Beliefs

October 26, 2006

1 Why Beliefs?

VnM Expected Utility:

There is a state space S (which, for the moment only, is finite) and an *objective probability distribution* p on S. There is also a set O of outcomes. An act is a map $f : S \to O$. A preference order $\succeq$ is a *VnM order* if acts are ranked by expected utility. That is, there is a payoff function $u : O \to \mathbf{R}$ such that

$$f \succeq g \qquad \text{iff} \qquad \sum_s u\big(f(s)\big)p(s) \geq \sum_s u\big(g(s)\big)p(s) \tag{1}$$

Why is this VnM? Where are the probability distributions on outcomes? To each act f we can associate a probability distribution p_f on O, such that $f \succ g$ iff $\sum_o u(o)p_f(o) \geq \sum_o u(o)p_g(o)$. To see this, define $p_f(o)$ to be the probability that outcome o is realized with act f. That is,

$$p_f(o) = \sum_{\{s:f(s)=o\}} p(s)$$

So,

$$\begin{aligned}\sum_s u\big(f(s)\big)p(s) &= \sum_o \sum_{\{s:f(s)=o\}} u\big(f(s)\big)p(s) \\ &= \sum_o \sum_{\{s:f(s)=o\}} u(o)p(s) \\ &= \sum_o u(o)p_f(o)\end{aligned}$$

This is just the familiar change of variables formula from basic stats. Where can probabilities come from?

1. Frequencies
2. Beliefs — expressed by asking
3. Beliefs — expressed through behavior

Empirical frequencies are fine for bets at Las Vegas. But not for lots of things. Savage wrote in 1954 that "I, personally, consider it more probable that a Republican president will be elected in 1996 than that it will snow in Chicago sometime in the month of May, 1994. But even this late spring snow seems to me more probable than that Adolf Hitler is still alive."

1. A Dem was elected in 1996.
2. I couldn't find 1994 data, but Chicago had May snow in both 2001 and in 2002.
3. Hitler?

Savage's idea is to derive beliefs from preferences. (And presumably preferences can be elicited by presenting people with choices.) Eliciting a DM's preferences over various acts — bets on events A and B — will tell us which event the DM thinks is more likely. That is, we will derive an ordering $\rhd$ on events in which $A \rhd B$ means that the DM believes A to be more likely than B. This order is called a *qualitative probability* or a *comparative* *probability*. He goes on to give conditions under which this order can be represented by a probability distribution p, and further conditions under which the preference ordering has a representation like that of equation (1) with respect to p. So "why beliefs"? Because in a well-structured theory of choice under uncertainty, subjective probability can play the same role that objective, which is to say frequentist, probability can play in the VnM choice theory. From preferences we derive a probability distribution p on $\mathcal{S}$ such that with an approprite payoff function $u : O \to \mathbf{R}$ also derived from preferences, the representation (1) holds.

2 Representing Beliefs

This section is just a list of ways that beliefs can be represented:

- qualitative probability
- probability (Kolmogorov)
  - Objective — frequentist
  - Subjective
- plausibility measure
- possibility measure
- belief functions
- capacities
- sets of beliefs
- conditional probability systems

There are reasons for eliciting belief that are distinct from decision theory, and the connection of some of these representations to decision theory is not known by me.

3 Qualitative Probability Defined

Subjective probability attempts to make precise the connection between *coherent views of uncertainty* and quantitative (Kolmogorov) probability. It accommodates the following views:

- Classical: Bayes, Laplace
- Intuitive: B.O. Koopman, I. J. Good
- Decision: Ramsey, De Finetti, Savage

Other writers view qualitative probability as an alternative to traditional probability, possibly weaker. This group includes Keynes and our own T. Fine. The notes should develop more on this.

We are given a set S of states, and an *algebra* of events $\mathcal{S}$. What does this mean?

1. $S \in \mathcal{S}$
2. $\emptyset \in \mathcal{S}$
3. if A and B are in $\mathcal{S}$, so is $A \cap B$.
4. if $A \in \mathcal{S}$, then $A^c \in \mathcal{S}$.

Why are these assumptions about what constitutes *observable events* natural?

Definition 1. *A* qualitative probability structure *is a triple* $(S, \mathcal{S}, \triangleright)$ *such that*

1. $\triangleright$ *is a preference order;*
2. $S \triangleright \emptyset$ *and for all* $A \in \mathcal{S}$, $A \trianglerighteq \emptyset$*;*
3. *if* A *is disjoint from both* B *and* C*, then* $B \triangleright C$ *iff* $A \cup B \triangleright A \cup C$*.*

4 Derivation of Qualitative Probabilities

A decision problem is described by

States S

Events $\mathcal{S}$

Outcomes O

Acts $f : S \to O$. Suppose that acts are *simple*, that the range of f is finite. The set of all acts is A.

Preferences A binary relation $\succ$ on A.

Savage introduces the following axioms, which will be sufficient to generate a qualitative probability.

Axiom 1. $\succ$ *is a preference relation.*

Without this axiom there is nothing to do.

Identify each outcome $o \in O$ with the *constant act* which pays out o regardless of the state. The preference order $\succ$ on A induces a preference order $\succ$ on O through its ordering of the constant acts. We use the same symbol for both preference orders because it really is the same order on both sets.

Axiom 2. *There are x and y in O such that $x \succ y$.*

Without this axiom the theory would be trivial.

Now we introduce some new notation. Given acts f and g and event A, define the act

$$f_A g(s) \begin{cases} f(s) & \text{if } s \in A, \\ g(s) & \text{otherwise.} \end{cases}$$

This act pays off according to f if $s \in A$ and according to g if $s \notin A$. Extending this idea, define for disjoint events $A_1, \ldots, A_{n-1}$ we will define the act $f^1_{A_1} \cdots f^{n-1}_{A_{n-1}} f^n$ to be the act which gives outcome f_k when $s \in A_k$, and f^n when $s \in (A_1 \cup \cdots \cup A_{n-1})^c$.

Axiom 3. *For all acts f, g and h $f_A h \succ g_A h$ iff for all k, $f_A k \succ g_A k$.*

This axiom states that if f and g differ only on A, then the comparison between them should depend only upon how they behave on A. So if we vary f and g in any way off of A, then so long as we vary them identically, the ranking between them should not change. This axiom seems very appealing, but we shall see that it can be violated in practice.

This axiom gives us a plausible way of defining preferences conditional on some event. Given f and g, how do I feel about them if I am told that event A will happen? To answer this question, modify f and g so that they behave identically off of A, and see how I feel about the modified acts. Axiom 3 states that the results of this comparison will be independent of the modification. This allows the derivation of a *conditional preference* given an event A.

Definition 2. *$f \succ_A g$ iff there is an act h such that $f_A h \succ g_A h$.*

Proposition 1. *$\succ_A$ is a preference order.*

Exercise 1. *Prove this.*

Some events simply do not matter. We can express this in several ways. For instance, if f and g are any acts, how I feel about them is determined by their behavior off of A. That is, $f \succ g$ iff $f \succ_{A^c} g$. An equivalent formulation of this idea is given in the following definition:

Definition 3. *An event A is* null *if for all f and g, $f \sim_A g$.*

The following properties of null events are easy to prove.

Proposition 2. *The following statements are true of null events:*

1. *S is not null,*
2. *$\emptyset$ is null,*
3. *if A is null and $B \subset A$, then B is null,*
4. *if A and B are disjoint, and null, then $A \cup B$ is null.*

Proof. 1. Axiom 2. Otherwise $x \sim y$ for all pairs $x, y \in O$.

2. $f_\emptyset h = h$ for all h. The claim now follows from the reflexivity of $\succeq$.

3. For any acts f and h, $f_B h = f_B h_{A/B} h$. So for any acts f, g and h, $f_B h = f_B h_{A/B} h \sim g_B h_{A/B} h = g_B h$. Thus $f \sim_B g$.

□

Exercise 2. *(a) Prove part 4. (b) Using (a) if necessary, prove that the conclusion of part 4 holds for all null events, and not just disjoint null events.*

The next proposition is known as the *sure thing principle.* Its content is intuitive: If I prefer f to g given A, and if I prefer f to g given A^c, then unconditionally I should prefer f to g.

Proposition 3. *If $B_1, \ldots, B_n$ is a partition of S by elements of $\mathcal{S}$ and $f \succeq_{B_i} g$ for all i, then $f \succeq g$. If in addition, for any one such i, $f \succ B_i$, then $f \succ g$.*

Proof. For the case of $n = 2$ and weak inequalities, $f = f_{B_1} f \succeq g_{B_1} f = f_{B_2} g \succeq g_{B_2} g = g$, and $f \succ g$ result follows from transitivity of $\succeq$. Replacing any weak preference by strict preference gives strict preference in the conclusion. □

Exercise 3. *The notes prove the sure thing principle for partitions of size $n = 2$. Complete the proof of the sure thing principle.*

The next axiom says that preferences over outcomes do not depend upon the state. Conditional preferences given any event over pure outcomes are indentical to unconditional preferences.

Axiom 4. *If A is not null, then $x \succ_A y$ iff $x \succ y$.*

A qualitative probability will be constructed by examining "bets" on events: If event A happens, the decisionmaker wins and gets a winning prize. If A does not happen, the decisionmaker loses and gets a losing prize (worse than the winning prize). It makes sense to interpret a preference for a bet on A over the same bet on B as a claim that A is more likely than B. In order for this to make work, preferences on bets must be independent of the particular winning and losing prizes. This is the content of the next axiom.

Axiom 5. *If $x \succ y$ and $w \succ z$, then for any pair of events A and B, $x_A y \succ x_B y$ iff $w_A z \succ w_B z$.*

Concretely, if a decisionmaker prefers a dollar bet on the event that Cornell will win the ECAC Hockey title this year to a dollar bet on the event that Cornell will finish last, then the DM should prefer $10, $100, etc., bets on Cornell winning to the same dollar bet that Cornell will finish last. If $x \succ y$, think of the act $x_A y$ as a *bet on A.*

Axiom 5 makes possible the following definition of a likelihood order on events:

Definition 4. *For all A and B in $\mathcal{S}$, $A \rhd B$ iff there are outcomes $x \succ y$ in O such that $x_A y \succ x_B y$.*

This likelihood ordering of acts is a qualitative probability on $\mathcal{S}$. Thus preferences encode (qualitative) beliefs.

Proposition 4. *$A \trianglerighteq B$ iff there is an $x \succ y$ such that $x_A y \succeq x_B y$.*

Proof. Suppose not $B \rhd A$. Then for all $x \succ y$, not $x_B y \succ x_A y$, and so for all such x, y pairs, $x_A y \succeq x_B y$.

Suppose $x_A y \succeq x_B y$. Then not $x_B y \succ x_A y$, and so Axiom 5 implies that for all $w \succ z$, not $w_B z \succ w_A z$. Consequently, not $B \rhd A$, and so $A \trianglerighteq B$. □

Theorem 1. *$\rhd$ is a qualitative probability order on $\mathcal{S}$.*

Proof. 1. $\rhd$ is asymmetric. If $A \rhd B$, then $x_A y \succ x_B y$, and Axiom 5 implies that for all $w \succ z$, $w_A z \succ w_B z$. From Axiom 1 it follows that for no $w \succ z$ is $w_B z \succ w_A z$. For negative transitivity, suppose not $A \rhd B$ and not $B \rhd C$. Then for all $x \succ y$, not $x_A y \succ x_B y$ and not $x_B y \succ x_C y$. Axiom 1 implies that for all $x \succ y$, not $x_A y \succ x_C y$. Therefore not $A \rhd C$.

2. We have already shown that, as a consequence of Axiom 3, $S \rhd \emptyset$. Choose $x \succ y$. First, suppose that A is null. Then $x_A y \equiv y = x_\emptyset y$, so according to Proposition 4, $A \unrhd \emptyset$. If A is not null, then for any $x \succ y$, $x \succ_A y$ according to Axioms 2 and 4. Thus $x_A y \succ y_A y = y_\emptyset y = x_\emptyset y$, so $A \rhd \emptyset$.

3. Suppose C is disjoint from A and B, and $A \rhd B$. Choose $x \succ y$. Then $x_A y \succ x_B y$. Since $C \subset (A \cup B)^c$, and since both acts take on the value y, on $(A \cup B)^c$, Axiom 3 says that if both acts are modified identically on C, then their ranking remains unchanged. Thus

$$x_{A \cup C} y = x_A x_C y \succ x_B x_C y = x_{B \cup C} y \, .$$

In the other direction, suppose $A \cup C \rhd B \cup C$. The same argument shows that $A \rhd B$.

□

Null events can be interpreted with the qualitative probability. We have already shown that if A is null, then $A \equiv \emptyset$. The other direction is true as well.

Proposition 5. *For any event $A \in \mathcal{S}$, $A \equiv \emptyset$ iff A is null.*

Proof. We need only prove that if $A \equiv \emptyset$, then A is null. If $A \equiv \emptyset$, then for any $x \succ y$, $x_A y \sim y = y$. From Axiom 4 infer that A is null. □

Recall the steps in the Savage program:

1. Derive beliefs from preferences.

2. Represent the beliefs by a probability distribution p.

3. Show that preferences depend only on distributions of outcomes under p. Specifically, if the two distributions on outcomes p_f and p_g are equal, then $f \sim g$. This means that preferences on acts generate preferences on probability distributions of outcomes.

4. Show that the preferences on probability distributions of outcomes are vNM.

We cannot do step 2 without more assumptions. Savage makes additional assumptions on preferences that essentially guarantee that for any n one can divide S up into n equally likely sets. In the notes on representations I have proved Suppe's Theorem:

Theorem 2 (Suppes Theorem). *Suppose $(X, \mathcal{S}, \succ)$ is a finite qualitative probability structure such that if $A \rhd B$, there is a $C \in \mathcal{S}$ such that $A \equiv B \cup C$. Then $\succ$ has a probability representation that puts equal weight on the smallest elements (atoms) of $\mathcal{S}$.*

The assumption on the qualitative probability can be rephrased directly in terms of preferences on bets:

Axiom 6. *For every $x \succ y$ and A, B such that $x_A y \succ x_B y$, there is an event C such that $x_A y \sim x_{B \cup C} y$.*

With this assumption, Suppe's Theorem completes part two of the Savage program. S is finite and $\rhd$ is represented by the probability distribution which puts equal weight on all points of S. (There is no loss in assuming that the atoms of $\mathcal{S}$ are the singleton sets.)

Theorem 3. *If f and g are two acts such that $p_f = p_g$, then $f \sim g$.*

Proof. Prove this by induction on the number of outcomes n that f and g take on. If $n = 1$ the acts are constant acts, and so if $p_f = p_g$ the acts are identical and the conclusion follows from the asymmetry of $\succ$.

For $n = 2$ the acts are of the form $f = x_A y$ and $g = x_B y$. The distributions p_f and p_g are equal iff $A \equiv B$, which is true iff $f \sim g$ by definition.

Suppose the conclusion is true for all functions taking on no more than $n - 1 \geq 2$ values. Suppose f and g take on the values $x^1, \ldots, x^n$, and that $p_f = p_g$. For all $k \leq n$, $\#g^{-1}(x^k) = \#h^{-1}(x^k)$, since p is uniform on S.

Choose any two states s' and s''. Define f' such that

$$f'(s) = \begin{cases} f(s'') & \text{if } s = s', \\ f(s') & \text{if } s = s'', \\ f(s) & \text{otherwise.} \end{cases}$$

The function f' is constructed from f by permuting its values at states s' and s''. First we show that $f' \sim f$. Suppose without loss of generality that $f(s'), f(s'') \neq x^n$, and let $A = \{s : f(s) \neq x^n\}$. Let $g = f_A x^1$ and $h = f'_A x^1$, so that both functions take on only the values $x^1, \ldots, x^{n-1}$. For all $k \leq n-1$, $\#g^{-1}(x^k) = \#h^{-1}(x^k)$, and all states are equally likely, so their distributions under p are identical. The induction hypothesis implies that $g \sim h$. Axiom 3 implies that $f \sim f'$ since f and f' agree on A^c, g agrees with f on A and h agrees with f' on A. By a finite sequence of such pairwise permutations, f can be transformed into g. Consequently Axiom 1 implies that $f \sim g$. □

5 Savage's Approach

Savage assumes the following:

Axiom 7. *If $f \succ g$ and $x \in O$, then there is a partition $B_1, \ldots, B_n$ of S such that $f_{B_i} x \succ g$ and $f \succ g_{B_i} x$ for all i.*

This axiom can only hold if S is not finite. It implies a similar statement about $\rhd$.

Proposition 6. *If Axiom 6 is satisfied and if $A \rhd B$, then there is a partition $C_1, \ldots, C_n$ of S such that for all i, $A \rhd B \cup C_i$.*

Exercise 4. *Prove this.*

Savage goes on to show the following

Theorem 4. *If a qualitative probability $\rhd$ satisfies the conclusion of Proposition 6, then there is a unique probability distribution p on $(S, \mathcal{S})$ which represents $\rhd$. Moreover, for any $0 \leq \alpha \leq 1$ and $B \in \mathcal{S}$ there is a $C \in \mathcal{S}$ such that $p(C) = \alpha p(B)$.*

6 Savage and Utility

In this section we assume that O is finite, and that Axioms 1 through 7 hold. Then $\rhd$ has a representation p. We also assume without proof the conclusion of Theorem 3; that if f and g are two acts such that $p_f = p_g$, then $f \equiv g$. Now how do we get to utility?

From Theorem 3 it is clear that $\succ$ on acts induces an order $\succ$ on probability distributions: $\mu \succ \nu$ iff there are acts f and g such that $\mu = p_f$, $\nu = p_g$, and $f \succ g$. The ordering on probability distributions is a preference order. We will also see that it satisfies the independence axiom and the Archimdean axioms of vNM. This guarantees the existence of a utility function $u : O \to \mathbf{R}$ such that $p_f \succ p_g$ iff $\sum_o u(o)p_f(o) > \sum_o u(o)p_g(o)$. Changing variables, $\int u\big(f(s)\big)dp(s) > \int u\big(g(s)\big)dp(s)$.

Theorem 5 (Independence). *If p_f, p_g and p_h are distributions induced by acts f, g and h respectively, and if $0 < alpha \leq 1$, then $\alpha p_f + (1-\alpha)p_h \succ \alpha p_g + (1-\alpha)p_h$ iff $p_f \succ p_g$.*

Proof. Let f_i index the values of f, etc. Let $B_i = f^{-1}(f_i)$, and $C_i = g^{-1}(g_i)$. Construct $D_{ij} \subset B_i \cap C_j$ such that $p(D_{ij}) = \alpha p(B_i \cap C_j)$, and let $D = \cup_{i,j} D_{ij}$. Then $p(D) = \alpha$, $p(B_i|D) = p(B_i)$ and $p(C_j|D) = p(C_j)$. The theorem says that $f \succ_D g$ iff $f \succ g$. (Rebuild h on D^c.) Theorem 3 says that the validity of this statement does not depend on the particular choice of D so long as the required distributional constraints are met. So we will say that $f \succ_\alpha g$ iff $f \succ_D g$ for some D constructed as above.

If $f \sim_\alpha g$ for all $0 < alpha \leq 1$ there is nothing to prove. Suppose that $f \succ_\alpha g$ for some α_0.

1. $f \succ_\alpha g$ iff $f \succ_{\alpha/2} g$. This follows from the sure thing principle. Partition D into two sets $D^1 \equiv D^2$ with corresponding D^k_{ij}'s so that the conditional distribtution of f given D_1 equals that given D_2, and similarly for g.

2. If $\alpha + \beta \leq 1$, and $f \succ g$ given both α and β, then $f \succ g$ given $\alpha + \beta$. Again, the sure thing principle.

3. Choose any α, and choose β such that $\alpha = \beta\alpha_0$. For all n $\beta \approx k/2^n$. Finally, by Axiom 6, if we modify f and g on a sufficiently small set the ranking does not change. So $f \succ g$ given any α.

$\square$