an expected utility form (2.1).

It is clear from (2.1) that if $U : \mathcal{P} \to \mathbb{R}$ has an expected utility form, then U is linear as a function of $p \in \mathcal{P}$, meaning that:

$$U(\alpha p + (1-\alpha)p') = \alpha U(p) + (1-\alpha)U(p') \quad \text{for} \quad \alpha \in [0,1]. \tag{2.2}$$

The converse also holds: if (2.2) holds for every p, p' and α, then U must have an expected utility form (exercise).

The next axiom, which is more controversial, will allow us to say a great deal about the structure of U. Note that a preference function is a different kind of function than the utility of wealth we discuss elsewhere in the course: its argument is a lottery, not a wealth level.

Definition 2.4. A preference relation satisfies the axiom of "independence of irrelevant alternatives" if for all $p, p', p'' \in \mathcal{P}$ and for all $\alpha \in [0,1]$, we have

$$p \succeq p' \quad \Leftrightarrow \quad \alpha p + (1-\alpha)p'' \succeq \alpha p' + (1-\alpha)p''.$$

For brevity we'll call this the "independence axiom" – not to be confused with the statistical notion of independence of random variables.

The independence axiom says that if I prefer p to p', I'll also prefer the possibility of p to the possibility of p', given that the alternative in both cases is the same. In particular, the axiom says that if I'm comparing

$$\alpha p + (1-\alpha)p'' \ \text{ to } \ \alpha p' + (1-\alpha)p''\,,$$

I should focus on the distinction between p and p' and hold the same preference independently of both α and p''. This axiom is sometimes also called the substitution axiom: the idea being that if p'' is substituted for part of p and part of p', this shouldn't change my ranking.

Definition 2.5. An agent whose preferences among lotteries are described by a relation $\succeq$ on $\mathcal{P}$ which is complete and transitive, and satisfies continuity and independence is said to be *von Neumann Morgenstern rational* or *vNM-rational*. The four properties of completeness, transitivity, continuity and independence are called the *von Neumann Morgenstern axioms* (see Von Neumann and Morgenstern (1945)).

Lemma 2.1. Let $\preceq$ be the preference relation of a vNM-rational agent. there exists at least one lottery $\overline{p} \in \mathcal{P}$ (called a "most preferred" lottery) such that

$$p \preceq \overline{p} \ \text{ for all } \ p \in \mathcal{P}.$$

There also exists at least one "least preferred" lottery, defined by reversing the inequality.

Proof. Consider the finite set of lotteries which correspond to deterministic outcomes. Specifically, to each outcome x_i we associate a lottery L_i in which outcome x_i is attained with certainty, or in other words the probability vector is $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ where the 1 appears in the i-th position.

The set $\{L_1, \ldots, L_n\}$ may be sorted according to the relation $\preceq$, where equivalent elements are placed in random order. Let L_j be the element that ends up as maximal with respect to this sort. Any arbitrary lottery p can be expressed as a linear combination of the extremal lotteries:

$$p = \sum_i p_i L_i = p_j L_j + \sum_{i \neq j} p_i L_i$$

Note that $\sum_{i \neq j} p_i = 1 - p_j$ so we can re-normalize the "not equal to j" probabilities so that they sum to 1. Recall that j is the index such that L_j is maximal among deterministic lotteries. Define

$$q_i := \frac{p_i}{\sum_{i \neq j} p_i} \ \text{ for } i \neq j$$

Rhen re-write the above as

$$p = \sum_i p_i L_i = p_j L_j + (1 - p_j) L_q, \ \text{ where } \ L_q := \sum_{i \neq j} q_i L_i$$

Note that $L_j \geq L_q$ since L_j is preferred-or-indifferent to every one of the component lotteries of which L_q is a mixture. Set $\alpha = p_j$; we then have

$$p = \alpha L_j + (1-\alpha)L_q \preceq \alpha L_j + (1-\alpha)L_j = L_j$$

where we used the independence axiom. Note that we can then take $\overline{p} = L_j$ and this will be preferred-or-equivalent to any other lottery. Similarly, one can show there also exists a lottery $\underline{p}$ such that $\underline{p} \leq p$ for all $p \in \mathcal{P}$ by the same argument, just reversing the directions of various preference relations. This completes the proof of Lemma 2.1.

3. The von Neumann-Morgenstern Theorem

Henceforth we shall always assume that $\overline{p} \succ \underline{p}$, because if $\overline{p} \sim \underline{p}$ then nothing is preferred to anything else, any constant function is an expected-utility representation, and all decisions are trivial. There may be multiple most-preferred lotteries, and/or multiple least-preferred lotteries; for everything that follows, we will choose $\overline{p}$ to be any one of the most-preferred, and none of our results depend on this choice.

Lemma 3.1. Assume a preference relation satisfying the von Neumann Morgenstern axioms. Let $\overline{p}, \underline{p} \in \mathcal{P}$ be most and least preferred lotteries in $\mathcal{P}$. If $1 > \beta > \alpha > 0$, then

$$\overline{p} \succ \beta\overline{p} + (1-\beta)\underline{p} \succ \alpha\overline{p} + (1-\alpha)\underline{p} \succ \underline{p}.$$

Proof. For the first inequality, write the left hand side as $\beta\overline{p} + (1-\beta)\overline{p}$. Because $\overline{p} \succ \underline{p}$, the independence axiom immediately generates the inequality. For the second inequality, write the left hand side as $(\beta-\alpha)\overline{p} + \alpha\overline{p} + (1-\beta)\underline{p}$ and the right hand side as $(\beta-\alpha)\underline{p} + \alpha\overline{p} + (1-\beta)\underline{p}$, and again invoke independence. A similar argument works for the third inequality. $\square$

We now come to the main theorem of this section.

Theorem 3.1 (Von Neumann and Morgenstern (1945)). Suppose given a complete and transitive preference relation on $\mathcal{P}$. Such a relation satisfies continuity and independence if and only if it admits an expected utility representation (in the sense of Definition 2.3).

Proof. Let $\succeq$ be given and assume it satisfies continuity and independence. We will construct a preference function U that represents $\succeq$ and show that it has an expected utility representation. Let $\overline{p}, \underline{p} \in \mathcal{P}$ be most and least preferred lotteries in $\mathcal{P}$ as in Lemma 2.1 and without loss of generality assume $\overline{p} \succ \underline{p}$.

Now, I claim that for any $p \in \mathcal{P}$, there exists a unique λ_p such that:

$$\lambda_p \overline{p} + (1 - \lambda_p)\underline{p} \sim p. \tag{3.1}$$

Because $\overline{p} \succeq p \succeq \underline{p}$, some such λ_p exists as a consequence of continuity. Lemma 3.1 implies that this λ_p which exists is also unique (exercise: check this).

Finally, I claim that the preference function

$$U(p) := \lambda_p$$

is an expected utility representation of $\succeq$. To see that U represents $\succeq$, observe that:

$$p \succeq q \quad \Leftrightarrow \quad \lambda_p \overline{p} + (1 - \lambda_p)\underline{p} \succeq \lambda_q \overline{p} + (1 - \lambda_q)\underline{p}$$

because we can replace p by the equivalent lottery $\lambda_p \overline{p} + (1 - \lambda_p)\underline{p}$ anywhere we want, and the same applies to q: we can replace q by the equivalent lottery $\lambda_q \overline{p} + (1 - \lambda_q)\underline{p}$. Furthermore,

$$\lambda_p \overline{p} + (1 - \lambda_p)\underline{p} \succeq \lambda_q \overline{p} + (1 - \lambda_q)\underline{p} \quad \Leftrightarrow \quad \lambda_p \geq \lambda_q$$

To see that U has an expected utility form, as noted above, it suffices to prove that for any $\alpha \in [0, 1]$ and $p, p' \in \mathcal{P}$:

$$U(\alpha p + (1 - \alpha)p') = \alpha U(p) + (1 - \alpha)U(p'). \tag{3.2}$$

This is just (2.2) again, and one of the accompanying exercises is to prove that if (2.2) holds for every p, p' and α, then U must have an expected utility form.

By definition of λ_p and because $U(p) = \lambda_p$, we know that:

$$p \sim U(p)\overline{p} + (1 - U(p))\underline{p}$$
$$p' \sim U(p')\overline{p} + (1 - U(p'))\underline{p}$$

Therefore, by the independence axiom

$$\begin{aligned} \alpha p + (1 - \alpha)p' &\sim \alpha \left[U(p)\overline{p} + (1 - U(p))\underline{p}\right] \qquad (3.3)\\ &\qquad + (1 - \alpha)\left[U(p')\overline{p} + (1 - U(p'))\underline{p}\right] \\ &\sim \left[\alpha U(p) + (1 - \alpha)U(p')\right]\overline{p} \\ &\qquad + \left[1 - \alpha U(p) - (1 - \alpha)U(p')\right]\underline{p}. \qquad (3.4) \end{aligned}$$

Our strategy for proving (3.2) is to calculate $U(\alpha p + (1 - \alpha)p')$, which requires us to find the λ corresponding to the lottery $\alpha p + (1 - \alpha)p'$. In other words, we have to represent $\alpha p + (1 - \alpha)p'$ as equivalent to some linear combination of $\overline{p}$ and $\underline{p}$, and the associated λ will be the coefficient of $\overline{p}$. But (3.4) has already done that. In fact (3.4) shows that the associated λ is just $\alpha U(p) + (1 - \alpha)U(p')$. In other words,

$$U(\alpha p + (1-\alpha)p') = \alpha U(p) + (1-\alpha)U(p').$$

This completes the proof in one direction. We leave as an exercise to show the reverse direction: if $\succeq$ admits an expected utility representation U, then $\succeq$ must satisfy continuity and independence. $\square$

4. Dutch books

Suppose we are interested in predicting the result of an upcoming tennis match, say between Fisher and Neyman. One might naturally wonder whether the bets offered by bookmakers contain any information, as a possible way of quantifying uncertainty about who will win. If the posted odds in favor of Fisher are, say, 1-2, one can bet one dollar, and win two dollars (plus the initial stake) if Fisher wins, and if Fisher does not win, the initial stake of \$1 is not returned. Many people also talk about the reciprocal of the odds, also called "odds against".

Definition 4.1 (Baxter and Rennie (1996)). When a price (of a bet) is quoted in the form $n : m$ against, such as $2 : 1$ against, it means that a successful bet of $\$m$ will be rewarded with $\$n$ plus stake returned. The implied probability of victory (were the price fair) is $m/(m+n)$.

To make this more formal, let θ be the indicator of the event "Fisher wins." We say, equivalently, that θ occurred or θ is true or that the truth value of θ is 1. A *bet* is defined to be a ticket that will be worth a stake S if θ occurs and nothing if θ does not occur. A bookmaker generally sells bets at a price $\pi_\theta S$. The price is expressed in units of the stake; when there is no ambiguity we will simply use π.

Note that $\pi > 1$ is nonsensical, because you can never win more than the stake, and $\pi < 0$ is similarly nonsensical because you would be getting paid to take the bet which, at worst, returns nothing. The ratio $\pi : (1-\pi)$ is the betting odds in favor of the event θ. In the Fisher-Neyman tennis example, where odds in favor are $1 : 2$, the stake S is three dollars, the price πS is one dollar, and π is $1/3$. Rather than writing the odds as $1/3 : 2/3$ you would multiply both ratios by 3 and express the odds as integers, hence $1 : 2$.

The action of betting on θ, or buying the ticket described above, will be denoted by $a_{S,\theta}$. What are the consequences of this action? If $\theta = 1$, the buyer will have a net gain of $(1-\pi)S$, that is, the stake S less the price πS. If $\theta = 0$, the buyer will net $-\pi S$, in other words he will lose the "initial stake" or the price paid for the ticket.

Now suppose the prices posted by the bookmaker are $1:4$ *in favor* for the event

$$\theta = \text{Fisher wins}$$

and $7:3$ in favor for bets on the event "Neyman wins". Is this good for the bookmaker?
Let's analyze all possible cases in turn.

(1) You buy a ticket for the event "Fisher wins"
    (a) Fisher wins: bookmaker nets $-0.8S$
    (b) Neyman wins: bookmaker nets $0.2S$
(2) You buy a ticket for the event "Neyman wins"
    (a) Fisher wins: bookmaker nets $0.7S$
    (b) Neyman wins: bookmaker nets $-0.3S$

If you buy both tickets (one for Fisher, plus one for Neyman), you will make $0.1S$ (and the bookmaker will lose $0.1S$) in either possible scenario.

This "arbitrage" can exist because there is an internal inconsistency in the prices posted by the bookmaker. In some bygone era, this used to be called "making Dutch Book against the bookmaker". We mean no offense to the Dutch, who are of course wonderful people.

Definition 4.2. A bookmaker's betting odds are said to be *coherent* if a client cannot place a bet or a combination of bets such that no matter what outcome occurs, the bookmaker will lose money.

A Dutch book is thus associated with probabilities implied by the odds not being coherent.

We have seen that, with badly-constructed bet prices, the bookmaker can make it possible for a clever adversary to make a Dutch book, also known as an arbitrage. This still leaves the question: can the bookmaker arrange things so that he/she will earn a profit irrespective of the outcome?

In another example, consider a horse race and suppose a bookmaker's odds on each horse are given by the second column in the table below. Suppose further that there is plenty of demand to stake every single horse at these odds, but the bookmaker limits the amount of betting in each horse differently, and, more specifically, limits them in such a way that the bookmaker always pays \$200 to whoever wins. So for example, for the horse that's 4-1 against, the bookmaker limits the total bet size to \$200 / (4 + 1) = \$40.

Horse	Offered odds	Implied probability	Stake	Bookie Pays
1	Even	$\frac{1}{1+1} = 0.5$	\$100	\$100 stake + \$100
2	3 to 1 against	$\frac{1}{3+1} = 0.25$	\$50	\$50 stake + \$150
3	4 to 1 against	$\frac{1}{4+1} = 0.2$	\$40	\$40 stake + \$160
4	9 to 1 against	$\frac{1}{9+1} = 0.1$	\$20	\$20 stake + \$180
		Total: 1.05	Total: \$210	Always: \$200

In this situation, the bookmaker takes in the sum of all the stakes, for a total of \$210, waits for the horses to run, and then pays \$200 to the winner, regardless of the winning horse. This is related to the fact that the implied probabilities add up to a number greater than 1. Crucially, we assumed that there is plenty of demand to stake every single horse at these odds, and so the bookmaker has the luxury of attracting the "right" amount of stake for each bet. In reality, if the odds are too far away from people's expectations, there might be no bets at all. The bookmaker must adjust the odds ratios so that there is enough demand to attract the necessary stake sizes.

There is a famous theorem regarding Dutch books, which states roughly that for any decision-maker, if the independence axiom is violated by the decision-maker's preferences, then there is a dutch book the decision-maker would agree to (ie. an arbitrage they would be the source of), *even if* the other three axioms are satisfied by the decision-maker's preferences. Combining this with the vN-M theorem, we infer that for any decision-maker, either: (a) their decision-making process has an expected-utility form (whether they know it or not), or (b) they would agree to a sequence of trades that forms an arbitrage opportunity for the counterparty.

References

Baxter, Martin and Andrew Rennie (1996). *Financial calculus: an introduction to derivative pricing*. Cambridge university press.

Von Neumann, John and Oskar Morgenstern (1945). *Theory of games and economic behavior*. Princeton University Press Princeton, NJ.

Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012

Department of Mathematics, MFE program, Baruch College, CUNY, New York, NY 10012