Savages' Subjective Expected Utility Model

Edi Karni

Johns Hopkins University

November 9, 2005

Savage's subjective expected utility model. In his seminal book, *The Foundations of Statistics,* Savage (1954) advanced a theory of decision making under uncertainty and used that theory to define choice-based subjective probabilities. He intended these probabilities to express the decision maker's beliefs, thereby furnishing Bayesian statistics with its behavioral foundations.

The interpretation of probability as a numerical expression of beliefs is as old as the idea of probability itself. According to Hacking (1984), the notion of probability emerged in the 1650s with a dual meaning; the relative frequency of a random outcome in repeated trials and a measure of a decision maker's degree of belief in the truth of propositions, or the likely realization of events. Both the "objective" and the "subjective" probabilities, as these interpretations are now called, played important roles in the developments that lead to the formulation of Savage's subjective utility model.

In the early stages of their respective evolutions, the notion of utility was predicated on the existence of objective probabilities, and the notion of subjective probabilities presumed the existence of some form of utility. The ideas of utility and expected utility-maximizing behavior were originally introduced by Bernoulli (1738). Bernoulli's preoccupation with resolving the famous St. Petersburg paradox justifies his taking for granted the existence of probabilities in the sense of relative frequencies. In the same vein, von Neumann and Morgenstern's (1944) axiomatic characterization of expected utility-maximizing players facing opponents who may employ a randomizing device to determine the choice of a pure strategy assumes that probabilities of these strategies are relative frequencies.

In the late 1920s and 1930s, Ramsey and de Finetti formalized the concept of choice-based subjective probability assuming that individual seek to maximize expected utility when betting on the truth of propositions. In the behaviorist tradition, they explored the possibility of inferring the degree of confidence a decision maker has in the truth of a proposition from his betting behavior and quantifying the degree of confidence, or belief, by probability. Invoking the axiomatic approach and taking the existence of utilities as given, Ramsey (1931) sketched a proof of the existence of subjective probabilities. de Finetti (1937) proposed a definition of subjective probabilities assuming linear utility and no arbitrage opportunities.

These developments culminated in the work of Savage. While synthesizing the ideas of de Finetti and von Neumann and Morgenstern, Savage introduced a new analytical framework and conditions that are necessary and sufficient for the existence and joint uniqueness of utility and probability, and the characterization of individual choice as expected utility-maximizing behavior.

Savage's analytical framework: Decision making under uncertainty pertains to situations in which a choice of a course of action, by itself, does not determine a unique outcome. To formalize this notion Savage (1954) introduced an analytical framework consisting of a set S, whose elements are *states of the world* (or states, for brevity); an arbitrary set C, of *consequences*; and the set F, of *acts* (that is, functions from the set of states to the set of consequences). Acts correspond to courses of action, consequences describe anything that may happen to a person, and states are possible resolutions of uncertainty, that is, "a description of the world so complete that, if true and known, the consequences of every action would be known" (Arrow [1971], p. 45). Implicit in this definition is the notion that there is a unique true state. Events are subsets of the set of states. An event is said to obtain if it includes the true state.

A decision maker is characterized by a preference relation, $\succcurlyeq$, on F. The statement $f \succcurlyeq f'$ has the interpretation "the course of action f is at least as desirable as the course of action f'." Given $\succcurlyeq$, the strict preference relation $\succ$ and the indifference relation $\sim$ are defined as follows: $f \succ f'$ if $f \succcurlyeq f'$ and not $f' \succcurlyeq f$; $f \sim f'$ if $f \succcurlyeq f'$ and $f' \succcurlyeq f$.

The preference structure: The evaluation of a course of action in the face of uncertainty involves the decision maker's taste for the possible consequences and his beliefs regarding their likely realization. Savages' subjective expected utility theory postulates a preference structure, depicted axiomatically, permitting the numerical expression of the decision maker's valuation of the consequences by a utility function, that of his beliefs by a (subjective) probability measure on the set of all events, and the evaluation of acts by the mathematical expectations of the utility with respect to the subjective probability.

To state Savage's postulates, I employ the following notation and definitions. Given an event E and acts f and h, let $f_E h$ be the act such that $(f_E h)(s) = f(s)$ if $s \in E$, and $(f_E h)(s) = h(s)$ otherwise. An event E is *null* if $f_E h \sim f'_E h$ for all acts f and f', otherwise it is *non-null*. A constant act is an act that assigns the same consequence to all the states. To simplify the exposition, I denote the constant acts by their values (that is, if $f(s) = x$ for all s, I denote the act f by x).

The first postulate asserts that the preference relation is transitive and that all acts are comparable.

P.1 A preference relation is a transitive and complete binary relation on F.

The second postulate, also known as the Sure Thing Principle, requires that the preference between acts depend solely on the consequences in states in which the payoffs of the two acts being compared are distinct. This implies that the valuation of the consequences of an act in one event is independent of the payoffs of the same act in the complementary event.

P.2 For all acts, f, f', h, h' and every event E, $f_E h \succcurlyeq f'_E h$ if and only if $f_E h' \succcurlyeq f'_E h'$.

The sure thing principle makes it possible to define conditional preferences as follows: For every event E, $f \succcurlyeq_E f'$ if $f \succcurlyeq f'$ and for every s not in E, $f(s) = f'(s)$.

The third postulate asserts that the ordinal ranking of consequences is independent of the event and the act that yield them.

P.3 For every non-null event E and all constant acts, x and y, $x \succcurlyeq y$ if and only if $x_E f \succcurlyeq y_E f$ for every act f.

In view of P.3, it is natural to refer to an act that assigns to an event E a consequence that ranks higher than the consequence it assigns to the complement of E as a *bet* on E. Ramsey (1931) was the first to suggest that a decision maker's belief that an event E is at least as likely to obtain as another event E' should manifest itself through preference for a bet on E over the same bet on E'. The fourth postulate which requires that the betting preferences be independent of the specific consequences that define the bets, formalizes this idea.

P.4 For all events E and E' and constant acts x, y, x' and y' such that $x \succ y$ and $x' \succ y'$, $x_E y \succcurlyeq x_{E'} y$ if and only if $x'_E y' \succcurlyeq x'_{E'} y'$.

Postulates P.1-P.4 imply the existence of a transitive and complete relation on the set of events that has the interpretation "at lease as likely to obtain as" representing the decision maker's beliefs as qualitative probabilities. Together with P.3 it also implies that the decision maker's risk attitudes are event independent.

The fifth postulate renders the decision making problem and the qualitative probabilities nontrivial by ruling out that the decision maker is indifferent among all acts.

P.5 For some constant acts x and x', $x \succ x'$.

The sixth postulate introduces a form of continuity of the preference relation. It asserts that no consequence is either infinitely better or infinitely worse than any other consequence. Put differently, the next postulate requires that there be no consequence that, if it were to replace the payoff of an act on a non-null even, no matter how unlikely, will reverse a strict preference ordering of two acts.

P.6 For all acts $f, g,$ and h satisfying $f \succ g$, there is a finite partition $(E_i)_{i=1}^n$ of the set of states such that, for all i, $f \succ f_{E_i} h$ and $h_{E_i} f \succ g$.

A probability measure is nonatomic if every non-null event may be partitioned into two non-null subevents. Formally, π is a nonatomic probability measure on the set of states if for every event E and number $0 < \alpha < 1$, there is an event $E' \subset E$ such that $\pi(E') = \alpha\pi(E)$. Postulate P.6 implies that there are infinitely many states of the world and that if there exists a probability measure representing the decision maker's beliefs, it must be nonatomic. Moreover, the probability measure is defined on the set of all events, hence it is finitely additive (that is, for every event E, $0 \leq \pi(E) \leq 1$, $\pi(S) = 1$ and for any two disjoint events, E and E', $\pi(E \cup E') = \pi(E) + \pi(E')$).

The seventh postulate is a monotonicity requirement asserting that if the decision maker considers an act strictly better (worse) than each of the payoffs of another act on a given non-null event, then the former act is conditionally strictly preferred (less preferred) than the latter.

P.7 For every event E and all acts f and f', if $f \succ_E f'(s)$ for all s in E then $f \succcurlyeq_E f'$ and if $f'(s) \succ_E f$ for all s in E then $f' \succcurlyeq_E f$.

Representation: Savage's theorem establishes an equivalence between a preference relation having the properties described by the seven postulates and a preference relation induced by the maximization of the expectations of a utility function on the set of consequences with respect to a probability measure on the set of all events. The utility function is unique up to a positive affine transformation and the probability measure is unique.

Savage's theorem: Let $\succcurlyeq$ be a preference relation on F. Then the following two conditions are equivalent:

(i) $\succcurlyeq$ satisfies postulates P.1-P.7.

(ii) There exists a unique, nonatomic, finitely additive, probability measure π on S such that $\pi(E) = 0$ if and only if E is null, and a bounded, unique up to a positive affine transformation, real-valued function u on C such that, for all acts f and g, $f \succcurlyeq f'$ if and only if $\int_S u(f(s))\, d\pi(s) \geq \int_S u(f'(s))\, d\pi(s)$.

Interpretation and criticism: In Savage's theory, consequences are assigned utilities that are independent of the underlying state of the world, and events are assigned probabilities that are independent of acts. These assignments, however, are not implied by the postulates. This observation merits elaboration.

The structure of the preference relation, in particular postulates P.3 and P.4, implies that the preference relation is state independent. In other words, the ordinal rankings of both consequences and bets are independent of the underlying events. This implies event-independent risk attitudes but does not, by itself, rule out that the states affect the decision maker's well-being, or that the utility of the consequences is state dependent. Put differently, Savage's model implies state-independent preferences but not a state-independent utility function. The utility and probability that figure in the representation of the preferences in Savage's theorem are unique as a pair, that is, the probability is unique given the utility and the utility is unique (up to a positive affine transformation) given the probability. It is possible, therefore, to define new probability measures and state-dependent utility functions – and thereby to obtain a new subjective expected utility representation

– without violating any of Savage's postulates. For instance, let γ be a bounded, positive, nonconstant, real-valued function on S, and let $\Gamma = \int_S \gamma(s)\, d\pi(s)$. For every event E, define $\hat{\pi}(E) = \int_E \gamma(s)\, d\pi(s) / \Gamma$ and let $\hat{u}(x, s) = \Gamma u(x) / \gamma(s)$ for all s in S and x in C. Then, for every act f, $\int_S u(f(s))\, d\pi(s) = \int_S \hat{u}(f(s), s)\, d\hat{\pi}(s)$. Because γ is arbitrary and nonconstant, $\pi \neq \hat{\pi}$. This shows that the uniqueness of the probability in Savage's theory is predicated on the convention that the utility function is state independent (that is, constant acts are constant utility acts). This convention is not implied by the postulates, it has no choice manifestation, and its validity is not subject to refutation in the context of Savage's analytical framework. Moreover, the employment of this convention renders the definition of probability in Savage's model arbitrary and the claim that it represents the decision maker's beliefs scientifically untenable. That said, it is noteworthy that, insofar as the theory of decision making under uncertainty is concerned, because all the representations obtained using the procedure outlined above are equivalent, the failure to correctly quantify the decision maker's beliefs is not critical. Insofar as providing choice-based foundations of Bayesian statistics is concerned, however, this failure is fatal.

A somewhat related aspect of Savage's model that is similarly unsatisfactory concerns the interpretation of null events. Ideally, an event should be designated as null and be ascribed zero probability if and only if the decision maker believes it to be impossible. In Savage's model an event is null if the decision maker displays indifference among all acts that agree on the payoff on the complement of the said event. However, this definition does not make a distinction between an event that the decision maker perceives as impossible and one whose possible outcomes he perceives as equally desirable. It is possible, therefore, that events that the decision maker believes possible, or even likely, are defined as null and assigned zero probability. Consider this example, a passenger who is indifferent to the size of his estate in the event that he dies is about to board a flight. For such a passenger, a plane crash is a null event and is assigned zero probability, even though he may believe that the plane could crash. This problem renders the representation of beliefs by subjective probabilities dependent on the implicit and unverifiable assumption that in every event some outcomes are strictly more desirable than others. If this assumption is not warranted, the procedure may result in a misrepresentations of beliefs.

The requirement that the preferences be state independent imposes significant limitations on the range of applications of Savage's theory. Choosing a disability insurance policy, for example, is an act whose consequences - the indemnities - depend on the realization of the decision maker's state of health. In addition to affecting the decision maker's well-being, it is conceivable that alternative states of disability influence his risk attitudes. Disability may also alter the decision maker's ordinal ranking of the consequences, which is a violation of P.3; For instance, a leg injury may reverse a decision maker's preferences between going hiking and attending a concert. Similar observations apply to the choice of life and health insurance policies.

Savage presented his seven postulates as principles that a rational individual ought to follow rather than an hypothesis describing how individuals actually choose among courses of action in the face of uncertainty. Indeed, almost from the moment of it inception, the descriptive validity of Savage's model - in particular, the sure thing principle, which is responsible for the specific functional form of the representation and the separability and linearity in the probabilities - has been questioned. It has repeatedly been shown, in experimental settings that the theory fails systematically to predict subjects choice. The most severe and remarkable criticism in this regard is due to Elsberg (1961), who demonstrated using simple thought experiments that individuals display choice patterns that are inconsistent with the existence of beliefs representable by a probability measure.

BIBLIOGRAPHY

Arrow, Kenneth J. 1971. *Essays in the Theory of Risk Bearing.* Chicago: Markham Publishing Co.

Bernoulli, Daniel 1738. "Specimen Theoriae Novae de Mensura Sortis," *Commentarii Academiae Scientiatatum Imperalas Petropolitanae*, 5; 175 - 192. (Translated as "Exposition of a New Theory on the Measurement of Risk." *Econometrica* 22 (1954): 23 - 26.)

de Finetti, Bruno. 1937. "La prévision: Ses lois logiques, ses sources subjectives," *Annals de l'Institute Henri Poincare,* Vol. 7, 1-68. (English translation, by H. E. Kyburg, appears in H. E. Kyburg and H. E. Smokler (eds.) (1964) *Studies in Subjective Probabilities.* New York. John Wiley and Sons.)

Ellsberg, Daniel 1961. "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75: 643-659.

Hacking, Ian. *The Emergence of Probabilities.* Cambridge University Press 1984.

Ramsey, Frank P. 1931. "Truth and Probability," In *The Foundations of Mathematics and Other Logical Essays.* R. B. Braithwaite and F. Plumpton (Eds.) London: K. Paul, Trench, Truber and Co.

Savage, L.J. 1954. *The Foundations of Statistics.* New York: John Wiley.

von Neumann John, and Oscar Morgenstern. 1944. *Theory Games and Economic Behavior* Princeton, NJ: Princeton University Press.