Northwestern University Marciano Siniscalchi
Fall 2005 Econ 410-1

RISK AND UNCERTAINTY

1. Intro

We now return to choice theory to analyze decisions under *risk* and *uncertainty.* In case you were wondering, the basic distinction between these two concepts was proposed by Frank Knight in 1921. *Decisions under risk* correspond to choice situations in which "probabilities are given"; for instance, games of chance. *Decisions under unertainty* are instead situations in which "the decision-maker must figure out what the probabilities are" [the quotation marks do not indicate actual quotes, just informal expressions!] Now, this is not very satisfactory, for two reasons. First, despite this distinction, 99.9% of "information economics" is carried out pretty much as if probabilities were always an objective element of the model, just like technology or commodities. This is often harmless, but sometimes (especially in multi-person situations, e.g. in game theory) it implies unwarranted (and perhaps undesired) restrictions.

Second, do we really know what probabilities are? You should at least be aware of the existence of three different interpretations. The first really has to do with games of chance, it is historically important, but it is basically self-referential and therefore not very useful: it maintains that "the probability of an event equals the number of realizations (of some experiment) consistent with it, divided by the total number of possible realizations". This implies that each realization is "equally likely"—a concept that itself refers to a notion of probability.

The second interpretation is called **frequentist**. It maintains that the probability of an event is the limit of its relative frequency in a sequence of repetitions of a suitable experiment (e.g. tossing a coin, measuring some physical quantity, etc.) This is much better, because it is not self-referential. However, it is basically useless for the purposes of economic theory; we can (almost) never *exactly* replicate a choice experiment (such as purchasing shares of a given asset at a certain date).

This leaves us with the third interpretation, called **Bayesian** or **subjectivist**. The basic idea is that *probability is a mathematical representation of an individual's subjective beliefs, as they manifest themselves in actual choice behavior.* More precisely: just like utility is not meaningful per se, but merely represents preferences, it is possible to consider a carefully constructed *choice situation*—typically, bets—that can reasonably be interpreted as "eliciting" the decision-maker's beliefs; it can then be shown that, under suitable assumptions, this choice behavior can be represented (in an appropriate sense) by a probability measure.

We won't spend much time on these issues (but we will point out how a relatively satisfactory theory of subjective probability can be constructed on the basis of our results). Rather, we shall focus on more "practical" matters:

Acknowlegement. My lecture notes for this course draw from a variety of sources, including MWG, David Kreps's "A Course in Microeconomic Theory", Eddie Dekel's own lecture notes for the Fall 2002 edition of 410-1, and, where appropriate, journal articles and papers. The latter are explicitly referenced in the text. However, I claim full ownership of any errors you may find in these notes... if you find one, please let me know!

- How do we represent choice under risk and uncertainty?
  Keywords: **Lotteries**, **State space**, **acts**.
- How *should* people behave in such situations?
  Keywords: **von Neumann-Morgenstern Theorem**, **Expected Utility**.
- What kind of interesting economic phenomena can we describe?
  Keywords: **Risk Aversion**, **Portfolio Choice**, **Insurance**...
- Do people *really* behave as we hope they do?
  Keywords: **Allais Paradox**, **Ellsberg Paradox**.

2. Frameworks for Risk and Uncertainty

There are economies of scale in describing the two frameworks we are interested in before actually analyzing choice. This allows us to highlight a common component: both feature preferences over a convex subset of a vector space. In the next section, we will use essentially the same result, the von Neumann-Morgenstern theorem, to provide a characterization of such preferences.

2.1. Choice under risk: Lotteries.

Let us go back to the Knightian distinction, and consider choice under risk first.

We consider the following setup. There is a set X of **prizes**, which could be anything: money, goods, commodity bundles, intertemporal consumption plans... In particular, X could be finite or infinite. The objects of choice are **lotteries** over the set X; formally, a lottery is just an X-valued random variable. [In your basic statistics class, you encountered real-valued random variables; but, there is no reason not to contemplate r.v.'s that, for example, yield different consumption bundles with different probabilities.] Actually, in applications, it is often convenient to consider continuous lotteries—er, random variables (think of the value of a portfolio, under the assumption that returns are normally distributed). However, for the purposes of axiomatic treatment, it is sufficient to focus on finite-valued lotteries. Hence, we can equivalently say that *a (finite) lottery is a probability distribution over X, with finite support.* We write $p = (x_1, p_1; \ldots; x_n, p_n)$ to denote the generic lottery that yields prize x_i with probability p_i, and $\mathcal{L}(X)$ to denote the set of (finite) lotteries over X. We are interested in studying a preference relation $\succcurlyeq$ over $\mathcal{L}(X)$.

An important feature of the latter set is that it *convex*. More precisely, if X is a finite set of cardinality, say, n, we can simply identify $\mathcal{L}(X)$ with the set

$$\Sigma = \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n_+ : \sum_i p_i = 1 \right\},$$

known as the unit simplex in $\mathbb{R}^n$. This set is clearly convex with respect to the usual vector space operations.

If X is not finite, we can proceed as follows. First, let $\mathcal{M}(X) = \mathbb{R}^X$, i.e. the set of all functions from X to $\mathbb{R}$. Notice that this is a vector space under the usual pointwise sum and scalar multiplication operations: that is, if $f, g \in M(X)$, then the function $f+g$ defined by $(f+g)(x) = f(x) + g(x)$ is also a function from X to $\mathbb{R}$, and so is the function αf given by $(\alpha f)(x) = \alpha f(x)$, for all $\alpha \in \mathbb{R}$.

Second, note that any $p \in \mathcal{L}(X)$ can be viewed as a point in $\mathcal{M}(X)$. Write $p = (x_1, p_1; \ldots; x_n, p_n)$; in particular, we can associate p with the function $f_p : X \to \mathbb{R}$ such that

$$\forall x \in X, \quad f_p(x) = \sum_{i: x_i = x} p_i.$$

Notice that this allows for the case of n distinct prizes $x_1, \ldots, x_n$, as well as for the possibility that some prizes may be repeated more than once; furthermore, note that if some $x \in X$ is *not* a prize delivered by p (or if the corresponding p_i is zero), then $f_p(x) = 0$.

With this stipulation, every $p \in \mathcal{L}(X)$ maps to some $f_p \in \mathcal{M}(X)$ that (i) takes non-zero value only for finitely many $x's$, (ii) is non-negative, and (iii) satisfies $\sum_x f_p(x) = 1$; conversely, it is clear that every $f \in \mathcal{M}(X)$ that satisfies properties (i), (ii) and (iii) uniquely defines a lottery $p_f \in \mathcal{L}(X)$. Furthermore, you should convince yourselves that the subset of $\mathcal{M}(X)$ of functions that satisfy (i), (ii), (iii) is convex. We have achieved the desired identification: $\mathcal{L}(X)$ can be viewed as a convex subset of a vector space.

[In the literature, you will encounter the expression *mixture space* to indicate a set with an abstract "convex combination operation" that has all the algebraic properties of "regular" convex combinations. For instance, you will find statements such as "the set of lotteries is a mixture space". Thus, a convex subset of a vector space is automatically a mixture space, but, in principle, the converse may not be true. However, in practice, the objects we are interested in are really convex subsets of a vector space, so there will always be a "honest" convex combination operation to consider. And furthermore,, it turns out that even the alleged generality of "abstract" mixture spaces is only fictitious: every mixture space can be viewed as a convex subset of a vector space, provided suitable vector-space operations are defined.]

2.2. **Choice under Uncertainty: Acts.** Turn now to choice under uncertainty. Here, the basic idea is that we do *not* "know what the probabilities are". However, we are able to identify the following elements of the choice problem at hand:

- A set of **states**, i.e. elementary events or basic ("atomic") realizations of the underlying uncertainty; denote the set of states by S.
- A set of **prizes**, as above, denoted X;
- A collection of **acts**, i.e. functions from S to X.

You should think of states exactly as elementary events in probability theory, if you are familiar with that. If we are interested in rolling a dice, the possible states are $S = \{1, \ldots, 6\}$, corresponding to the outcome of the dice roll. But states need not be numerical quantities; an example will be provided momentarily.

Acts are the basic object of choice. The idea is that, in a concrete choice problem, the decision-maker (DM) will be able to in principle take a variety of "actions"; however, we are not interested in the actual, physical description of such actions; rather, we content ourselves with a description of how actions associate prizes (i.e. ultimate outcomes) to the possible realizations of the underlying uncertainty. Hence, we identify physical "actions"

This conceptualization is largely due to Leonard J. Savage, who is considered one of the founding fathers of decision theory, as well as of the subjectivist approach to probability and statistics. He also provided a really cute example of his framework. Suppose you are preparing dinner; you are making an omelette, and you have already used one egg. There is a second egg in you refrigerator, and you are considering whether or not to add it to the omelette. The problem is that you don't rememember how long the egg has been sitting in your fridge. Now, if the egg is still fresh and you use it, you get a large omelette (which is supposedly good for you, disregarding other health considerations...); if it is rotten and you use it, you spoil the omelette and have nothing to eat for dinner; and if you choose not to use it, you have to settle for a smaller omelette.

In the Savage framework, this situation can be described as follows. First, let $S = \{s_r, s_f\}$, where s_r stands for rotten and s_f for fresh; next, let $X = \{0, 1, 2\}$, corresponding to no omelette, a 1-egg omelette, or a 2-egg omelette; finally, the acts we are considering are f and g, where f corresponds to adding the extra egg and is defined by

$$f(s_r) = 0, \quad f(s_f) = 2$$

and g corresponds to not adding the extra egg, so

$$g(s_r) = g(s_f) = 1.$$

Preferences $\succcurlyeq$ are defined over the set of relevant acts. Now, in the framework as described so far, Savage was able to provide axioms that lead to the intended characterization of subjective expected utility [see below for details]. However, this result requires a lengthy and tedious proof; furthermore, there are certain unpleasant elements about the setting he considers, such as the absolute necessity of assumptions that imply that the state space S be uncountably infinite.

F. Anscombe and R. Aumann provided a "shortcut", which we are also going to follow. The basic idea is to allow acts to associate *lotteries over prizes* to states, instead of just prizes. The big advantage is that this turns the set of acts into (you guessed it!) a convex subset of a vector space. To see this, recall that $\mathcal{L}(X)$ is itself a convex subset of a vector space; for any two acts f, g mapping X to $\mathcal{L}(X)$, we can then define the "convex-combination act" $\alpha f + (1-\alpha)g$ as follows: for every state $s \in S$, let

$$[\alpha f + (1-\alpha)g)](s) = \alpha f(s) + (1-\alpha)g(s).$$

Since $f(s), g(s) \in \mathcal{L}(X)$ and we already know that convex combinations of lotteries are well-defined, the above is a well-posed definition.

A little more formally, we can view the set of all "Anscombe-Aumann acts" as a convex subset of the space $[\mathcal{M}(X)]^S$, consisting of functions from S to the set of functions from X to $\mathbb{R}$ (which we denoted $\mathcal{M}(X)$ above). If S is finite, this is just a finite Cartesian product of $\mathcal{M}(X)$; otherwise, we typically need to introduce measurability and other technical restrictions, but there is no conceptual difference. The key point is that we are still considering objects in some convex subset of a linear space.

As we will see, this allows us to provide a quick proof of the expected-utility theorem, essentially as a further corollary to the von Neumann-Morgenstern theorem, and without any assumption about S. However, you should be warned that this does have a cost in terms of the interpretation of the model. Specifically, we need to assume that *lotteries exist and are perceived as "objective"*. This is OK if all we are interested is a characterization of "rational" choice under uncertainty; however, if we wish to invoke this result to provide a foundation for the subjective approach to probability (which, after all, was Savage's original goal), this is *not* good. We use a mathematical argument to construct subjective probabilities, but, in order to do so, we need to assume the existence of objective probabilities in lotteries; the question is, where do *these* come from? [It turns out that there are ways around this; in particular, as you will see, there is *nothing* in the von Neumann-Morgenstern theorem that requires probabilities—the only necessary feature of the model is convexity. If you can guarantee that the set of objects of choice is convex with respect to a suitable operation, you can still use that theorem, and everything goes through. And, it turns out that you *can* construct suitable convex combinations in an entirely subjective way. So, everything is fine in the end.]

3. Expected Utility and the Von Neumann- Morgenstern Theorem

How *should* people choose among lotteries, or acts? Let's talk about lotteries first. One obvious criterion, at least if prizes are monetary, is *expected value maximization*: when faced with a choice of lotteries, pick the one that yields the maxmimum expected monetary prize. But, while this criterion seems plausible, it does not capture "risk aversion".

The classical example, due to J. Bernoulli, is the **St. Petersburg Paradox**. Suppose you are offered the following bet: a coin is flipped, and if Head obtains, you get 1 dollar. Otherwise, it is flipped again, and if Head obtains, you get 2 dollars. In general, we keep flipping the coin until Head obtains; if this happens on the n-th coin flip, you get 2^n dollars. How much would you be willing to pay for the privilege of participating in this bet?

If you rank lotteries according to the expected value criterion, you should be willing to pay an *infinite* entry fee, because the expected value of this bet is

$$1\frac{1}{2} + 2\frac{1}{4} + 4\frac{1}{8} + \ldots = \frac{1}{2}(1 + 1 + \ldots) = \infty.$$

Formally, this lottery can be written as $p = (1, \frac{1}{2}; 2, \frac{1}{4}; 4, \frac{1}{8}; \ldots)$. While this is not finite, the main point goes through if there is a pre-specified maximum number of coin flips, say N, and we stipulate that if N tails obtain, you get nothing. If N is large, this does not change the substance of the argument: you should be willing to pay an extremely large (if not infinite) entry fee.

The idea of **expected utility** was introduced precisely to avoid such difficulties. Suppose you value each successive dollar less, so your valuation for sums of money is *concave*. Then you can justify the "normal" response to the St. Petersburg Paradox. For instance, suppose that the value, or "utility", of x dollars for you is $u(x) \equiv \log_2 x$ [whatever this means]. Suppose further that the value of a lottery $(x_1, p_1; \ldots x_n, p_n)$ for you is its "expected utility"

$$U(p) \equiv \sum_{i=1}^{n} p_i u(x_i) = \sum_{i=1}^{n} p_i \log_2 x_i.$$

Then you should see that the amount you would be willing to pay for this bet is finite (and small).

Now, this is all well and good, and may remind you of our development of utility in Chapter 1 of MWG. But you will also remember that our mantra then was: *utility is just a representation of preferences*, so you should ask whether *this* notion of "utility" can be viewed as representing preferences, under suitable assumptions.

You will also remember that, in the context of choice under certainty, utility differences, and hence the concavity or convexity of a particular representation of preferences, do *not* matter. So, why is it that they matter here?

The latter question has a simple, but not very informative answer. It is true that, if some function φ represents preferences over some abstract set of choices $\mathcal{X}$, then so does any increasing transformation $\psi(\varphi)$ of φ. Here, for instance, if U as defined above represents the DM's prefs, so does $U^2(p) = (\sum_i p_i \log_2 x_i)^2$. What is *not* true is that you can replace $\log_2 x$ with some increasing transformation—think of replacing $\log_2 x$ with $2^{\log_2 x} = x$, for instance.

Thus, the confusion is mainly terminological: in Chapter 1 of MWG, we called *utility* any *representation function* φ over a set of alternatives $\mathcal{X}$; here, on the other hand, "utility" is the little-u, whereas big-U is called "expected utility". *We cannot take arbitrary increasing transformations of little-u*, even though we still can do whatever we want to big-U.

In fact, in order to emphasize this aspect of little-u, the "official" convention is to call $U(\cdot)$ the **expected utility functional** and $u(\cdot)$ the **von Neumann-Morgenstern utility function**, or vNM utility. MWG uses the expression **Bernoulli Utility** to denote u, which is just as good. I tend to use vNM utility, but you can take your pick.

Still, this answer is not complete. Where do we get all these properties of big-U and little-u—er, vNM utility—from? We need to turn to axiomatics. However, let us first encode the notion of expected-utility representation in two definitions—one for risk, and one for uncertainty.

Definition 1. *A relation $\succcurlyeq$ on $\mathcal{L}(X)$ is an **expected-utility preference**, with **von Neumann-Morgenstern utility** $u : X \to \mathbb{R}$, iff, for all $p = (x_1, p_1; \ldots, x_n, p_n), q = (y_1, q_1; \ldots; y_m, q_m) \in \mathcal{L}(X)$, $p \succcurlyeq q$ iff $\sum_i p_i u(x_i) \geq \sum_i q_i u(y_i)$.*

Before we state the corresponding definition for uncertainty, one technical bit must be clarified. This is only an issue when the state space is infinite, so if you content yourself with EU on finite state spaces, just skip the next couple of paragraphs and disregard any mention of "algebra" and "measurability" in Definition 2 below.

An *algebra* of subsets of S is a collection of subsets of S that is closed under complements and finite unions, and also contains S. A σ*-algebra* is an algebra that is also closed under *countable* unions. Given a σ-algebra $\mathcal{S}$, a *probability measure* is a set function $P : \mathcal{S} \to [0, 1]$ such that, for $A_1, A_2, \ldots \in \mathcal{S}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, $P(\bigcup_i A_i) = \sum_i P(A_i)$: that is, P is a countably additive set function. In the current setting, without additional assumptions, the "standard" axiomatizations of subjective EU only deliver a *finitely additive probability*, or *probability charge*: a set function P defined on an *algebra* $\mathcal{A}$ of subsets of of S, and such that $P(A \cup B) = P(A) + P(B)$ whenever $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$. On the other hand, it should be noticed that charges can be defined on the (sigma-) algebra 2^S, which is (roughly speaking) generally not possible if one insists on preserving countable additivity.

It turns out that one can define integrals for both probability measures and charges (with countable additivity, the relevant notion is the standard one of *Lebesgue* integration, of course). In both cases, for *simple functions*, i.e. linear combinations of indicator functions, the integral does what it should: if $a = \sum_{i=1}^n 1_{E_i} a_i$, where $E_1, \ldots, E_n$ is a measurable partition of S and $a_i \in \mathbb{R}$, then $\int a dP = \sum_{i=1}^n a_i P(E_i)$. For simplicity, we restrict attention to "simple acts", i.e. acts that yield finitely many non-indifferent prizes. As for the case of lotteries, the theory has a (straightforward and unique) extension to more general acts.

Definition 2. *Fix an algebra $\mathcal{A}$ on S. A relation $\succcurlyeq$ on $\mathcal{A}$–measurable simple acts from states S to prizes X is an **expected-utility preference**, with **von Neumann-Morgenstern utility** $u : X \to \mathbb{R}$ and **probability charge** $P : \mathcal{A} \to \mathbb{R}$, iff, for all $\mathcal{A}$-measurable simple acts f, g, $f \succcurlyeq g$ iff $\int u(f(s))dP \geq \int u(g(s))dP$.*

3.1. **Axioms and vNM Theorem for Lotteries.** Let us consider preference over lotteries first. The following axioms turn out to be necessary and sufficient for the existence of an EU representation.

Axiom 1 (Weak Order)**.** *$\succcurlyeq$ is complete and transitive.*

Axiom 2 (Continuity)**.** *For all $p, q, r \in \mathcal{L}(X)$ such that $p \succ q \succ r$, there exist $\alpha, \beta \in (0, 1)$ such that $\alpha p + (1 - \alpha) r \succ q$ and $q \succ \beta p + (1 - \beta) r$.*

Weak Order is standard (i.e. we cannot say anything more about it than we did when we talked about abstract choice theory). Continuity is essentially technical: it states that, when taking convex combinations of lotteries, there are no holes or jumps in preferences. This can plausibly fail if e.g. r is an exceedingly "bad" lottery (e.g. "you get shot with probability one"). But otherwise it is a reasonable continuity assumption.

Here is the big one:

Axiom 3 (Independence)**.** *For all $p, q, r \in \mathcal{L}(X)$, and all $\alpha \in (0,1)$: if $p \succ q$, then $\alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)r$.*

I won't be able to summarize the on-going debate on this central axiom, but here's the gist of the argument. *Suppose* that we can interpret the convex combinations $\alpha p + (1-\alpha)r$ and $\alpha q + (1-\alpha)r$ as "two-stage lotteries": first, some random device determines which of the lotteries p or r (respectively q and r) is selected; then, the lottery selected in the first stage is played out. Under this interpretation, Independence says that preferences over two-stage lotteries should not be reversed due to the presence of the "irrelevant alternative" r [which is irrelevant in the sense that it is the same for both two-stage lotteries]. This makes sense. *However*, the fact is that $\alpha p + (1-\alpha)r$ and $\alpha q + (1-\alpha)r$ are *not* really two-stage lotteries: they are more boring convex combinations of lotteries—i.e. they are one-stage objects. Thus, either we think of a different interpretation (it's not easy), or we just *assume that the DM reduces two-stage lotteries to convex combinations.* This is a rather heroic assumptions, and it is fairly easy to produce violations: the Allais paradox does just that.

Be that as it may, here's the main result.

Theorem 1. *A preference $\succcurlyeq$ on $\mathcal{L}(X)$ satisfies Weak Order, Continuity and Independence if and only if there exists a function $U : \mathcal{L}(X) \to \mathbb{R}$ such that, for all $p, q \in \mathcal{L}(X)$,*
(i) $p \succcurlyeq q$ iff $U(p) \geq U(q)$; and
(ii) for all $\alpha \in (0,1)$, $U(\alpha p + (1-\alpha)q = \alpha U(p) + (1-\alpha)U(q)$.
Furthermore, if a function $V : \mathcal{L}(X) \to \mathbb{R}$ satisfies (i) and (ii) above, then there exist $\alpha, \beta \in \mathbb{R}$ with $\alpha > 0$ such that, for all $p \in \mathcal{L}(X)$, $V(p) = \alpha U(p) + \beta$.

The last statement of the theorem ("Furthermore...") is a *uniqueness* claim: it asserts that every function satisfying (i) and (ii) is "the same" as U, up to a positive affine transformation. Often, this is abbreviated by saying that U is "cardinally unique".

Notice that the function U in the above theorem (also known as the *von Neumann-Morgenstern* or *expected-utility functional* is a representation of $\succcurlyeq$, not a vNM utility function. So, where's vNM utility? You get it immediately from property (ii) and the structure of lotteries:

Corollary 1. *Consider a function $U : \mathcal{L}(X) \to \mathbb{R}$ that satisfies (ii) in Theorem 1. Define a function $u : X \to \mathbb{R}$ by letting $u(x) = U((x,1))$ for all $x \in X$. Then, for all $p = (x_1, p_1; \ldots; x_n, p_n) \in \mathcal{L}(X)$, $U(p) = \sum_{i=1}^n p_i u(x_i)$.*

The proof is straightforward, so I'll leave it to you [hint: argue by induction, noting that $p = (x_1, p_1; \ldots; x_n, p_n)$ can be written as $p_1(x_1, 1) + (1-p_1)q$, where $q = (x_2, \frac{p_2}{1-p_1}; \ldots; x_n, \frac{p_n}{1-p_n})$.]. Clearly, since U is unique up to positive affine transformations, *so is u*. This clarifies the above discussion about *not* allowing arbitrary increasing transformations of u: if we consider anything other than a positive affine transformation, *we are changing the underlying preferences.*

The reason for this roundabout statement of the vNM theorem for lotteries will be clear momentarily. For now, let's turn to the

Proof. First of all, note that the Theorem is trivially true (including the "uniqueness" claim) if $p \sim q$ for all $p, q \in \mathcal{L}(X)$. Thus, assume there exist $p^*, p_* \in \mathcal{L}(X)$ with $p^* \succ p_*$.

We argue in a sequence of steps.

<u>Step 1</u>. $p \succ q$ and $1 \geq \alpha > \beta \geq 0$ imply $\alpha p + (1-\alpha)q \succ \beta p + (1-\beta)q$.

This is almost immediate, once you notice that $\beta p + (1-\beta)q = \gamma[\alpha p + (1-\alpha)q] + (1-\gamma)q$ for $\gamma = \frac{\beta}{\alpha}$ and apply Independence.

Step 2a If $p \succeq q \succeq r$ and $p \succ r$, there exists a unique $\alpha^* \in [0,1]$ such that $q \sim \alpha^* p + (1-\alpha^*)r$.

From Step 1, if α^* as above exists, it must be unique; thus, let's focus on existence. Also, existence is obvious in the trivial cases $q \sim p$ and $q \sim r$, so let's assume that $p \succ q \succ r$. Let $B = \{\alpha \in [0,1] : \alpha p + (1-\alpha)r \succcurlyeq q\}$ and $W = \{\alpha \in [0,1] : \alpha p + (1-\alpha)r \preccurlyeq q\}$. We will show that there exists $\alpha^* \in (0,1)$ such that $\alpha^* \in B \cap W$.

Both sets are non-empty, because $1 \in B$ and $0 \in W$; furthermore, Continuity implies that B contains a point other than 1, and W contains a point other than 0. Finally, any element of W is a lower bound for B: to see this, suppose there exist $\alpha \in B$ and $\beta \in W$ such that $\beta > \alpha$. By Step 1, since $p \succ r$, this implies that $\beta p + (1-\beta)r \succ \alpha p + (1-\alpha)r$. But, by the choice of α and β, $q \succcurlyeq \beta p + (1-\beta)r$ and $\alpha p + (1-\alpha)r \succcurlyeq q$, so by Transitivity we get $q \succ q$, a contradiction. Thus, for all $\alpha \in B$ and $\beta \in W$, $\alpha \geq \beta$.

It follows that, in particular, B is a non-empty set that is bounded below, and hence has a greatest lower bound, denoted $\alpha^* \in [0,1]$. This is the only "high-tech" (?) statement in this proof.

Since B contains a point other than 1, $\alpha^* < 1$; and since W contains a non-zero point that, as argued above, is also a lower bound for B, $\alpha^* > 0$.

Next, $\alpha^* \in B$; to see this, argue by contradiction. If $\alpha^* \notin B$, then $\alpha^* \in W$, so $\alpha^* p + (1-\alpha^*)r \prec q \prec p$. Continuity then yields $\gamma \in (0,1)$ such that $\gamma[\alpha^* p + (1-\alpha^*)r] + (1-\gamma)p \prec q$, i.e. $[\gamma\alpha^* + (1-\gamma)]p + \gamma(1-\alpha^*)r \prec q$. This means that $\gamma\alpha^* + (1-\gamma) \in W$, and so this quantity is also a lower bound to B; but, since $\alpha^* \in (0,1)$ and $\gamma \in (0,1)$, $\gamma\alpha^* + (1-\gamma) > \alpha^*$, which contradicts the fact that $\alpha^* = \inf B$. Finally, $\alpha^* \in W$. The argument is similar to the one just given: if $\alpha^* \notin W$, then $\alpha^* p + (1-\alpha^*)r \succ q \succ r$, and Continuity yields $\gamma \in (0,1)$ such that $\gamma[\alpha^* p + (1-\alpha^*)r] + (1-\gamma)r \succ q$, i.e. $\gamma\alpha^* p + [\gamma(1-\alpha^*) + (1-\gamma)]r \succ q$. But this implies that $\gamma\alpha^* \in B$, and since $\alpha^*, \gamma \in (0,1)$, $\gamma\alpha^* < \alpha^*$, which contradicts the fact that α^* is a lower bound to B.

Step 2b If $p \sim q$, then $\alpha p + (1-\alpha)r \sim \alpha q + (1-\alpha)r$ for all $r \in \mathcal{L}(X)$ and $\alpha \in [0,1]$.

This is a bit tricky. First, suppose that, for all $s \in \mathcal{L}(X)$, $s \sim p \sim q$: then in particular this is true for $s = \alpha p + (1-\alpha)r$ and $s = \alpha q + (1-\alpha)r$, so there is nothing to prove.

Next, suppose that there exists $s \in \mathcal{L}(X)$ with $s \succ p \sim q$, and that, for definiteness, $\alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)r$ for some $\alpha \in (0,1)$ and $r \in \mathcal{L}(X)$ [To clarify, the latter assumption is really only for definiteness: if the opposite strict preference holds, just rename p to q and q to p. Also note that s is not necessarily the same as r.] We will derive a contradiction in three steps:

(a) For all $\beta \in (0,1)$, Independence and $s \succ q$ implies that $\beta s + (1-\beta)q \succ \beta q + (1-\beta)q = q \sim p$.

(b) For all $\beta \in (0,1)$, Independence and $\beta s + (1-\beta)q \succ p$, which we got from (a), imply that also $\alpha[\beta s + (1-\beta)q] + (1-\alpha)r \succ \alpha p + (1-\alpha)r$.

(c) Now fix $\beta = \frac{1}{2}$. Then, from (b) and the initial assumption, we have $\alpha[\frac{1}{2}s + \frac{1}{2}q] + (1-\alpha)r \succ \alpha p + (1-\alpha)r \succ \alpha q + (1-\alpha)r$. But then Continuity implies that there exists $\gamma \in (0,1)$ such that

$$\begin{aligned} \alpha p + (1-\alpha)r &\succ \gamma\left\{\alpha\left[\frac{1}{2}s + \frac{1}{2}q\right] + (1-\alpha)r\right\} + (1-\gamma)[\alpha q + (1-\alpha)r] \\ &= \alpha\left\{\frac{1}{2}\gamma s + \left(1 - \frac{1}{2}\gamma\right)q\right\} + (1-\alpha)r, \end{aligned}$$

which constitutes a violation of (b) for $\beta = \frac{1}{2}\gamma \in (0,1)$.

Finally, the case $p \sim q \succ s$ is handled similarly: the details are omitted.

Step 3. We can now conclude the proof by constructing a function $U : \mathcal{L}(X) \to \mathbb{R}$ that satisfies properties (i) and (ii) in the statement of Theorem 1.

By Step 2a and the assumption that $p^* \succ p_*$, for all p with $p^* \succcurlyeq p \succcurlyeq p_*$ we can let $U(p)$ be the unique α such that $p \sim \alpha p^* + (1-\alpha)p_*$. For $p \succ p^*$, let $U(p) = \frac{1}{\alpha}$, where α is the (unique, again by Step 1) α such that $\alpha p + (1-\alpha)p_* \sim p^*$. [Intuitively, the "better" p is, the "smaller" must α be, and thus the "larger" $\frac{1}{\alpha}$ is.] Similarly, for $p \prec p_*$, let $U(p) = -\frac{\alpha}{1-\alpha}$, where α is the unique weight for which $\alpha p^* + (1-\alpha)p \sim p_*$. [similar intuition] We now verify that properties (i) and (ii) hold. Both are boring, but conceptually straightforward.

(i) Consider $p, q \in \mathcal{L}(X)$; we must show that $p \sim q$ implies $U(p) = U(q)$ and $p \succ q$ implies $U(p) > U(q)$.

If $p^* \succcurlyeq p \succcurlyeq q \succcurlyeq p_*$, the claim follows from Steps 1 and 2a and the definition of U: specifically, $U(p)p^* + [1-U(p)]p_* \sim p \succcurlyeq q \sim U(q)p^* + [1-U(q)]p_*$ implies $U(p) \geq U(q)$ [consider the contrapositive of Step 1]; furthermore, $p \succ q$ and $U(p) = U(q)$ would yield $U(p)p^* + [1-U(p)]p_* \sim p \succ q \sim U(q)p^* + [1-U(q)]p_* = U(p)p^* + [1-U(p)]p_*$, a contradiction: hence, in this case, $U(p) > U(q)$.

If $p \succ p^* \succ q$, then by construction $U(p) > 1$ and $U(q) < 1$, so again the claim holds, and indeed the strict preference is seen to yield a strict inequality; similarly if $p \succ p_* \succ q$.

There are two remaining cases: $p \succcurlyeq q \succ p^*$ and $p_* \succ p \succcurlyeq q$; I only analyze the former, as the latter is similar. By definition, $\frac{1}{U(p)}p + \left(1 - \frac{1}{U(p)}\right)p_* \sim p^* \sim \frac{1}{U(q)}q + \left(1 - \frac{1}{U(q)}\right)p_*$. Suppose $p \sim q$: by Step 2b, $p^* \sim \frac{1}{U(p)}p + \left(1 - \frac{1}{U(p)}\right)p_* \sim \frac{1}{U(p)}q + \left(1 - \frac{1}{U(p)}\right)p_*$, so by Step 1, $\frac{1}{U(p)}$ is the unique number $\alpha^* \in [0,1]$ such that $\alpha^* q + (1-\alpha^*)p_* \sim p^*$; but by definition this means that $U(q) = \frac{1}{\alpha^*} = U(p)$, as required. Suppose instead $p \succ q$: then Independence implies that $p^* \sim \frac{1}{U(p)}p + \left(1 - \frac{1}{U(p)}\right)p_* \succ \frac{1}{U(p)}q + \left(1 - \frac{1}{U(p)}\right)p_*$; by Step 1, this implies that the unique $\alpha^* \in [0,1]$ such that $\alpha^* q + (1-\alpha^*)p_* \sim p^*$ must satisfy $\alpha^* > \frac{1}{U(p)}$,[1] so $U(q) = \frac{1}{\alpha^*} < U(p)$, as required.

(ii) Consider $p, q \in \mathcal{L}(X)$ and $\alpha \in (0,1)$. Suppose first that $p^* \succcurlyeq p \succcurlyeq q \succcurlyeq p_*$. By definition, $U(p)p^* + [1-U(p)]p_* \sim p$ and $U(q)p^* + [1-U(q)]p_* \sim q$. By Step 2b, $U(p)p^* + [1-U(p)]p_* \sim p$ implies

$$\alpha\{U(p)p^* + [1-U(p)]p_*\} + (1-\alpha)\{U(q)p^* + [1-U(q)]p_*\} \sim \alpha p + (1-\alpha)\{U(q)p^* + [1-U(q)]p_*\};$$

similarly, $U(q)p^* + [1-U(q)]p_* \sim q$ implies

$$(1-\alpha)\{U(q)p^* + [1-U(q)]p_*\} + \alpha p \sim (1-\alpha)q + \alpha p.$$

Therefore, by Transitivity, and rearranging terms,

$$\{\alpha U(p) + (1-\alpha)U(q)\}p^* + \{1 - \alpha U(p) - (1-\alpha)U(q)\}p_* \sim \alpha p + (1-\alpha)q,$$

which, by Step 1, means that $\alpha U(p) + (1-\alpha)U(q)$ is the unique number $\alpha^* \in [0,1]$ such that $\alpha^* p^* + (1-\alpha^*)p_* \sim \alpha p + (1-\alpha)q$; but, by definition, this number is of course just $U(\alpha p + (1-\alpha)q)$, which proves the claim in this case.

The other cases are too boring to do in full detail. Trust me, (ii) holds for all p, q.

[1] We know $\alpha^* \neq \frac{1}{U(p)}$ because $p^* \succ \frac{1}{U(p)}q + \left(1 - \frac{1}{U(p)}\right)p_*$: Step 1 tells us that any smaller value of α^* yields an even less preferred mixture of q and p^*.

Finally, let us verify the uniqueness claim. Unfortunately, we need to do cases once again. Suppose V satisfies (i) and (ii): then, in particular, since by definition $U(p)p^* + [1-U(p)]p_* \sim p$, it must satisfy

$$V(p) = V(U(p)p^* + [1-U(p)]p_*) = U(p)V(p^*) + [1-U(p)]V(p_*),$$

which means that $V(p) = \alpha U(p) + \beta$ for $\alpha = U(p^*) - U(p_*) > 0$ and $\beta = U(p_*)$. If $p \succ p^*$, then by definition $\frac{1}{U(p)}p + (1 - \frac{1}{U(p)}p_* \sim p^*$, so

$$V(p^*) = \frac{1}{U(p)}V(p) + \left(1 - \frac{1}{U(p)}\right)V(p_*),$$

which leads to the same conclusion. *Please* spare me the case $p \prec p_*$, which is similar! □

3.2. **The General von Neumann-Morgenstern Theorem.** The following observation is now crucial to extend the analysis to choice under uncertainty (and, it turns out, beyond!) The preceding proof of Theorem 1 does *not* use the fact that the objects of choice are lotteries: they simply involve the careful application of the axioms. The lottery nature of the objects $p, q, \ldots$ was only used in Corollary 1, where we established the existence of vNM utility on the set X of prizes.

Furthermore, from a formal standpoint, the axioms stated above make sense whenever the set of objects of choice is convex, which ensures that we know how to take convex combinations. In other words, whenever we are given a convex set Z of objects, we can consider a relation $\succcurlyeq^*$ on Z and at least write down the counterpart of the Weak Order, Continuity and Independence axioms above. The argument given in the proof of the preceding theorem still applies. Hence, we have:

Theorem 2 (The True vNM Theorem)**.** *Let Z be a convex subset of a vector space and $\succcurlyeq^*$ a binary relation on Z. The following are equivalent:*

(1) $\succcurlyeq^*$ *(and the corresponding strict preference $\succ^*$) satisfy*
  (a) *Weak Order: $\succcurlyeq^*$ is complete and transitive;*
  (b) *Continuity: If $z \succ^* z' \succ^* z''$, then there exist $\alpha, \beta \in (0,1)$ such that $\alpha z + (1-\alpha)z'' \succ^* z' \succ^* \beta z + (1-\beta)z''$;*
  (c) *Independence: If $z \succ^* z'$ and $\alpha \in (0,1)$, then for all $z'' \in Z$, $\alpha z + (1-\alpha)z'' \succ^* \alpha z' + (1-\alpha)z''$.*

(2) *There exists a function $U : Z \to \mathbb{R}$ such that, for all $z, z' \in Z$,*
  (a) *$z \succcurlyeq^* z'$ iff $U(z) \geq U(z')$, and*
  (b) *for all $\alpha \in [0,1]$, $U(\alpha z + (1-\alpha)z') = \alpha U(z) + (1-\alpha)U(z')$.*

Furthermore, if a function $V : Z \to \mathbb{R}$ satisfies (2) above, there exist real numbers $\alpha > 0$ and β such that $V(z) = \alpha U(z) + \beta$ for all $z \in Z$.

Of course, Corollary 1 does not generalize to arbitrary convex sets Z—simply because there may not be "prizes" to speak about!

3.3. **Choice under Uncertainty.** Now comes the big payoff. Turn to the setting of choice under uncertainty. You may have noticed that I have not introduced notation for the set of all acts. This is because, depending on the required level of generality, we may need to consider different sets. For the purposes of these lecture notes, I focus on the case of a finite state space S; below I provide some indications as to how to extend the analysis.

As long as there are finitely many states, acts are simply vectors of lotteries, indexed by the states. In this environment, we need to consider the set of *all* such vectors, which we might denote by $[\mathcal{L}(X)]^S$. If we write $S = \{s_1, \ldots, s_N\}$, we can also denote the set of all acts by $[\mathcal{L}(X)]^N$; however, I prefer the former notation, because it reminds us that acts are functions in general.

If you really need to know, if S is infinite, we need to worry about measurability issues; furthermore, it turns out that we can restrict attention to simple acts, i.e. acts that take up finitely many distinct values (lotteries), much like we can and do restrict attention to finite lotteries in the setting of choice under risk. It turns out that the EU representation over simple acts has a *unique continuous extension* to the set of all bounded, measurable acts (where continuinty and boundedness can be suitably defined). But, to formalize all this, additional notation is required. Let's not bother here.

In light of the preceding discussion, let us first restate the three key axioms in the Von Neumann-Morgenstern theorem for acts.

Axiom 4 (Weak Order). *$\succcurlyeq$ is complete and transitive.*

Axiom 5 (Continuity). *For all $f, g, h \in [\mathcal{L}(X)]^S$ such that $f \succ g \succ h$, there exist $\alpha, \beta \in (0,1)$ such that $\alpha f + (1-\alpha)h \succ g$ and $g \succ \beta f + (1-\beta)h$.*

Axiom 6 (Independence). *For all $f, g, h \in [\mathcal{L}(X)]^S$, and all $\alpha \in (0,1)$: if $f \succ g$, then $\alpha f + (1-\alpha)h \succ \alpha g + (1-\alpha)h$.*

Note that we can identify an individual lottery $p \in \mathcal{L}(X)$ with the *constant act* that assigns the lottery p to each $s \in S$—much like we can identify individual prizes with degenerate lotteries. With this identification, the axioms above apply a fortiori to $\mathcal{L}(X)$, so Theorem 1 implies that the restriction of $\succcurlyeq$ to $\mathcal{L}(X)$ has an EU representation. The question is, how do we extend this representation to arbitrary acts?

We need one additional axiom. Notice that the statement below explicitly uses the fact that $\succcurlyeq$ is defined over lotteries as well, via the identification with constant acts.

Axiom 7 (Monotonicity/State Independence). *For all $f, g \in [\mathcal{L}(X)]$: if $f(s) \succcurlyeq g(s)$ for all $s \in S$, then $f \succcurlyeq g$.*

What this axiom *seems* to be saying is that, if f is better than g pointwise, then f should be considered better than g. But this would not make much sense if *the same prize had a different value in different states.* So, the axiom implicitly says that this is not the case: prizes or lotteries are equally valuable in every state. If you think this is trivially true in most cases, consider the states "healthy" and "sick", and the "intrinsic value" of a huge chocolate tarte (or pumpkin pie, or whatever) in the two states... [It turns out that there is a vast literature on *state-dependent preferences.* Interesting, but subtly difficult. The main contributor/proponent is Edi Karni, along with his coauthors.]

With this additional assumption, we have:

Theorem 3 (Anscombe-Aumann). *Consider a finite state space S, a set X of prizes, and a preference relation $\succcurlyeq$ on $[\mathcal{L}(X)]^S$. The following statements are equivalent:*

1. *$\succcurlyeq$ satisfies Weak Order, Continuity and Independence;*
2. *There exists a function $u : X \to \mathbb{R}$ and a probability P on S such that, for all $f, g \in [\mathcal{L}(X)]^S$,*

$$f \succcurlyeq g \quad \Leftrightarrow \quad \sum_{s \in S} U(f(s))P(\{s\}) \geq \sum_{s \in S} U(g(s))P(\{s\}),$$

where, for every $p = (x_1, p_1; \dots; x_n, p_n) \in \mathcal{L}(X)$, $U(p) = \sum_{i=1}^n p_i u(x_i)$.

Furthermore, if there exist $f, g \in [\mathcal{L}(X)]^S$ *such that* $f \succ g$, *then* P *is unique, and* u *is unique up to a positive affine transformation.*

Proof. Of course, we only need to show that (1) implies (2): the other direction is easy to check. Also, we can assume that $\succcurlyeq$ is not trivial, i.e. there exist f, g such that $f \succ g$, otherwise the implication is true, but uninteresting. Notice that this implies that $x \succ x'$ for some pair of prizes x, x'.

Note first that $\succcurlyeq$, restricted to constant acts, satisfies the axioms of the vNM Theorem; hence, there is an affine function $U : \mathcal{L}(X) \to \mathbb{R}$, unique up to a positive affine transformation, such that, for all $f, g \in [\mathcal{L}(X)]^S$ such that $f(s) = p$ and $g(s) = q$ for all $s \in S$, $f \succcurlyeq g$ iff $U(p) \geq U(q)$. Renormalize this U so that its image contains $[-1, 1]$: this will turn out to be very convenient [this can obviously be done via a positive affine transformation]. In particular, note that there are lotteries $p_0, p_1, p_{-1} \in \mathcal{L}(X)$ such that $U(p_0) = 0$, $U(p_1) = 1$, and $U(p_{-1}) = -1$. Furthermore, for all $i = 1, \dots, N$, there exist $f_i, \bar{f}_i \in [\mathcal{L}(X)]^S$ such that $U(f_i(s_i)) = 1$, $U(\bar{f}(s_i)) = -1$, and $U(f_i(s_j)) = U(\bar{f}_i(s_j)) = 0$ for $i \neq j$. Finally, let $f_0(s) = p_0$ for all s.

Now turn to $\succcurlyeq$ on general acts. As noted above, this preference satisfies the axioms of Theorem 2, so there exists an affine $\mathcal{U} : [\mathcal{L}(X)]^S \to \mathbb{R}$ that represents it. In particular, by the uniqueness property of the vNM representation, $\mathcal{U}$ restricted to constant acts must be a positive affine transformation of U, and therefore we can renormalize it so that, indeed, $\mathcal{U}(f) = U(p)$ whenever $f(s) = p$ for all $s \in S$.

We next record two immediate consequences of Monotonicity and the fact that $\mathcal{U}$ is affine. First, if $U(f(s)) = \gamma U(g(s))$ for all $s \in S$ and some $\gamma \in [0, 1]$, then $U(f(s)) = U(\gamma g(s) + (1 - \gamma) f_0(s)$ for all s, i.e. $f(s) \sim \gamma g(s) + (1 - \gamma) f_0$; Monotonicity then implies that $f \sim \gamma g + (1 - \gamma) f_0$, so $\mathcal{U}(f) = \gamma \mathcal{U}(g)$ [because $\mathcal{U}(f_0) = U(p_0) = 0$.] Second, if $U(f(s)) = -U(g(s))$ for all s, then $U(\frac{1}{2} f(s) + \frac{1}{2} g(s)) = U(p_o) = U(f_0(s))$ for all s: that is, $\frac{1}{2} f(s) + \frac{1}{2} g(s) \sim f_0(s)$ for all s; again, by Monotonicity, $\frac{1}{2} f + \frac{1}{2} g \sim f_0$, so $\frac{1}{2} \mathcal{U}(f) + \frac{1}{2} \mathcal{U}(g) = 0$, and therefore $\mathcal{U}(f) = -\mathcal{U}(g)$. In particular, $\mathcal{U}(f_i) = -\mathcal{U}(\bar{f}_i)$ for all i.

Consider an act $g \in [\mathcal{L}(X)]^S$ such that, for all $s \in S$, $U(g(s)) \in [-1, 1]$, and $\sum_s |U(g(s))| = 1$. Consider the mixture $h = \sum_{i=1}^N U(g(s_i)) g_i$, where $g_i = f_i$ if $U(g(s_i)) \geq 0$, and $g_i = \bar{f}_i$ if $U(g(s_i)) < 0$. Thus, for every $i = 1, \dots, N$, $h(s_i) = |U(g(s_i))| p_1 + [1 - |U(g(s_i))|] p_0$ if $U(g(s_i)) \geq 0$, and $h(s_i) = |U(g(s_i))| p_{-1} + [1 - |U(g(s_i))|] p_0$ if $U(g(s_i)) < 0$, because $f_j(s_i) = \bar{f}_j(s_i) = p_0$ for all $j \neq i$. This implies that, regardless of the sign of $U(g(s_i))$, $U(h(s_i)) = U(g(s_i))$, and therefore $h(s_i) \sim g(s_i)$ for all $i = 1, \dots, N$. Monotonicity then yields $h \sim g$. But then, since $\mathcal{U}$ is affine and $\mathcal{U}(f_i) = -\mathcal{U}(\bar{f}_i)$, a simple induction shows that

$$\mathcal{U}(g) = \mathcal{U}(h) = \sum_{i=1}^N |U(g(s_i)) \mathcal{U}(g_i) = \sum_{i: U(g(s_i)) \geq 0} U(g(s_i)) \mathcal{U}(f_i) + \sum_{i: U(g(s_i)) < 0} [-U(g(s_i))] \mathcal{U}(\bar{f}_i) = \sum_{i=1}^n U(g(s_i)) \mathcal{U}(f_i).$$

Next, consider an act g such that $M \equiv \sum_s |U(g(s))| > 1$. The act $h = \frac{1}{M} g + \left(1 - \frac{1}{M}\right) f_0$ clearly satisfies $U(h(s)) \in [-1, 1]$ and $\sum_s |U(h(s))| = 1$, so by the result just proved $\mathcal{U}(h) = \sum_s U(h(s)) \mathcal{U}(f_i) = \frac{1}{M} \sum_s U(g(s)) \mathcal{U}(f_i)$. Since $\mathcal{U}(h) = \frac{1}{M} \mathcal{U}(g)$, it follows that $\mathcal{U}(g) = \sum_s U(g(s)) \mathcal{U}(f_i)$ as well.

Next, consider an act g such that $M \equiv \sum_s |U(g(s))| \in (0, 1)$. For every s, there is a lottery p_s such that $U(p_s) = \frac{1}{M} U(g(s))$, because $U(g(s)) \in [-M, M]$ and the range of U contains $[-1, 1]$. Define $h(s) = p_s$ for all s, so $U(Mh(s) + (1 - M) f_0) = U(Mp_s + (1 - M) p_0) = U(g(s))$ for all s: thus,

$Mh(s)+(1-M)f_0(s) \sim g(s)$ for all s, and by Monotonicity $Mh+(1-M)f_0 \sim g$, or $\mathcal{U}(h) = \frac{1}{M}\mathcal{U}(g)$. As above, $U(h(s)) \in [-1,1]$ and $\sum_s |U(h(s))| = 1$, so $\mathcal{U}(g) = M\mathcal{U}(h) = M\sum_i U(h(s_i))\mathcal{U}(f_i) = \sum_i U(g(s_i))\mathcal{U}(f_i)$.

Finally, f_0 is the unique act for which $\sum_s |U(f(s))| = 0$, and it is clearly true that $\mathcal{U}(f_0) = \sum_i U(f_0(s_i))\mathcal{U}(f_i)$.

To conclude the proof, notice that $\mathcal{U}(f_i) \geq 0$ for all i, because $f_i \succcurlyeq f_0$ by Monotonicity; also, $\sum_i \frac{1}{N} f_i(s) = \frac{1}{N}p_1 + (1-\frac{1}{N})p_0$ for all s: thus,

$$\sum_i \mathcal{U}(f_i) = N\sum_i \frac{1}{N}\mathcal{U}(f_i) = N\mathcal{U}(\sum_i \frac{1}{N}f_i) = NU(\frac{1}{N}p_1 + (1-\frac{1}{N})p_0) = N\frac{1}{N} = 1.$$

Therefore, the required probability P can be defined by letting $P(\{s_i\}) = \mathcal{U}(f_i)$ for all $i = 1, \ldots, N$.

Uniqueness of P follows from non-triviality and the fact that, for every state s and $x, x' \in X$ such that $x \succ x'$, there is a unique $\alpha \in [0,1]$ such that $\alpha x + (1-\alpha)x' \sim f$, where $f(s) = x$ and $f(s') = x'$ for all $s' \neq s$. □

4. Key Concepts and Applications of EU

The economics literature is a vast repository of applications of EU theory; MWG Chap. 6 offers a brief overview of "classical" topics such as portfolio choice and insurance. Read it up!

I do wish to emphasize a few key concepts and techniques. These apply to settings characterized by risk or uncertainty. To provide a unified treatment, it is useful to adopt the language of statistics and refer to *preferences over random variables*, in lieu of lotteries or acts. But you should realize that this is merely out of terminological convenience. After all, lotteries can be viewed as probability distributions of random variables: $(x_1, p_1; \ldots; x_n, p_n)$ can be viewed either as a lottery that yields the prize x_i with probability p_i, or as the probability distribution of a random variable—call it $\tilde{x}$—that takes the value x_i with probability p_i. Furthermore, recall that a random variable is nothing but a (measurable) function from one (measurable) space to another; in our case, the domain is S, the state space, and the range is X, the set of prizes; so, "act" is really just a fancy name for "random variable"!

For simplicity, we will also *assume throughout that X is a convex subset of* $\mathbb{R}$. [However, note that people can and do talk about random *vectors*, random *consumption bundles*, random *consumption streams*, etc.; we can think of these objects as random variables that take values in some set of vectors, consumption bundles, consumption streams, etc., as the case may be. Thus, in fact, the concept of random variable is quite a bit more general than one might think, as are most of the definitions and results in this section.]

Bottom line: we consider preferences $\succcurlyeq$ over random variables taking values in a convex subset X of $\mathbb{R}$. We typically interpret prizes as monetary quantities. Again, each random variable, denoted $\tilde{x}$, $\tilde{y}$, etc., may correspond to a lottery-as-probability-distribution, or it may be an act, for some suitably defined state space. As usual, we abuse notation and write things like $\tilde{x} \succcurlyeq x$ to mean that the r.v. $\tilde{x}$ is weakly preferred to the prize (or degenerate r.v.) x.

We *will* be interested in making probabilistic statements about these random variables. To do so, we will assume that each random variable is characterized by a **cumulative distribution function**, or **c.d.f**, denoted F, G, etc. For continuous random variables, we will use **density functions**, denoted f, g, etc. We may also need to refer to joint c.d.f.'s or density functions.

If you don't recall the relevant definitions, *seek immediate help!* However, I do wish to point out the connections between lotteries, acts, and c.d.f.'s, in the case $X = \mathbb{R}$.

First, if $p = (x_1, p_1; \ldots; x_n, p_n)$ is a lottery and $\tilde{x}$ denotes the corresponding r.v. [short for "random variable", not "recreational vehicle"], then the c.d.f. of $\tilde{x}$, denoted $F : \mathbb{R} \to [0, 1]$, can be obtained by letting

$$F(x) = \sum_{i:x_i \leq x} p_i.$$

Note that this will be a step function; in particular, *if* $x_1 < x_2 < \ldots < x_n$, then $F(x_1) = p_1$ and $F(x_i) - F(x_{i-1}) = p_i$ for $i = 2, \ldots, n$.

Second, if P is a probability measure on the state space $(S, \mathcal{A})$, and $f : S \to \mathbb{R}$ is an $\mathcal{A}$-measurable act, then of course

$$F(x) = P(\{s : f(s) \leq x\}).$$

In particular, if f is a simple function, so that it can be written as $f(s) = \sum_{i=1}^{n} 1_{E_i}(s) x_i$ for some partition $E_1, \ldots, E_n$ of S, then $F(x) = \sum_{i:x_i \leq x} P(E_i)$. Again, if $x_1 < \ldots < x_n$, then $F(x_1) = P(E_1)$ and $F(x_i) - F(x_{i-1}) = P(E_i)$ for $i = 2, \ldots, n$.

As these examples suggest, *c.d.f.'s are non-decreasing, right-continuous functions that satisfy the normalization conditions* $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$.

As far as notation is concerned, again if $X = \mathbb{R}$ and F is right-continuous, non-decreasing function (e.g. the c.d.f. of a r.v., but not necessarily), $\int_A g(x) F(dx)$ or $\int_A g(x) dF$ denote the integral of the function $g : \mathbb{R} \to \mathbb{R}$ over the measurable set $A \subset \mathbb{R}$ with respect to F. You can think of this as either a Riemann or a Lebesgue integral (or something even fancier).

As usual, the region of integration is omitted if it equals the domain of F, i.e. all of $X = \mathbb{R}$. If F is the identity, then $\int g(x) dF$ is also denoted $\int g(x) dx$. Finally, the expectation of a r.v. $\tilde{x}$ with c.d.f. F is denoted $\mathrm{E}[\tilde{x}] \equiv \int x dF$. Expectations of functions of r.v.'s are also denoted by $\mathrm{E}[g(\tilde{x})] = \int g(x) F(dx)$, where F is the c.d.f. of $\tilde{x}$. Thus, the expected utility of a random variable $\tilde{x}$ will be typically denoted by $\mathrm{E}[u(\tilde{x})]$.

Let me conclude with a key fact about the expectation of functions of random variables.

Lemma 1. *Suppose* $X = \mathbb{R}$. *For all random variables* $\tilde{x}$ *and all concave functions* $g : \mathbb{R} \to \mathbb{R}$, $\mathrm{E}[g(\tilde{x})] \leq g(\mathrm{E}[\tilde{x}])$.

Proof. Consider first a random variable that takes up finitely many values $x_1 < \ldots < x_n$. Let F be its c.d.f. and define $p_1 = F(x_1)$ and $p_i = F(x_i) - F(x_{i-1})$ for all $i = 2, \ldots, n$. Then clearly $\mathrm{E}[\tilde{x}] = \sum_i p_i x_i$ and $g(\mathrm{E}[\tilde{x}]) = \sum_i g(x_i) p_i$. Since g is concave, $g(\lambda x + (1-\lambda) y) \geq \lambda g(x) + (1-\lambda) g(y)$ for all $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}$. A simple induction argument then proves the claim for this r.v..

Without providing the details, this implies that the result is true for all *bounded* r.v.'s, because they can be approximated by a uniformly convergent sequence of step (i.e. simple) r.v.'s. Finally, the claim can be extended to arbitrary r.v.'s. □

4.1. **Certainty Equivalents and Risk Aversion.** Begin with a basic definition:

Definition 3. *Consider a random variable* $\tilde{x}$. *Then a prize* $x \in X$ *is a* ***certainty equivalent*** *of* $\tilde{x}$ *if* $x \sim \tilde{x}$.

The concept is typically applied in settings where $X = \mathbb{R}$ or $X = \mathbb{R}_+$, representing monetary outcomes. But the idea is general, and the term "certainty equivalent" self-explanatory.

Note that certainty equivalents may fail to exist in general settings: for instance, consider an EU DM whose Bernoulli utility exhibits a discontinuity at some x_0, and a r.v. $\tilde{x}$ such that $\lim_{x\uparrow x_0} u(x) < \mathrm{E}[u(\tilde{x})] < \lim_{n\downarrow x_0} u(x)$. On the other hand, a r.v. may have multiple certainty equivalents: consider e.g. an EU DM whose Bernoulli utility has flat portions. But, *with monetary outcomes, we typically assume that vNM utility is continuous and strictly increasing in money, so the certainty equivalent exists and is unique.* Under these conditions, the certainty equivalent x of a r.v. $\tilde{x}$ must satisfy

$$x = u^{-1}\left(\mathrm{E}[u(\tilde{x})]\right).$$

The notation $\mathrm{CE}[\tilde{x}]$ (or similar) is sometimes used to denote the certainty equivalent of p.

The main reason why EU theory is useful is probably related to the following definition.

Definition 4. *A preference relation $\succcurlyeq$ is* ***risk-averse*** *(resp. -neutral, -loving) iff, for all r.v.'s $\tilde{x}$,*

$$\mathrm{E}[\tilde{x}] \succcurlyeq \tilde{x}$$

(resp. $\sim \tilde{x}$, $\preccurlyeq \tilde{x}$).

That is: a DM is risk-averse if she (weakly) prefers to receive the expected value of a r.v $\tilde{x}$ to the r.v. itself. Equivalently, under the assumption that certainty equivalents exist, a DM is risk-averse iff, for any r.v. $\tilde{x}$, $\mathrm{E}[\tilde{x}] \geq \mathrm{CE}[\tilde{x}]$.

Many characterizations of risk-aversion are available. The most basic one relies on Jensen's Inequality:

Proposition 1. *An EU preference relation $\succcurlyeq$ with Bernoulli utility u is risk-averse if and only if u is concave.*

Proof. The "if" part follows from Lemma 1. For the "if" part, fix $x, y \in X$ and $\lambda \in [0,1]$. For definiteness, assume $x > y$; consider a r.v. $\tilde{x}$ with c.d.f. F such that $F(z) = 0$ for $z < y$, $F(z) = 1 - \lambda$ for $z \in [y, x)$, and $F(z) = 1$ otherwise. Then clearly $\mathrm{E}[\tilde{x}] = \lambda x + (1-\lambda)y$ and $\mathrm{E}[u(\tilde{x})] = \lambda u(x) + (1-\lambda)u(y)$, and concavity of u yields the required conclusion. □

It is also possible to *compare different individuals' risk attitudes.* The basic idea is simple. Suppose an individual (Ann) weakly prefers a random variable $\tilde{x}$ to a certain prize x. Then, an individual (Bob) who is less risk-averse than Ann should also prefer $\tilde{x}$ to x. This intuition of course requires that the ranking of prizes be the same across individuals. The following definition formalizes these ideas, under the usual assumption that preferences are strictly increasing in money.

Definition 5. *Suppose that $\succcurlyeq_1$ and $\succcurlyeq_2$ are such that, for all $i = 1, 2$ and $x, y \in X$, $x \succcurlyeq_i y$ iff $x \geq y$. Then $\succcurlyeq_1$ is at least as risk-averse as $\succcurlyeq_2$ iff, for all r.v.'s $\tilde{x}$ and prizes x, $\tilde{x} \succcurlyeq_1 x$ implies $\tilde{x} \succcurlyeq_2 x$.*

For EU preferences, it is relatively simple to compare risk attitudes:

Proposition 2. *Under the conditions of Def. 5, assume further that, for $i = 1, 2$, $\succcurlyeq_i$ is an EU preference with continuous Bernoulli utility u_i. Then $\succcurlyeq_1$ is at least as risk-averse as $\succcurlyeq_2$ iff there exists a concave, strictly increasing function $g : u_2(X) \to \mathbb{R}$ such that $u_1(x) = g(u_2(x))$.*

Proof. Suppose such a function g can be found. Then $\tilde{x} \succcurlyeq_1 x$ iff $\mathrm{E}[u_1(\tilde{x})] \geq u_1(x)$, i.e. iff $\mathrm{E}[g(u_2(\tilde{x}))] \geq g(u_2(x))$. By Jensen's Inequality (Lemma 1), since g is concave, the l.h.s. is not greater than $g(\mathrm{E}[u_2(\tilde{x})]$; and since g is strictly increasing, this implies that $\mathrm{E}[u_2(\tilde{x})] \geq u_2(x)$, i.e. $\tilde{x}_2 \succcurlyeq_2 x$, as required.

In the opposite direction, assume each $\succcurlyeq_i$ is monotonic w.r.to prizes and $\succcurlyeq_1$ is at least as risk-averse as $\succcurlyeq_2$. Define $g : u_2(X) \to \mathbb{R}$ by letting $g(r) = u_1(x)$ whenever $u_2(x) = r$. Notice that g is well-defined, because $u_2(x) = u_2(y)$ implies $u_1(x) = u_1(y)$; furthermore, $u_2(x) = r > r' = u_2(x')$ imply $u_1(x) > u_1(x')$, so g is strictly increasing. To show that it is also concave, fix $r, r' \in u_2(X)$ such that $r > r'$ and $r = u_2(x)$, $r' = u_2(x')$. Consider a r.v. $\tilde{x}$ with probability distribution $(x, \lambda; x', 1-\lambda)$. Since u_1 is continuous and strictly increasing, there exists a unique certanity equivalent y of $\tilde{x}$ for $\succcurlyeq_1$: hence, $\tilde{x} \sim_1 y$. By assumption, this implies that $\tilde{x} \succcurlyeq_2 y$. Therefore, $\lambda u_2(x) + (1-\lambda)u_2(x') \geq u_2(y) = u_2(u_1^{-1}(\lambda u_1(x) + (1-\lambda)u_1(x')))$. But $u_2(x) = r$, $u_2(x') = r'$, $u_1(x) = g(u_2(x)) = g(r)$ and similarly $u_1(x') = g(r')$, so we get $\lambda r + (1-\lambda)r' \geq u_2(u_1^{-1}(\lambda g(r) + (1-\lambda)g(r')))$. Finally, by definition $g(t) = u_1(u_2^{-1}(t))$, so $u_2(u_1^{-1}(t)) = g^{-1}(t)$, so the above inequality reduces to $\lambda r + (1-\lambda)r' \geq g^{-1}(\lambda g(r) + (1-\lambda)g(r'))$, or $g(\lambda r + (1-\lambda)r') \geq \lambda g(r) + (1-\lambda)g(r')$, as required. □

We also have an equivalent characterization in terms of certainty equivalents:

Corollary 2. *Under the same assumptions, $\succcurlyeq_1$ is at least as risk-averse as $\succcurlyeq_2$ iff $\mathrm{CE}_1[\tilde{x}] \geq \mathrm{CE}_2[\tilde{x}]$ for every r.v. $\tilde{x}$, where $\mathrm{CE}_i[\cdot]$ denotes certainty equivalents w.r.to $\succcurlyeq_i$.*

Proof. Clearly, the above assumption implies that $\succcurlyeq_1$ is at least as risk-averse as $\succcurlyeq_2$ [provide the simple details]. In the other direction, $u_1(x) = g(u_2(x))$ for all x implies that $\mathrm{CE}_1(\tilde{x}) = u_1^{-1}(\mathrm{E}[u_1(\tilde{x})]) = u_2^{-1}(g^{-1}(\mathrm{E}[g(u_2(\tilde{x}))]))$ for all r.v. $\tilde{x}$; by Jensen's inequality, $\mathrm{E}[g(u_2(\tilde{x}))] \leq g(\mathrm{E}[u_2(\tilde{x})])$, and since u_2^{-1} and g^{-1} are strictly increasing, we get $u_1^{-1}(\mathrm{E}[u_1(\tilde{x})]) \leq u_2^{-1}(\mathrm{E}[u_2(\tilde{x})])$, as required. □

4.2. **Measures of Risk Aversion.** Without further ado, here are the relevant definitions.

Definition 6. *Consider an EU DM with twice continuously differentiable Bernoulli utility u. The **Arrow-Pratt measure of absolute risk aversion** is the function*

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

*The **Arrow-Pratt measure of relative risk aversion** is the function*

$$R(x) = -\frac{xu''(x)}{u'(x}) = xA(x).$$

Remark 1. *An EU DM with twice continuously differentiable Bernoulli utility u is risk-averse (-loving, -neutral) if and only if the corresponding Arrow-Pratt measure $A(\cdot)$ is non-negative (resp. non-positive, zero).*

If you really like taking derivatives, you can even prove the following:

Proposition 3. *Consider two EU DM's with twice continuously differentiable Bernoulli utilities u_1, u_2. Then $\succcurlyeq_1$ is more risk-averse than u_2 iff the respective Arrow-Pratt measures A_1 and A_2 satisfy $A_1(x) \geq A_2(x)$ for all $x \in X$.*

Thus, the Arrow-Pratt measures provide a *partial* ordering of EU preferences in terms of risk-aversion. This, in itself, is not so exciting. What's more interesting is the source of these quantities. MWG provides an answer, but I'm going to give you a slightly different (and somewhat more traditional) one.

The basic observation, in either case, is that the notion of (comparative) risk aversion considered in the previous subsection is *global*. It would be nice if we had a *local* measure. For instance, this would make it possible for us to say that an individual "becomes less risk-averse as she gets wealthier"—which certainly sounds true.

That's where the Arrow-Pratt measures kick in. One way to address this issue is to ask the following question. Consider an individual whose *initial wealth* is W dollars. A "risk" is a r.v. with zero expectation. Now suppose that this individual is subject to a "risk" $\tilde{x}$, so that his *terminal wealth*, absent any "protective measure", will be $W - \tilde{x}$. Clearly, regardless of W, a risk-averse individual will dislike being subject to this risk: this much we know from our previous analysis. Question: how much would this individual be willing to pay in order to avoid it? This leads to the following definition [which is independent of the concavity of u].

Definition 7. *Consider an EU DM with strictly increasing and continuous Bernoulli utility. The* ***(insurance) risk premium at wealth level*** W ***for the risk*** $\tilde{x}$ *is the quantity* $\pi(W, \tilde{x})$ *such that*

$$W - \pi(W, \tilde{x}) \sim W - \tilde{x}.$$

The idea is that the DM is just indifferent between paying $\pi(W, \tilde{x})$ and avoiding the risk, and not paying this premium while being subject to the risk.

Note that the risk premium obviously also depends upon u, but this is not indicated for notational simplicity. Also, under the stated assumptions about u, the risk premium always exists and is unique: in particular, note that we must have

$$W - \pi(W, \tilde{x}) = \mathrm{CE}[W - \tilde{x}].$$

The basic idea is now that, for "small" risks, $\pi(W, \tilde{x})$ is proportional to $A(W)$, up to second-order terms. The basic trick is to define "small" in a meaningful way. Fix a r.v. $\tilde{x}$ with $\mathrm{E}[\tilde{x}] = 0$, and define

$$g(k) = \pi(W, k\tilde{x}).$$

We are going to assume that g is twice differentiable. This is true provided u is twice continuously differentiable [I think: under "suitable regularity conditions" otherwise...] Here, we consider a risk $k\tilde{x}$ "small" if k is small; ultimately, we will Taylor-expand $g(k)$ around 0, and for k small we will get a reasonable approximation.

Now note that $g(0) = 0$: if there is no risk, the DM does not wish to pay any premium. Second, consider the equation that determines $g(k)$, namely $u(W - g(k)) = \mathrm{E}[u(W - k\tilde{x})]$. Assuming that it's OK to differentiate under the integral sign, differentiate w.r.to k to get

$$-u'(W - g(k))g'(k) = \mathrm{E}[-u'(W - k\tilde{x})\tilde{x}]$$

[it is useful to do these calculations for the case of a discrete r.v., just to make sure nothing funny is going on!] For $k = 0$, the l.h.s. equals $-u'(W)g'(0)$ and the r.h.s. equals $-u'(W)\mathrm{E}[\tilde{x}] = 0$, because $\mathrm{E}[\tilde{x}] = 0$. Since u is strictly increasing, $u'(W) > 0$, and we can conclude that $g'(0) = 0$.

Now differentiate again the above equation w.r.to k:

$$-u''(W - g(k))[-g'(k)]^2 - u'(W - g(k))g''(k) = \mathrm{E}[u''(W - k\tilde{x})\tilde{x}^2].$$

Again consider $k = 0$. We have shown that $g'(0) = 0$, so the l.h.s. reduces to $-u'(W)g''(0)$. The r.h.s., on the other hand, equals $u''(W)\mathrm{E}[\tilde{x}^2] = u''(X)\mathrm{Var}[\tilde{x}]$, again because $\mathrm{E}[\tilde{x}] = 0$. Therefore, we get $g''(0) = -\frac{u''(W)}{u'(W)}\mathrm{Var}[\tilde{x}] = A(W)\mathrm{Var}[\tilde{x}]$.

Finally, consider a second-order Taylor expansion of g around 0: we get

$$\pi(W, k\tilde{x}) = g(k) \approx g(0) + kg'(0) + \frac{1}{2}k^2 A(W)\mathrm{Var}[\tilde{x}] = \frac{1}{2}A(W)\mathrm{Var}[k\tilde{x}].$$

Thus, as claimed, for "small" risks $\tilde{\epsilon}$, $\pi(W, \tilde{\epsilon}) \approx \frac{1}{2}A(W)\mathrm{Var}[\tilde{\epsilon}]$.

[Notice that this argument is not exactly a paragon of rigor, but it gets the job done]

4.3. **Stochastic Dominance.** The last useful set of definitions pertains to orderings of r.v.'s that are *consistent with, but coarser* than EU. The basic idea is that, in a variety of situations, we wish to make predictions without introducing Bernoulli utility functions. You can look at the details in MWG, but there are two main definitions of interest. The first is straightforward.

Definition 8. *A r.v. with c.d.f. F is said to* ***first-order stochastically dominate*** *another r.v. with c.d.f. G iff, for all $x \in X$, $F(x) \leq G(x)$.*

This means the following: for every "target" prize x, the first r.v. is never more likely to yield x or less than the second. Equivalently, the first r.v. is always at least as likely as the second to yield strictly more than x. This is clearly a good thing!

The following characterization is nice and useful—and you should prove this result for your personal edification!

Proposition 4. *A r.v. $\tilde{x}$ with c.d.f. F first-order stochastically dominates a r.v. $\tilde{y}$ with c.d.f. G iff, for all strictly increasing Bernoulli utility functions u, $\mathrm{E}[u(\tilde{x})] \geq \mathrm{E}[u(\tilde{y})]$.*

Thus, any DM who likes more money to less will agree that $\tilde{x}$ is at least as good as $\tilde{y}$. Indeed, first-order stochastic dominance is a good thing! Note that it yields a partial ordering, but one that is still useful e.g. for comparative statics...

There is a second, stronger (i.e. more discriminating) notion. For this, we need to consider jointly distributed r.v.'s. Consider $\tilde{x}, \tilde{y}, \tilde{z}$, and assume that (1) $\tilde{y} = \tilde{x} + \tilde{z}$, and (2) $\mathrm{E}[\tilde{z}|\tilde{x}] = 0$. This means that $\tilde{y}$ differs from $\tilde{x}$ by the addition of a "conditional risk"—a random variable whose conditional distribution given any possible realization of $\tilde{x}$ has zero mean.

A simple example: $\tilde{x}$ and $\tilde{z}$ are independent, with $\mathrm{E}[\tilde{z}] = 0$. But, for a more interesting example, suppose that $\tilde{x}$ has distribution, say, $(10, \frac{1}{2}; 5, \frac{1}{2})$ and $\tilde{z}$ has conditional distribution $(5, \frac{1}{2}; -5, \frac{1}{2})$ given $\tilde{x} = 10$, and $(0, 1)$ given $\tilde{x} = 5$. Then $\tilde{y} = \tilde{x} + \tilde{z}$ has distribution $(15, \frac{1}{4}; 5, \frac{3}{4})$. Notice that $\mathrm{E}[\tilde{y}] = \mathrm{E}[\tilde{x}]$, but in an intuitive sense $\tilde{y}$ is "more risky"—its values are more spread out, for one thing.

In this case, $\tilde{x}$ should be seen as "better" than $\tilde{y}$:

Definition 9. *Consider two r.v.'s $\tilde{x}$ and $\tilde{y}$ with $\mathrm{E}[\tilde{x}] = \mathrm{E}[\tilde{y}]$. Then $\tilde{x}$* ***second-order stochastically dominates*** *$\tilde{y}$ iff there is a r.v. $\tilde{z}$ such that (1) $\tilde{y} = \tilde{x} + \tilde{z}$, and (2) $\mathrm{E}[\tilde{z}|\tilde{x}] = 0$.*

Since a notion of risk is involved, risk-averse individuals should prefer second-order stochastically dominating r.v.'s. Indeed, this is the case, as is relatively easy to prove integrating by parts. But the converse is also true (although the proof is much harder):

Proposition 5. *Consider two r.v.'s $\tilde{x}$ and $\tilde{y}$ with $\mathrm{E}[\tilde{x}] = \mathrm{E}[\tilde{y}]$. Then $\tilde{x}$* ***second-order stochastically dominates*** *$\tilde{y}$ if and only if $\mathrm{E}[u(\tilde{x})] \geq \mathrm{E}[u(\tilde{y})]$ for all strictly increasing and concave Bernoulli utilities u.*

Thus, indeed, for r.v.'s with the same mean, second-order dominance refines the ordering given by first-order dominance.

4.4. **Terminal Wealth Levels.** I conclude with a comment on the interpretation of X, the set of "prizes". Recall that one motivation behind the Arrow-Pratt measure of risk aversion was to provide a "local", i.e. *wealth-dependent* measure of risk aversion. For this to even be possible, it must be the case that **prizes represent terminal wealth levels**. Under this assumption, it is meaningful to consider some initial wealth W and, for instance, calculate the risk premium corresponding to a risk $\tilde{x}$. As we have seen, this premium is a function of initial wealth. If prizes represented only *changes* in wealth, i.e. "gains and losses", this exercise would be pointless. It would not even be possible [within the standard EU framework] to talk about concepts such as decreasing risk aversion—one would have to ask, "decreasing relative to what?"

I wish to emphasize that the axiomatic treatment of the previous section does not attach any meaning at all to prizes. They *could* be gains and losses, and the von Neumann-Morgenstern proof wouldn't even notice!

Yet, it is typical in 99% of economic analysis to assume that prizes are terminal wealth levels. Just to be clear, this means that, unless instructed to do otherwise, you *must* do the same in your classwork! Example: if you are asked to calculate how much an investor would be willing to pay for a risky asset, or what fraction of her wealth she is willing to invest in risky assets, your answer *must* take her initial wealth into account. It would be **wrong** not to do this! And, again, there are huge practical benefits to this assumption.

However, there are phenomena, such as the widely documented fact that people are "risk-seeking in losses and risk-averse in gains", that cannot be modeled if prizes correspond to final wealth levels. So, there is a price to be paid. Most of the time, it's worth paying it.