A Bayesian Approach to Decision Making in Applied Meteorology[1]

EDWARD S. EPSTEIN

Department of Engineering Mechanics, The University of Michigan

(Manuscript received 14 July 1961, in revised form 29 January 1962)

ABSTRACT

The problem of decision making in applied meteorology is approached from the point of view of decision theory and subjectivist statistics. The modern concept of "utility" is discussed, and optional rules for decision making based on the availability of a limited amount of meteorological data are presented and discussed. Bayes' theorem forms the basis for the statistical estimation of the frequencies of various alternative weather events. The method is applied to a single example for the purpose of illustration, but it is emphasized that the generality of these techniques is great and that they warrant further study.

1. Introduction

In any discussion of the role of weather prediction in the decision making process, it must be kept clear that there is a duality of roles. The meteorologist analyzes and evaluates the present and past weather, and estimates the future state of the weather; the entrepreneur or other user of the meteorological service must be able to evaluate these predictions and analyses and translate them into the most favorable or most desirable course of action. It has been pointed out quite early that probabilistic forecasts are better suited to economic decision making than are categorical forecasts (Thompson, 1952). This is especially true in the domain of public forecasts, or wherever there is no close rapport between the meteorologist and the decision maker. However, given a maximum of communication between these two groups, as is becoming increasingly common in many segments of the economy, much more efficient use could be made of weather forecasts if their stochastic aspects were to be appreciated more adequately.

The basis for translating forecasts into optimum decisions has been taken by numerous authors to be the economic gain to be derived from particular combinations of actions and subsequent weather. For example Thompson (1952), Thompson and Brier (1955), Gringorten (1958, 1959), Borgman (1960), Gleeson (1960), and Nelson and Winter (1960) have all carried out analyses based on cost or profit matrices. In all of these analyses, a single optimum strategy is selected for any given current state of the weather (or forecast), based on consideration of how to maximize economic expectation, or in Gleeson's analysis, on how to maximize minimum economic expectation. There are two aspects of these analyses on which we shall comment in this paper: the inadequacy of the cost or profit matrix as the principal guide to choosing the proper decision; and the manner in which our meteorological knowledge and experience have been utilized in assessing the probabilities of the future states of the weather.

The analysis which will follow is based very largely on recent advances in the theory of decision making. Numerous books have been written on the subject. Luce and Raiffa (1957) offer a comprehensive and critical review of the subject, while a relatively mathematical, axiomatic treatment of the subject is available in a recent book by Raiffa and Schlaifer (1961). Another treatment of the subject, by Schlaifer (1959), while simple mathematically, is quite penetrating intellectually.

Many schemes have been devised to aid the decision maker; some are applicable to situations involving complete certainty of the outcome, some to complete ignorance (a concept that not all authorities consider meaningful), and some to partial ignorance of future events. The first of these need not concern us insofar as the future states of the weather are concerned; nor, I should hope, should the second. Gleeson's Method B, which employs the theory of games, is one of several available techniques which fall in this second category. The third situation, referred to as decision making under risk, considers that one of several future events may occur, each with specified probability. In order to be applicable to the meteorological situation we must apply this last case to situations in which the frequencies of the various future states are to be estimated on the basis of accumulated data.

In the following section we will consider the recommended procedures for decision making under risk and

[1] Publication No. 54 from The Meteorological Laboratories, The University of Michigan.

also the Bernoulli–Ramsey (1931)—von Neumann–Morgenstern (1947) concept of utility, which is an integral part of decision theory. The third section will deal with methods of computing the probabilities of future states of the weather through application of Bayes' theorem [see, for example, Savage (1954) or Miller (1961)]. We will then proceed to suggest several methods for incorporating the probabilities of section 3 with the decision theory described in section 2. Finally an example taken from the literature will be used to illustrate the application of the methods developed in this paper.

2. Decision making under risk, and utility

In this section I propose to discuss briefly, and in very general terms, for those who may not already be familiar with the subject, the basic concepts of utility and decision making under risk. For those desiring a more complete exposition of the subject, the treatments by Luce and Raiffa (1957) and Savage (1954) are recommended. Savage, in particular, offers an informative historical review of the concept of utility.

The subject of utility is adaptable to a careful and axiomatic treatment; indeed Ramsey, and later von Neumann and Morgenstern, give a proof of the existence of utility in the sense in which we shall use it. The present treatment, however, shall be a heuristic one.

Consider that you are given a choice between (i) an outright grant of $1000 and (ii) a lottery in which you would receive $10,000 with probability p or receive nothing with probability $1-p$. Certainly, if p is sufficiently near zero, practically anyone would choose (i); while for p sufficiently near one, almost all would choose (ii). It is clear that for any individual there exists some value of p, $0<p<1$, such that he is completely indifferent between the two alternatives. If this indifference occurs for $p=p_0$, then we say that the utility of $1,000,

$$U(\$1{,}000)=p_0U(\$10{,}000)+(1-p_0)U(\$0).$$

Note that a person who needs $1,000 (but no more) for some particular purpose (say, an investment in a "sure thing") may prefer (i) even if p is as large as $\frac{3}{4}$. Another person may be in the position that $1,000 would not be of very great personal value, but that $10,000 would rescue him from severe difficulty (pay a ransom, perhaps, to use an example due to Bernoulli); he might very well be willing to accept the lottery (ii) even if there were but one chance in fifty of receiving the $10,000.

In this simple example, the utilities which are finally assigned to the various alternative events are expressions of an individual's preference; different individuals will have different preferences. (I, for one, have an aversion to gambles and my personal point of indifference would occur for some $p_0>0.1$.) The actual values of the utilities assigned to the alternatives are arbitrary as to zero point and scale. Thus, if the utilities of $0, $1,000, and $10,000 are taken to be U_0, U_1, and U_{10}, respectively, then the triplet of utilities aU_0+b, aU_1+b, and $aU_{10}+b$, $a>0$, would express the very same preferences among alternative lotteries involving only these three sums of money.

This general notion may be extended to the more general situation in which an individual is offered a choice among a set of lotteries, L_i, each of which consists of the same set of alternative prizes (not necessarily cash) A_j $(j=1,\cdots,n)$ but with probabilities p_{ij}, $0<p_{ij}<1$, and $\sum_{j=1}^{n} p_{ij}=1$. By considerations such as those given above, the utilities of each of the alternatives A_j can be determined in terms of the (arbitrary) utilities of the most desirable and least desirable of the A_j, assuming an element of consistency in the preferences. Then one can assign a utility to each of the lotteries:

$$U(L_i)=\sum_{j=1}^{n} p_{ij}U(A_j),$$

where $U(A_j)$ is the utility of A_j. Then, if one selects that lottery L_i such that $U(L_i)\geqq U(L_j)$ for all $j\neq i$, he is acting in consistency with his own preferences.

It is important to note here that no consideration is given, or need be given, to repetitive trials or to an outcome over the "long haul." The significant point is that the rational and consistent individual, when faced with a choice from among alternatives, or lotteries of alternatives of specified probability, will act as though he were *maximizing the expected value of his utility*.[2] It remains to be determined how one would act if only estimates of the frequencies of the alternatives were available, and how, especially in the meteorological context of this paper, the available data would be used to estimate these frequencies.

3. Bayes' theorem and the estimation of the probabilities of future states of the weather

The applicability of Bayes' theorem to the problem of meteorological forecasting has been recognized by Nelson and Winter (1960); but to the author's knowledge, the only case in which it has been adequately employed in this field is the study of Miller (1961). Bayes' theorem has appeared in texts on probability theory [e.g., Feller (1950)] for many years, but only since the rise of the "subjectivist" school of statistics, and especially the work of Savage (1954), and Schlaifer (1959), has it been respectably applied to problems of prediction.

We shall consider that the current state of the

[2] According to Schlaifer (1959) this is "the *practical* way of choosing the 'best' act."

atmosphere has been categorized in a fashion such as described by Gringorten (1955, 1958) and Gringorten, Lund and Miller (1956), or, alternatively, that the forecaster has issued one of a finite set of permissible forecasts. This current state of the weather, or the issuance of the particular forecast, we shall describe as the condition S. In this manner we limit somewhat the scope of the study from a conceptual point of view. On the other hand, we are in no way limited to considering either "objective" or "subjective" forecasting procedures. In the past, the situation S has been observed to occur n times, and the subsequent states of the weather have been recorded. From these records it is possible to count the numbers of times (x_i) that each of the several mutually exclusive and exhaustive weather categories (W_i) of operational significance to the decision maker has occurred. We assume that there exist conditional probabilities of occurrence (p_i) of each of the W_i, given that the situation S has prevailed previously. The x_i will then be samples drawn from a multinomial distribution

$$F(x|p)=\frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}. \quad (1)$$

Our initial knowledge of the values of the p_i is extremely limited, although we may have some *a priori* judgments (the "personal" probability discussed by Savage (1954)) as to what may or may not be reasonable values. Thus, we feel quite certain that the temperature in December will not reach 100F in Fairbanks, Alaska; we consider it not likely that more than 3 inches of precipitation will fall on Washington, D. C., on any given day. These notions may be expressed in terms of $h(p)$, the *a priori* joint probability density of the p_i.

What we seek, however, is not an *a priori* distribution of the p's, but it is rather the *a posteriori* probability distribution. We want to know the probability that the p's have a particular set of values, given the available information as to the events subsequent to n other occurrences of S. In other words, we wish to know the probability density function $f(p|S,x)$.[3] According to Bayes' theorem, this probability density may be written as

$$f(p|x)=\frac{F(x|p)h(p)}{H(x)},$$

where $F(x|p)$, the conditional probability density of x given p (Eq 1), is proportional to the likelihood function of p for x. The term $H(x)$, the *a priori* probability density of x is

$$H(x)=\int_p F(x|p)h(p)dp$$

but the details of its evaluation need not concern us. Note that $H(x)$ is independent of p.

Before proceeding with the development let me first comment in general on the use of *a priori* distributions. Although the formal use of *a priori* distributions has not been widely recognized in the meteorological literature, it is evident that they have been used on an informal basis, indeed frequently without the user's awareness. Thus many studies of a statistical nature have involved the editing of data before any statistics are computed. This editing, or the elimination of data which do not appear to be reasonable, is an implicit manifestation of the *a priori* distribution of the analyst. He has assigned a subjective *a priori* probability of zero, to data points that fall outside of some "acceptable" region.

As will be pointed out again later, the initial, or *a priori* distribution is important where few additional data are available. However given plentiful data, the final impact of the initial distribution is vanishing small. Indeed if one is given no data, or very few data, how else is one to act except on the basis of the meteorologist's professional judgment—namely his *a priori* distribution?

Given below are three plausible *a priori* distributions. They do not form an exhaustive set; there are advantages and disadvantages to each.

Prior distribution I: Climatology. It would not seem unreasonable to anticipate that the most likely values of the p_i should be near the climatological relative frequencies of the weather events W_i. If we let these climatological relative frequencies be represented by $a_i(\sum_{i=1}^{n} a_i=1)$, then an *a priori* distribution which accomplishes this end is

$$h(p)=\frac{\Gamma(k+1)}{\Gamma(a_1+1)\Gamma(a_2+1)\cdots\Gamma(a_k+1)}p_1^{a_1}p_2^{a_2}\cdots p_k^{a_k}. \quad (2)$$

The term involving the gamma functions is a normalizing factor to meet the requirement that $\int h(p)dp=1$. Eq (2) is plotted in Fig. 1 for the case $k=3$ and $a_1=a_2=a_3=\frac{1}{3}$. Note that there are always $k-1$ independent p's since $\sum p_i=1$. The case $k=3$ is chosen because it is the most complicated that can be represented graphically.

Given this distribution, the *a priori* maximum likelihood estimate of p_i is a_i. The effect of this distribution is to weight the subsequent *a posteriori* distribution toward the climatological mean. Yet this is not a wholly satisfactory choice of a prior distribution for it ignores the fact expressed earlier that the proba-

[3] Following normal usage the vertical bar denotes a conditional distribution of the term(s) preceding the bar, given the occurrence of those following the bar. Throughout this paper, all distributions are conditional on the occurrence of S, and hereafter this condition will be omitted from the notation. Thus $h(p)$ implies the distribution of the p's given S, and $f(x|p)$ the distribution of the x's given p and S.

bilities are conditional upon the occurrence of the event S. The question arises as to whether it would be reasonable to expect, on an *a priori* basis, the weather subsequent to a particular kind of weather situation, S, to be the same as that which we would expect regardless of the initial conditions. The use of this particular *a priori* distribution would seem to represent a lack of faith in the method used for classifying the present weather. On the other hand, one might recognize that it is difficult to devise a classification scheme under which the probabilities of the various subsequent weather events will deviate very markedly from the climatological means. In this sense perhaps Prior Distribution I does represent a reasonable choice in spite of its apparent disregard for the given condition S.

Prior distribution II: Forecaster's confidence. The point was raised above that Eq (2) does not take account of any skill which the forecaster or analyst may introduce into his categorization of the present weather. A prior distribution which does accomplish this is

$$h(p)=\frac{\Gamma(k\epsilon)}{[\Gamma(\epsilon)]^k}(p_1p_2\cdots p_k)^{-1+\epsilon}, \tag{3}$$

where ϵ is a small positive number which may be interpreted as a measure of the degree of confidence one places in the forecaster or forecast scheme. The smaller ϵ is, the greater is the implied confidence. Eq (3) is plotted in Fig. 2 for the case $k=3$, $\epsilon=0.05$. Note that $h(p)$ takes on its largest values (indicating greatest *a priori* likelihood) when any of the p_i are near 1, and is smallest when the p_i take on intermediate values. Thus this distribution expresses the forecaster's or analyst's intent and desire that the weather event S which he has defined foretells the occurrence of just one of the various alternative subsequent weather events, although Eq (3) does not say which one. In this sense it may be regarded as just the opposite to Prior Distribution I, which in effect is saying that we have no confidence in the meteorologist's ability to distinguish extreme or unusual conditions from the climatological norm.

Whereas Prior Distribution I represents a surface which is peaked somewhere in the middle of the permissible range of values of the probabilities, and Prior Distribution II is peaked at the extremes of these permissible ranges, the third prior distribution represents a compromise between these two extremes.

Prior distribution III: Maximum entropy.

$$h(p)=\text{const}=(k-1)!=\Gamma(k). \tag{4}$$

This particular distribution may be thought of as a minimal assumption [equivalent to the "principle of insufficient reason" which may be suitable when one is "completely ignorant" (Luce and Raiffa, 1957)]. It represents a maximum degree of uncertainty, or entropy, within the framework of information theory. It may also be considered a compromise between Prior Distributions I and II. However it must be pointed out that there are certain logical inadequacies in the selection of this particular prior distribution. Consider the trichotomous subdivision into clear, cloudy and rain. Use of the uniform distribution (Eq 4) implies that one's *a priori* expected values (defined by $\int p_i h(p)dp$) of the relative frequencies are $\frac{1}{3}$ for each category. If

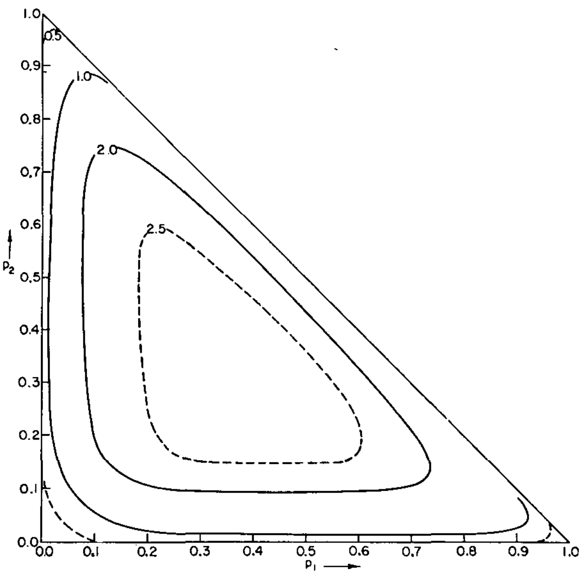

FIG. 1. Prior distribution I, Climatology, for the case $k=3$, $a_1=a_2=a_3=\frac{1}{3}$.

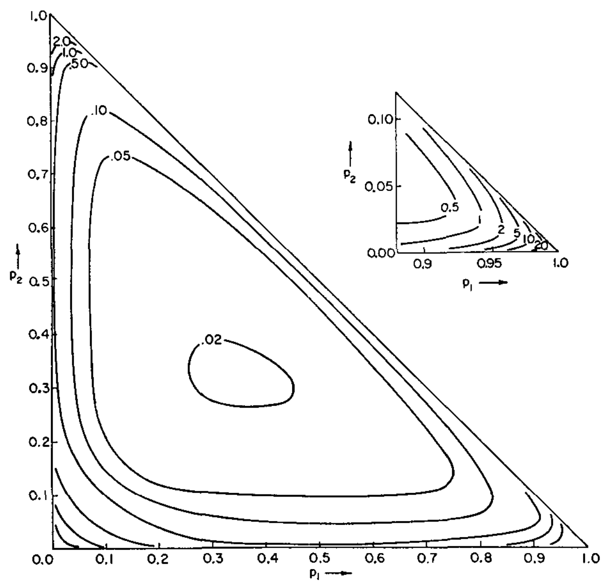

FIG. 2. Prior Distribution II, Forecaster's Confidence, for the case $k=3$, $\epsilon=0.05$.

one then subdivided one of these categories, say cloudy into partly cloudy and overcast, and retained the uniform distribution over the four categories, one is in effect reducing his judgment of the relative frequency of clear skies and of rain to $\frac{1}{4}$ for each.

In the example which will be presented in section 5, each of these *a priori* distributions will be employed and the resulting *a posteriori* distributions will be shown. The reader will then be in a position to judge the influences which these distributions have on the final results.

4. Rules for decision making given stochastic estimates of the probabilities of future events

We have seen, in section 2, that when the probabilities of the several alternative future states are known, it is consistent with one's preferences to select that action which maximizes the expected value of utility. However, when we deal with a frequency distribution of these probabilities the selection of the optimal action is less clear.

In the discussion that follows we will assume that the decision maker must choose among the decisions D_j $(j=1,\cdots,N)$, and that the subsequent weather will fall into one and only one of the categories W_i $(i=1,\cdots,k)$. The result of making the decision D_j, when the subsequent weather falls into the category W_i, has a utility given by U_{ij}. Then $U(D_j)=\sum_{i=1}^{k} p_iU_{ij}$ is a random variable whose distribution is related to that of the p_i.

Listed below are three options, one of which the decision maker may choose when he is operating in this framework. Under some conditions each one will lead to the selection of the same course of action; at other times they could lead to two or three different courses of action and then the final decision will have to be based on consideration of the somewhat subtle differences between the statistical implications of the options.

Option I: Choose that course of action which maximizes the expected value of utility,

$$\tilde{U}(D_j)=\sum_{i=1}^{k} \tilde{p}_iU_{ij}.$$

Option II: Choose that course of action for which the maximum likelihood estimate of the utility is maximum, i.e., maximize

$$\hat{U}(D_j)=\sum_{i=1}^{k} \hat{p}_iU_{ij}.$$

Option III: Choose that course of action which maximizes the probability that the decision will be consistent with the decision maker's preferences as expressed by the utilities. To accomplish this one selects the decision which maximizes, for all $i\neq j$,

$$\text{Prob}\{U(D_j)>U(D_i)\}=\int_{\Gamma_j} f(p|x)dp,$$

where Γ_j is that region of p-space such that $U(D_j)>U(D_i)$ for all $i=1,\cdots,M$, $i\neq j$.

One can see the differences between these three options by referring to Fig. 3. Here we have plotted a hypothetical surface $f(p_1,p_2|x)$, and indicated the regions of the two-dimensional p-space in which the utilities corresponding to each of four possible decisions is maximum. Point A is intended to locate the intersection of the expected values of p_1 and p_2; point B is the location of the joint maximum likelihood estimates, $\hat{p}_1$ and $\hat{p}_2$.

Since point A is in Γ_1, Option I dictates the selection of decision D_1; since point B is in Γ_2, Option II dictates the selection of decision D_2. Use of Option III requires the selection of Γ_4, since the volume under the surface $f(p_1,p_2|x)$ and over the region Γ_4 is larger, apparently, than that over any other region.

Some discussion of these options is certainly called for here, but it should first be emphasized that some of the distinctions among them are not yet completely understood by the author and must be subjected to further study. In the sense that the subjectivist statistician would choose that decision which maximizes the expected value of utility, where the probabilities involved in the various options are his personal probabilities—i.e., the expected values of his *a priori* distribution—he would certainly act similarly given an

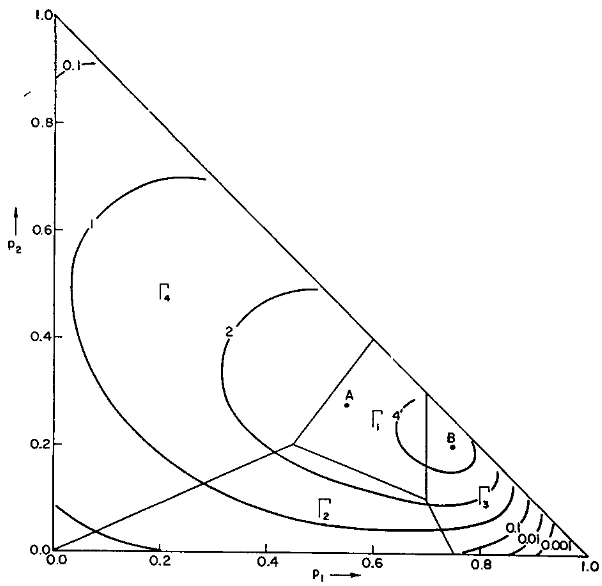

FIG. 3. Schematic *a posteriori* distribution of two probabilities, showing regions of the p-space in which each of four decisions yields maximum utility. Point A locates the expected values, and point B the maximum likelihood estimates of p_1 and p_2.

a posteriori distribution, and select Option I. Indeed, it would appear that one need not be a confirmed subjectivist to come to the same conclusion. Given reasonable *a priori* distributions and/or many data, Option I will, *in the long run*, maximize the average utility.

Option II apparently suffers from the inadequacy that only a small portion of the final distribution surface is of importance in determining the course of action. Yet maximum likelihood estimators have advantages, viz., sufficiency and efficiency [cf. Mood (1950), pp. 158 ff], which may, in some manner not clear to me, work to the advantage of the decision maker. Also, as the data become increasingly plentiful, the difference between the expected value and the maximum likelihood estimate of a parameter becomes negligible. Finally, it must be pointed out that if one selects Eq (3) (Forecaster's Confidence) as the *a priori* distribution, then the *a posteriori* expected values, and the maximum likelihood estimators considering only the data and not the *a priori* distribution, are almost equal.

Option III, although initially attractive, is difficult to justify logically. Would one choose a decision which, in 99 cases out of 100, would provide some inconsequential improvement over another; while in the one remaining case it would result in disaster? This is an extreme, but the logical extreme, of Option III.

5. A specific application of the method

To demonstrate the method I will use an example employed by Gleeson (1960). Although Gleeson's example is originally given in terms of a profit matrix, I will assume, for the sake of argument, that within the range of outcomes permitted by this matrix, utility is a linear function of profit, so that the numbers he gives may be interpreted as "utiles." Gleeson's example consists of four possible decisions (D_1, D_2, D_3, D_4) the outcomes of which are dependent on which of three possible subsequent weather events (W_1, W_2, W_3) occurs. Although Gleeson does not give values for the x_i (on the basis of which the *a posteriori* frequency distributions of the probabilities (p_1, p_2, p_3) are to be determined), he does give confidence limits on the occurrences of the weather states. The x_i may be deduced from these. If it is assumed that the values given represent 99 per cent confidence limits, reasonable correspondence is obtained with $n=23$, $x_1=12$, $x_2=4$, and $x_3=7$. This information, along with Gleeson's cost (or our utility) matrix, are shown in Table 1.

Each of the three prior distributions presented in section 3 will be used to determine *a posteriori* joint distributions of p_1 and p_2 ($p_3=1-p_1-p_2$). In the case of Prior Distribution I, Climatology, we shall assume that each of the categories is equally likely on a climatological basis, i.e., $a_1=a_2=a_3=\frac{1}{3}$. (See Fig. 1.)

TABLE 1. Gain matrix* and *a posteriori* expected values of the probabilities of weather states.

		W_1	W_2	W_3
Decisions	D_1	4	1	−2
	D_2	1	2	0
	D_3	−1	2	3
	D_4	0	0	0
Confidence limits*		0.84	0.49	0.69
		0.21	0.01	0.09
x_i		12	4	7

* After Gleeson (1960).

For Prior Distribution II, Forecaster's Confidence, the parameter ϵ has been set equal 0.05. This prior distribution is plotted in Fig. 2.

The three *a posteriori* joint distributions of p_1 and p_2, corresponding to the three prior distributions, are plotted in Figs. 4, 5 and 6. The equations for these distributions are, in Fig. 4, using Prior Distribution I, Climatology:

$$f(p|x)=\frac{\Gamma(27)}{\Gamma(40/3)\Gamma(16/3)\Gamma(25/3)} \times p_1^{37/3}p_2^{13/3}(1-p_1-p_2)^{22/3}; \quad (5)$$

in Fig. 5, using Prior Distribution II, Forecaster's Confidence:

$$f(p|x)=\frac{\Gamma(23.15)}{\Gamma(12.05)\Gamma(4.05)\Gamma(7.05)} \times p_1^{11.05}p_2^{3.05}(1-p_1-p_2)^{6.05}; \quad (6)$$

and in Fig. 6, using prior Distribution III, Maximum Entropy:

$$f(p|x)=\frac{\Gamma(26)}{\Gamma(13)\Gamma(5)\Gamma(8)}p_1^{12}p_2^{4}(1-p_1-p_2)^{7}. \quad (7)$$

Also shown on these graphs are those regions of the triangular p-space where each of the individual decisions dominate. The boundaries of those regions are determined by solving the six equations $U(D_1)=U(D_2)$, $U(D_3)=U(D_4)$, etc. Note that there is no region in which D_4 maximizes utility; this is because it is "dominated" (Gleeson, 1960) by D_2. The decision D_4 will then be considered no further, since there is never any reason to select it in preference to D_2.

A comparison of Figs. 4, 5 and 6 illustrates quite clearly the remarkable similarity between the *a posteriori* distributions, in spite of the rather marked differences among the *a priori* distributions. Thus even as few data as are used in this example are sufficient to overcome differences among the various *a priori* distributions. Indeed, it is concluded that with even a reasonable number of data, almost any reasonable

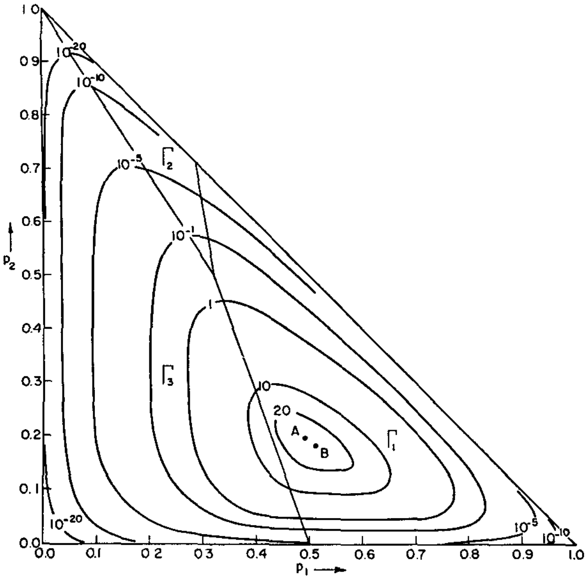

FIG. 4. *A posteriori* distribution, using Prior Distribution I, Climatology, for example described in text [see Eq (5)].

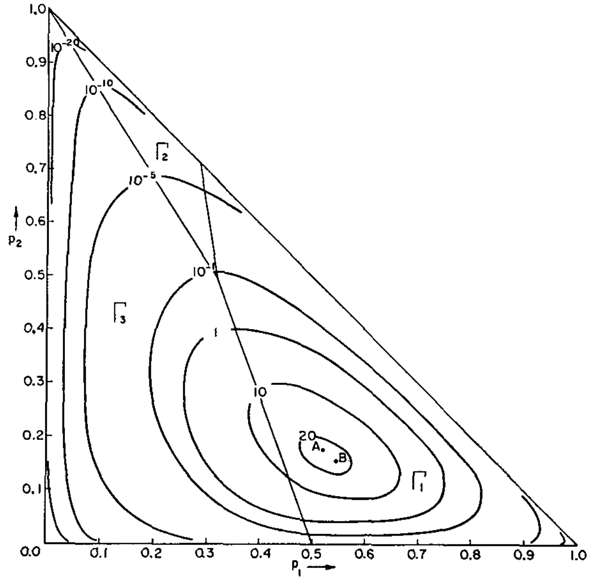

FIG. 5. *A posteriori* distribution, using Prior Distribution II, Forecaster's Confidence, for example described in text [see Eq (6)].

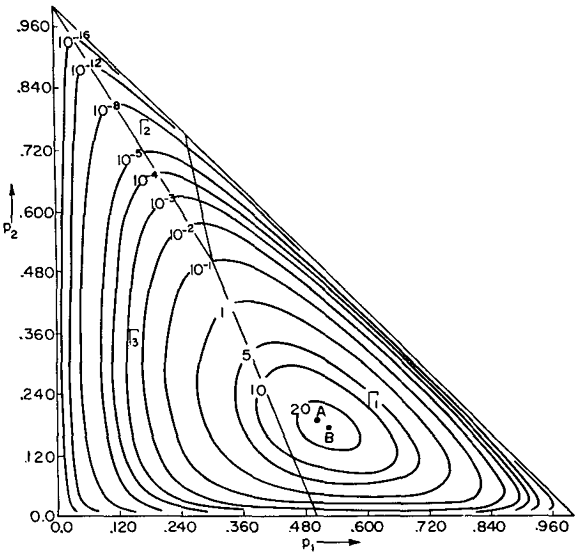

FIG. 6. *A posteriori* distribution, using Prior Distribution III, Maximum Entropy, for example described in text [see Eq (7)].

prior distribution would in general lead to equivalent results. This emphasizes the point made earlier that only in the case of no or very few data are the prior distributions of very great significance.

The similarities among the several subsequent distributions may further be seen in Table 2, where the expected values and maximum likelihood estimates of the utility of the several pertinent decisions are shown for each of the prior distributions. The probability that

TABLE 2. Expected values and maximum likelihood estimates of utility for the several decisions and prior distributions.

Prior distribution		D_1	D_2	D_3
I	$\tilde{U}$	1.56	0.89	0.83
	$\hat{U}$	1.62	0.88	0.76
II	$\tilde{U}$	1.65	0.87	0.74
	$\hat{U}$	1.74	0.85	0.65
III	$\tilde{U}$	1.59	0.89	0.80
	$\hat{U}$	1.65	0.87	0.74
	Prob$\{U(D_i)>U(D_j)\}$	0.813	0.187	0.00004

each decision will yield greater utility than any other decision has been computed only for Prior Distribution III, but it is evident from the graphs and from the other numbers shown in Table 2, that similar results would have been obtained for the other prior distributions. Fig. 7 gives the cumulative *a posteriori* frequency distributions of the utilities corresponding to the three pertinent decisions, and for Prior Distribution III.

To assist in our analysis of these results let me first briefly describe the conclusions pertaining to this example which were reached by Gleeson. Using his Method A, in which he considers the confidence limits, and estimates for each decision the minimum expectation by allowing the most unfavorable events to occur to the limit of the confidence intervals, Gleeson concludes that the best decision would be D_2. This gives a minimum expected profit of 0.32 unit. His Method B,

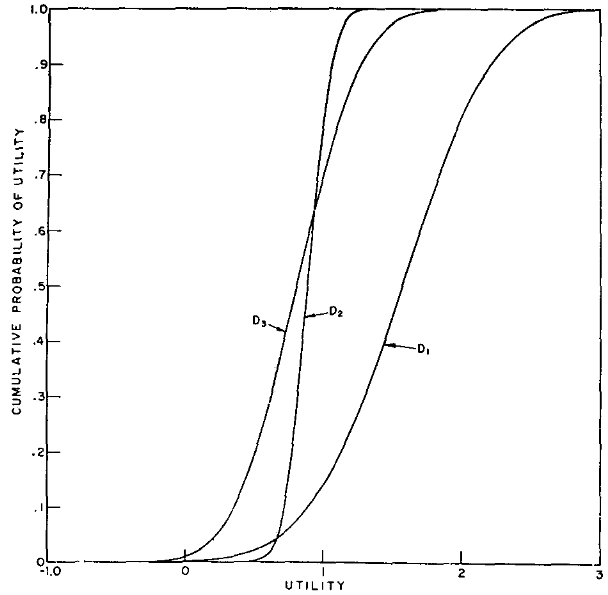

FIG. 7. The cumulative *a posteriori* frequencies of the utilities for decisions D_1, D_2, and D_3, using Prior Distribution III.

in which he employs the Theory of Games, and assumes that nature will always choose its best strategy, yields the result that decisions D_1 and D_3 should be employed in the ratio 6:4. This yields a minimum expectation of 1.00 units, the "value" of the game. By the methods described by Nelson and Winter (1960) the decision D_1 would be chosen, since this yields the largest expected value of gain.[4]

The reason that D_2 is selected by Gleeson's Method A is apparent from Fig. 7. If this decision is made, the probability of a very small gain is minimized; gains less than 0.5 would occur only about 1 time in 1,000. Note that this is still larger than the minimum expected gain for this decision by Gleeson's method. The reason for this is the extremely small probability of p_3 being near the upper limit of its confidence interval (0.69) while p_1 is at the lower limit of its range (0.21), leaving $p_2=0.10$. (See Figs. 4–6). On the other hand the probability that a gain greater than 1.1 will occur using D_2 is only about 3 per cent. Yet using D_1 the probability of a gain exceeding 1.5 is 55 per cent, and there is still about a 20 per cent probability of a gain greater than 2.0.

On the basis of Fig. 7, one would always choose D_1 in preference to D_3, unless one feared the almost infinitesimally small ($\sim$0.005 per cent) possibility of a loss greater than 0.4. Yet by Gleeson's Method B, D_3 would be selected 40 per cent of the time. This is because the theory of games presupposes that the opponent (here, nature), upon "observing" one's strategy of always selecting D_1, would change its strategy and always select W_3 (see Table 1). It would hardly appear that this is an event to be feared.

Employing the options described in the previous section, and returning to the utility concept, we find (see Table 2) that all three lead to the selection of D_1 as the optimum decision. In Figs. 4, 5 and 6, points A and B both fall in the regions in which D_1 yields the largest utility, and the probability associated with that decision in Table 2 is also the largest. Thus, in this example, D_1 provides the greatest expected value and maximum likelihood estimate of utility, and also the greatest probability of obtaining the maximum utility.

6. Summary and conclusions

The particular approach which one takes in reaching decisions should vary with the nature of the problem. I have attempted to describe a methodology which appears to be eminently suitable to problems of applied meteorology. From the point of view of the theory of games, one might say that we assume our opponent (nature) is unaware of the rules of the games (i.e., the system of awards, the utilities of the outcomes) but has chosen a strategy nevertheless and will adhere to that strategy in the future. We do not know what his strategy is, but must estimate that strategy on the basis of his past performance. Also, since we are confident that our opponent will not (or cannot) alter his strategy when ours is put into effect, our "best" strategy can take the form of a "pure" strategy. That is, we may always make the same decision; we need not keep our nonexistent opponent guessing.

It should also be evident that weather information, plus the potential for profit or loss, is not in general sufficient to determine what this "best" strategy is. Other considerations, such as the availability of working capital, or the effect of any actions on one's customers or competitors, are allowed to influence the decision through the use of the utility concept.

The use of Bayes' Theorem to evaluate the frequency distribution of the probabilities of future weather events has applications much broader than those discussed in this paper. For example, there is nothing implicit in the method which requires us to restrict it to multinomial distributions and categorized subsequent weather events. Miller [1961] has employed it to estimate the probabilities of subsequent weather as an adjunct to multiple discriminant analysis. Furthermore, as additional data, beyond those employed in the initial application of the method, become available, the method may be applied again. In the additional applications of the method, however, the *a posteriori* distributions of the previous application become the *a priori* distributions of the next application. Indeed, as large amounts of data become available, the *a priori*

[4] This statement is true even though the usages of the term "expected value" differ between Nelson and Winter (1960) and myself. In effect Nelson and Winter use x_i/n as the expected value of p_i. In our examples, we determine $\bar{p}_i=\int_p p_i f(p|x)dp$.

distribution of the very first application has ever diminishing influence on the final distributions.

As a final note, suggestive of a possible direction for further study, I would introduce the general problem of forecast verification. Would not an optimum forecasting (or analysis) procedure provide maximum discrimination among, not necessarily the types of weather which may occur, but the utilities of the decisions from which one must choose?

Acknowledgments. Draft versions of this manuscript were read by D. A. Leabo, Associate Professor of Statistics, and by L. J. Savage, Professor of Mathematics, both of The University of Michigan. Their comments and suggestions were most helpful. The author is especially appreciative of several enlightening discussions with Professor Savage.

REFERENCES

Borgman, L. E., 1960: Weather-forecast profitability from a client's viewpoint. *Bull. Amer. meteor. Soc.*, **41**, 347–356.

Feller, W., 1950: *An introduction to probability theory and its applications.* New York, John Wiley & Sons, Inc., 419 pp.

Gleeson, T. A., 1960: A prediction and decision method for applied meteorology and climatology, based partly on the theory of games. *J. Meteor.*, **17**, 116–121.

Gringorten, I. I., 1955: Methods of objective forecasting. In *Advances in geophysics, Vol. 2*, New York, Academic Press, 286 pp.

——, 1958: On the comparison of one or more sets of probability forecasts. *J. Meteor.*, **15**, 283–287.

——, 1959: Probability estimates of the weather in relation to operational decisions. *J. Meteor.*, **16**, 663–671.

——, I. A. Lund and M. A. Miller, 1956: The construction and use of forecast registers. Geophysical Research Papers No. 53, Geophysics Research Directorate, Air Force Cambridge Research Center.

Luce, R. D., and H. Raiffa, 1957: *Games and decisions.* New York, John Wiley & Sons, Inc., 509 pp.

Miller, R. G., 1961: An application of multiple discriminant analysis to the probabilistic prediction of meteorological conditions affecting operational decisions. Technical Memorandum No. 4, Travelers Research Center, Inc., Hartford, Conn., 138 pp.

Mood, A. McF., 1950: *Introduction to the theory of statistics.* New York, McGraw-Hill Book Co., Inc., 433 pp.

Nelson, R. R., and S. G. Winter, Jr., 1960: Weather information and economic decisions a preliminary report. Report of the RAND Corporation on Contract No. NASw-103, National Aeronautics and Space Administration, 121 pp.

Raiffa, H., and R. Schlaifer, 1961: *Applied statistical decision theory.* Division of Research, Graduate School of Business Administration, Harvard University, Boston, 356 pp.

Ramsey, F. P., 1931: Truth and probability; and further considerations. In *The foundations of mathematics and other logical essays*, New York, Harcourt, Brace and Co., 292 pp.

Savage, L. J., 1954: *The foundations of statistics.* New York, John Wiley & Sons, Inc., 294 pp.

Schlaifer, R., 1959: *Probability and statistics for business decisions.* New York, McGraw-Hill Book Co., Inc., 732 pp.

Thompson, J. C., 1952: On the operational deficiencies in categorical weather forecasts. *Bull Amer. meteor. Soc.*, **33**, 223–226.

——, and G. W. Brier, 1955: The economic utility of weather forecasts. *Mon. Wea. Rev.*, **83**, 249–254.

von Neumann, J., and O. Morgenstern, 1947: *Theory of games and economic behavior.* Princeton, N. J., Princeton University Press, 641 pp.