BAYESIAN DECISIONS UNDER TOTAL AND PARTIAL IGNORANCE

by

Dean Jamison

and

SUBJECTIVE PROBABILITIES UNDER TOTAL UNCERTAINTY

by

Dean Jamison and Jozef Kozielecki

TECHNICAL REPORT NO. 121

September 4, 1967

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

TABLE OF CONTENTS

BAYESIAN DECISIONS UNDER TOTAL AND PARTIAL IGNORANCE[1]

by

Dean Jamison

Introduction

1. A triple $P = < D, \Omega, U >$ may be considered a finite decision problem when: (i) D is a finite set of alternative courses of action available to a decision-maker, (ii) Ω is a finite set of mutually exclusive and exhaustive possible states of nature, and (iii) U is a function on $D \times \Omega$ such that $u(d_i, \omega_j)$ is the utility to the decision-maker if he chooses d_i and the true state of nature turns out to be ω_j. A decision procedure (solution) for the problem P consists either of an ordering of the d_i according to their desirability or of the specification of a subset of D that contains all d_i that are in some sense optimal and only those d_i that are optimal.

If there are m states of nature, a vector $\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)$ is a possible probability distribution over Ω (with $\text{prob}(\omega_j) = \xi_j$) iff $\sum_{j=1}^{m} \xi_j = 1$ and $\xi_j \geq 0$ for $1 \leq j \leq m$. The set of all possible probability distributions over Ω, that is, the set of all vectors whose components satisfy the above equation and set of inequalities, will be denoted by $\approx$. Atkinson, Church, and Harris (1964) assume our knowledge of $\vec{\xi}$ to be completely specified by asserting that $\vec{\xi} \in \approx_0$ where $\approx_0 \subseteq \approx$. If $\approx_0 = \approx$, they say we are in complete ignorance of $\vec{\xi}$. In the manner of Chernoff (1954) and Milnor (1954), Atkinson, et al, give axioms stating desirable properties for decision procedures under complete ignorance. A class of decision procedures that isolate an optimal subset of D is shown to exist and satisfy the axioms. Efron (1965) extends the range of applicability of the decision procedures suggested by Atkinson, et al, and dubs them "iterated minimax regret rules". These rules are non-Bayesian in the sense that the criterion for optimality is not maximization of expected utility. Other non-Bayesian decision procedures for complete ignorance (that fail to satisfy some axioms that most people would consider reasonable) include the following: minimax regret, minimax risk (or maximin utility), and Hurwicz's α procedure for extending the minimax risk approach to non-pessimists.

The Bayesian alternative to the above procedures attempts to order the d_i according to their expected utility; the optimal act is, then, simply the one with the highest expected utility. Computation of the expected utility of d_i, $\mathcal{E}u(d_i)$, is straightforward if the decision-maker <u>knows</u> that $\approx_o$ is a set with but one element $- \vec{\xi}^*$: $\mathcal{E}u(d_i) = \sum_{j=1}^{m} u(d_i, \omega_j)\xi_j^*$. Only in the rare instances when considerable relative frequency data exist will the decision-maker be able to assert that $\approx_o$ has only one element. In the more general case the decision-maker will be in "partial" or "total" ignorance concerning the probability vector $\vec{\xi}$. It is the purpose of this paper to characterize total and partial ignorance from a Bayesian point of view and to show that decision procedures based on maximization of expected utility extend readily to these cases.

<u>Decisions under Total Ignorance</u>

2. Rather than saying that our knowledge of the probability vector $\vec{\xi}$ is specified by asserting that $\vec{\xi} \in \approx_o$ for some $\approx_o$, I suggest that it is natural to say that our knowledge of $\vec{\xi}$ is specified by a density, $f(\xi_1, \xi_2, \ldots, \xi_m)$, defined on $\approx$. If the probability distribution over

Ω is known to be $\vec{\xi}^*$, then f is a δ function at $\vec{\xi}^*$ and computation of $\mathcal{E}u(d_i)$ proceeds as in the Introduction. At the other extreme from precisely knowing the probability distribution over Ω is the case of total ignorance. In this section a meaning for "total ignorance" of $\vec{\xi}$ will be suggested and decision-making under total ignorance will be discussed. In the following section decisions under partial ignorance--anywhere between knowledge of $\vec{\xi}$ and total ignorance--will be discussed.

3. If $H(\vec{\xi})$ is the Shannon (1949) measure of uncertainty concerning which ω in Ω occurs, then $H(\vec{\xi}) = \sum_{i=1}^{m} \xi_i \log_2 (1/\xi_i)$, where $H(\vec{\xi})$ is measured in bits. When this uncertainty is a maximum, we may be considered in total ignorance of ω and, as one would expect, this occurs when we have no reason to expect any one ω more than another, i.e., when $\xi_i = 1/m$ for $1 \leq i \leq m$. By analogy, we can be considered in total ignorance of $\vec{\xi}$ when $H(f) = \iint_{\approx} \cdots \int f(\vec{\xi}) \log_2(1/f(\vec{\xi}))\, d \approx$ is a maximum. This occurs when f is a constant, that is, when we have no reason to expect one value of $\vec{\xi}$ to be more probable than any other (see Shannon (1949), Ch. 3). If there is total ignorance concerning $\vec{\xi}$, then it is reasonable to expect that there be total ignorance concerning ω --and this is indeed true (if we substitute the expectation of ξ_i, $\mathcal{E}(\xi_i)$, for $\xi_i)^2$. Let me now prove this last assertion, which is the major result of this section.

4. Proving that under total ignorance $\mathcal{E}(\xi_i) = 1/m$ involves, first, determination of the appropriate constant value of f, then determination of the probability density functions for the ξ_is and, finally, integration to find $\mathcal{E}(\xi_i)$.

Let the constant value of $f(\xi_1, \xi_2, \ldots, \xi_m)$ be equal to K; since f is a density, the integral of K over $\approx$ must be unity:

$$\iint \underset{\approx}{\cdots} \int K \, d \approx = 1 \qquad (1)$$

where $d \approx = d\xi_1 d\xi_2 \ldots d\xi_m$. Our first task is to solve this equation for K. Since f is defined only on a section of a hyperplane in m-dimensioned space, the above integral is a many dimensioned "surface" integral. Figure 1 depicts the 3-dimensional case. As $\sum_{i=1}^{m} \xi_i = 1$, ξ_m

---

Insert Figure 1 about here

---

is determined given the previous m-1 ξ_is and the integration need only be over a region of m-1 dimensioned space, the region A in Figure 1. It is shown in advanced calculus (see, for example, Crowell & Williamson (1962), pp. 409-419) that $d \approx$ and dA are related in the following way:

$$d \approx = \sqrt{\left(\frac{\partial(x_2, \ldots, x_m)}{\partial(\xi_1, \ldots, \xi_{m-1})}\right)^2 + \ldots + \left(\frac{\partial(x_1, \ldots x_{m-1})}{\partial(\xi_1, \ldots, \xi_{m-1})}\right)^2} \, dA ,$$

where x_i is the function of $\xi_1, \ldots, \xi_{m-1}$ that gives the i^{th} component of $\vec{\xi}$, that is, $x_i(\xi_1, \ldots, \xi_{m-1}) = \xi_i$ if $1 \leq i \leq m-1$ and $x_i(\xi_1, \ldots, \xi_{m-1}) = 1-\xi_1-\ldots-\xi_{m-1}$ if $i = m$. It can be shown that each of the m quantities that are squared under the radical above is equal to either plus or minus one; thus, $d \approx = \sqrt{m}\, dA$. Therefore, equation (1) may be rewritten as follows:

$$\iint \underset{A}{\cdots} \int K\sqrt{m}\, dA = 1 ,$$

or

$$\int_0^1 \int_0^{1-\xi_1} \cdots \int_0^{1-\xi_1-\ldots\xi_{m-2}} d\xi_{m-1} d\xi_{m-2} \ldots d\xi_1 = 1/K\sqrt{m} . \qquad (2)$$

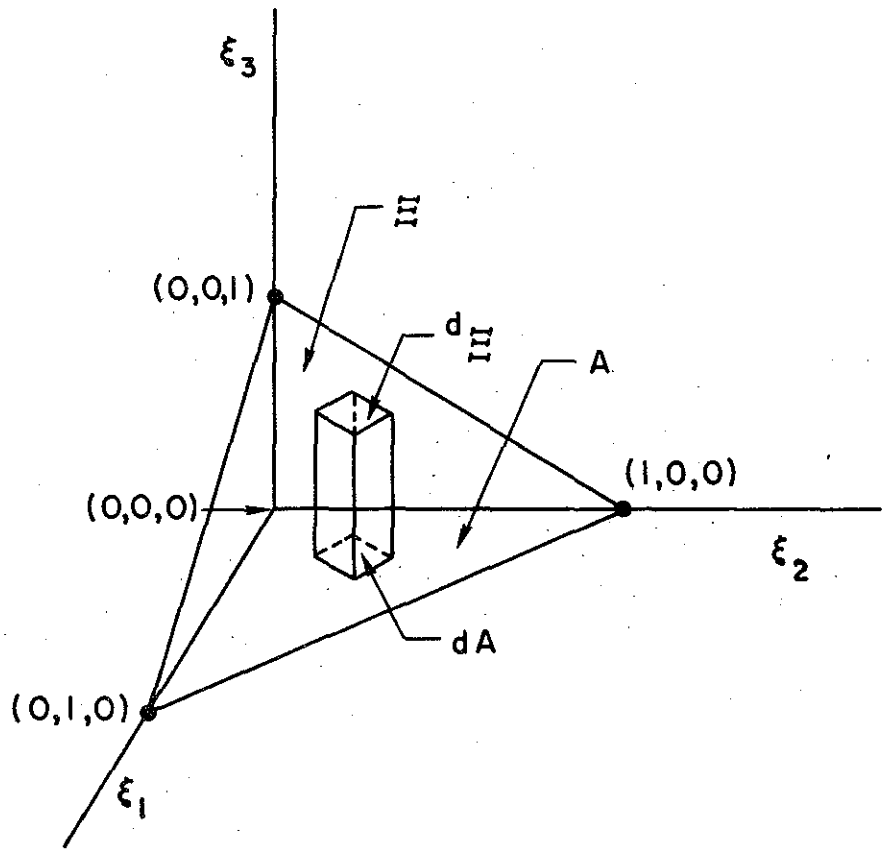


Figure 1. Ξ, the set of possible probability distributions over Ω.

The multiple integral in (2) could conceivably be evaluated by iterated integration; it is much simpler, however, to utilize a technique devised by Dirichlet. Recall that the gamma function is defined in the following way: $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx$ for $n \geq 0$. If n is a positive integer, $\Gamma(n) = (n-1)!$ and $0! = 1$. Dirichlet showed the following (see Jeffreys and Jeffreys (1956), pp. 468-470): If A is the closed region in the first ortant bounded by the coordinate hyperplanes and by the surface $(x_1/c_1)^{p_1} + (x_2/c_2)^{p_2} + \ldots + (x_n/c_n)^{p_n} = 1$, then

$$\iint_A \ldots \int x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \ldots x_n^{\alpha_n - 1}\, dA = \frac{c_1^{\alpha_1} c_2^{\alpha_2} \ldots c_n^{\alpha_n}}{p_1 p_2 p_3 \ldots p_n} \cdot \frac{\Gamma(\frac{\alpha_1}{p_1})\Gamma(\frac{\alpha_2}{p_2}) \ldots \Gamma(\frac{\alpha_n}{p_n})}{\Gamma(1 + \frac{\alpha_1}{p_1} + \frac{\alpha_2}{p_2} + \ldots + \frac{\alpha_n}{p_n})} . \quad (3)$$

For our purposes, $c_i = p_i = \alpha_i = 1$, for $1 \leq i \leq m$ and the $m-1$ ξ_is replace the n xs. The result is that the integral in (2) becomes $1/\Gamma(m) = 1/(m-1)!$. Therefore $K = (m-1)!\sqrt{m}/m$.

Having determined the constant value, K, of f we must next determine the densities $f_i(\xi_i)$ for the individual probabilities. By symmetry, the densities must be the same for each ξ_i. The densities are the derivatives of the distribution functions which will be denoted $F_i(\xi_i)$. $F_1(c)$ gives the probability that ξ_1 is less than c; denote by $F_1^*(c)$ the probability that $\xi_1 \geq c$, that is, $F_1(c) = 1 - F_1^*(c)$. $F_1^*(c)$ is simply the integral of f over $\approx_c$, where $\approx_c$ is the subset of $\approx$ including all points such that $\xi_1 \geq c$. See Fig. 2. $F_1^*(c)$ is given by:

Insert Figure 2 about here

$$F_1^*(c) = \iint_{\underset{\approx}{z}_c} \cdots \int f(\vec{\xi})\, d\underset{\approx}{z} = \iint_{A_c} \cdots \int K\sqrt{m}\; dA_c \; . \qquad (4)$$

Since $K = (m-1)!\,\sqrt{m}/m$, (4) becomes (after inserting the limits of integration):

$$F_1^*(c) = (m-1)! \int_c^1 \int_0^{1-\xi} \cdots \int_0^{1-\xi_1-\ldots-\xi_{m-2}} d\xi_{m-1}\, d\xi_{m-2} \ldots d\xi_1 \; . \qquad (5)$$

A translation of the ξ_1 axis will enable us to use Dirichlet integration to evaluate (5); let $\xi_1' = \xi_1 - c$. Then $\xi_1' + \xi_2 + \ldots + \xi_{m-1} = 1-c$, or $\xi_1'/(1-c) + \xi_2/(1-c) + \ldots + \xi_{m-1}/(1-c) = 1$ (since $\sum_{i=1}^{m-1} \xi_i = 1$ is the boundary of the region A). Referring back to equation (3) it can be seen that the c_is in that equation are all equal to $1-c$ and that, therefore, the integral on the r.h.s. of (5) is $(1-c)^{m-1}/\Gamma(m)$. Thus $F_1^*(c) = (m-1)!\,(1-c)^{m-1}/\Gamma(m) = (1-c)^{m-1}$. Therefore $F_1(c) = 1 - (1-c)^{m-1}$. Since this holds if c is set equal to any value of ξ_1 between 0 and 1, ξ_1 can replace c in the equation; differentiation gives the probability density function of ξ_1 and hence of all the ξ_is:

$$f_i(\xi_i) = (m-1)(1-\xi_i)^{m-2} \; . \qquad (6)$$

From (6) the expectation of ξ_i is easily computed-- $\mathcal{E}(\xi_i) = \int_0^1 \xi_i(m-1)(1-\xi_i)^{m-2}$. Recourse to a table of integrals will quickly convince the reader that $\mathcal{E}(\xi_i) = 1/m$. Figure 3 shows $f_i(\xi_i)$ for several values of m.

Insert Figure 3 about here

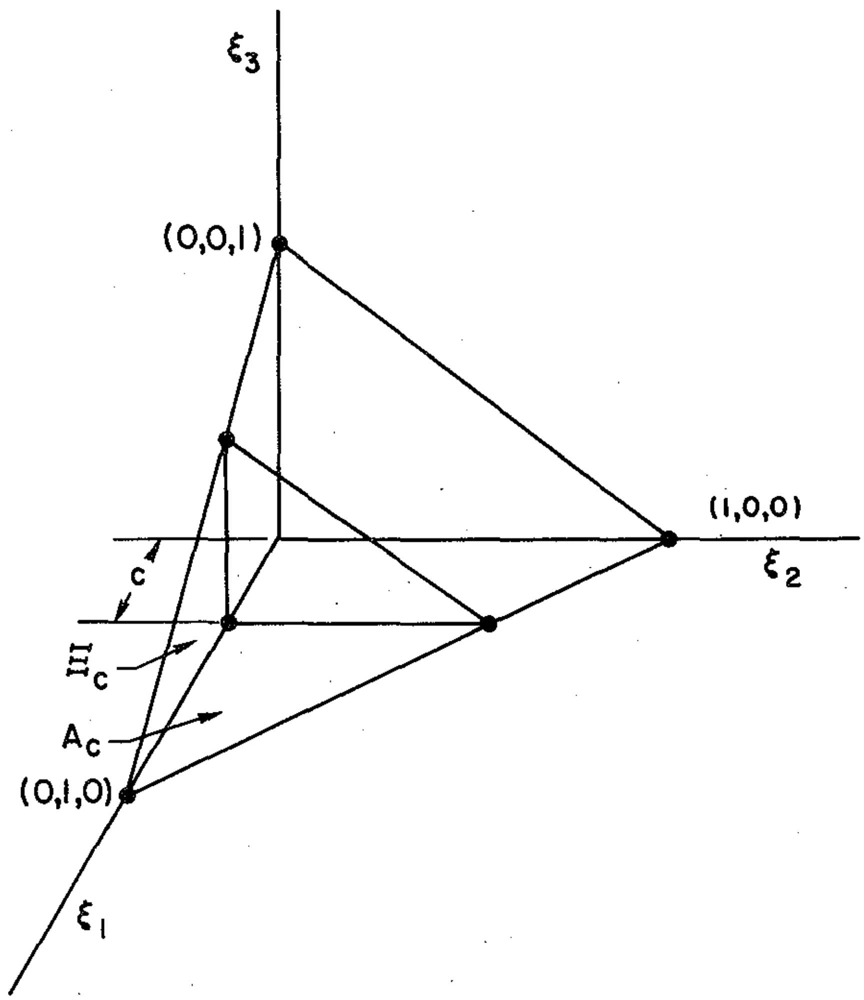


Figure 2. Ξ_c, the subset of Ξ such that $\xi_1 \geq c$.

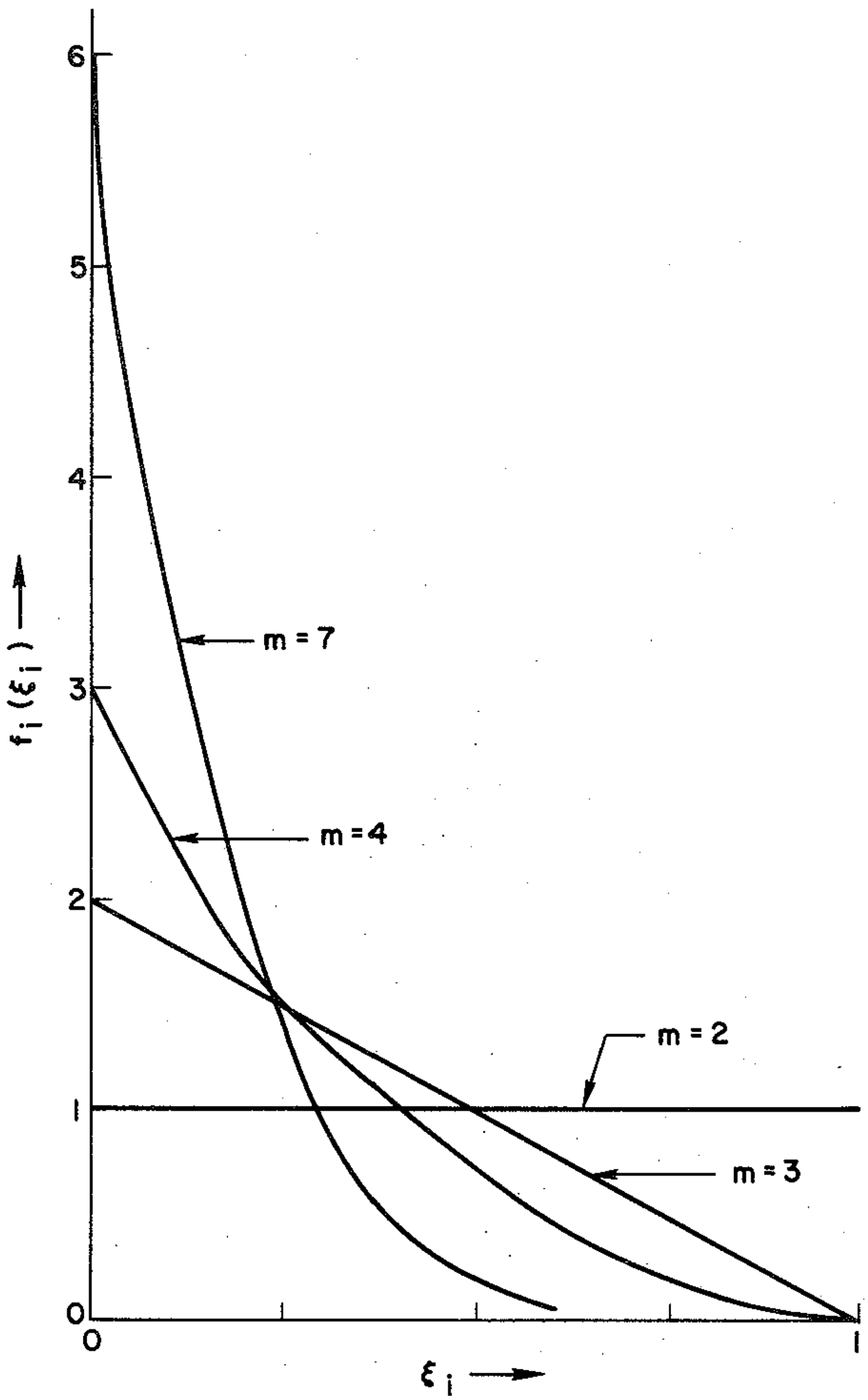


Figure 3. Marginal densities under total uncertainty.

5. Let $u(d_i, \vec{\xi}) = \sum_{j=1}^{m} \xi_j u(d_i, \omega_j)$. Then the expected utility of d_i is given by:

$$\mathcal{E} u(d_i) = \iint \cdots \int K \, u(d_i, \vec{\xi}) d\underset{\approx}{\xi} . \qquad (7)$$

This is equal to $\sum_{j=1}^{m} \mathcal{E}(\xi_j)\, u(d_i, \omega_j) = (1/m) \sum_{j=1}^{m} u(d_i, \omega_j)$, since $u(d_i, \vec{\xi})$ is a linear function of the random variables ξ_i. (See Lindley (1965), p. 126ff.)

Thus, taking the view of total ignorance adopted herein, we arrive at the decision rule advocated by Bernoulli and Laplace and axiomatized in Chernoff (1954). This decision rule is claimed to have two defects. Firstly, what to call a distinct state of nature is often an arbitrary choice, and, secondly, the rule cannot deal with an infinite number of states of nature. I think these defects are unimportant.

Assume that all utilities are normalized to the interval between zero and one and that ϵ represents a utility so small as to be of no practical importance to the decision-maker. Then all utilities can be rounded off to the nearest of a total number of $1/\epsilon$ distinct utility points between zero and one. If there are a finite number, n, of decisions possible, there can be no more than $(1/\epsilon)^n$ different utility columns. For example if $1/\epsilon = 3$ and $n = 2$, the matrix below contains all the different utility columns.

d_1	u_1	u_1	u_1	u_2	u_2	u_2	u_3	u_3	u_3
d_2	u_1	u_2	u_3	u_1	u_2	u_3	u_1	u_2	u_3

Adding another column would only duplicate a previous one. By grouping together all "states of nature" that yield identical utility columns, the decision-maker has a minimum of at most $(1/\epsilon)^n$ (and usually far fewer) states of nature to consider. This is a practical solution to the problem of an infinite number of states of nature. The decision-maker may choose to consider more than the minimum number of states of nature if he wishes to regard as different two distinct sets of physical circumstances that yield identical utility columns. That seems to me to be a matter of individual taste.

The last sentence in the preceding paragraph suggests that the same decision problem may justly be formulated in many ways and that the choice among the alternatives is up to the decision-maker. Clearly if Bayesian decision procedures for total ignorance were used in each formulation of the problem, then the d_i chosen could depend on the formulation. However, changing the formulation of the problem gives us information about the new set of outcomes. For example, if we collapse two exclusive states of nature into one, the probability of the new state must be considered the sum of the probabilities of the old ones, and we can no longer be considered in total ignorance of the outcome. In fact, very rarely will we be in a decision situation of total ignorance; usually there will be some partial information. To this class of problems we now turn.

Decisions under Partial Ignorance

6. Partial ignorance exists in a given formulation of a decision if we neither know the probability distribution over Ω nor are in total ignorance of it. If we are given $f(\xi_1, \xi_2, \ldots \xi_m)$, the density over

$\approx$, computation of $\mathcal{E}u(d_i)$ under partial ignorance is in principle straightforward and proceeds along lines similar to those developed in the previous section. Equation (7) is modified in the obvious way to:

$$\mathcal{E}u(d_i) = \iint_{\approx}\cdots\int f(\vec{\xi})u(d_i, \vec{\xi})d \approx . \tag{8}$$

If f is any f the large variety of appropriate forms indicated just prior to equation (3), the integral in (8) may be easily evaluated using Dirichlet integration; otherwise, more cumbersome techniques must be used.

In practice it seems clear that unless the decision-maker has remarkable intuition, the density f will be impossible to specify from the partial information at hand. Fortunately there is an alternative to determining f.

7. Jeffrey (1965, pp. 183-190), in discussing degree of confidence of a probability estimate, describes the following method for obtaining the distribution function, $F_i(\xi_i)$, for a probability. Have the decision-maker indicate for numerous values of ξ_i what his subjective estimate is that the "true" value of ξ_i is less than the value named. To apply this to a decision problem the distribution function--and hence $f_i(\xi_i)$--for each of the ξ_is must be obtained. Next, the expectations of the ξ_is must be computed and, from them, the expected utilities of the d_is can be determined. In this way partial information is processed to lead to a Bayesian decision under partial ignorance.

It should be clear that the decision-maker is <u>not</u> free to choose the f_is subject only to the condition that for each f_i, $\int_0^1 f_i(\xi_1)d\xi_i = 1$. Consider the example of the misguided decision-maker who believed himself to be in total ignorance of the probability distribution over 3 states of nature. Since he was in total ignorance, he reasoned, he must have a uniform p.d.f. for each ξ_i. That is, $f_1(\xi_1) = f_2(\xi_2) = f_3(\xi_3) = 1$ for $0 \leq \xi_i \leq 1$. If he believes these to be the p.d.f.s, he should be willing to simultaneously take even odds on bets that $\xi_1 > 1/2$, $\xi_2 > 1/2$, and $\xi_3 > 1/2$. I would gladly take these three bets, for under no conditions could I fail to have a net gain. This example illustrates the obvious -- certain conditions must be placed on the f_is in order that they be coherent. A necessary condition for coherence is indicated below; I have not yet derived sufficient conditions.

Consider a decision, d_k, that will result in a utility of 1 for each ω_j. Clearly, then, $\mathcal{E}u(d_k) = 1$. However, $\mathcal{E}u(d_k)$ also equals $\mathcal{E}(\xi_1)u(d_k,\omega_1) + \ldots + \mathcal{E}(\xi_m)u(d_k,\omega_m)$. Since for $1 \leq i \leq m$, $u(d_k,\omega_i) = 1$, a necessary condition for coherence of the f_is is that $\sum_{i=1}^{m} \mathcal{E}(\xi_i) = 1$, a reasonable thing to expect. That this condition is not sufficient is easily illustrated with two states of nature. Suppose that $f_1(\xi_1)$ is given. Since $\xi_2 = 1 - \xi_1$, f_2 is uniquely determined given f_1. However it is obvious that infinitely many f_2s will satisfy the condition that $\mathcal{E}(\xi_2) = 1 - \mathcal{E}(\xi_1)$, and if a person were to have two distinct f_2s it would be easy to make a book against him; his beliefs would be incoherent.

If m is not very large, it would be possible to obtain conditional densities of the form $f_2(\xi_2|\xi_1)$, $f_3(\xi_3|\xi_1,\xi_2)$, etc., in a manner analogous to that discussed by Jeffrey. If the conditional densities were obtained, then $f(\vec{\xi})$ would be given by the following expression:

$$f(\underset{\sim}{\xi}) = f_1(\xi_1)f_2(\xi_2|\xi_1) \ldots f_m(\xi_m|\xi_1,\xi_2, \ldots ,\xi_{m-1}). \tag{9}$$

A sufficient condition that the $f_i s$ be coherent is that the integral of f over $\approx$ be unity; if it differs from unity, one way to bring about coherence would be to multiply f by the appropriate constant and then find the new $f_i s$. If m is larger than 4 or 5, this method of insuring coherence will be hopelessly unwieldy. Something better is needed.

8. At this point I would like to discuss alternatives and objections to the theory of decisions under partial information that is developed here. The notion of probability distributions over probability distributions has been around for a long time; Knight, Lindall, and Tintner were among the first to explicitly use the notion in economics (see Tintner (1941)).[3] This work has not, however, been formulated in terms of decision theory. Hodges and Lehmann (1952) have proposed a decision rule for partial ignorance that combines the Bayesian and minimax approaches. Their rule chooses the d_i that maximizes $\mathcal{E}u(d_i)$ for some best estimate (or expectation) of $\vec{\xi}$, subject to the condition that the minimum utility possible for d_i is greater than a preselected value. This preselected value is somewhat less than the maximum utility; the amount less increases with our confidence that $\vec{\xi}$ is the correct distribution over Ω. Ellsberg (1961), in the lead article of a spirited series in the Quarterly Journal of Economics, provides an elaborate justification of the Hodges and Lehmann procedure, and I will criticise his point of view presently.

Hurwicz (1951) and Good (1950) (discussed in Luce and Raiffa (1956), p. 305) have suggested characterizing partial ignorance in the same fashion that was later used by Atkinson, et al (1964). That is, our knowledge of $\vec{\xi}$ is of the form $\vec{\xi} \in \underset{\sim}{\approx}_0$ where $\underset{\sim}{\approx}$ is a subset of $\underset{\sim}{\approx}$, Hurwicz then proposes that we proceed as if in total ignorance of where $\vec{\xi}$ is in $\underset{\sim}{\approx}_0$. In the spirit of the second section of this paper, the decision rule could be Bayesian with $f(\vec{\xi}) = K$ for $\vec{\xi} \in \underset{\sim}{\approx}_0$ and $f(\vec{\xi}) = 0$ elsewhere. Hurwicz suggests instead utilization of non-Bayesian decision procedures; difficulties with non-Bayesian procedures were alluded to in the Introduction.

The reader interested in a more thorough discussion and bibliography concerning decisions under total and partial ignorance is referred to Arrow (1951), Luce and Raiffa (1956), and Luce and Suppes (1965). I will now try to conter some objections that have been raised against characterizing partial ignorance as probability distributions over probabilities.

On page 659 Ellsberg (1961) takes the view that since representing partial ignorance (ambiguity) as a probability distribution over a distribution leads to an expected distribution, ambiguity must be someting different from a probability distribution. I fail to understand this argument; ambiguity is high, it seems to me, if f is relatively flat over $\underset{\sim}{\approx}$, otherwise not. The "reliability, credibility, or accuracy" of one's information simply determines how sharply peaked f is. Even granted that probability is somehow qualitatively different from ambiguity or uncertainty, the solution devised by Hodges and Lehmann (1952) and advocated by Ellsberg relies on the decision-maker's completely arbitrary judgment of the amount of ambiguity present in the decision situation. Ellsberg would have us hedge against our uncertainty in $\vec{\xi}$ by rejecting a decision that maximized utility against the expected distribution but that has a possible outcome with a utility below an arbitrary minimum. By the same reasoning one could "rationally" choose d_1 over d_2 in the non-ambiguous problem below if, because of our uncertainty in the outcome, we said (arbitrarily) that we would reject any decision with a minimum gain of less than 3.

	ω_1	ω_2
d_1	5	5
d_2	1	25

$\xi_1 = \xi_2 = .5$

I would reject Ellsberg's approach for the simple reason that its pessimistic bias leads to decisions that fail to fully utilize one's partial information.

Savage (1954, pp. 56-60) raises two objections to second-order probabilities. The first, similar to Ellsberg's, is that even with second-order probabilities expectations for the primary probabilities remain. Thus we may as well have simply arrived at our best subjective estimate of the primary probability, since it is all that is needed for decision-making. This is correct as far as it goes, but, without second-order probabilities, it is impossible to specify how the primary probability should change in the light of evidence. This will be discussed in more detail in the final section.

Savage's second objection is that "...once second order probabilities are introduced, the introduction of an endless hierarchy seems inescapable. Such a hierarchy seems very difficult to interpret, and it seems at best to make the theory less realistic, not more." Luce and Raiffa (1956) express much the same objection on page 305. An endless hierarchy does not seem inescapable to me; we simply push the hierarchy back as far as is required to be "realistic". In making a physical measurement we could attempt to specify the value of the measurement, the probable error in the measurement, the probable error in the probable error, and on out the endless hierarchy. But it is not done that way; probable errors seem to be about the right order of realism. Similarly, I suspect that second-order probabilities will suffice for most circumstances.[4]

The Role of Evidence in Changing Probabilities

9. The preceding discussion has been limited to situations in which the decision-maker has no option to experiment or buy information. When the possibility of experimentation is introduced, the number of alternatives open to the decision-maker is greatly increased, as is the complexity of his decision problem, for the decision-maker must now decide which experiments to perform and in what order, when to stop experimenting, and which course of action to take when experimentation is complete. (See discussion in Chap. 3 of Blackwell and Girshick (1954).) A crucial question here is to specify how evidence is to be used to change our probability estimates.

If we are quite certain that $\vec{\xi}$ is very nearly the true probability distribution over Ω, additional evidence will little change our beliefs. If, on the other hand, we are not at all confident about $\vec{\xi}$ -- if f is fairly flat -- new evidence can change our beliefs considerably. (New evidence may leave the expectations for the ξ_is unaltered even though it changes beliefs by making f more sharp. In general, of course, new evidence will both change the sharpness of f and change the expectations of the ξ_is.) Without second-order probabilities there appears to be no answer to the question of exactly how new evidence can alter probabilities. Suppes (1956) considers an important defect of both his and Savage's (1954) axiomatizations of subjective probability and utility to be their failure to specify how prior information is to be used. Let us consider an example used by both Suppes and Savage.

10. A man must decide whether to buy some grapes which he knows to be either green (ω_1), ripe (ω_2), or rotten (ω_3). Suppes poses the

following question: If the man has purchased grapes at this store 15 times previously, and has never received rotten grapes, and has no information aside from these purchases, what probability should he assign to the outcome of receiving rotten grapes the 16th time?

Prior to his first purchase, the man was in total ignorance of the probability distribution over Ω. Thus from equation (6) we see that the density for ξ_3, the prior probability of receiving rotten grapes, should be $f_3(\xi_3) = 2 - 2\xi_3$. Let X be the event of receiving green or ripe grapes on the first 15 purchases; the probability that X occurs, given ξ_3, is $p(X|\xi_3) = (1 - \xi_3)^{15}$. What we desire is $f_3(\xi_3|X)$, the density for ξ_3 given X, and this is obtained by Bayes' theorem in the following way:

$$f_3(\xi_3|X) = p\ X|\xi_3)f_3(\xi_3)/\int_0^1 p(X|\xi_3)f_3(\xi_3)d\xi_3\ . \qquad (10)$$

After inserting the expressions for $f_3(\xi_3)$ and $p(X|\xi_3)$, equation (10) becomes:

$$f_3(\xi_3|X) = (1 - \xi_3)^{15}(2 - 2\xi_3)/\int_0^1 (1 - \xi_3)^{15}(2 - 2\xi_3)d\xi_3\ .$$

Performing the integration and simplifying gives $f_3(\xi_3|X) = 17(1 - \xi_3)^{16}$; from this the expectation of ξ_3 given X can be computed -- $\mathcal{E}(\xi_3|X) = 17\int_0^1 \xi_3(1 - \xi_3)^{16} = 1/18$. (Notice that this result differs from the 1/17 that Laplace's law of succession would give. The difference is due to the fact that the Laplacian law is derived from consideration of only two states of nature -- rotten and not rotten[5].)

Let us consider another example discussed by Savage (1954, p.65): "...it is known of an urn that it contains either two white balls, two black balls, or a white ball and a black ball. The principle of insufficient reason has been invoked to conclude that the three possibilities are equally probable, so that in particular the probability of one white and one black ball is concluded to be 1/3. But the principle has also been applied to conclude that there are four equally probable possibilities, namely, that the first is white and the second black, etc. On that basis, the probability of one white and one black is, of course, 1/2." Let us consider the second case more carefully. Analogously to the problem of purchasing grapes, we may consider this a sequential problem with two states of nature -- ω_1, the event of drawing black and ω_2, the event of drawing white. Before the first draw we are in total ignorance of ξ_1 and ξ_2, i.e., $f_1(\xi_1) = f_2(\xi_2) = 1$. Denote by $\omega_1\omega_2$ the event of black followed by white; then the probability of $\omega_1\omega_2$ given ξ_1 is $\xi_1(1 - \xi_1)$ and, therefore, $p(\omega_1\omega_2) = \int_0^1 \xi_1(1 - \xi_1)f_1(\xi_1)d\xi_1 = 1/6$. Likewise $p(\omega_2\omega_1) = 1/6$ so that the probability of one black and one white is 1/3. Thus the apparent paradox is resolved. However, there were two critical assumptions in deriving this result: (i) There was total ignorance of the prior distribution over Ω, and (ii) The decision-maker chooses to utilize evidence in accord with Bayes' theorem. Drop either assumption and you can easily obtain a value of 1/2 (or just about anything else) for the probability of one black and one white.

11. In this section I have tried to show why second-order probability distributions are necessary in order to appropriately change one's probability estimates in the light of new evidence (or to adequately

utilize one's prior information). My discussion is primarily suggestive; it fails to deal with any of the difficult features of sequential decision-making alluded to at the beginning of this section. Yet the relevance of second-order probabilities to these problems should be clear. A more complete theory of Bayesian decisions under partial ignorance will be required to deal with decision situations where experimentation is allowed. Suppes' (1965) suggestion for introducing a theory of confirmation or inductive logic at the foundations of decision theory should be considered in the light of second-order probabilities and Bayes' theorem. The relation between evidence and information should be clarified and a theory of the value of evidence as based on its information content should be developed.

References

Arrow, K. Alternative approaches to the theory of choice in risk taking situations. Econometrica, 1951, 19, 404-437.

Atkinson, F., Church, J., & Harris, B. Decision procedures for finite decision problems under complete ignorance. Annals of Mathematical Statistics, 1964, 35, 1644-1655.

Blackwell, D., & Girshick, M. A. Theory of Games and Statistical Decisions. New York: John Wiley, 1954.

Chernoff, H. Rational selection of decision functions. Econometrica, 1954, 22, 422-443.

Crowell, R. H., & Williamson, R. E. Calculus of Vector Functions. Englewood Cliffs, New Jersey: Prentice Hall, 1962.

Efron, B. Note on decision procedures for finite decision problems under complete ignorance. Annals of Mathematical Statistics, 1965, 35, 691-697.

Ellsberg, D. Risk, ambiguity, and the savage axioms. Quarterly Journal of Economics, 1961, 75, 643-669.

Good, I. J. Probability and the Weighing of Evidence. London: Chas. Griffin and Co., 1950.

Good, I. J. The Estimation of Probabilities: An Essay on Modern Bayesian Methods. Cambridge: MIT Press, 1965.

Hodges, J., & Lehmann, E. The use of previous experience in reaching statistical decisions. Annals of Mathematical Statistics, 1952, 23, 396-407.

Howard, R. A. Prediction of replacement demand. In G. Kreweras and G. Morlat (Eds.), Proceedings of the Third International Conference on Operational Research. London: English Universities Press, Ltd., 1963. Pp. 905-918.

Hurwicz, L. Some specification problems and applications to econometric models (abstract). Econometrica, 1951, 19, 343-344.

Jeffrey, R. The Logic of Decision. New York: McGraw-Hill, 1965.

Jeffreys, H., & Jeffreys, B. S. Mathematical Physics. Cambridge: The University Press, 1956.

Lindley, D. V. Probability and Statistics from a Bayesian Viewpoint, Vol. I. Cambridge: The University Press, 1965.

Luce, R. & Raiffa, H. Games and Decisions: Introduction and Critical Survey. New York: John Wiley, 1956.

Luce, R. & Suppes, P. Preference, utility, and subjective probability. In R. Bush, E. Galanter, and R. Bush (Eds.), Handbook of Mathematical Psychology, Vol. III. New York: John Wiley, 1965. Pp. 249-410.

Milnor, J. Games against nature. In C. Coombs, R. Davis, and R. Thrall (Eds.), Decision Processes. New York: John Wiley, 1954. Reprinted in M. Shubik (Ed.), Game Theory and Related Approaches to Social Behavior. New York: John Wiley, 1964. Pp. 120-131.

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

Shannon, C. E., & Weaver, W. The Mathematical Theory of Communication. Urbana: University of Illinois Press, 1949.

Suppes, P. The role of subjective probability and utility in decision-making. In R. Bush, E. Galanter, and R. Luce (Eds.), Readings in Mathematical Psychology, Vol. II. New York: John Wiley, 1965. Pp. 503-515. Reprinted from the Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954-1955.

Tintner, G. The theory of choice under subjective risk and uncertainty. Econometrica, 1941, 9, 298-304.

Footnotes

[1] The author is deeply indebted to Professors Patrick Suppes and Howard Smokler for helpful advice and comments concerning this paper. The author has had a number of helpful conversations with Professor John Breakwell, Professor Donald Dunn, Professor Ronald Howard, Mr. Joe Good, and Mr. Michael Raugh. The advice of these persons does not, of course, necessarily imply their consent to the conclusions of the paper.

[2] Usually we can characterize the uncertainty in a decision situation as the sum of $H(\mathcal{E}(\vec{\xi}))$ and $H(f)$. If, however, f itself is not precisely known, the uncertainty associated with alternative possible fs must be added in, and so on.

[3] R. A. Howard (1963) utilizes what are essentially probability distributions over probability distributions by considering a probability density function for the parameters of another probability density function.

[4] Professor Suppes points out to me that, though there is a rich body of results in meta-mathematics, mathematicians apparently feel no need to derive formal results concerning meta-mathematics in a meta-meta-mathematics.

[5] Laplace's law of succession is derived from Bayes' theorem and the assumption of a uniform prior for ξ_i. If the uniform prior is changed to any of the possibilities given in equation (6), the following generalization of the law of succession can be derived: $p_{r+1}(\omega_i) = (n + 1)/(r + m)$, where $p_{r+1}(\omega_i)$ is the (expectation of) the probability

that on the $r + 1$st trial ω_i will occur, n is the number of times it has occurred in the previous r trials, and m is the number of states of nature. Since I completed this paper, Mr. R. Toumela has pointed out to me that Good (1965) has discussed notions that are formally analogous to $f(\vec{\xi})$. Good mentions that this generalized version of the law of succession was known to Lidstone in 1925.

SUBJECTIVE PROBABILITIES UNDER TOTAL UNCERTAINTY[1]

by

Dean Jamison and Jozef Kozielecki

Introduction

Humans must frequently choose among several courses of action under circumstances such that the outcome of their choice depends on an unknown "state of nature". Let us denote the set of possible states of nature by Ω and consider Ω to have m members that are mutually exclusive and collectively exhaustive--$\omega_1, \omega_2, \ldots, \omega_m$. The vector $\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_m)$ is a probability distribution over Ω if and only if $\sum_{i=1}^{m} \xi_i = 1$ and $\xi_i \geq 0$ for $i = 1, \ldots, m$. ξ_i corresponds to the probability that ω_i will occur.

Edwards (1954), Luce and Suppes (1965), and others, dichotomize experimental situations involving choice behavior in the following way. If the decision-maker's choice determines the outcome with probability 1 (i.e., one of the ξ_i's is equal to 1), then the experimental situation is one with <u>certain</u> outcomes; otherwise, the outcome is <u>uncertain</u>. Further distinctions are in common use regarding the word "uncertain". If the subject knows the probability distribution over the outcomes, i.e., if he knows $\vec{\xi}$, his choice is <u>risky</u>; if he only has "partial knowledge" or "no knowledge" of $\vec{\xi}$ his choice is partially or totally <u>uncertain</u>. We shall use "total uncertainty" in this last way; our purpose is to examine the structure of a subject's beliefs when he has no knowledge of $\vec{\xi}$, that is, when the S is totally uncertain. Jamison (1967) has proposed a definition of total uncertainty that is an extension of the Laplacian principle of insufficient reason.

This definition and some of its implications will be described briefly here as theoretical background for our experimental results.

Consider the set of all possible probability distributions over Ω, that is, the set of all vectors $\vec{\xi}$. Let us denote this set by $\underset{\sim}{\approx}$ and describe the decision-maker's knowledge of $\vec{\xi}$ by a density $f(\xi_1, \xi_2, \ldots, \xi_m) = f(\vec{\xi})$ defined on $\underset{\sim}{\approx}$. If $f(\vec{\xi})$ is an impulse (δ function) at $\vec{\xi} = (1, 0, \ldots, 0)$ or $\vec{\xi} = (0, 1, 0, \ldots, 0)$, or $\ldots$, $\vec{\xi} = (0, 0, \ldots, 1)$, then decision-making is under certainty. If $f(\vec{\xi})$ is an impulse elsewhere in $\underset{\sim}{\approx}$, the decision-making is risky. If $f(\vec{\xi})$ is a constant, the decision-maker is, by definition, totally uncertain of $\vec{\xi}$. The intuitive motivation for this definition is that if $f(\vec{\xi})$ is a constant, no probability distributions over Ω are more likely than any others. Partial uncertainty occurs when $f(\vec{\xi})$ is neither an impulse nor a constant.[2]

If K is the constant value of $f(\vec{\xi})$ under total uncertainty, then:

$$\iint \underset{\underset{\sim}{\approx}}{\cdots} \int K d\xi_m d\xi_{m-i} \ldots d\xi_1 = 1. \qquad (1)$$

Evaluating this definite integral enables us to find K, which turns out to be $(m - 1)!\sqrt{m}/m$. The probability that ξ_1 is greater than some specific value, say C, is given by:

$$\text{prob}(\xi_1 > C) = \int_C^1 \int_0^{1-\xi_1} \cdots \int_0^{1-\xi_1-\xi_2-\ldots-\xi_{m-2}} \sqrt{m}\, K d\xi_{m-2} d\xi_{m-3} \ldots d\xi_1$$

$$= (1 - C)^{m-1}. \qquad (2)$$

One minus $\text{prob}(\xi_1 > C)$ is simply the probability that $\xi_1 \leq C$ or the cumulative marginal for ξ_1, which we shall denote by $F_1(C)$. By symmetry $F_1(C) = F_2(C) = \ldots = F_i(C) = \ldots F_m(C)$; thus we have:

$$F_i(C) = 1 - (1 - C)^{m-1}. \qquad (3)$$

Fig. 1 shows $F_i(C)$ for several values of m.

Insert Fig. 1 about here

The derivative of the marginal cumulative is the marginal density, which we shall denote $f_i(C)$:

$$f_i(C) = \frac{dF_i(C)}{dc} = (m - 1)(1 - C)^{m-2}. \qquad (4)$$

Fig. 2 shows $f_i(C)$ for several values of m.

Insert Fig. 2 about here

The purpose of our experiment was to determine if the normative model just described for belief under total uncertainty approximates the actual structure of Ss beliefs. To achieve this purpose we placed Ss in a situation of total uncertainty and then empirically determined the cumulative $F_i(C)$ for a number of values of m.

Method

Subjects

The Ss were 30 students from Stanford University fulfilling course requirements for introductory psychology. Each participated in one experimental session of approximately 30 minutes duration. Ss were run individually.

Experimental design and procedure

At the onset of the experiment the Ss were told that the experimenter wished to examine his beliefs concerning the outcome of a hypothetical scientific experiment about which the Ss would be given very little information. A particle measuring device would be placed into an environment in which there were m distinct types of particles. The Ss were told that the particle measuring device counted the number of each type of particle striking it in any given time interval and that it was left in the environment until a total of 1000 particles of the m types had been detected. A copy of the instructions is included as an Appendix.

The experiment consisted of three series run with 10 subjects each; in Series I $m = 2$, in Series II $m = 4$ and in Series III $m = 8$. For $m = 2$, the particles were named ω and ϵ; for $m = 4$ they were named ω, ϵ, δ, and ψ; and for $m = 8$ they were named ω, ϵ, δ, ψ, ξ, ζ, χ, and θ. The experimenter asked the Ss a list of questions of the following form: "What do you think the probability is that the particle measuring device counted less than 500 ϵ-particles among the 1000 total"? The Ss were asked to write their responses as a two-digit decimal on a 3" x 5" card, then to turn the card over. Each S was given all the time he wished to answer. With $m = 2$ and $m = 4$, Ss were asked for each particle what he thought the probability was that less than 25, 100, 200, 350, 500, 650, 800, 900, and 975 of that type of particle would be among the 1000 counted. The question order was random.

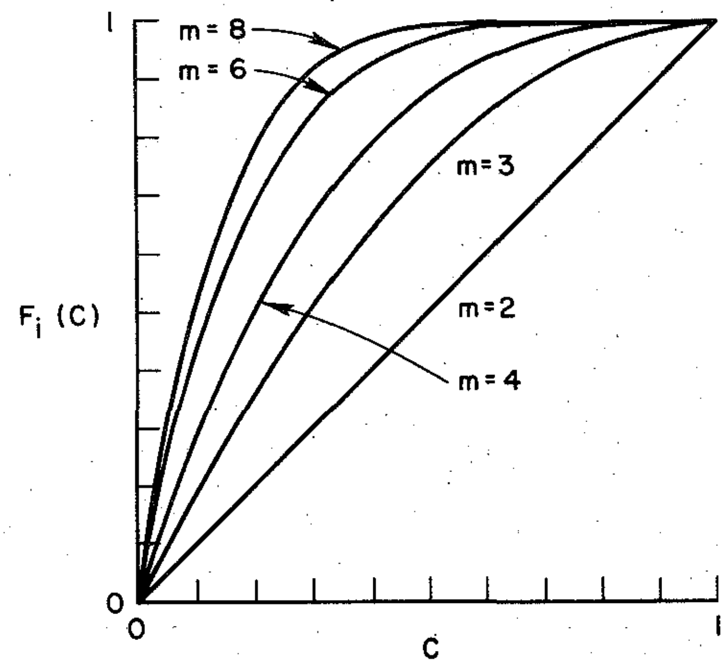


Figure 1. Marginal cumulatives under total uncertainty.

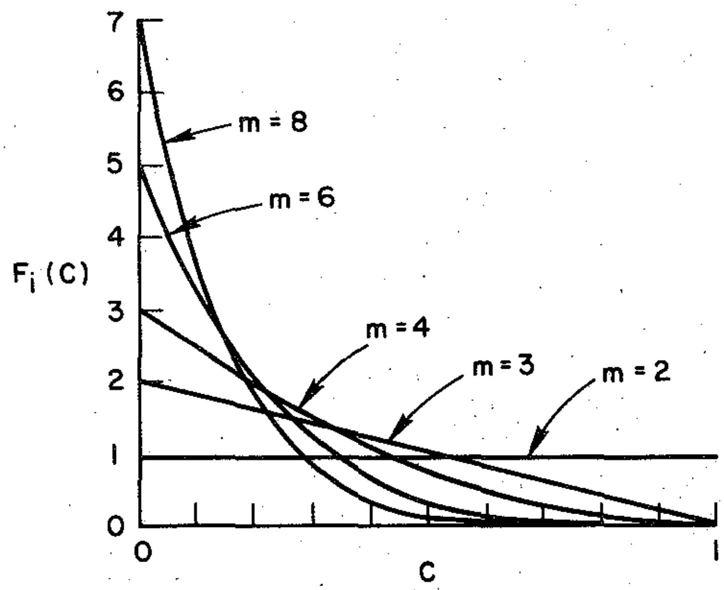


Figure 2. Marginal densities under total uncertainty.

For $m = 8$ the 350, 650, and 975 questions were deleted. After the experiment Ss were asked questions concerning their method of answering.

Results

The results were a number of discrete values of $F_i(C)$ for each particle and for each subject. For each particle we pooled the results of the 10 subjects who were tested for each value of m. We then did a standard analysis of variance test to ascertain whether any significant differences existed in Ss' responses for the different particles. As Table 1 shows, there were no significant differences among particles at the .05 level.

Table 1 - Analysis of Variance on Differences Among Particles

Series	df	F	Significance Level
$m = 2$	1/162	.10	$p > .05$
$m = 4$	3/324	.35	$p > .05$
$m = 8$	7/432	1.03	$p > .05$

What Table 1 indicates is that Ss accepted Laplace's principle of insufficient reason; they showed no preference for any particular particles. The Ss' answers to questions after experimentation confirmed this result. Since Ss accepted the principle of insufficient reason, results were also pooled across particles. Figs. 3a, 3b, and 3c show the normative cumulatives $F_i(C)$ as well as our data points pooled across Ss and particles for each of the three different values of m. The median responses shown in the figures correspond closely to the means.

Insert Figs. 3a, 3b, and 3c about here

Fig. 3a clearly indicates that for $m = 2$ the normative model fits the data very well, whereas for $m = 4$ and $m = 8$ there is some relation between the normative model and the data but not a fit.

The variance analysis of the data that is displayed in Table 2 indicates that when $m = 2$ there is no significant difference between the normative curve and the data at the .05 level. For $m = 4$ and $m = 8$ the difference between the normative curve and the data is significant at the .001 level.

Table 2 - Analysis of Variance on Differences between Normative Models and Data

Series	df	F	Significance Level
$m = 2$	1/162	1.36	$p > .05$
$m = 4$	1/162	100.33	$p < .001$
$m = 8$	1/108	229.52	$p < .001$

Since the normative curves fit the data so poorly when $m = 4$ and $m = 8$, we decided to use a one-parameter curve of the same form as the normative model and fit it to the data by least squares techniques. That is, we wished to describe the data by a curve of the following nature:

$$F_i^*(C) = 1 - (1 - C)^{m^*-1}. \qquad (5)$$

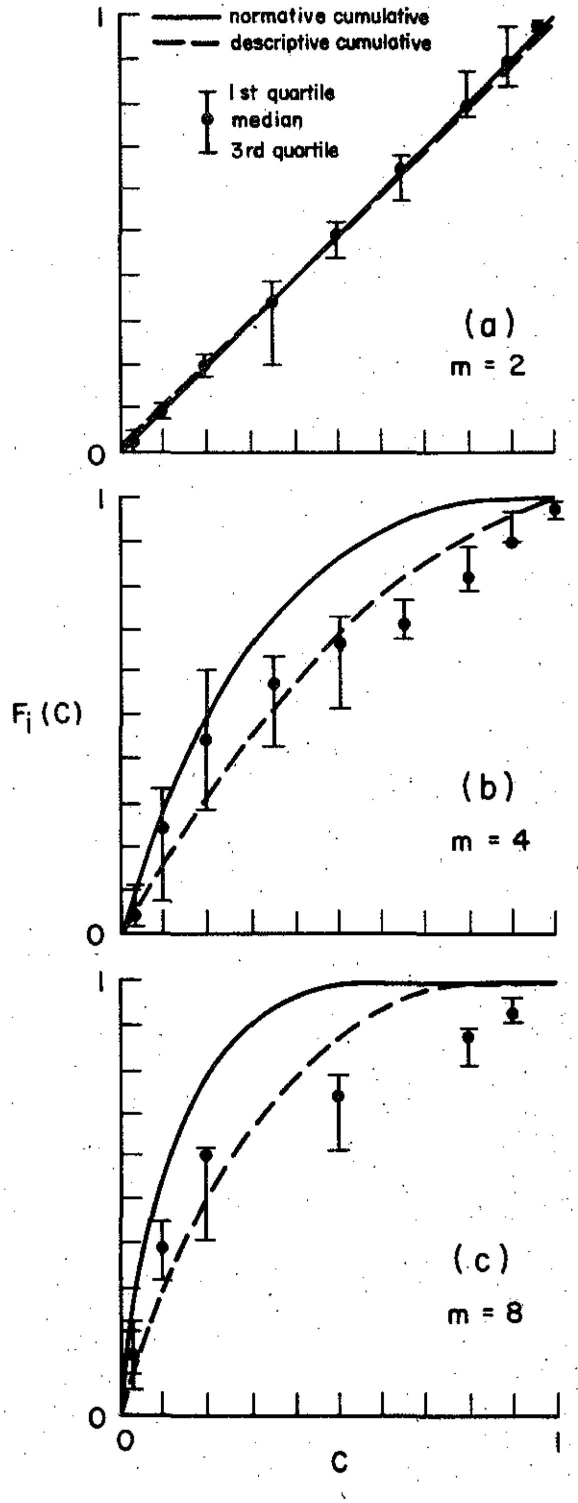


Figure 3. Normative and descriptive cumulatives for (a) m = 2, (b) m = 4, and (c) m = 8.

The $*$ superscript indicates that $F_i^*(C)$ and m^* are descriptive rather than normative. The least squares estimate of m^* is that value of m^* which minimizes the Δ given in equation (6).

$$\Delta = \sum_{j=1}^{9} \left\{ \left[1 - (1 - C_j)^{m^*-1} \right] - P_{j,obs} \right\}^2 , \tag{6}$$

where $C_1 = 25/100$, $C_2 = 100/1000$, etc., and $P_{j,obs}$ is the mean probability estimate of the Ss. Table 3 shows the least squares estimates of m^* computed numerically on Stanford's IBM 7090.

Table 3 - Least Squares Estimates of m^*

Series	m^*	Δ
m = 2	1.98	.00
m = 4	2.63	.04
m = 8	4.05	3.07

Figs. 2b and 2c show $F_i^*(C)$ based on the values of m^* given in Table 3.

Our data indicate that Ss' beliefs are quite close to the normative model for $m = 2$, scarcely a surprising result. For $m > 2$ Ss' beliefs shift toward the normative model, but not sufficiently far. The reason for this is suggested in Figs. 4a, 4b, and 4c where $f_i(C)$ and $f_i^*(C)$ are plotted. ($f_i^*(C)$ is the descriptive density based on the value of m^* given in Table 3 inserted into equation (4).)

Insert Figures 4a, 4b, and 4c about here

Fig. 4 shows that Ss underestimate probability density when the density is relatively high and overestimate the density when the density is relatively low. When the density is constant (m = 2), they neither underestimate nor overestimate it. This is a generalization to situations involving total uncertainty of the well-known work of Preston and Baratta (1948) and others who have shown that Ss tend to underestimate high probabilities and overestimate low ones.

Discussion

Our findings corroborate the results of Cohen and Hansel (1956) that Ss tend to apply the principle of insufficient reason if they are given no information. In addition, the phenomenon of underestimating high probabilities and overestimating low is shown to have a direct analog in situations involving probability densities. Here Ss underestimate regions of high density and overestimate regions of low density.

Our results have an important bearing on the question of the consistency of Ss' beliefs. An individual's beliefs (subjective probability estimates) are said to be incoherent if an alert bookmaker can arrange a set of bets based on the person's probabilities such that the person can win in no eventuality. When the probabilities are well known (i.e., when $f(\vec{\xi})$ is an impulse at some particular $\vec{\xi}$) a necessary and sufficient condition for coherence is that the sum of the

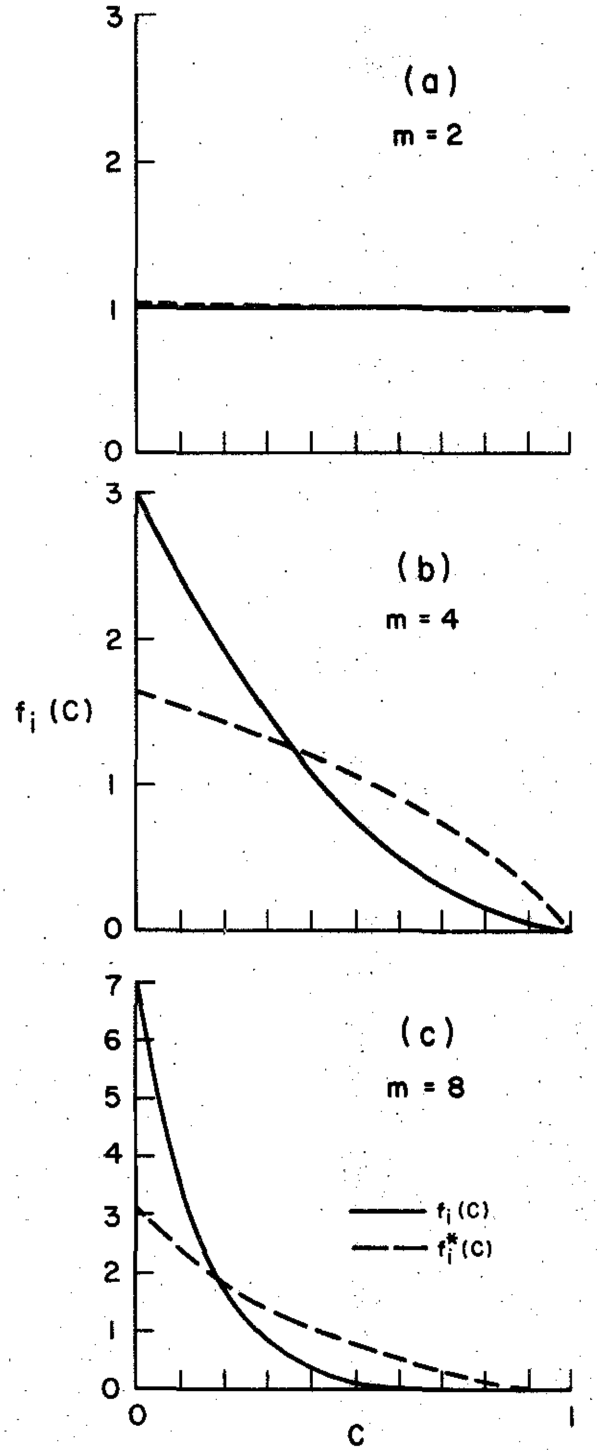
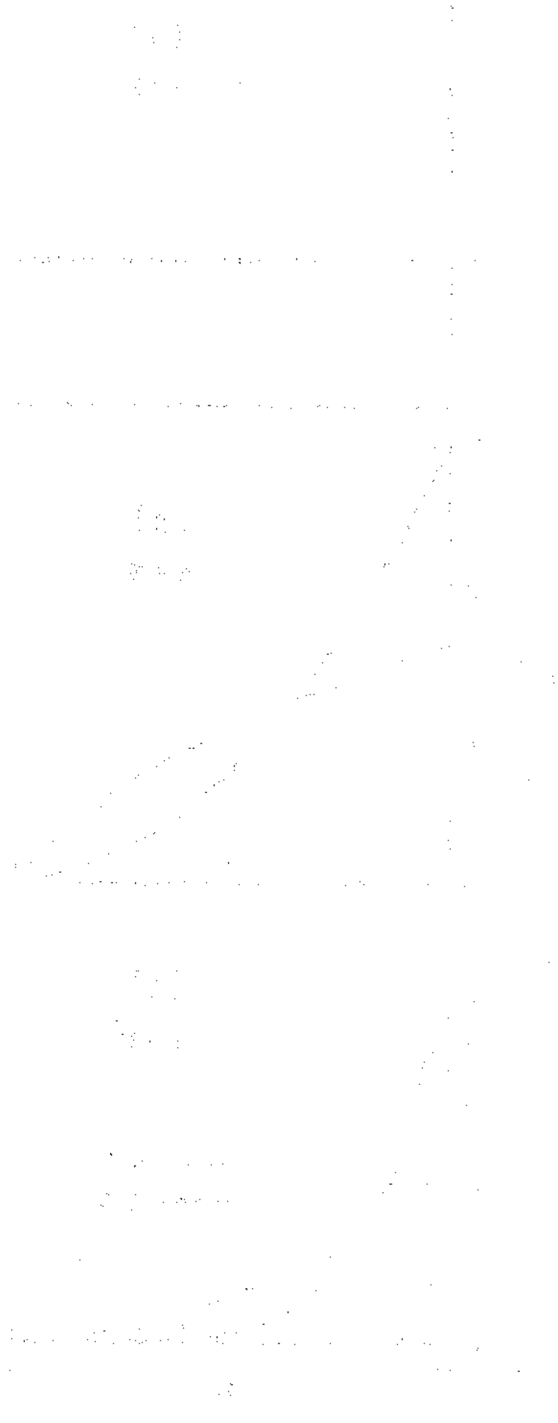


Figure 4. Normative and descriptive densities for (a) m = 2, (b) m = 4, and (c) m = 8.



probabilities of mutually exclusive and collectively exhaustive events be unity (see Shimony (1955)). Analogously, a necessary (but <u>not</u> sufficient) condition for coherence when probabilities are not well known is that the sum of the expectations of the probabilities be unity. That is, the R defined below must equal one.

$$R = \sum_{i=1}^{m} \int_{0}^{1} C\ f_i(C)dC. \tag{7}$$

Since all the f_is are equal (from the insignificance of the differences among particles),

$$R = m\int_{0}^{1} C(m^* - 1)(1 - C)^{m^*-2}dC = \frac{m}{m^*}. \tag{8}$$

Thus $R = 1$ only when $m^* = m$. It is clear from Table 3 that when $m = 4$ and $m = 8$, the Ss in our experiment had beliefs that were strongly incoherent.

Our study is an examination of the <u>static</u> structure of a person's beliefs when he is in a situation of total uncertainty. The natural extension of this work is to examine the <u>kinematics</u> of belief change when the S is given information relevant to the situation. Work on the kinematics of belief change when probabilities are well known is reported in a number of papers in a volume edited by Edwards (1967).

Appendix

Instruction to Subjects

The instructions that were read to the Ss when $m = 8$ are given below. The instructions for $m = 2$ and $m = 4$ are the same except for obvious modifications.

* * * * * * * *

We are running an experiment to examine the nature of a person's intuitions concerning situations where he has little or no concrete evidence to guide him. You will be asked to estimate the likelihood of certain propositions concerning a hypothetical scientific experiment. While there are no absolutely "right" or "wrong" answers, some answers <u>are</u> better than others. Your response will be evaluated against a hypothetical ideal subject.

Let me now describe the hypothetical scientific situation about which we wish to examine your beliefs. A particle measuring device is placed into an environment where there are 8 distinct types of particles which we shall designate by letters of the Greek alphabet-- ω, ϵ, ψ, δ, ζ, ξ, χ, θ. What the particle measuring device does is count the number of each type of particle that hits it in a given time interval. We leave the counter in the environment until it has been struck by a total of 1000 particles of the 8 types. Do you remember what the 8 types were? Prior to the experiment you are assumed to have absolutely no knowledge about the relative numbers of the 8 types of particles except that some of each may exist and that no other type of particle is in the environment. Given this scant information, and nothing else, we want to examine your intuitions concerning how many of each type of particle will be included in the 1000 measured by the detector.

The questions we ask you concerning your beliefs will be of the following form: What do you think the probability is that there are *less* than some specific number of, say, ϵ-particles *among the 1000 counted*? This statement would be true, of course, if there were 0, 1, 2, 3, ... , or any number up to that number of ϵ-particles among those counted but it would *not* be true if there were more than that many ϵ-particles. What you are being asked is how *likely* is it that there are less than that number of ϵ-particles? If you believed that there were *certainly* less than that number of ϵ-particles, you would tell us that the probability of there being less than that number is ..... If, on the other hand, you believed that there were certainly more than that number of ϵ-particles, you would tell us that the probability that there is less than that number is ..... If you believe that it is equally likely that there are more than that number as less, you would say the probability is .5. You can give us any probability between zero and one.

Perhaps a more concrete example will help make things clear. Consider an ordinary die such as this one. What do you think the probability is that if I roll this die a number *less* than 2 will be on the upturned face? What do you think the probability is of less than 5? Clearly, the probability of less than 2 must be smaller than the probability of less than 5. Well, you see, this is exactly the same type of question that we shall be asking concerning particles counted by our counter. The only difference is that with a die you already; have a good idea of the probability asked for, whereas in this experiment we are asking for your intuitions concerning unknown probabilities.

Let me now ask you a few sample questions before we begin. First, what do you think the probability is that there are less than 1001 ψ-particles among those counted? [Explain if answer is wrong.] What do you think the probability is that there are less than 950 ω- particles among the 1000 counted? Less than 75 ϵ? Remembering, again, that there are 8 types of particles, what do you think the probability is of less than 500 ϵ-particles? Less than 950 δ? [No feedback is given on last 4 questions.]

In front of you is a stack of 3" x 5" cards that you will write your replies on. Could you write your replies as a two-digit decimal ... like so.

Before we begin, please feel free to ask any questions you might have.

References

Atkinson, F., Church, J. & Harris, B. Decision procedures for finite decision problems under complete ignorance. Ann. Math. Statist., 1965, 35, 1644-1655.

Cohen, J. & Hansel, M. Risk and Gambling. London: Longmans Green.

Edwards, W. The theory of decision making. Psychol. Bull., 1954, 5, 280-417.

Edwards, W. (Ed.) Revision of opinions of men and man-machine systems. Special issue of IEEE Trans. on Human Factors in Electron., 1967, 7, 1-63.

Jamison, D. Bayesian decisions under total and partial ignorance. This technical report.

Luce, R. D. & Raiffa, H. Games and Decisions: Introduction and Critical Survey. New York: John Wiley & Co., 1956.

Luce, R. D. & Suppes, P. Preference, utility, and subjective probability. In R. Luce, R. Bush & E. Galanter (Eds.), Handb. Math. Psychol., Vol. 3. New York: John Wiley & Co., 1965. Pp. 250-410.

Preston, M. G. & Baratta, P. An experimental study of the auction value of an uncertain outcome. Amer. J. Psychol., 1948, 61, 183-193.

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

Shimony, A. Coherence and the axioms of confirmation. J. Symbolic Logic, 1955, 20, 1-28.

Footnotes

[1]The work reported here was performed at the Institute for Mathematical Studies in the Social Sciences, Stanford University. The authors wish to thank the Institute's Director, Professor Patrick Suppes, for his support of and helpful comments concerning this work. Professor Richard Smallwood made some helpful comments on an early draft of this paper.

[2]Luce and Raiffa (1956) review normative theories of decision making under total uncertainty. Extensions of these other theories may be found in Atkinson, Church, & Harris (1965). Savage (1954) presents a number of objections to the probability of probabilities approach used here. These alternatives and objections are all discussed in Jamison (1967).