Economics 696, Lecture Note 3: Choice under Uncertainty

1 Introduction

Recall that in the previous lecture note we developed the result that the class of admissible decision rules more or less coincides with the class of Bayes rules. This still leaves open the question of which prior distribution to use, or whether the prior can be interpreted as corresponding to some beliefs on the part of the researcher. There are other questions as well. Where does the loss function come from? What justifies taking its expectation?

The theory we will survey in this lecture note provides partial answers to these questions. We will start by reviewing expected utility theory, recalling some results from the von-Neumann Morgenstern theory, and then the joint development of utility and subjective probability in Savage's theory of decision-making. Then we will discuss the concept of exchangeability, which essentially justifies the use of i.i.d. models with prior distributions. For the most part the discussion will be heuristic, as some of the technical details are fairly involved.

2 Expected Utility

Let $\mathcal{P}$ denote a set of *prizes* or *consequences*, and let $\mathcal{P}^*$ denote the set of all lotteries (probability distributions over $\mathcal{P}$) that put all their mass on a finite number of points in $\mathcal{P}$. An element $p \in \mathcal{P}^*$ is written $p = (f_1 b_1, \ldots, f_n b_n)$, indicating that the lottery p chooses b_i with probability f_i.

Definition 1 *A* preference relation *on $\mathcal{P}^*$ is a linear ordering, $\precsim$, of $\mathcal{P}^*$ that is: (a) complete: if p and $q \in \mathcal{P}^*$, then either $p \precsim q$ or $q \precsim p$ or both; and (b) transitive: if p,q, and $r \in \mathcal{P}^*$, and if $p \precsim q$ and $q \precsim r$, then $p \precsim r$.*

We can then define a strict preference relation $\prec$ as $p \prec q$ if $p \precsim q$ and not $q \precsim p$. Similarly, we say that p and q are equivalent and write $p \sim q$ if $p \precsim q$ and $q \precsim p$.

Definition 2 *A* utility *on $\mathcal{P}^*$ is a real-valued function, u, defined on $\mathcal{P}^*$, which is linear on $\mathcal{P}^*$; that is, if p and $q \in \mathcal{P}^*$ and $0 \leq \lambda \leq 1$,*

$$u(\lambda p + (1-\lambda)q) = \lambda u(p) + (1-\lambda)u(q).$$

Definition 3 *A preference relation $\precsim$, and a utility u, on $\mathcal{P}^*$ are said to* agree *if for all p and $q \in \mathcal{P}^*$,*

$$p \precsim q \qquad \text{if and only if} \qquad u(p) \leq u(q).$$

Hypothesis 1 *(Independence) If p, p', and $q \in \mathcal{P}^*$ and $0 < \lambda \leq 1$, then $p \precsim p'$ if and only if*

$$\lambda p + (1-\lambda)q \precsim \lambda p' + (1-\lambda)q.$$

Hypothesis 2 *(Continuity) If p, q, and $r \in \mathcal{P}^*$ are such that $p \prec q \prec r$, then there exist numbers λ and μ with $0 < \lambda < 1$ and $0 < \mu < 1$, such that*

$$\lambda r + (1-\lambda)p \prec q \prec \mu r + (1-\mu)p.$$

Theorem 1 *If a preference relation $\precsim$ on $\mathcal{P}^*$ satisfies Hypotheses 1 and 2, then there exists a utility, u, on $\mathcal{P}^*$, which agrees with $\precsim$. Furthermore, u is uniquely determined up to a positive affine transformation.*

Proof: See Ferguson, p. 15.

Example 1 Recall the portfolio allocation problem discussed in Lecture Note 1. If $\theta = (\mu, \Sigma)$ is known, then each portfolio allocation $a \in [0, 1]$ results in a distribution over terminal wealth W_{T+H}. A preference ordering on the space of lotteries indexed by $[0, 1]$ has a utility representation by Theorem 1. So there is a function $U(W)$ such that

$$a \precsim a' \qquad \text{if and only if} \qquad E(U(W_{T+H})|a, \mu, \Sigma) \leq E(U(W_{T+H})|a', \mu, \Sigma).$$

Choosing U to have the isoelastic form $U(W) = W^{1-\gamma}/(1-\gamma)$ is convenient because then the preferences exhibit constant relative risk aversion.

Now, suppose θ was not known but had a known distribution. Then we could modify the analysis by averaging the expected utilities over the distribution of θ. (This is called "marginalizing out" θ.)

However, in this case it isn't clear that θ actually arises from some sort of randomization by nature. In the statistical decision theory approach of Lecture Note 2, we used the prior distribution $p(\theta)$ as a device to generate admissible rules. It would be nice if we could interpret the prior distribution as corresponding to "beliefs" about which values for θ are more likely than others, and justify such an interpretation within an axiomatic framework, much as we did with the loss and risk functions. This is the goal of the theory of "subjective" expected utility, which is discussed next.

3 Subjective Expected Utility

In the standard expected utility theory just outlined, the probabilities associated with a lottery are known. In most situations, however, these probabilities can only be guessed at, and different decision-makers might even disagree in their assessments of such probabilities.[1] Indeed, one of the main reasons for analyzing data is to learn about unknown distributions. So it would be nice if we had a theory that allowed for mathematical probabilities to be based on judgments on the part of the decision-maker, or better yet, if the probabilities could arise in a natural way from a preference relation, in a manner analogous to the way the utility function was shown to represent the preference relation in the von-Neumann–Morgenstern theory.

The theory of subjective expected utility does just this, providing an axiomatic derivation of *both* probability distributions and expected utilities, where the expectation is with respect to the probability distribution, from a decision problem involving only preferences, actions, and outcomes. Savage (1954) developed the most general of the classical results on subjective expected utility.[2] However, there is a slightly simpler approach, due to Anscombe and Aumann (1963), which is somewhat less general but easier to fit into our discussion of statistical problems.

As before, let $\mathcal{P}$ denote a set of payoffs and let $\mathcal{P}^*$ be the set of finite probability distributions on $\mathcal{P}$. Suppose that the preference pattern $\precsim$ satisfies Hypotheses 1 and 2. Then from Theorem 1 there is a utility u on $\mathcal{P}^*$ that agrees with $\precsim$.

We now consider an experiment with a finite number of outcomes $\theta_1, \ldots, \theta_k$, one and only one of which is bound to occur, and we consider the gamble whose payoff is $p_1 \in \mathcal{P}^*$ if θ_1 occurs, $p_2 \in \mathcal{P}^*$ if θ_2 occurs, and so on. Denote such a gamble by $[p_1, \ldots, p_k]$. The set $\mathcal{G}$ of such gambles may be considered as a set of payoffs. Let $\mathcal{G}^*$ be the set of all finite probability distributions over $\mathcal{G}$, and assume that there is a preference relation $\precsim_g$ that satisfies Hypotheses 1 and 2. Then there will be a utility u_g on $\mathcal{G}^*$ that agrees with $\precsim_g$, by Theorem 1.

An analogy may help in interpreting this setup. Think of an element of $\mathcal{P}^*$ as a roulette wheel—some kind of randomizing device that chooses payoffs according to fixed, known probabilities. Next, suppose there is a horse race between the horses $\theta_1, \ldots, \theta_k$. Which roulette wheel the statistician gets to play depends on which horse θ_i wins the race. Now the objects of choice are gambles, which associate a roulette wheel with each horse but do *not* specify which horse will win or even a distribution over the winning horse. We assume that the statistician has preferences over gambles that satisfy the basic expected utility axioms. The statistician also has preferences over the original lotteries in $\mathcal{P}^*$ (the roulette wheels), so it is natural to connect the two preference relations together.

---

[1] This could reflect differences in the information sets of the decision-makers, or just plain disagreement.

[2] Kreps (1988) calls the Savage theory "as much as anything, the crowning achievement of single-person decision theory."

Hypothesis 3 *For all i, if $p_i \precsim p_i'$, then*

$$[p_1, \ldots, p_i, \ldots, p_k] \precsim_g [p_1, \ldots, p_i', \ldots, p_k].$$

Hypothesis 4 *If $p \prec p'$, then $[p, \ldots, p] \prec_g [p', \ldots, p']$.*

Hypothesis 5

$$(f_1[p_{11}, \ldots, p_{1k}], \ldots, f_m[p_{m1}, \ldots, p_{mk}]) \sim_g [(f_1 p_{11}, \ldots, f_m p_{m1}), \ldots, (f_1 p_{1k}, \ldots, f_m p_{mk})].$$

Hypotheses 3 and 4 basically assert that the two preference relations are compatible. Hypothesis 5 essentially says that only ultimate outcomes matter: the individual is indifferent whether the roulette wheel is spun before or after the horse race.

Theorem 2 *If both preference relations, $\precsim$ on $\mathcal{P}^*$ and $\precsim_g$ on $\mathcal{G}^*$, satisfy Hypotheses 1 and 2, and if Hypotheses 3, 4, and 5 are satisfied, then there exist utilities u on $\mathcal{P}^*$ and u_g on $\mathcal{G}^*$ and there exist nonnegative numbers $\pi_1, \ldots, \pi_k$ with $\sum_{j=1}^k \pi_j = 1$ such that u agrees with $\precsim$, u_g agrees with $\precsim_g$, and*

$$u_g[p_1, \ldots, p_k] = \sum_{j=1}^{k} \pi_j u(p_j).$$

Proof: See Ferguson, p. 19.

The Anscombe-Aumann approach relies on the existence of external, "objective" randomizing devices—the roulette wheels of the analogy. The Savage theory does not require even this, but entails a number of additional mathematical issues.

4 Connections to Statistical Decision Theory

Recall that in the statistical decision setup of Lecture Note 2, the decision-maker is allowed to use a decision rule d that maps values for Z into actions $a \in \mathcal{A}$. Suppose that if nature chooses θ and the decision-maker chooses action a, then the decision-maker receives a prize $b(\theta, a)$, which needs not be monetary or have a natural ordering. To begin, assume θ is known and fixed, and write $b_\theta(a) := b(\theta, a)$. Then a decision rule d results in a distribution over prizes b. To see this, suppose that the decision rule uses only a finite subset of actions, so that d has the form

$$d(z) = a_j \in \mathcal{A} \qquad \text{if} \qquad z \in \mathcal{C}_j \qquad (j = 1, \ldots, J),$$

where the $\mathcal{C}_j$ form a (measurable) partition of the sample space $\mathcal{Z}$. Then the payoff to the decision-maker from using the decision rule d is a lottery where the prize is $b_\theta(a_j)$ with probability $f_\theta(j) := \Pr_\theta(Z \in \mathcal{C}_j)$.

Assume that the decision-maker has a preference relation $\precsim$ over such lotteries that satisfies Hypotheses 1 and 2. Then the preference relation agrees with a utility

$$u(\theta, d) = \sum_{j=1}^{J} f_\theta(j) u(b_\theta(a_j)).$$

Notice that if we define the loss function as $L(\theta, a) = -u(b_\theta(a_j))$, then the risk function corresponding to decision rule d is $R(\theta, d) = -u(\theta, d)$. Thus the risk function agrees with the inverse preference ordering over lotteries. In other words, minimizing risk is compatible with utility maximization.

If it is not known which value of $\theta \in \Theta \equiv \{\theta_1, \ldots, \theta_k\}$ will obtain, then there are k possible lotteries that could arise from decision rule d. So a choice of d corresponds to a choice of a set of lotteries, $[(\theta_1, d), \ldots, (\theta_k, d)]$. If there is a preference relation $\precsim_g$ over such gambles, and if Hypotheses 1, 2, 3, 4, and 5 are satisfied, then there exists a function u_g and nonnegative numbers $\pi_1, \ldots, \pi_k$ with $\sum_{j=1}^k \pi_j = 1$ such that u_g agrees with $\precsim_g$ and

$$u_g[(\theta_1, d), \ldots, (\theta_k, d)] = \sum_{j=1}^{k} \pi_j u(\theta_j, d).$$

So a rational statistician acts as if she were minimizing the Bayes risk associated with the prior distribution π.

5 Exchangeability

The Anscombe-Aumann theory assumed the existence of the objective randomizing devices associated with each θ_i. Likewise, in the statistical decision theory approach of the previous lecture note, the existence of a set of probability measures indexed by $\theta \in \Theta$ was taken as a primitive feature of the problem. In most real estimation problems, the specification of the model $\{P_\theta : \theta \in \Theta\}$ is as important and difficult as deciding which procedure to use once the experiment is well-defined.

In the more general Savage theory of expected utility, we end up with a probability distribution over a mutually exclusive, exhaustive set of "states of nature." For example, consider the experiment of flipping a possibly unfair coin repeatedly. Let $Y_1 = 1$ if the first coin comes up heads, and 0 if the first coin comes up tails. Likewise let $Y_2 = 1$ if the second coin comes up heads, and so on. We can think of a state of nature as any possible string of outcomes of the coin-tossing process, e.g. (0,1,0,0,...). A rational decision-maker will act as if she had placed a probability distribution $P(y_1, y_2, ...)$ on such strings. What can we say about this distribution?

It might seem appropriate to assume that the distribution has the Y_i independent, but the assumption of independence is actually quite strong: it says that whether you observe that the first 100 tosses of the coin are all heads or all tails is irrelevant for predicting the 101'st toss of the coin, so that there is no point in collecting data about the coin. This would be fine if we knew that the coin was fair, or knew how biased the coin was, but in the absence of such knowledge, the assumption of independence is unattractive.

But suppose that we are willing to assume that the distribution displays symmetry, so that the *order* in which the observations are recorded does not matter.

Definition 4 *(Finite Exchangeability) The random variables $Y_1, \ldots, Y_n$ with probability distribution P are said to be finitely exchangeable if*

$$P(y_1, \ldots, y_n) = P(y_{\pi(1)}, \ldots, y_{\pi(n)})$$

for all permutations π defined on the set $\{1, \ldots, n\}$.

Definition 5 *(Infinite Exchangeability) A sequence of random variables $Y_1, Y_2, \ldots$ is said to be infinitely exchangeable if every finite subset is exchangeable in the sense of Definition 4.*

Theorem 3 *Suppose $Y_1, Y_2, \ldots$ is an infinitely exchangeable sequence of 0-1 random variables, with probability distribution P. Then there exists a probability distribution Q on $[0, 1]$ such that for any vector of 0's and 1's, $(y_1, \ldots, y_n)$,*

$$p(y_1, \ldots, y_n) = \int_0^1 \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} dQ(\theta),$$

where

$$Q(\theta) = \lim_{n \to \infty} P(\sum_{i=1}^{n} y_i/n \leq \theta)$$

and $\theta = \lim_{n \to \infty} \sum_{i=1}^{n} y_i/n$.

Proof: See Bernardo and Smith, p. 172.

This result, known as *de Finetti's Theorem*, says any probabilistic beliefs about the joint distribution of the Y_i's which satisfy infinite exchangeability are equivalent to assuming that the Y_i's are i.i.d. Bernoulli, with some prior distribution Q placed on the Bernoulli parameter θ. Remarkably, the theorem can be extended to real-valued random variables as well. In that case the resulting representation says that the y_i are i.i.d. with distribution $f(y_i|\theta)$, and a prior distribution is placed on θ. The parameter θ completely specifies the distribution of y_i, and so might have to be a very high-dimensional object, possibly infinite-dimensional.

De Finetti's Theorem, therefore, justifies the common device of assuming that a sample is i.i.d., in a situation where there are *no* objectively given probabilities—just data and actions. However, it is silent about the many cases where exchangeability, does not apply. For example, if data are indexed by time, Y_t, then the ordering of data could be very important.

6 Notes

The idea of developing both utilities and subjective probabilities from preferences seems to have been introduced by Ramsey (1931) (which was actually written in 1926); like Ramsey's other work on optimal taxation and growth, this paper was not given much attention until many years after being published.

References

Anscombe, F. J., and R. J. Aumann (1963): "A Definition of Subjective Probability," *Annals of Mathematical Statistics*, 34, 199–205.

Kreps, D. (1988): *Notes on the Theory of Choice*. Westview Press, Boulder.

Ramsey, F. P. (1931): "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, ed. by R. B. Braithwaite, pp. 156–98. Routledge and Kegan Paul, London.

Savage, L. J. (1954): *The Foundations of Statistics*. Wiley, New York, reissued in 1972 by Dover.