Von Neumann – Morgenstern & Savage Expected Utility

I. Introduction, Definitions, and Applications

Microeconomic Theory I
Fall 2024

Why Study Decision Theory?

> What is, then, the use of setting up the above propositions of arithmetic and logic? The use is twofold: to describe approximately the behavior of men who, it is believed, cannot be "all fools all the time," and to give advice on how to reach "correct" conclusions. These two aspects of the rules of logic and arithmetic can be called, respectively, the descriptive [positive] and the recommendatory [normative] aspect.
>
> Marschak (1950)

TOC

Models of Preferences

von Neumann – Morgenstern

Subjective Probability

The Savage Framework

Anscombe–Aumann

Beyond Expected Utility

Appendix

Models of Preferences

Models of Preferences

- Specialize the general choice model to the case of modeling uncertain prospects.
- Tasks:
  - Define representations of the object of choice.
  - Take advantage of these representations to define special classes of preferences that express considerations about uncertainty.

Models of Uncertainty

- von Neumann and Morgenstern (1947)
  Objects of choice are probability distributions on *outcomes*.
- Savage (1954)
  Objects of choice are outcome-valued random variables; functions from *states* to outcomes.
- Anscombe and Aumann (1963)
  Objects of choice are functions from states to probability distributions over outcomes.

Expected Utility Preferences on Simple Lotteries

X is a set of outcomes or prizes.

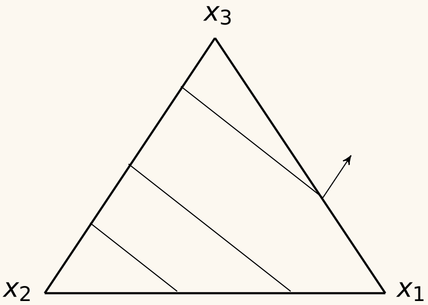
P is the set of probability distributions on X, *lotteries*.

$\operatorname{supp} p$ is the support of p, the set of $x \in X$ such that $p(x) > 0$.

A preference relation $\succ$ on P has an expected utility representation if there is a real-valued function on X, $u : X \to \mathbb{R}$, such that

$$p \succ q \text{ iff } \sum_{x \in \operatorname{supp} p} u(x)p(x) > \sum_{y \in \operatorname{supp} q} u(y)q(y)$$

Expected Utility for Lotteries — Finite X



For fixed prizes, indifference curves are linear in probabilities.

Expected Utility Preferences in Savage Models

X is a set of outcomes or prizes.

S is a set of *states of the world.*

F is the set of *acts,* functions from $S \to X$.

A preference relation $\succ$ on F has an expected utility representation if there is a probability distribution p on S and a real-valued function on X, $u : X \to \mathbb{R}$, such that

$$f \succ g \text{ iff } \sum_{s \in S} u\big(f(s)\big)p(s) > \sum_{s \in S} u\big(g(s)\big)p(s)$$

Expected Utility Preferences in Anscombe-Aumann Models

S is a set of states of the world.

X is the set of outcomes.

P is the set of probability distributions on X,

A is the set of Anscombe-Aumann acts, functions from S to P.

A preference relation $\succ$ on F has an expected utility representation if there is a probability distribution p on S and a real-valued function on X, $u : X \to \mathbb{R}$, such that

$$a \succ b \text{ iff } \sum_{s \in S} \sum_{x \in X} u(x)a(s)(x)p(s) > \sum_{s \in S} \sum_{x \in X} u(x)b(s)(x)p(s)$$

von Neumann-Morgenstern Preferences

Origins

Blaise Pascal, 1623 – 1662

- Early inventor of the mechanical calculator
- Invented Pascal's Triangle
- Invented expected utility, hedging strategies, and a cynic's argument for faith in God all at once.

	God exists	God does not exist
live as if he does	$-C+\infty$	$-C$
live for yourself	$U-\infty$	U

Pascal's Wager

Origins

Daniel Bernoulli

- Mechanics
- Hydrodynamics — Kinetic Theory of Gases
- Bernoulli's Principle

The St. Petersburg Paradox

A coin is tossed until a tails comes up. How much would you pay for a lottery ticket that paid off 2^n dollars if the first tails appears on the n'th flip?

Extrait d'une Lettre de M. N. Bernoulli à M. de M… du 9 Septembre 1713.

Le Livre de feu mon Oncle vient de ſortir de la preſſe, le Libraire m'a dit qu'il en a envoyé un Exemplaire par la Poſte à M. Koenig; ſi vous êtes curieux de le voir, vous pourrés le faire retirer par quelqu'un de chés M. Koenig, à qui j'en donnerai avis, en attendant que je puiſſe lui envoyer quelqu'autres Exemplaires pour vous & pour mes autres amis de Paris. Il n'y aura gueres rien de nouveau pour vous. J'ai été empêché depuis quelque temps de faire de nouvelles recherches ſur la matiere du hazard, c'eſt pourquoi je ne puis rien vous communiquer; cependant en revanche des Problêmes que vous m'avés propoſés, & dont j'examinerai les ſolutions quand j'aurai du loiſir, je vous en propoſe quelqu'autres qui meritent votre application. *Premier Problème.* *A* & *B* jouent alternativement avec un dé à quatre faces marquées de o, 1, 2, 3, *A* met une certaine ſomme d'écus au jeu, & commence à jouer; & après avoir amené ou o, ou 1, ou 2, ou 3 points, il reprend autant d'écus du jeu qu'il a amené de points, & cede le cornet à *B*, qui prend auſſi du reſte autant d'écus qu'il a amené de points; mais s'il amene la face marquée de o, il paye un écu à *A*; & s'il amene un plus grand nombre de points qu'il ne reſte d'écus au jeu, non ſeulement il ne prend rien, mais il met autant d'écus au jeu qu'il a amené de points de trop, & ils continuent ainſi juſqu'à ce qu'il ne reſte plus rien au jeu; je demande quelle eſt la ſomme que *A* doit mettre au jeu pour que leurs ſorts ſoient égaux. *Second Problème.* Si *B* au lieu de payer un écu à *A* quand il n'amene rien, met un écu au jeu, trouver ce qu'alors *A* doit mettre au jeu. *Troiſiéme Problème.* Deux Joueurs *A* & *B* jouent alternativement avec un dé ordinaire, *A* met un écu au jeu, *B* commence à jouer; s'il amene un nombre pair, il prend cet écu; s'il amene un nombre impair, il met un écu au jeu, enſuite

c'eſt *A* qui joue, lequel en amenant un nombre pair prend un écu au jeu comme *B*; mais il ne met rien au jeu quand il amene un nombre impair, & ils continuent juſqu'à ce qu'il ne reſte plus rien au jeu, toujours avec cette condition, qu'ils prennent l'un & l'autre un écu du jeu quand ils amenent un nombre pair; mais que *B* ſeul met un écu au jeu quand il amene un nombre impair, on demande leurs ſorts. *Quatriéme Problème.* *A* promet de donner un écu à *B*, ſi avec un dé ordinaire il amene au premier coup ſix points, deux écus s'il amene le ſix au ſecond, trois écus s'il amene ce point au troiſiéme coup, quatre écus s'il l'amene au quatriéme, & ainſi de ſuite; on demande quelle eſt l'eſperance de *B*. *Cinquiéme Problème.* On demande la même choſe ſi *A* promet à *B* de lui donner des écus en cette progreſſion 1, 2, 4, 8, 16, &c. ou 1, 3, 9, 27, &c. ou 1, 4, 9, 16, 25, &c. ou 1, 8, 27, 64, &c. au lieu de 1, 2, 3, 4, 5, &c comme auparavant. Quoique ces Problêmes pour la plûpart ne ſoient pas difficiles, vous y trouverés pourtant quelque choſe de fort curieux : je vous ai déja propoſé le premier dans ma derniere Lettre. Vous me ferés plaiſir de me communiquer enfin votre ſolution du Her, afin que je puiſſe vous donner l'explication de mon Anagramme. Au reſte, Monſieur, je me réjouis de ce que votre ſanté eſt meilleure; mais je vous plains de ce que vous avés perdu votre Princeſſe. J'ai l'honneur d'être avec un attachement inviolable,

MONSIEUR,

Votre très humble & très obéïſſant Serviteur
N. BERNOULLY.

SPECIMEN
THEORIAE NOVAE
DE
MENSVRA SORTIS.
AVCTORE
Daniele Bernoulli.

1.

Ex eo tempore, quo Geometrae conſiderare coeperunt menſuras ſortium, affirmarunt omnes, *valorem expectationis obtineri, cum valores ſinguli expectati multiplicentur per numerum caſuum quibus obtingere poſſunt, aggregatumque productorum diuidatur per ſummam omnium caſuum: caſus autem conſiderare iubent, qui ſint inter ſe aeque procliues*: Hacque poſita regula, quodcunque reliquum eſt in iſta doctrina huc redit, vt caſus omnes enumerentur, in aeque procliues reſoluantur atque in debitam claſſem diſponantur.

§. 2. Demonſtrationes huius propoſitionis, quarum quidem in lucem prodierunt multae, ſi recte examines, omnes videbis hac inniti hypotheſi, *quod cum nulla ſit ratio, cur expectanti plus tribui debeat vni quam alteri, vnicuique aequae ſint adiudicandae partes*; rationes autem nullas conſiderari, quae

The St. Petersburg Paradox

This doesn't solve the problem. Suppose wealth brows exponentially. Average payoff from paying c when initial wealth is w:

$$\begin{aligned} E &= \frac{1}{2} \cdot \log^{+}(w + 2^{0} - c) + \frac{1}{4} \cdot \log^{+}(w + 2^{1} - c) \\ &\quad + \frac{1}{8} \cdot \log^{+}(w + 2^{2} - c) + \cdots \\ &= \infty \end{aligned}$$

Finite X Axioms

A1. $\succeq$ is complete and transitive.

A2. (Independence axiom) For all $0 < \alpha \leq 1$ and all $r \in P$,

$$p \succeq q \text{ iff } \alpha p + (1-\alpha)r \succeq \alpha q + (1-\alpha)r.$$

A.3. (Archimedean axiom) If $p \succ q \succ r$ then there exists $0 < \alpha, \beta < 1$ such that

$$\alpha p + (1-\alpha)r \succ q \succ \beta p + (1-\beta)r.$$

Finite X Representation Theorem

Theorem. If $\succeq$ satisfies $A.1-3$, then $\succeq$ has an EU representation; there is a function $u : X \to \mathbb{R}$ such that

$$p \succeq q \text{ iff } \sum_{x \in X} u(x)p(x) \geq \sum_{x \in X} u(x)q(x).$$

Furthermore, if $v : X \to \mathbb{R}$ is another EU representation, then there are constants $a > 0$ and b such that $v(x) \equiv au(x) + b$.

Proofs of everything can be found in Fishburn (1970) except where noted otherwise.

The Meaning of u

Suppose $\succeq$ on P is represented by the utility function

$$V(\rho) = \sum_x u(x)\rho(x)$$

Theorem. If $W(\rho) = \sum_x w(x)\rho(x)$ also represents $\succeq$, then $w(\cdot) = \alpha v(\cdot) + \beta$ with $\alpha > 0$.

Preferences on P have ordinal representations. A cardinal measure is a representation of a relation that is unique up to positive affine transformations. It is often said that the theorem states that u is a cardinal utility on X. This reasoning is false. Cardinality is a property of the relation being measured. Any increasing transformations of u represent $\succeq$ restricted to X just as u does. For instance, $V(\rho)^3$ also represents $\succeq$.

Simple Lotteries on Countable X

A *simple lottery* is $p = (p_1 : x_1, \ldots, p_K : x_K)$ where $x_1, \ldots, x_K$ are prizes in $\mathbf{R}$ and $p_1, \ldots, p_K$ are probabilities. Let $\mathcal{L}$ denote the set of simple lotteries. Let

$$u : X \to \mathbf{R}$$

$$\text{and } V(p) = \sum_k u(x_k) p_k.$$

This is the expectation of the random variable $u(x)$ when the the random variable x is described by the probability distribution p.

How do we see that this is “linear” in lotteries? How do we “mix” lotteries?

- For lotteries with common support, mixing is just the convex combination of the probabilities.

- What about lotteries with different support?

  Consider the lotteries $p = (p_1 : x_1, p_2 : x_2)$ and $q = (q_1 : y_1, q_2 : y_2, q_3 : y_3)$.

$$p \oplus_\alpha q = (\alpha p_1 : x_1, \alpha p_2 : x_2, (1-\alpha) q_1 : y_1, (1-\alpha) q_2 : y_2, (1-\alpha) q_3 : y_3)$$

  This is not a convex combination. It combines objects of different sizes.

Expected utility is linear:

$$\begin{aligned} U(p \oplus_\alpha q) &= \sum_{k=1}^{2} \alpha p_k u(x_k) + \sum_{k=1}^{3} (1-\alpha) q_k u(y_k) \\ &= \alpha \sum_{k=1}^{2} p_k u(x_k) + (1-\alpha) \sum_{k=1}^{3} q_k u(y_k) \\ &= \alpha U(p) + (1-\alpha) U(q) \end{aligned}$$

Herstein and Milnor (1953) provide a general definition of mixing that enables the development of representation theorems.

Definition. A mixture space is a set of objects Π, with typical elements π, ρ, μ, ν and a family of functions for $0 \leq \alpha \leq 1$ $\oplus_\alpha : \Pi \times \Pi \to \Pi$ such that

i) $\pi \oplus_1 \rho = \pi$,

ii) $\pi \oplus_\alpha \rho = \rho \oplus_{1-\alpha} \pi$, and

iii) $\left(\pi \oplus_\beta \rho\right) \oplus_\alpha \rho = \pi \oplus_{\alpha\beta} \rho$.

Example 1: Convex sets with the operations of convex combination.

Example 2: Simple probability distributions on a convex sets.

Example 3. S and X are sets, and let M denote the set of functions from S to probability distributions on X. The $\oplus_\alpha$ are the (pointwise) convex combinations of these functions.

Suppose that Π is a mixture space and $\succeq$ is a preference relation on Π. Suppose that $A.1$ holds, and the mixture-space versions of A.2 and A.3:

A.2. (Independence axiom) For all $0 < \alpha \leq 1$ and all $r \in P$,

$$p \succeq q \text{ iff } p \oplus_\alpha r \succeq q \oplus_\alpha r.$$

A.3. (Archimedean axiom) If $p \succ q \succ r$ then there exists $0 < \alpha, \beta < 1$ such that

$$p \oplus_\alpha r \succ q \succ p \oplus_\beta r.$$

vNM Theorem. If $\succeq$ satisfies A.1–3, the $\succeq$ has an EU representation; there is a a function $u : X \to \mathbb{R}$ such that

$$p \succeq q \text{ iff } \sum_{x \in X} u(x)p(x) \geq \sum_{x \in X} u(x)q(x).$$

Furthermore, if $v : X \to \mathbb{R}$ is another EU representation, then there are constants $a > 0$ and b such that $v(x) \equiv au(x) + b$.

What the theorem really says is: If M is a mixture space and $\succeq$ satisfies A.1–3 then there is a linear function $U : M \to \mathbb{R}$. Any other linear representation V is a positive affine transformation of U.

This specializes to the vNM Theorem when M is the set of probability distributions.

Archimedes?

Suppose that there are three outcomes x, y, and z occurring with probabilities ρ_x, ρ_y, and ρ_z, respectively. Suppose that x is *infinitely better* than y and z. Suppose that $(\rho_x, \rho_y, \rho_z) \succ (\rho'_x, \rho'_y, \rho'_z)$ if $\rho_x > \rho'_x$ or if $\rho_x = \rho'_x$ and $\rho_y > \rho'_y$.

These preferences fail the Archimedean assumption. Let $p = (1, 0, 0)$, $q = (0, 3/4, 1/4)$, and $r = (0, 1/4, 3/4)$. Then $p \succ q \succ r$. For all $0 < \alpha < 1$,

$$\alpha p + (1 - \alpha) r = (\alpha, (1 - \alpha)/4, 3(1 - \alpha)/4) \succ q$$

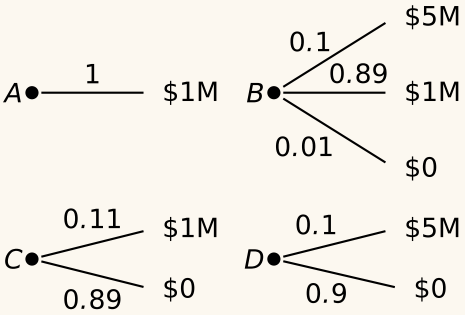
since $\alpha > 0$.

A.1 and A.2 alone imply that $\succeq$ has a lexicographic expected utility representation. See Fishburn (1982).

The property of having no infinitely large or infinitely small elements first appears as Axiom V of Archimedes *On the Sphere and Cylinder.*

Independence?

Are preferences linear in probabilities?



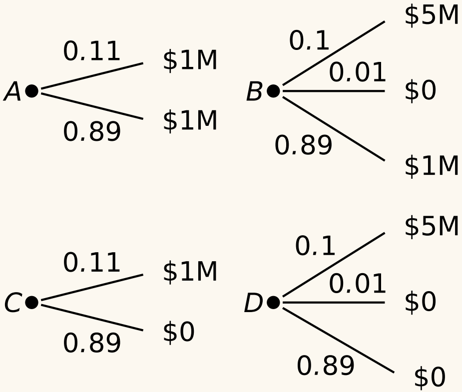
- Compare A to B.
- Compare C to D.

Many people prefer A to B and D to C. This violates the independence axiom.

Why?

This example is called the Allais Paradox after Maurice Allais (1953).

Redrawing the gambles makes it clear that the preference reversal, $A \succ B$ but $D \succ C$, violates independence.



Lotteries on $\mathbb{R}$

Dealing with general probability distributions on the real line requires topological considerations beyond this course. Basically, one has to strengthen the Archimedean Axiom. The conclusion is:

Theorem. Assume A.1, A.2, and a new Archimedean Axiom A.3'. Then $\succeq$ has an EU representation; there is a a bounded and continuous function $u : X \to \mathbb{R}$ such that

$$p \succeq q \text{ iff } \sum_{x \in X} u(x)p(x) \geq \sum_{x \in X} u(x)q(x).$$

Furthermore, if $v : X \to \mathbb{R}$ is another EU representation, then there are constants $a > 0$ and b such that $v(x) \equiv au(x) + b$.

Resolutions of the St. Petersburg Paradox

Nikolaus Bernoulli suggests a resolution to the St. Petersburg Paradox:

> *From all this I conclude that the just value of a certain expectation is not always the average that one finds by dividing by the sum of all the possible cases the sum of the products of each expectation by the number of the case which gives it; that which is against our fundamental rule. The reason for this is that the cases which have a very small probability must be neglected and counted for nulls, although they can give a very great expectation. For this reason one is able yet to doubt if the value of the expectation of B in the case of the 4th problem such as I have found above, is not too great. Similarly in Lotteries where there are one or two quite great Lots, the just value of a single ticket is smaller than the sum of all the money of the Lottery divided by the sum of all the tickets, supposing that the number of those here is also very great. This is a remark which merits to be well examined.*
>
> *Letter from N. Bernoulli to Montmort, 20 Feb. 1714*

In the hands of Kahneman and Tversky, this idea becomes *prospect theory*.

Gabriel Cramer suggests Expected Utility! The first use of a utility function!!

. . . the mathematicians value money in proportion to its quantity, and men of good sense in proportion to the usage that they may make of it. That which renders the mathematical expectation infinite, is the prodigious sum that I am able to receive, if the side of Heads falls only very late, the 100th or 1000th toss. Now this sum, if I reason as a sensible man, is not more for me, does not make more pleasure for me, does not engage me more to accept the game, than if it would be only 10 or 20 million coins. Let us suppose therefore that the total sum beyond 20 millions or (for more ease) beyond 224 = 16777216 coins, is equal to him or rather that I am never able to receive more than 224 coins, however late comes the side of Heads. And my expectation will be

[. . . Calculations. . .]

Therefore *speaking morally* my expectation is reduced to 13 coins, and my equivalent to as much, which would seem much more reasonable than to make it infinite.

If one wishes to suppose that the *moral value* of goods was as the square root of the mathematical quantities,. . . .

Letter from G. Cramer to N. Bernoulli, 21 May. 1728

But there's a hole. For any unbounded utility function u, define $x_k = u^{-1}(2^k)$
Then the expected utility of $p = (2^{-1} : x_1, 2^{-2} : x_2, \ldots)$ is

$$\begin{aligned} U(p) &= \frac{1}{2}u(x_1) + \frac{1}{4}u(x_2) + \cdots \\ &= \frac{1}{2}u\big(u^{-1}(2)\big) + \frac{1}{4}u\big(u^{-1}(4)\big) + \cdots \\ &= 1 + 1 + \cdots \end{aligned}$$

Which lottery is better: p or $q = (2^{-1} : u^{-1}(3), 2^{-2} : u^{-1}(9), \ldots$.

Only in 1934 does Karl Menger observe this and note that a St. Petersburg gamble will fail to exist if and only if u is bounded.

What Have We Accomplished?

Why Axioms?

> *By an axiom system we understand the representation of a theory in such a way that certain sentences of this theory (the axioms) are placed at the beginning, and from then further sentences (the theorems) are derived by means of logical deduction.*
>
> *Carnap (1954, p. 171)*

But in EU theory we will have a sentence of the form:

$P1$ and $P2$ and $P3$ iff $\succeq$ has an EU representation.

Which is a more fundamental description? Red or Blue?

In contrast, the point of view of the Moderns is hypothetico-deductive; . . . . According to this conception, which has since imposed itself, it is not necessary for the principles and derived propositions to be true for a deductive relationship to be established between the former and the latter. At the same time as the idea of truth, that of the superior certainty of the principles falls, and ultimately, to justify the privileged role of the principles, only internal considerations to the deductive strategy remain: thus, the axioms must be coherent, simple to formulate, and relatively few compared to the mass of propositions they generate.

Mongin (2003, pp. 102–3)

- Axioms capture our intuitions about some subject matter.
- Axioms offer an explanation. Axiomatic explanations are not causal; rather they are intuitions.
- Axioms facilitate classification, and thus organization, of an epistemic field.

Subjective Probability

Where Do Probabilities Come From?

- "The measure of the probability of an event is the ratio of the number of cases favourable to that event, to the total number of cases favourable or contrary, and all equally possible, or all of which have the same chance."
- "The probability of an event is the reason we have to believe that it has taken place, or that it will take place."

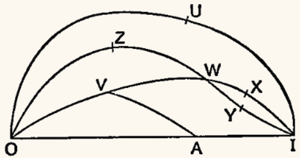
Both quotes from Poisson (1837), who recognizes objective and subjective sources of probability.

- Frequentist probability is defined through a thought experiment. Repeatedly flip a fair coin and the fraction of heads realized converges to 1/2. Probabilities are limit frequencies of infinite sequences of trials.
- Subjective probability is an individual person's measure of belief that an event will occur.

Qualitative Probability

Keynes (1921), not a subjectivist, was nonetheless the first to introduce a measurement framework for discussing subjective probability. *O* and *I* are, respectively, the empty element and the universal element, and receive numerical probabilities 0 and 1. The element on any path from *O* to *I* can be ordered by likelihood, increasing from left to right, but they cannot be compared to elements on other paths. They do not have numerical probabilities. *A*, on the other hand, although comparable only to *O* and *I*, does have a numerical assignment.

Keynes' figure is the graph of a partial order.



Keynes (1921, p. 30)

Numerical Representations for Qualitative Probability

Suppose S is finite. Suppose that $\mathcal{S}$ is a collection of subsets of S such that

i) $\emptyset \in \mathcal{S}$,

ii) if $A \in \mathcal{S}$, then $A^c \in \mathcal{S}$,

iii) if $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.

$\mathcal{S}$ is a (Boolean) algebra of events.

When S is finite, we can take $\mathcal{S} = 2^S$. When $S = \mathbb{R}$ more care is needed.

Definition. A probability on $\mathcal{S}$ is a function $p : \mathcal{S} \to [0, 1]$ such that

i) $p(S) = 1$,

ii) if $A \cap B = \emptyset$ then $p(A \cup B) = p(A) + p(B)$.

Definition. A qualitative probability is a binary relation $\sqsupseteq$ on $\mathcal{S}$ such that

i) $\sqsupseteq$ is complete and transitive,

ii) $S \sqsupset \varnothing$,

iii) for all $A \in \mathcal{S}$, $\varnothing \sqsubseteq A \sqsubseteq S$,

iv) if $A, B, C \in \mathcal{S}$ and $A \cap C = B \cap C = \varnothing$, then $A \sqsupseteq B$ iff $A \cap C \sqsupseteq B \cap C$.

Check that if p is a probability on $\mathcal{S}$ and $A \sqsupset B$ iff $p(A) \geq p(B)$, then $\sqsupset$ is a qualitative probability.

If $\sqsupset$ is a qualitative probability, is there a probability p that represents $\sqsupseteq$? No! *i*)–*iv*) are necessary but not sufficient for $\succeq$ to have a probability representation. See Kraft, Pratt and Seidenberg (1959).

Why should a qualitative probability have a probability representation? *i*) implies that it has a numerical representation q, and it can always be scaled so that $q(\varnothing) = 0$ and $q(S) = 1$. It may fail disjoint additivity.

Let $\mathcal{A} = \{A_1, ..., A_k\}$ and $\mathcal{B} = \{B_1, ..., B_k\}$ be lists of events, allowing repetitions. The lists $\mathcal{A}$ and $\mathcal{B}$ are balanced if for each state $s \in S$ the number of events containing s in $\mathcal{A}$ equals that in $\mathcal{B}$.

A.1. $\succeq$ on $\mathcal{S}$ is complete.

A.2. (Positivity) For all $A \in \mathcal{S}$, $A \succeq \emptyset$.

A.3. (Non-triviality) $S \succ \emptyset$.

A.4. (Finite Cancellation) For all pairs of balanced lists $\mathcal{A}$ and $\mathcal{B}$, if for all $1 \leq j \leq k-1$, $A_j \succeq B_j$, then $B_k \succeq A_k$.

FC implies that $\succeq$ is transitive. Consider the lists $\mathcal{A} = \{A, B, C\}$ and $\mathcal{B} = \{B, C, A\}$, where $A \succeq B$ and $B \succeq C$.

Theorem. $\succeq$ satisfies A.1–4 iff there is a probability ρ on $\mathcal{S}$ such that $A \succeq B$ iff $\rho(A) \geq \rho(B)$.

Sources of Probability

- Frequentist

  Probability is something physical, and can be in principle measured my repeated experiments.

  > [I]f we constitute a fraction whereof the numerator be the number of chances whereby an event may happen, and the denominator the number of all the chances whereby it may either happen or fail, that fraction will be a proper designation of the probability of happening.
  >
  > de Moivre (1718, p. 1-2)

- Logical Probability

  Probability measures the weight of objective support for a proposition. It captures the support evidence E lends to hypothesis H. It is intended to generalize deductive logic with 0-1 truth values to a more graded measure of support.

  > Part of our knowledge we obtain direct; and part by argument. The Theory of Probability is concerned with the part which we obtain by argument, and it treats of the different degrees in which the results so obtained are conclusive or inconclusive.
  >
  > Keynes (1921, p. 3)

- Bayesian Probability

Bayesian probabilists interpret probability as *degree of belief*.

> One can, however, also give a direct, quantitative, numerical definition of the degree of probability attributed by a given individual to a given event, ... it is a question simply of making mathematically precise the trivial and obvious idea that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on it.
>
> de Finetti (1937, p. 62)

What does Probability Mean?

The first generation of probabilistic works, from Huygens through Jakob Bernoulli, already interpreted probability in a variety of distinct senses that pertained to different kinds of subjects. Degrees of certainty (or degrees of probative weight) were states of mind - or rather, states of minds, for they were intersubjective, if not objective. The probative weight of this witness' testimony or that piece of circumstantial evidence was assumed to be the same for all competent judges, and therefore has closer affinities in the twentieth century to John Maynard Keynes' logical probabilities than to Leonard Savage's personal probabilities. These latter should also be distinguished from the psychological probabilities of real (as opposed to ideal) subjects, which may not be internally consistent, or even numerically continuous.

The early probabilists also spoke of degrees of facility, predicated of physical objects: physically symmetric gambling devices are the classic but not the sole example; Jakob Bernoulli sometimes writes as if human bodies had such facilities or propensities with respect to susceptibility to various mortal diseases. Quetelet's "penchants" toward crime or marriage resemble these physical facilities, but apply to averages of collectives, "I'homme moyeri rather than to single physical objects. Karl Popper's propensity interpretation of the probability that an atomic nucleus will decay comes quite close to the original facility notion. Probabilities were also early and long understood as frequencies applied to individual objects insofar as they are members of collectives; the original instances were predictions of individual longevity on the basis of mortality statistics. During the heyday of descriptive statistics in the nineteenth century, this frequentist interpretation drove all others from the field, and is still strongly represented in current philosophy and applications of probability, such as in the theory of random genetic drift.

. . .

There were also "practical probabilities" in which the probabilities were in effect invisible, since the situation called for a combined judgment of probabilities and outcome value. These sorts of expectations, which concern the preferences of agents, were the backbone of the first formulations of mathematical probability, and are familiar to today’s probabilists from de Finetti’s and Savage’s system of personal probabilities. True to their name, practical probabilities surface only in cases that call for concrete action, like betting. They are subjective in that they express an individual’s conviction that an event will or will not come to pass, but unlike degrees of certainty, they are neither intersubjective, nor do they stand alone. They are an indissoluble part of the expectation in which they arise.

Gigerenzer *et al.* (1989, Sec. 8.2)

Savage’s Subjective Expected Utility

The Savage Framework

Savage observes that if certain properties of preferences hold, than the decision-maker chooses as if he is maximizing expected utility with respect to some belief. Probabilities emerge from choices among bets on events.

- X is the set of outcomes.
- S is the set of states of the world.
  A state s is description of particular realizations of all the things the decisionmaker is uncertain about, and which together determine the outcome of every possible choice the DM is considering.
- $\mathcal{A}$ is the algebra of *all* subsets of S.
- F is the set of acts, functions $f : S \to X$. Acts are the objects of choice.
- $\succeq$ A preference order on F.

The key idea is that probabilities are revealed by choices. People choose as if their choice is informed by probabilities.

Some of Savage's Axioms

For an act h, define $f|_A h$ — get $f(s)$ for $s \in A$, else $h(s)$. Let xAy denote the bet that pays off x on A and y otherwise.

P2. If $f|_A h \succ g|_A h$ then for all k $f|_A k \succ g|_A k$

Definition. $f \succeq g$ given A ($f \succeq_A g$) if f' and g' are actions such that f' agrees with f on B, g' agrees with g on B, and f' and g' agree with each other outside of A.

Definition. An event A is null if for all f and g, $f \succeq_A g$.

P3. For outcomes x, y and non-null A, $x \succeq_A y$ iff $x \succeq y$. ($x|_A f \succeq y|_A f$ iff $x \succeq y$.)

P4. For outcomes $x \succ y$ and $x' \succ y'$, and sets A, B, $xAy \succeq xBy$ iff $x'Ay' \succeq x'By'$.

P5. There exist outcomes $x \succ y$.

Definition. Sets are ordered $A \succeq B$ iff there are outcomes $x \succ y$ such that $xAy \succeq xBy$.

- **P6.** (small-event continuity) If $f \succ g$ then for any consequence x there is a partition of S such that on each S_i $f|_{S_i}h \succ g$ and $f \succ g|_{S_i}h$.
- **P7.** If f and g are acts and A is an event such that $f(s) \succeq_A g$ for every $s \in A$, then $f \succeq_A g$; and if $f \succeq_A g(s)$ for every $s \in A$, then $f \succeq_B g$.

For a discursive discussion of how these work, read this blog.

Savage Implies Bayes

$f \succ_A g$ iff for any act h $f|_A h \succ g|_A h$.

$$E_\mu\{u(f|_A h\} > E_\mu\{u(g_A h\} \text{ iff}$$

$$\int_S u\big(f|_A h(s)\big)d\mu > \int_S u\big(g|_A h(s)\big)d\mu \text{ iff}$$

$$\int_A u\big(f(s)\big)d\mu + \int_{A^c} u\big(h(s)\big)d\mu > \int_A u\big(g(s)\big)d\mu + \int_{A^c} u\big(h(s)\big)d\mu \text{ iff}$$

$$\int_A u\big(f(s)\big)d\mu > \int_A u\big(g(s)\big)d\mu \text{ iff}$$

$$\int_A u\big(f(s)\big)d\mu/\mu(A) > \int_A u\big(g(s)\big)d\mu/\mu(A)$$

$$E_\mu\{u(f)|A\} > E_\mu\{u(g)|A\}$$

Anscombe-Aumann Expected Utility

Anscombe-Aumann Expected Utility

Start with some probabilities, use bets on them to get more probabilities.

X outcomes,

P probability distributions on outcomes, roulette wheels,

S states of the world,

H functions $f : S \to P$, horse races.

An AA representation of $\succeq$ is a function $u : X \to \mathbb{R}$ and a probability distribution μ on S such that

$$f \succeq g \text{ iff } \sum_{S} \sum_{X} u(x) f(s)(x) \mu(s) \geq \sum_{S} \sum_{X} u(x) f(s)(x) \mu(s).$$

A.1. $\succeq$ on H is transitive and complete.

A.2. (Independence.) $f \succeq g$ and $0 < \alpha < 1$ imply that $\alpha f + (1-\alpha)h \succeq \alpha g + (1-\alpha)g$.

A.3. (Archimedean.) $f \succ g \succ h$ implies that there are $0 < \alpha, \beta < 1$ such that $\alpha f + (1-\alpha)h \succ g \succ \beta f + (1-\beta)h$.

Theorem. If $\succeq$ satisfies A.1–3 then there are functions $u_s : X \to \mathbb{R}$ such that

$$f \succeq g \text{ iff } \sum_s \sum_x u_s(x) f(s)(x) \geq \sum_s \sum_x u_s(x) g(s)(x). \qquad (*)$$

Furthermore, if $(v_s)_{s \in S}$ is another such representation, then there are constants $a > 0$ and b_s such that each $v_s = a u_s + b_s$.

Getting to an expected utility representation theorem requires additional assumptions.

A.4. (non-triviality) For some $f, g \in H$ $f \succ g$.

A state s is null if for all $f, g, h \in H$, $f|_s h \sim g|_s h$. A.4 guarantees that some non-null state exists. A state s is null iff u_s in $(*)$ is a constant function.

A.5. (state independence) If for some $s \in S$, $h \in H$, and $p, q \in P$, $h|_{\{s\}^c} p \succ h|_{\{s\}^c} q$, then for all non-null states t, $h|_{\{s\}^c} p \succ h|_{\{s\}^c} q$.

Theorem. If $\succeq$ satisfies A.1–5, then there is a function $u : X \to \mathbb{R}$ and a probability distribution ρ on S such that

$$f \succeq g \text{ iff } \sum_s \sum_x u(x) f(s)(x) \rho(s) \geq \sum_s \sum_x u(x) g(s)(x) \rho(s).$$

Beyond Expected Utility

Issues

- Preferences on P need not be linear in probabilities.
  - Allais paradox
  - Decoupling of intertemporal preferences and risk aversion
    Epstein-Zinn preferences

- Beliefs might not be representable by a probability distribution.
  - partial orders
  - Ellsberg paradox
    non-additive probabilities
    Gilboa Schmeidler
    the smooth model

Violations

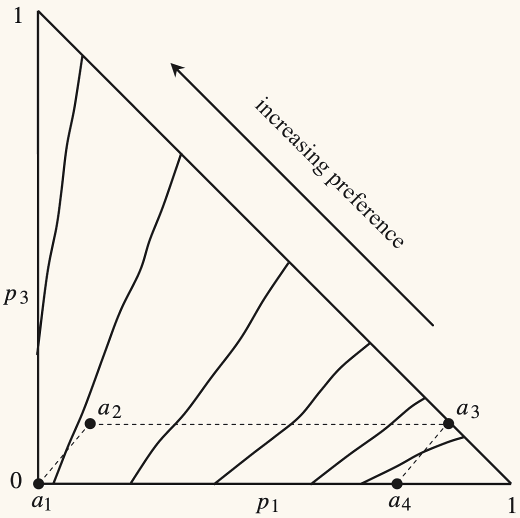
P3, P4. I prefer a T-shirt when it's hot and a flannel shirt when it's cold. These axioms are about the separation of tastes and beliefs.

Dreze (1961)Was the first to suggest state dependent expected utility:

$$U(f) = \sum_{s} u\big(f(s), s\big)\mu(s).$$

Define $v(x,s) = u(x,s)/\alpha_s, \alpha_s > 0$ and $\nu(s) = \alpha_s \mu(s)/\lambda_s$. If (u, μ) represents $\succeq$, so does (v, ν).

Allais (again)



Marschak triangle representation of the allais paradox

The four gambles A, B, C, D, respectively, are represented by the four points $a_1, \dots, a_4$. Since the a_1–a_2 is parallel to the a_3–a_4 line, parallel indifference curves would have to cut the dotted lines the same way, so that $A \succ B$ iff $C \succ D$.

Daniel Ellsberg’s (1961) Experiment

There is a single urn with three balls. One ball is red and the other two are blue or green. one ball is drawn from the urn and the bettor bets on its color. Winning bets pay $100. Available bets are **red**, **blue**, **not red** and **not blue**.

Typical preferences are inconsistent with probabilistic beliefs, e.g. both **red** and **not red** are preferred.

red $\succ$ **blue**

not red $\succ$ **not blue**

	Red	blue	green
red	100	0	0
blue	0	100	0
not blue	100	0	100
not red	0	100	100

Weighted EU

How to generalize EU to account for Allais etc.? One idea is that individuals overweight small-probability events. Thus imagine a weighting function $w : [0,1] \to [0,1]$ with $w(p) > p$ for small p and $w(p) < p$ for large p.

Problem: WEU will not respect FOSD. Suppose wlog both $w(1) = 1$ and there is a $0 < p < 1$ such that $w(p) + w(1-p) < 1$. Suppose $x < y$. The lottery $(x, p; y : 1-p)$ dominates x. But if y is sufficiently near to x and u is continuous, then $u(1)w(1) > u(x)w(p) + u(y)w(1-p)$.

The problem is that weights do not sum to 1. Constructing other similar examples leads to the conclusion that one must have $w(p_1 + p_2) = w(p_1) + w(p_2)$. It is easily proved that this implies $w(p) = cp$ for some constant c.

Rank-Dependent Expected Utility

Instead of weighing probabilities, apply probability weights to the CDF.

$$U(p) = \sum_n w_n(p) u(x_n)$$

where $x_1 \leq x_2 \leq \cdots \leq x_n$ and

$$w_n(p) = q\left(\sum_{k=1}^{n} p_k\right) - q\left(\sum_{k=1}^{n-1} p_k\right)$$

where $q : [0,1] \to [0,1]$ transforms probabilities, and $q(0) = 0 = 1 - q(1)$. If q is strictly increasing, then $\succeq$ respects FOSD.

RDEU is an example of Choquet Expected Utility. See Karni, Maccheroni, and Marinacci (2015) for more on Nonexpected Utility Theory.

Maxmin EU

Ambiguity is the idea that individuals may be uncertain about which probability distribution they face. In the Ellsberg example 1/3 of the balls are red. Thus up to 2/3 of the balls could be blue, and up to 2/3 could be green. The possible probability distribution that could arise are

$$P = \{(1/3, q, 2/3 - q) : 0 \leq q \leq 2/3\}.$$

If one were a worst-case bettor, the bet on red with \$99 stakes is \$33 for a bet on **red**, 0 for a bet on **blue**, \$33 for **not blue** and \$66 for **non red**. Thus **red** beats **blue** and **not red** beats **not blue**. This is a kind of ambiguity aversion.

This too is a kind of Choquet Expected Utility.

Capacities

Choquet Expected Utility is expected utility where the expectation is taken with respect to a non-additive probability, also called a capacity.

Suppose S is finite and $\mathcal{S}$ is the collection of all subsets of S.

A function $\mu : \mathcal{S} \to [1, 0]$ is a capacity if

1. $\mu\{\varnothing\} = 0$,
2. $\mu\{\Omega\} = 1$,
3. for all $A \subset B \in \mathcal{S}$, $\mu\{B\} \geq \mu\{A\}$.

Suppose beliefs are not additive. e.g. state weights are

$$\mu\{red\} = 1/3, \quad \mu\{blue\} = \mu\{green\} = 1/4 \quad \mu\{blue \cup green\} = 2/3.$$

Then μ-weighted expected utility generates the given preferences. Notice

$$\mu\{blue \cup green\} > \mu\{blue\} + \mu\{green\}$$

How to integrate with non-additive beliefs?

$$\int 1 d\mu = \mu\{S\} \cdot 1 = \mu(A) \cdot 1 + \mu(A^c) \cdot 1$$

but perhaps $\mu\{S\} \neq \mu\{A\} + \mu\{A^c\}$.

Choquet Expected Utility for Anscombe-Aumann Acts

A.1. $\succeq$ is complete and transitive.

A.2. An Archimedean axiom.

A.3. The independence axioms for all acts f, g, h which are comonotonic.

A.4. There are $f, g \in H$ such that $f \succ g$.

A.5. If for all $s \in S$, $f(s) \succ g(s)$, then $f \succ g$.

Theorem. If $\succeq$ on H satisfies A.1–5, then there is a function $u : X \to \mathbb{R}$ and a capacity μ on S such that

$$f \succ g \text{ iff } \int \sum_{x} u(x) f(s)(x)\, d\mu > \int \sum_{x} u(x) g(s)(x)\, d\mu$$

Ambiguity Aversion

A capacity μ is convex if for all $A, B \in \mathcal{S}$,

$$\mu\{A \cup B\} - \mu\{A\} \geq \mu\{B\} - \mu\{A \cap B\}.$$

The core of a capacity μ is $C(\mu) = \{\rho \in P : \rho(A) \geq \mu(A)\}$.

Lemma. Every convex capacity has a core.

Example. If $S = \{0, 1\}$ and $\mu\{0\} = \mu\{1\} = 0.3$, then

$$C(\mu) = \left\{\rho : 0.3 \leq \rho\{0\} \leq 0.7\right\}.$$

Fact: If P is a convex set of probability distributions, then

$$\mu(A) = \inf_{\rho \in P} \rho(A)$$

is a capacity.

Definition. Let $\succeq$ be binary relation on H . Then, $\succeq$ is said to be **uncertainty-averse** if $f, g \succeq h$ and $\alpha \in [0,1]$ implies $\alpha f + (1-\alpha) g \succeq h$.

Theorem. Suppose that $\succeq$ satisfies axioms A.1–5. Then the following are equivalent.

1. $\succeq$ is uncertainty-averse,
2. μ is convex,
3. $\int f \, d\mu = \inf_{\rho \in C(\mu)} \int f \, d\rho$.

So this is a characterization of **maxmin expected utility**. Gilboa and Schmeidler (1989) explains how to get maxmin EU from preferences.

Bibliography

Allais, M. 1953). “Le comportement de l’homme rationnel devant le risque: critique des postulats et axiomes de l’école Américaine”. *Econometrica* 21 (4): 503–546.

Anscombe, Francis J. and Robert J. Aumann. 1963. “A definition of subjective probability” *Annals of Mathematical Statistics* 34(1): 199–205.

Carnap, Rudolph. 1954 [1958]. *Introduction to Symbolic Logic.* Mineola NY: Dover Publications.

de Finetti, Bruno. 1937/1980. “La Prévision: Ses Lois Logiques, Ses Sources Subjectives.” *Annales de l’Institut Henri Poincaré* 7: 1–68; translated as “Foresight. Its Logical Laws, Its Subjective Sources” in *Studies in Subjective Probability.*. H. E. Kyburg, Jr. and H. E. Smokler (eds.), Robert E. Krieger Publishing Company, 1980, 55–118.

de Moivre, Abraham. 1718/1967. *The Doctrine of Chances: or, A Method of Calculating the Probability of Events in Play.* London: W. Pearson, 1718; reprinted 1967, New York, NY: Chelsea.

Dreze, Jacques. 1961. “Les Fondements Logiques de l’Utilite Cardinale et de la Probabilité Subjective.” in *La Décision.* Paris: Colloques Internaux du CNRS.

Dutka, Jacques. 1988. “On the St. Petersburg Paradox.” *Archive for History of Exact Sciences* 39(1): 13–39.

Ellsberg, Daniel. 1961. “Risk, Ambiguity, and the Savage Axioms.” *Quarterly Journal of Economics* 75(4): 643–69.

Fishburn, Peter. 1970. *Utility Theory for Decision Making.* New York: John Wiley and Sons.

Fishburn, Peter. 1982. *The Foundations of Expected Utility.* Dordrecht: D. Reidel.

Gilboa, Itzak and David Schmeidler. 1989. “Maximin expected utility with a non-unique prior”. *Journal of Mathematical Economics* 18(2): 141–53.

Gigerenzer, Gerd, Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty and Lorenz Krüger. 1989. *The Empire of Chance.* Cambridge: Cambridge University Press.

Gilboa, Itzhak and David Schmeidler. 1989. "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics* 18(2): 141–53.

Herstein, Israel N. and John Milnor. 1953. "An axiomatic approach to measurable utility" *Econometrica* 21(2): 291–297.

Karni, Edi, Fabio Maccheroni, and Massimo Marinacci. 2015. "Ambiguity and Nonexpected Utility". Ch. 17 in Young, H.P and S. Zamir, eds. *Handbook of Game Theory with Economic Applications,* v.4.

Keynes, John M. 1921. *A Treatise on Probability* London: Macmillan and Co.

Kraft,Charles, John W. Pratt and Abraham Seidenberg. 1959. "Intuitive probability on finite sets". *Annals of Mathematical Statistics* 30(2): 408–419.

Mackenzie, Andrew. 2019. "A foundation for probabilistic beliefs with or without atoms". *Theoretical Economics* 14: 709–778.

Marschak, Jacob. 1950. “Rational behavior, uncertain prospects, and measurable utility.” *Econometrica* 18:2: 111–141.

Menger, Karl 1934. “Das Unsicherheits-moment in der Wertlehre”. *Zeitschrift für Nationalokonomie* 5(4):459-85. English translation (1967) by Wolfgang Schoellkopf, “The role of uncertainty in economics”, ch. 16 in *Essays in Honor of Oskar Morgenstern*. ed. Martin Shubik (Princeton, 1967) 1-31.

Mongin, Philippe. 2003. “L’axiomatisation et les théories économiques” *Revue Économique* 54(1): 99–138.

de Montmort, Pierre Rémond. 1713. *Essay d’Analyse sur les Jeux de Hazard.* Paris: Chez Jacques Quillau. (2nd edition.)

von Neumann, John and Oskar Morgenstern. 1947. *Theory of Games and economic Behavior.* Princeton: Princeton University Press. (2nd revised edition.)

Poisson, Simon-Denis. 1837. *Recherches Sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile.* Paris: Bachelier.

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York: John Wiley & Sons. Revised and enlarged edition, Mineola NY: Dover Publications, 1972.

Scott, Dana. 1964. “Measurement structures and linear inequalities.” *Journal of Mathematical Psychology* 1(2):233–247.

Appendix

Proof of the Qualitative Probability Representation Theorem

Here is another Theorem of the Alternative.

Motzkin's Theorem. One and only one of the following two inequality systems has a solution.

$$
\begin{array}{ccc}
Ax \gg 0 & \text{or} & uA + vB = 0 \\
Bx \geq 0 & & u > 0 \\
 & & v \geq 0 \\
\text{(I)} & & \text{(II)}
\end{array}
$$

The proof of Motzkin's theorem is a consequence of Farkas lemma.

To prove the representation theorem, consider the equation system

$$\begin{aligned}
(1_A - 1_B)r &\gg 0 \\
&\vdots \\
(1_C - 1_D)r &\geq 0 \\
&\vdots \\
(1_D - 1_C)r &\geq 0 \\
&\vdots
\end{aligned} \tag{$**$}$$

where the A, B event pairs range over all those where $A \succ B$, and C, D range over all those where $C \sim D$. Any solution r is non-negative since each $\{s\} \succeq \emptyset$. Since $S \succ \emptyset$ w.l.o.g. the r_s sum to 1. Any such solution is a probability representation for $\succeq$.

If $(**)$ does not have a solution, then there are vectors q_{AB} non-negative and not all 0, and $q_{CD}, q_{DC} \geq 0$ such that for all s,

$$\begin{aligned}\sum_{AB} q_{AB}(1_A(s)) + \sum_{CD} q_{CD} 1_C(s) + \sum_{DC} q_{DC} 1_D(s) = \\ \sum_{AB} q_{AB}(1_B(s)) + \sum_{CD} q_{CD} 1_D(s) + \sum_{DC} q_{DC} 1_C(s)\end{aligned} \tag{$\dagger$}$$

The indicator functions take on only the values 0 and 1, so this system has a rational solution (e.g. by Fourier-Motzkin elimination). Therefore, w.l.o.g we can take the q coefficients to be integers.

Enumerate the pairs $(A_1, B_1), \ldots (A_k, B_k)$ and $(C_1, D_1), \ldots, (C_J, D_J)$. Construct two lists as follows:

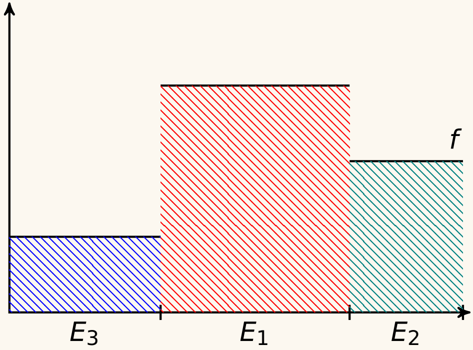
$$\mathcal{A} = \{\underbrace{A_1, \ldots, A_1}_{q_{A_1B_1} \text{ times}} \ldots, \underbrace{C_1, \ldots, C_1}_{q_{C_1D_1} \text{ times}} \ldots, \underbrace{D_1, \ldots, D_1}_{q_{D_1C_1} \text{ times}} \ldots\}$$

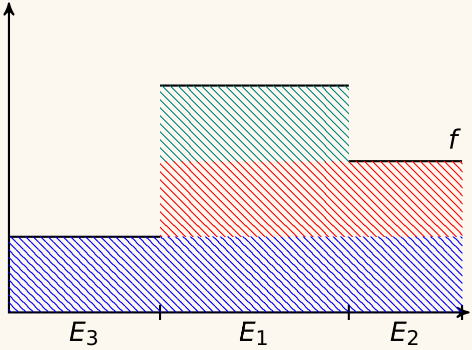
$$\mathcal{B} = \{\underbrace{B_1, \ldots, B_1}_{q_{A_1B_1} \text{ times}} \ldots, \underbrace{D_1, \ldots, D_1}_{q_{C_1D_1} \text{ times}} \ldots, \underbrace{C_1, \ldots, C_1}_{q_{D_1C_1} \text{ times}} \ldots\}.$$

Because q solves the equation system (†), this pair of lists is balanced. And for each set E in the $\mathcal{A}$ list, the corresponding F element in the $\mathcal{B}$ list is ordered $E \succeq F$. The $\mathcal{A}$ and $\mathcal{B}$ together violate finite cancellation. ■

Integrating with Capacities

Two Ways of Computing an Integral





Integrating with Non-Additive Beliefs

Integrating with respect to a capacity does the computation on the right. Suppose S is partitioned into sets $E_1, \ldots, E_n$ such that $f(s) = v_i$ for $s \in E_i$, and suppose $v_1 \leq v_2 \leq \cdots \leq v_n$. Take $v_{n+1} = 0$.

$$\int f\, d\mu = \sum_{i=1}^{n} v_i \big(\mu\{\cup_{j=1}^{i} E_j\} - (\mu\{\cup_{j=1}^{i-1} E_j\}\big)$$

$$= \sum_{i=1}^{n} (v_i - v_{i+1}) \mu\{\cup_{j=1}^{i} E_j\}$$

where $\cup_{j=1}^{0} E_j = \emptyset$.

This is the integral for non-negative step functions. It extends. . .

This integral is linear in beliefs μ. It is **not** linear in f

but ...

f and g are comonotonic if for all $s, t \in S$ if $f(s) > f(t)$ then $g(s) \geq g(t)$.

Theorem. If f and g are comonotonic, then

$$\int \alpha f + (1-\alpha) g \, d\mu = \alpha \int f \, d\mu + (1-\alpha) \int g \, d\mu.$$