72-5959

MORRIS, Peter Alan, 1945-
BAYESIAN EXPERT RESOLUTION.

Stanford University, Ph.D., 1971
Engineering, general

University Microfilms, A XEROX Company, Ann Arbor, Michigan

BAYESIAN EXPERT RESOLUTION

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ENGINEERING-ECONOMIC SYSTEMS

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

By

Peter Alan Morris

May 1971

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ronald A. Howard
(Principal Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard D. Smallwood

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James E. Matheson

Approved for the University Committee
on Graduate Studies:

Lincoln E. Moses
Dean of Graduate Studies

BAYESIAN EXPERT RESOLUTION

Peter Alan Morris, Ph.D.
Stanford University, 1971

This dissertation introduces a conceptual and methodological solution to the problem of how to use experts in decision situations. The approach taken is consistent with the theory of decision analysis and rests philosophically on the Bayesian or subjectivist viewpoint regarding matters of uncertainty. The results form fundamental practical tools for solving basic expert resolution problems.

The first development is the creation of a theoretical structure with which to view the problem in a logical consistent way. The revelation of an expert's prior probability assessment is viewed as an addition to a decision maker's current state of knowledge. Application of the tools of Bayesian inference provides a means whereby a decision maker can update his own prior assessment consistent with his subjective appraisal of the expert. By natural extension a multi-expert inferential structure is obtained.

A major part of the dissertation is the development of a practical methodology for expert use. The basic result is that a set of expert prior probability assessments (priors) may be processed in two steps:

<u>Step 1</u> A surrogate expert prior is formed by multiplying the product of the given expert priors by a subjectively assessed calibration function and normalizing.

<u>Step 2</u> The decision maker's posterior probability assessment is calculated as the normalized product of his own prior and the surrogate prior.

The calibration function is a representation of the decision maker's appraisal of the experts. A detailed analysis of this function is presented, demonstrating that it may be computed from a set of conceptually direct assessments specified by the decision maker. Important for physical insight and practical simplification is the introduction of the technique of trajectory analysis.

In the single expert case the surrogate prior represents the translation of the expert's probability assessment into the decision maker's probability language. In the multi-expert situation the surrogate prior is thought of as the calibrated prior of a conceptual surrogate expert, whose state of information represents (to the decision maker) the combined knowledge of the group of experts.

The theory developed also provides insights into aspects of the group decision making problem. Conditions are specified under which a group of decision makers should agree regarding the implications of expert opinion. Also presented are normative conditions for consensus among a group of assessors.

Three important aspects of expert use are focused on and modeled in a detailed way. These are the treatment of expert bias, the use of experts in experimental situations, and the relationship between the use of expert judgments and the use of formal analytic models. Included is a study, important in its own right, of the proper use of formal models.

An entire chapter is devoted to the derivation of new results in subjective probability theory. We introduce a means by which an assessor can assign probabilities consistent with his knowledge about his ability as a probability assessor. The Method of Equivalent Intervals is presented as a new probability assessment technique which ensures that an assessor's assigned prior is calibrated. The results form a fundamental link with the theory of expert use.

Approved for publication:

By ______________________
For Major Department

By ______________________
Dean of Graduate Studies

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the advice and encouragement of Professor Ronald A. Howard, under whose direction this dissertation was completed. His support and guidance were invaluable throughout my tenure as a graduate student.

Thanks are also due Professor Richard D. Smallwood and Dr. James E. Matheson for reviewing the manuscript and offering many useful comments.

My friends and colleagues in the Engineering-Economic Systems department provided important moral support. I would especially like to thank Dr. Carl-Axel von Holstein and Verne G. Chant for their time spent in many helpful discussions pertaining to the research.

A special debt of gratitude is owed to Professor William K. Linvill who motivated my entrance into the Ph.D. program, and whose direction and stimulation helped make my stay an enjoyable one.

The contributions of my parents throughout my education make this dissertation an achievement of their own. For this I am especially thankful.

To my wife Leanne, my special thanks for her illustrations and, above all, for her constant encouragement and patience.

I am grateful for the partial financial support provided by the National Science Foundation under Grant NSF-GK-16125.

CONTENTS

Contents (Continued)

Contents (Concluded)

LIST OF FIGURES

List of Figures (Continued)

List of Figures (Concluded)

LIST OF TABLES

Chapter I

INTRODUCTION AND OVERVIEW

1.0 Introduction

This dissertation introduces a new philosophical and methodological framework for the use of experts in decision situations. The normative model introduced provides a logical connection with the modern theory of decision analysis. The results obtained form fundamental practical tools directly applicable to real world problems.

A theory on expert use is of no practical import without a theory of decision. Because of the intimate link between decision analysis and the present work, a brief description of the theory of decision analysis follows. Many more detailed descriptions exist in the literature [12,18].

Decision analysis

Decision analysis is a practical theory based on the rudiments of logic, applicable in principle to any problem concerning the allocation of scarce resources. The foundations of the approach are a set of non-controversial axioms allowing the decision maker to encode his state of knowledge about an uncertain world via a quantitative probability measure, and his attitude toward uncertainty by an independent utility measure. Of fundamental importance to practical applications of decision analysis is the methodological integration of theory into practice through systematic structuring and modeling (systems analysis). A conceptual model offered by Howard [12] decomposes the analysis into three components: the deterministic phase, the probabilistic phase, and the informational phase. We shall focus on aspects of decision problems commonly associated with the probabilistic phase of analysis, although we occasionally discuss matters pertaining to deterministic and informational questions. Also included is a chapter providing new insights into the basic theory of subjective probability. A detailed guide to the dissertation will follow at the end of this chapter.

Definition of an expert

In general we shall adopt for a frame of reference the situation where a decision maker, for purposes of gaining additional knowledge, consults one or more persons on a variable or set of variables of interest. Any person who provides the decision maker with an opinion will be termed an expert. The most detailed and most interesting expert input pertaining to a variable is its entire probability distribution. This distribution represents the expert's complete state of information about the variable of interest. Notice, however, that the definition is completely general. It essentially adopts as an expert any person giving a judgment to the decision maker in any form.

Perhaps the most obvious example of such an expert is the weather forecaster whose job it is to assess the relative likelihoods of various types of weather. The type of information obtained from the weatherman is fundamentally different from the empirical data gathered from a past history of climate. A central task will be to explore and provide insights into this difference.

1.1 Applications

The practical implications of a rich theory on expert use are many and diverse. It is a rare decision maker who makes an important decision concerning the allocation of his resources without advice. Countless numbers of formal and informal institutions created for the express purpose of expediting decision maker access to experts attest to this fact. There is an example from almost every professional discipline and decision making institution.

For instance, in the field of Law there is expert testimony. Stockbrokers serve as experts to private investors. In medicine the role of the consulting physician is a particularly interesting example. Typically an entering patient has a set of symptoms which is then expanded by further testing. In many cases a team of doctors are asked to make a diagnosis based on the total set of data, after which a decision is made on the appropriate treatment to be applied. When the consulting specialists have conflicting opinions the decision maker (who is usually a designated physician or the patient himself) must, in some sense, resolve the controversy either implicitly, by choice of action, or explicitly, by specification of numerical probability.

In general, the theory is especially pertinent to decisions affecting systems whose characteristics include significant uncertainty, complexity, or both--precisely the practical domain of decision analysis. When such situations have the additional complication that experimentation is impossible or extremely expensive, the importance of the theory is further amplified. A good example is the formation of public policy decisions. Experiments on society are extremely expensive. Furthermore, much of the present knowledge pertaining to such systems rests only in the minds of experts, and is not transcribable into formal analytic models.

The theory presented has important side benefits in terms of insights and developments into other areas. We shall focus briefly on two in this paper: the aforementioned topic of subjective probability and the area of modeling in general.

1.2 Related Theories

The number of existing methods for expert use is surprisingly small in view of the importance of the problem. Historically this may be explained in large part by the only recent re-emergence of the subjective or Bayesian view of probability. This point of view is singular in that it admits explicitly the concept of uncertainty and thus the notion of personal judgment and assigned probabilities. The classical approach, still widely held, regards a person's likelihood judgments as irrelevant since probabilities are defined not as assignments, but as functions of observed data. A study of the use of subjective information of the type provided by experts is inconsistent for the classical probabilist to undertake, since it underlines the importance of subjective judgments in general.

The present techniques for using expert judgments are reviewed below. The relationship of each to the present paper will become much clearer in hindsight. Therefore only brief criticism will be presented.

Informal techniques

Informal techniques for expert use, while admittedly ad hoc, are nevertheless the most common in practice. The aim of these techniques is generally to achieve unanimity among two or more experts. The derived consensus opinion is then used directly for decision making purposes.

A common method is to choose the best expert (in some unspecified sense) and use the opinion he delivers directly. This engenders many objections. For instance there is no general criterion for selecting the best expert. There is no justification for throwing away information available from the other experts. Additionally, there is no good reason to directly use an expert's opinion, even if it is the only one given. Finally, there is an embarrassing question of how to treat the case of two equally trusted experts with conflicting opinions.

Another technique is to allow for some sort of group interaction such as committee, conference, or panel, where the group is directed to give as an output one opinion. Again there is no justification for using the consensus opinion. Furthermore, the derived group opinion has the disturbing property of being dependent on the mechanism, defined within the group, by which individual judgments are combined.

The general quest for a consensus, both in the literature and in practice, seems to stem largely from confusion about what to do if consensus is not achieved. There is no demanding logical reason why experts *should* agree. In fact we shall show (Chapter V) that a consensus may be inconsistent in general.

Delphi

The Delphi technique is probably the most widely known method for using experts. As defined by N. Dalkey, one of its founders, "Delphi is the name of a set of procedures for eliciting and refining the opinions of a group of people" [4]. Delphi is usually done operationally through an iterative questionnaire in which each expert, on a given cycle, purveys his opinion based on a summary of the opinions of each expert from the previous cycle. At the end of some cycle the results are summarized by a summary statistic (median, mean, etc.) of the group response. This statistic is the output of the technique.

Perhaps the most succinct way to characterize Delphi from our point of view is as a classical statistics treatment of an inherently Bayesian problem. Its adherents claim no more than a procedure with certain interesting statistical properties (see for example [3,9]). As such, most of the conceptual questions addressed in this paper are left unanswered. At best, Delphi is useful as a way to gather information to be used by a more basic theory stating how that information should be used. As we shall see, the method of combining the observed expert opinions is generally invalid.

Bayesian approaches

In recent years a number of approaches have been suggested by Bayesian theorists. These indicate a growing awareness in the fundamental character of the problem.

Two similar techniques are described by Winkler [25]. The first of these is called the Weighted Average Method, where a posterior distribution is derived by taking a linear combination of the given expert distributions. The weights used to form the weighted average are assigned by the decision maker, with the constraint that they be greater than zero and sum to one for guaranteed consistency. The second technique, the Natural-Conjugate Method, determines a posterior distribution by computing its parameters as weighted sums of the parameters of each expert prior. Thus, for example, if the i^{th} expert's prior is fitted to a member of the Beta family with parameters r_i and n_i, the decision maker's posterior after observing each expert's prior would be a Beta distribution with parameters r and n where

$$r = \sum_i w_i r_i, \quad n = \sum_i w_i n_i,$$

w_i being the assigned weight of the i^{th} expert's judgment. The method is, of course, exact if each expert's experience is perceived equivalent to r_i successes in n_i Bernoulli trials. The weights would then represent the degree to which the experts observed the same data. (If the degree of common experience is uncertain the weights amount to point estimates.)

The above methods were developed more out of an operational necessity for Bayesians to use experts than from any theoretical model. In this sense they are interesting as special limiting cases of the theory to be developed. As fundamental approaches themselves they are subject to criticism for being extremely restrictive. For example, in general, there is no *a priori* reason to constrain the decision maker's posterior to be in the set of linear combinations of the expert's priors. The assumption that each expert's experience is equivalent to observations of a fixed sampling process is also clearly rigid. Finally, the two methods give different answers in general.

A third approach in the literature is described by Raiffa [17]. We shall term this approach, for descriptive purposes, the *Method of Empirical Calibration*. Using this method an expert is asked to specify certain fractiles of a large number of random variables. The long-run frequency with which the true values fall below a given fractile is then ascribed as the probability of the corresponding interval defined by the expert's prior on the variable of interest. Since we shall later discuss this method extensively, we defer comment here except to observe that it does not address the more interesting multi-expert case.

1.3 A Conceptual Model

The conceptual model presented in Figure 1.1 provides an overview of the dissertation as well as a framework in which to compare the methodologies discussed in the previous section. The figure defines the intuitive characteristics of a complete theory of expert use.

Consider the large outer box to represent an ideal theory whose purpose is to guide the decision maker in the making of consistent inferences based on the revelation of N expert prior probability assessments on a given random variable. The output of the proposed theory is representatively termed the decision maker's posterior probability assessment. Clearly, in the general case, the theory must accept each expert's prior as an input.

The ideal theory must also take into account the decision maker's own prior state of knowledge on the random variable. To neglect to do so would conflict with the most basic Bayesian tenets. Also intuitive is that the theory should include the decision maker's subjective assessment of each expert. Certainly experts vary as to their expertise and therefore on how they should be "weighted."

A useful and natural distinction to make is between an expert's expertise on the random variable and his probability assessment ability

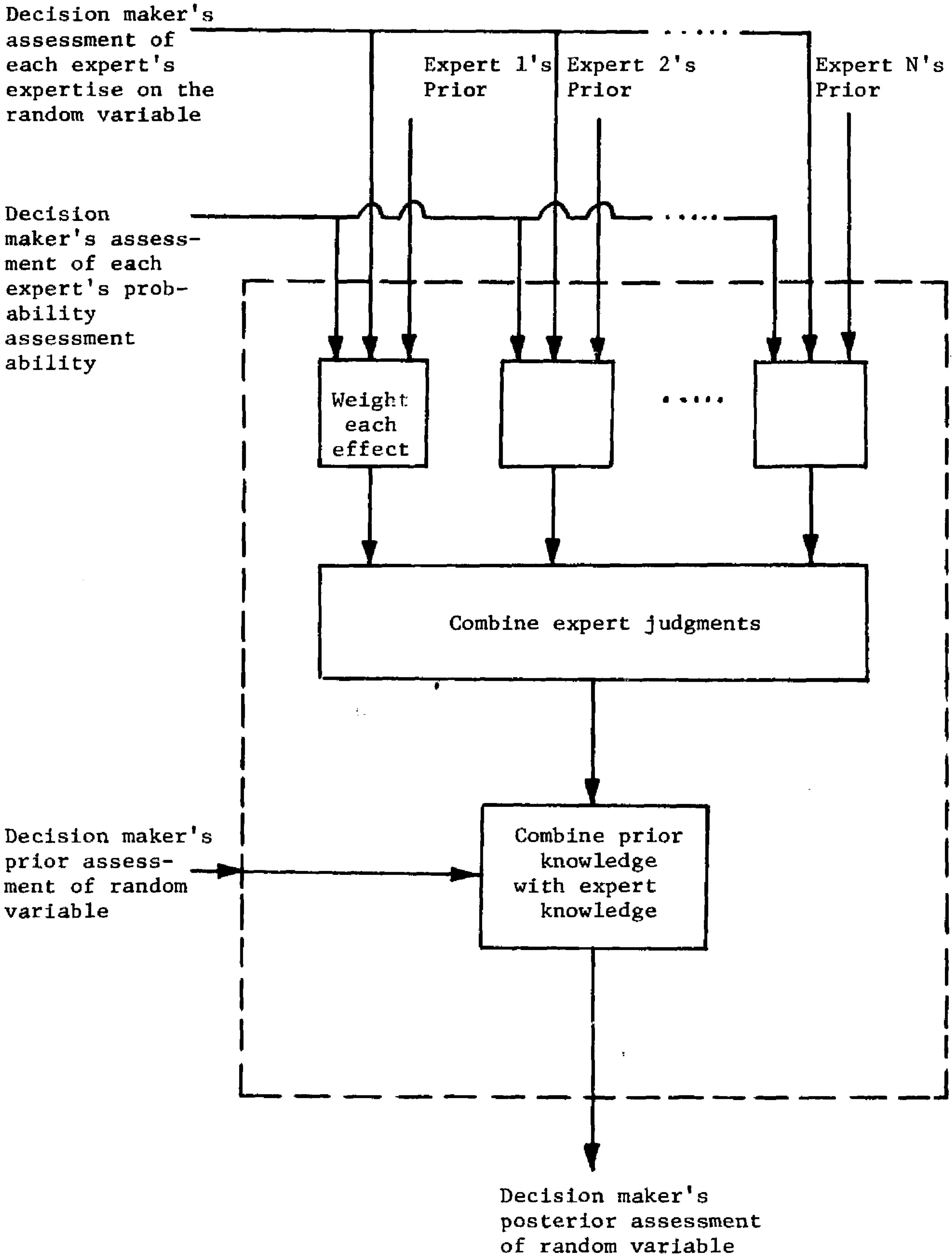


Figure 1.1 A Conceptual Model

in general. Each of these components should be included, yet it is not clear how to characterize either. Recall that both the Subjective Weights Method and the Natural-Conjugates Method partially attempted to structure this question. However, the subjective assignments were on variables of no well-defined physical meaning (except in a very special case). The problem is not trivial--a major task will be to characterize the subjective appraisal of experts.

Given a logical theory, able to admit the proposed inputs in consistent fashion, it is advantageous, although not theoretically necessary, to demand certain additional operational characteristics. The inside of the large box in Figure 1.1 shows a desirable decomposition within the theory itself. The two conceptual components of each expert's assessment ability are combined to provide N distinct inputs to a mechanism for combining the expert judgments. The output of this mechanism characterizes in a sense (yet unspecified) the total combined knowledge of the experts. This is then mixed separately with the decision maker's own prior knowledge of the random variable to form the posterior output. It will be a further aim of this dissertation to achieve such a decomposition.

1.4 Overview and Survey of Results

The dissertation is roughly structured into four components. The first is the development of a basic theoretical structure. The second component concerns the creation of a methodology for the practical application of the theory. These results are probably the most interesting and provocative, and serve ideally to compare this theory with others.

The third component of the dissertation is a detailed modeling and analysis of important aspects of the decision maker-expert relationship, and the fourth component is the derivation of results important independently of the central topic. These results are, however, also indispensable for our present purposes and are therefore embedded in the main work.

Chapters II and III develop a conceptual framework for the single expert and multi-expert resolution problem. Basic definitions are given and preliminary analysis is begun. The key idea in these chapters is that the revelation of expert probability assignments changes the decision maker's state of information. We apply the tools of Bayesian inference to develop a means by which a decision maker can update his prior probability assignment consistent with his assessment of the experts. An important development is the characterization of an assessment of experts as a likelihood function over possible expert responses.

Chapter IV introduces new insights and methodological advances in a controversial area in the field of subjective probability theory. We shall argue that information pertaining to the accuracy of an assessor's perception of the world is as basic as information about the world itself. Thus we shall be led to an explicit characterization of probability assessment ability. A basic result is a procedure by which an assessor can guarantee that his probability assignments are consistent with his knowledge about himself as an assessor. The chapter ends with the development of a new probability assessment technique: the Method of Equivalent Intervals. This technique allows an individual to self-calibrate himself, in the sense that his probability assignments are invariant to further information about his probability assessment performance. The analysis presented will form a fundamental prelude to the detailed study of expert use presented in the next chapter.

Chapter V develops a methodology for practical expert use. By use of a subjectively assessed calibration function, a set of expert priors can be converted into a single density function called the "surrogate prior." To obtain a consistent posterior distribution, the decision maker can multiply the surrogate prior by his own prior and normalize.

The surrogate prior in the single expert case reduces to the expert's prior if the decision maker knows the expert to be self-calibrated. If this is not the case, the surrogate prior represents what the decision maker expects the expert would submit if he self-calibrated himself. The multi-expert surrogate prior has an interesting interpretation. It is the probability assessment of a conceptual "surrogate expert," whose state of knowledge concerning an uncertain variable represents the combined knowledge of the group of experts.

The calibration function is calculable from a number of conceptually direct expected value assessments provided by the decision maker. In the multi-expert case a technique called trajectory analysis reduces a many dimensional assessment problem to an assessment on a path. The trajectory is derived from certain aspects of the experts' given probability assessments, and allows immediate physical insight into a problem without requiring a detailed mathematical analysis.

The theory presented in Chapter V also provides important practical comment on group decision making problems. A normative condition for agreement on probability assignments is derived. Also presented are conditions under which a group of experts should achieve a consensus.

The results provide working tools for the immediate solution of previously unresolved practical problems in decision situations. The methodology is designed for physical insight and provides a framework in which intuition is easily exercised.

In Chapter VI a detailed analysis of three important special topics is presented. The first study is of expert bias. It is argued that the fundamental problem relative to a biased expert is not the expert's degree of bias, but the decision maker's own state of uncertainty about the extent to which the expert is biased. The direct approach to the problem has many practical complexities. By restructuring the problem in an unusual way we discover a natural means of characterizing expert bias, out of which follows practical tools for analysis. The next subject in the chapter concerns the use of experts relative to experimental situations. We demonstrate by detailed structuring, that the key factor in such situations is the expert's model of an experiment, as revealed by his likelihood function. The final topic is the relationship of expert use to the use of formal mathematical models. In particular, we derive some basic results in model theory, demonstrating the conceptual similarity of models to experts.

Chapter II

THE USE OF ONE EXPERT

2.0 Introduction

Our basic task in this chapter will be to present a fundamental solution to the problem of how a decision maker should use the advice of one expert when it comes in the form of probability distributions on variables relevant to the decision of interest. The work here will provide the foundation for developments in the remainder of the dissertation. Many of the interesting implications and applications of the material to be presented will be deferred to later chapters for complete treatment.

In Section 2.1 we present in detail the basic notation. This will be an important section as previous attempts to handle the problem of interest have been severely hampered by inflexible notation. The solution for one expert giving advice on one variable will be presented in Section 2.2. Important for insight and understanding will be two illustrative examples presented in Section 2.3. Section 2.4 will show the relationship of this development to some of Raiffa's work and will serve as a further example of its applicability. In Section 2.5 we display a precise answer to the question of "what is a good expert" when we calculate the value of an expert. The implications of the theory for problems concerning many variables will be studied in Section 2.6. In particular we will explore the interesting relationships between various independence assumptions. We shall also give a preliminary discussion of the interesting concept of calibration. Finally, in Section 2.7 we shall review the idea of exchangeability and derive an immediate application of its use.

2.1 Notation

An uncumbersome notation is required for clear thinking about the problem. Such a notation has been developed by Howard [11] which explicitly displays the state of information a probability assignment is based upon. It also provides an especially convenient tool in the manipulation of inference problems.

Let x be arbitrary random variable. The probability density (or mass) function based on the state of information $\mathcal{S}$ is denoted $\{x|\mathcal{S}\}$. Similarly the probability of the event E conditional on $\mathcal{S}$ is represented by $\{E|\mathcal{S}\}$. When it is necessary to emphasize the functional character of a given assessment $\{x|\mathcal{S}\}$ we write $\{x \leftarrow x_o|\mathcal{S}\}$ for the density function evaluated at the point x_o.

All of the common results in the study of continuous random variables may be presented in analogous form to the corresponding results for discrete random variables. We shall exploit this property in our notation and define the generalized summation operator $\int_x$ as

$$\int_x = \begin{cases} \sum\limits_{\text{all } x} & x \text{ discrete} \\ \int\limits_{\text{all } x} dx & x \text{ continuous} \end{cases} \tag{2.1}$$

The expectation and variance of x are then defined as

$$\langle x|\mathcal{S}\rangle = \int_x x\ \{x|\mathcal{S}\}$$

$$\overset{v}{\langle}\, x|\mathcal{S}\rangle = \langle x^2|\mathcal{S}\rangle - \langle x|\mathcal{S}\rangle^2 \tag{2.2}$$

The two symbols ε and ρ are reserved for the states of information of the decision maker and his expert <u>prior</u> to the problem under consideration. For simplicity we refer to a probability distribution based only on this prior knowledge as a "prior."

All probability assignments will be conditional on at least ε or ρ. Included in ε and ρ are such concepts as experience, intuition, judgment and insight. Note that information of this type may not in general be totally transcribable into clear English sentences or partitioned into unambiguous events. The term "state of information" is a conceptual tool rather than a well defined object.

Two probability operations will be of particular interest. These are referred to as <u>Bayes' Theorem</u> and the <u>Expansion Rule</u>. Given the arbitrary random variables x and y we have

$$\text{Bayes' Theorem:} \quad \{x|y,\mathcal{S}\} = \frac{\{y|x,\mathcal{S}\}\{x|\mathcal{S}\}}{\{y|\mathcal{S}\}} \tag{2.3}$$

$$= k\{y|x,\mathcal{S}\}\{x|\mathcal{S}\}$$

$$\text{Expansion Rule:} \quad \{x|\mathcal{S}\} = \int_y \{x|y,\mathcal{S}\}\ \{y|\mathcal{S}\} \tag{2.4}$$

The symbol k will be reserved to represent a constant relative to the initial variable being assessed in a given development. This notational device allows k to be a different number in the same derivation (always calculable in a given expression by normalization).

2.2 The Basic Problem: One Expert, One Variable

The formal problem we wish to solve is:

Given 1) $\{x|\varepsilon\}$ the decision maker's prior on the random variable x

2) $\{x|\rho\}$ the expert's prior on the random variable x

How should the decision maker's prior be altered upon reception of the expert's prior?

We will treat the problem in the most general way, constraining only the decision maker to abide by the axioms of decision theory. Thus our approach is prescriptive for the decision maker only. The only rationality constraint on the expert will require $\{x|\rho\}$ to be, in fact, a probability distribution. Thus $\int_x \{x|\rho\} = 1$ and $\{x|\rho\} \geq 0$ for all x. In this chapter we will further constrain the expert's advice to reflect his actual state of knowledge on x. However in Chapter VI even this assumption will be relaxed to allow for gaming situations, biased experts and, in general, situations where the distribution on x the decision maker receives from the expert is not $\{x|\rho\}$.

As stated, the expert use problem is one of inference. The decision maker must infer how to use the information given to alter his prior distribution. Below we review the standard Bayesian solution to ordinary inference problems. This is also a convenient time to highlight other features of the Bayesian philosophy relevant to the present topic.

Subjective probability

The Bayesian or subjectivist uses probability theory as a logical way of reasoning about the unknown. Probabilities are regarded as basic measures of uncertainty. Since by definition uncertainty is a characteristic of the human mind, probability must be a function of a person's state of mind. Thus a probability is defined as an assignment rather than a characteristic of a physical process or object.

The key word in the development of subjective probability theory is consistency. It can be shown that all the usual probabilistic operations are derivable from a basic set of elementary axioms [19]. In particular, the use of Bayes' Theorem as a means of inference is necessary for consistency with the axioms. Bayes' Theorem (equation 2.3) produces a posterior assessment of a variable, conditional on observed data, that is consistent with both our prior knowledge about the variable and our assessment of how the information obtained is related to it.

In many situations the likelihood function $\{y|x,\varepsilon\}$ is derived from a particular probabilistic model. It is important to emphasize, however, that $\{y|x,\varepsilon\}$ is always a subjective assessment or the result of other basic subjective assessments. For instance, in the case of Bernoulli sampling, independence must be assessed and so must the probability of success given the Bernoulli model.

A possible approach

The immediate temptation for the Bayesian is to attempt a determination of $\{x|\rho,\varepsilon\}$, the density function on x given both the decision maker's and the expert's prior states of knowledge. It is useful to pursue this attempt since other expert use theories (such as Delphi) implicitly regard this as the underlying problem of interest. The Bayesian applies Bayes' Theorem to obtain:

$$\{x|\rho,\varepsilon\} = k\ \{\rho|x,\varepsilon\}\ \{x|\varepsilon\} \tag{2.5}$$

In light of our previous discussion in Section 2.1 we see that this approach is fruitless. First of all, it is meaningless to assign a probability to something other than an event, and evaluating $\{\rho|x,\varepsilon\}$ requires us to do this. Secondly and even more fundamental, the expression $\{x|\rho,\varepsilon\}$ implies that, in addition to being able to define a state of mind in terms of events, the decision maker knows what the expert's state of mind is. But the decision maker did not receive the expert's state of mind. He only received the expert's probability assignment on x. A further danger is to regard the union of ρ and ε as a joint state of information. This would require a whole new notion of probability.

<u>Solution to the basic problem</u>

The problem may be resolved in a deceptively simple way; simple in that the solution is the logical application of Bayes' Theorem, deceptive in that it will lead us to basic insights and results unobtainable by other methods. Our approach will be based on the observation that prior to receiving advice $\{x|\rho\}$ is an uncertain quantity to the decision maker. We will thus treat $\{x|\rho\}$ as an random variable whose value has been revealed to the decision maker. Therefore we wish to compute $\{x|\{x|\rho\},\varepsilon\}$.

In words, we want the density function on x conditional on the event that the expert's prior is $\{x|\rho\}$ and the decision maker's prior state of information is ε. To evaluate this we use Bayes' Theorem to write

$$\{x|\{x|\rho\},\varepsilon\} = \frac{\{\{x|\rho\}|x,\varepsilon\}\{x|\varepsilon\}}{\int_x \{\{x|\rho\}|x,\varepsilon\}\{x|\varepsilon\}}$$

$$= k\{\{x|\rho\}|x,\varepsilon\}\{x|\varepsilon\} \tag{2.6}$$

The problem can thus be solved if the decision maker can specify the likelihood function $\{\{x|\rho\}|x,\varepsilon\}$ for all values of x. In ordinary applications of Bayes' Theorem, the likelihood function is a functional mapping of elements from the vector space RxR into R.* Thus if x and y are elements of R, $\{y|x,\varepsilon\}$ is a real valued function of x and y. In our case let F denote the (convex) set of all probability functions. Our likelihood function is a mapping from RxF into R. (We will discuss in more detail the mathematical problems associated with this mapping in Chapter V.)

We can speak of a function of the variable x as taking on the value $\{x|\rho\}$ just as easily as we can speak of the variable x taking on the value x_o. $\{x|\rho\}$ is an uncertain variable to the decision maker whose value may be assessed conceptually like any other random variable. It is important to realize, however, that $\{\{x|\rho\}|x,\varepsilon\}$ is not the calculation in the classical sense of a "probability of a probability." Such a concept is meaningless in itself. $\{\{x|\rho\}|x,\varepsilon\}$ is (in the discrete case) the probability of the <u>event</u> that the expert's prior is $\{x|\rho\}$, given x and ε.

*R is the vector space of real numbers.

RxR is the cartesian production of R. It is the vector space of ordered pairs of real numbers.

For example, suppose that the set of possible values of $\{x|\rho\}$ is $[f_1(x),f_2(x),f_3(x),f_4(x),f_5(x)]$, where each $f_i(x)$ is drawn in Figure 2.1. We can assess the probability that the function $\{x|\rho\}$ takes on the value f_3; this is the probability that $\{x|\rho\} = f_3(x)$ for all x. We will show in Chapter V that this assessment can be done practically as well as conceptually.

The likelihood function also confirms the intuitive notion that we should weigh our use of advice with our feelings about our advisor. As stated before, the primary objective of any Bayesian analysis is consistency. The likelihood assessment allows the decision maker to be consistent with his own appraisal of the expert.

In fact, the likelihood function is our model of the expert. We can determine $\{\{x|\rho\}|x,\varepsilon\}$ as formally or informally as we like. The unique nature of this type of assessment presents an interesting conceptual challenge. Much of the remainder of this dissertation will be concerned with modeling the likelihood function.

In general the likelihood function is a subjective measure of the expert's accuracy in assessing x. The decision maker is given the true value x_o. If he feels that the expert is highly competent, he will assign high probability that the expert will respond with a narrow distribution about x_o. Therefore the likelihood function will be highly dependent on x_o and will considerably alter the decision maker's prior. Notice that if, in varying x_o, the decision maker's assessment of the expert is unchanged, the updated distribution will be equal to the original prior. Intuitively this means that the expert is, in the decision maker's mind, completely unresponsive to the random variable.

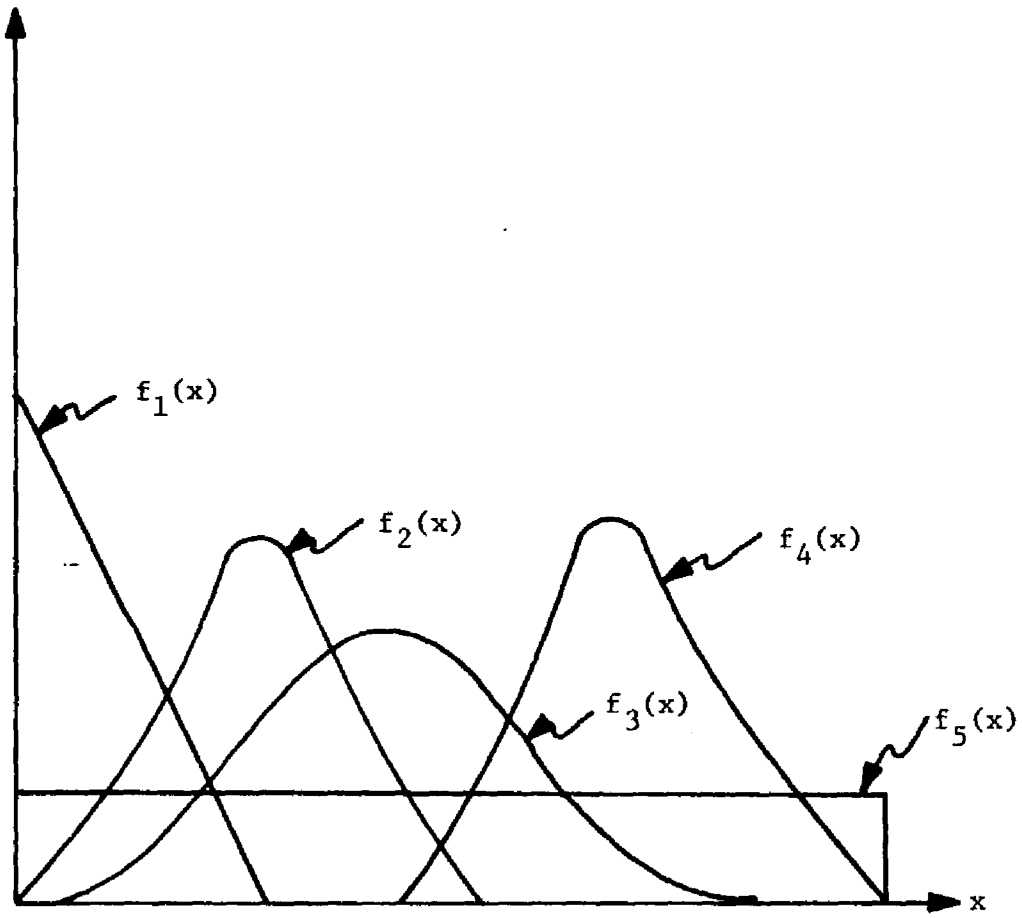


Figure 2.1 Set of Possible Expert Priors

Clearly he can learn nothing about x from such an expert. The following examples illustrate further the application of the above development in simple situations.

2.3 Examples

The first example will treat the simple, yet often realistic, case involving lotteries with only two outcomes. This example will be built on in later sections to give intuitive continuity to the theoretical developments. The second example will be a bit more complex and will display more clearly the assessment problem associated with equation 2.6.

Example 2.1--Thumbtack Flipping

Suppose that the decision maker has a decision to make based on the outcome of the flip of a thumbtack. If the thumbtack lands on its side we will say that it turned up "heads"; if it lands on its back we will say that it turned up "tails." We also suppose that the decision maker has never tossed a thumbtack and upon questioning reveals that he considers heads and tails equally likely. Let

$H \equiv$ the event thumbtack turns up heads,

$T \equiv$ the event thumbtack turns up tails.

The decision maker's prior is (Figure 2.2):

$$\{H|\varepsilon\} = 1/2 \ , \quad \{T|\varepsilon\} = 1/2 \ . \tag{2.7}$$

Let us assume now that the decision maker hires an expert to give him advice on the likelihood of heads and tails. The expert's prior can

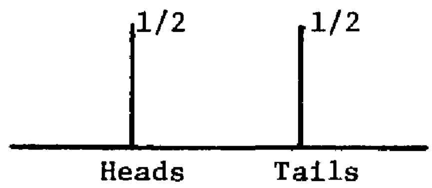


Figure 2.2 Decision Maker's Prior

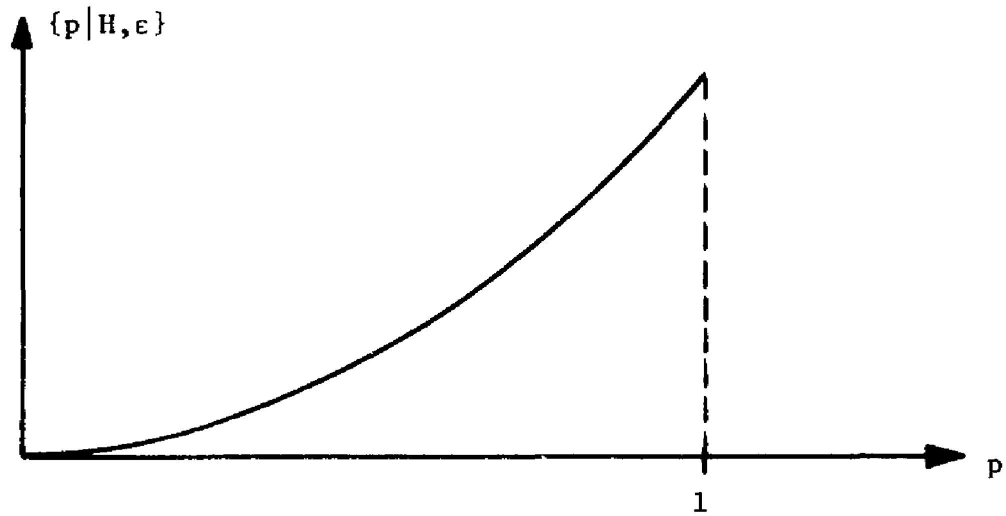


Figure 2.3 Decision Maker's Likelihood Assessment Given Heads

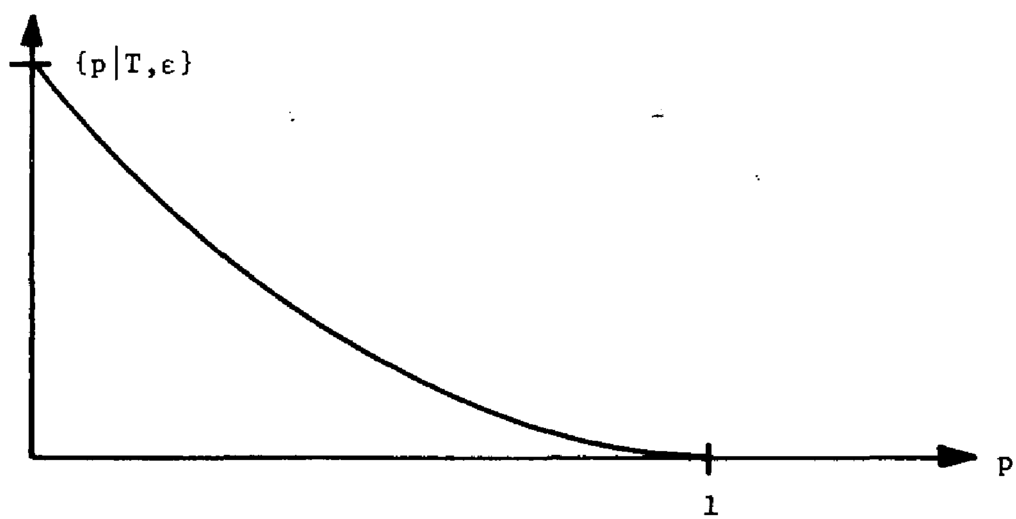


Figure 2.4 Decision Maker's Likelihood Assessment Given Tails

be characterized by the single parameter p where

$$\begin{aligned} \{H|\rho\} &= p \\ \{T|\rho\} &= 1 - p \end{aligned} \qquad 0 \leq \rho \leq 1 \tag{2.8}$$

Of course prior to receiving the expert's advice the decision maker is uncertain as to what the advice will be. Thus he assesses a likelihood function based on his subjective appraisal of the dependence between the expert's advice and the actual outcome. Specifically we ask him what his assessment of p would be if a clairvoyant told him that the thumbtack will surely come up heads. He feels that the expert is more likely to give a high probability of heads than a high probability of tails in this case and assigns:

$$\{p|H,\varepsilon\} = \begin{cases} 3p^2 & 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

He also feels congruently about tails; if the clairvoyant told him that the outcome will be tails, he believes that it is more likely that the expert will assign a higher probability of tails than heads:

$$\{p|T,\varepsilon\} = \begin{cases} 3(1-p)^2 & 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.10}$$

A graphical display of these distributions is given in Figures 2.3 and 2.4. We can roughly say the decision maker feels that the expert is responsive to the outcome. Next we can calculate the decision maker's posterior probability of heads, given that the expert responds by saying that the probability of heads is p:

$$\{H|p,\varepsilon\} = \frac{\{p|H,\varepsilon\}\{H|\varepsilon\}}{\{p|\varepsilon\}} \tag{2.11}$$

By summing over the numerator we obtain

$$\begin{aligned}\{p|\varepsilon\} &= \{p|H,\varepsilon\}\{H|\varepsilon\} + \{p|\ T,\varepsilon\}\{T|\varepsilon\} \\ &= 3p^2 - 3p + 3/2\end{aligned} \tag{2.12}$$

which is the decision maker's prior assessment of the expert (Figure 2.5). Substituting the appropriate expressions in equation 2.11 we obtain:

$$\{H|p,\varepsilon\} = \begin{cases} \dfrac{p^2}{2p^2 - 2p+1} & 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

and

$$\{T|p,\varepsilon\} = \begin{cases} \dfrac{(1-p)^2}{2p^2 - 2p+1} & 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.14}$$

The decision maker's posterior on the probability of heads as a function of the expert's prior is plotted in Figure 2.6. Note that there are only three possible assignments by the expert that should lead the decision maker to adopt without revision the expert's prior: when the expert is certain of the outcome either way or when the expert responds

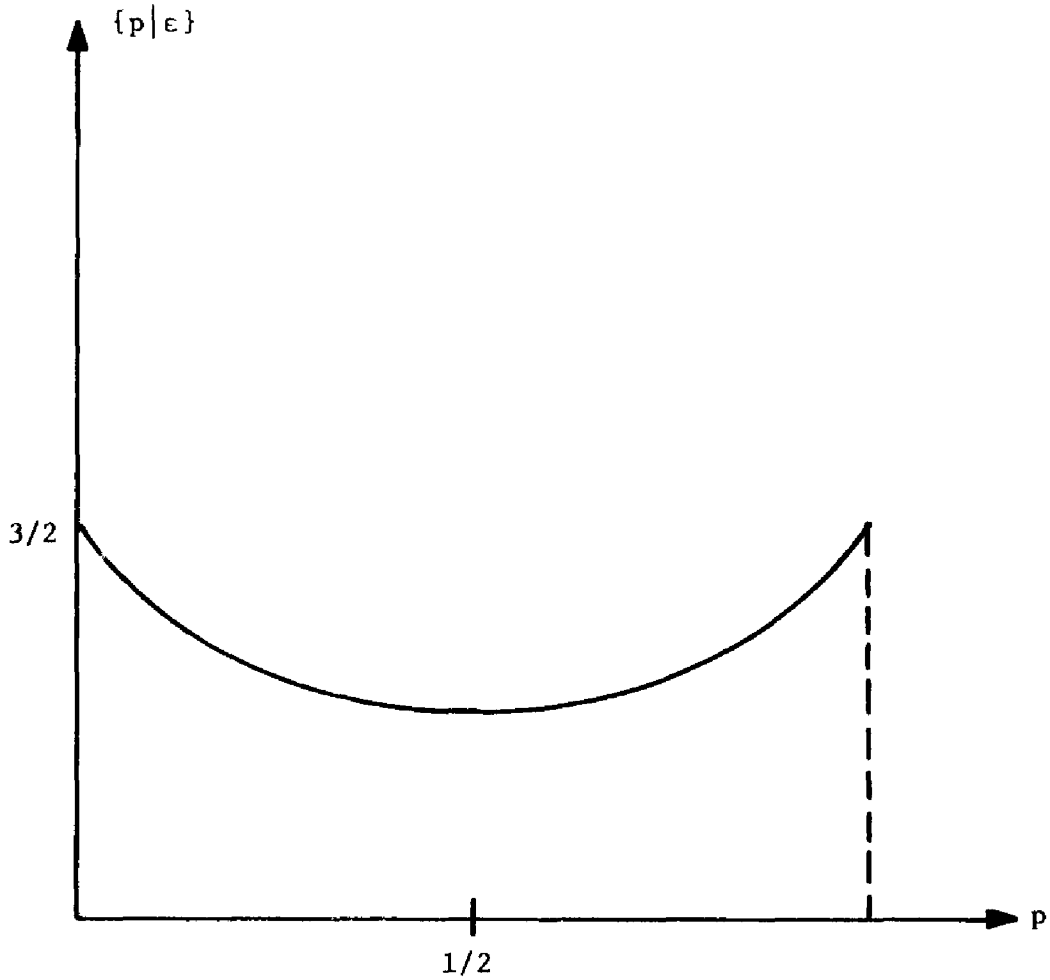


Figure 2.5 Decision Maker's Prior Assessment of the Expert

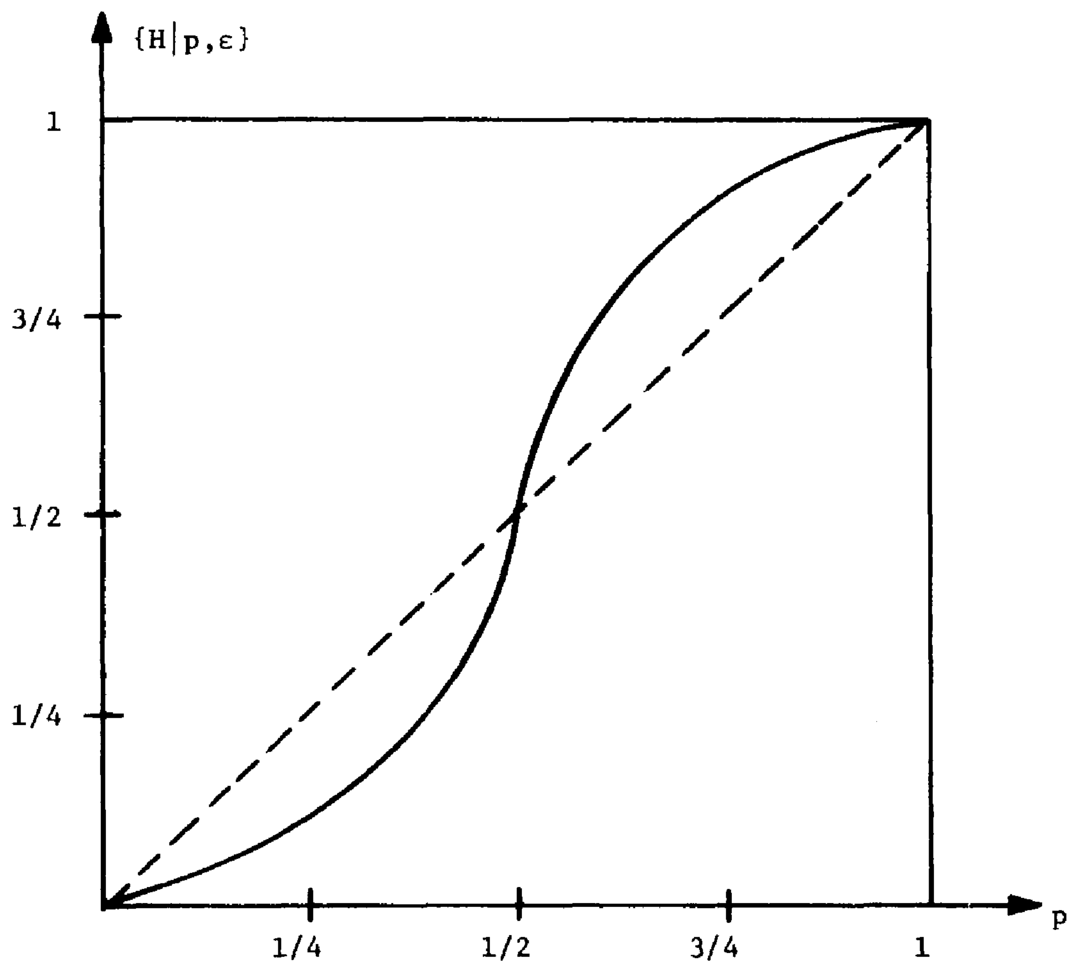


Figure 2.6 Decision Maker's Posterior as a Function of Expert's Prior

with the same prior as the decision maker's prior. No generalization should yet be drawn, however, as we shall see more clearly in Chapter V. Also notice that, aside from the three cases discussed above, the expert's advice always helps the decision maker to have a better idea of how the thumbtack will act: he will be able to assign a higher probability to either heads or tails. In fact, the decision maker feels that he is able to predict more closely than the expert himself the prospective outcome. This is because of, not in spite of, the decision maker's relatively high opinion of the expert. Further insight into this type of phenomenon will also be developed in Chapter V.

We shall continuously discover that our intuition needs much practice in regarding an expert's probability assignment as an observed bit of data, whose meaning to the decision maker may be quite independent of the generic meaning of the assignment to the expert himself. Without such practice intuition will be a poor guide to understanding many of the developments to follow.

Example 2.2--Recreation in Appalachia

Suppose that our decision maker is considering the possibility of investing in a large recreational complex in the Appalachian region of the United States. Appalachia presents a unique opportunity for such an investment as it has ideal recreational land that is virtually unused. Our decision maker feels that he could tell us for certain whether this investment is a good one, if he knew what the yearly demand will be. Unfortunately it is unclear to him whether a large complex could draw people away from other major vacation spots such as California and

Florida. In fact the decision maker considers it equally likely that the yearly demand will fall in any two equal segments between one hundred thousand and 1.1 million. Thus his prior density function on x, the yearly demand, is (Figure 2.7):

$$\{x|\varepsilon\} = \begin{cases} 1/10 & 1 \leq x \leq 11 \\ 0 & \text{otherwise} \end{cases} \tag{2.15}$$

where x is measured in units of hundred thousands.

Since the decision hinges on the decision maker's appraisal of x, he hires an expert who has worked with the Appalachian Regional Commission, has had experience in forecasting demand, and who is familiar with the specific region of interest.

Let us, for simplicity, constrain the expert's response to be in the set of uniform priors characterized by width w and center c. We assume that knowing the true value of x gives the decision maker information about the width of the expert's prior.

$$\{w|x,\varepsilon\} = \{w|\varepsilon\} \tag{2.16}$$

However the given width is still significant because it affects the assessment of c. Specifically the decision maker assesses (Figure 2.8)

$$\{c|w,x,\varepsilon\} = \begin{cases} 2/w^2(c - x + w/2) & x-w/2 \leq c \leq x+w/2 \\ 0 & \text{otherwise} \end{cases} \tag{2.17}$$

reflecting his confidence that the expert's prior will contain the true value within its bounds. This assessment also indicates that for a given x, the expert's prior mean (c) is more likely to be greater than x than below.

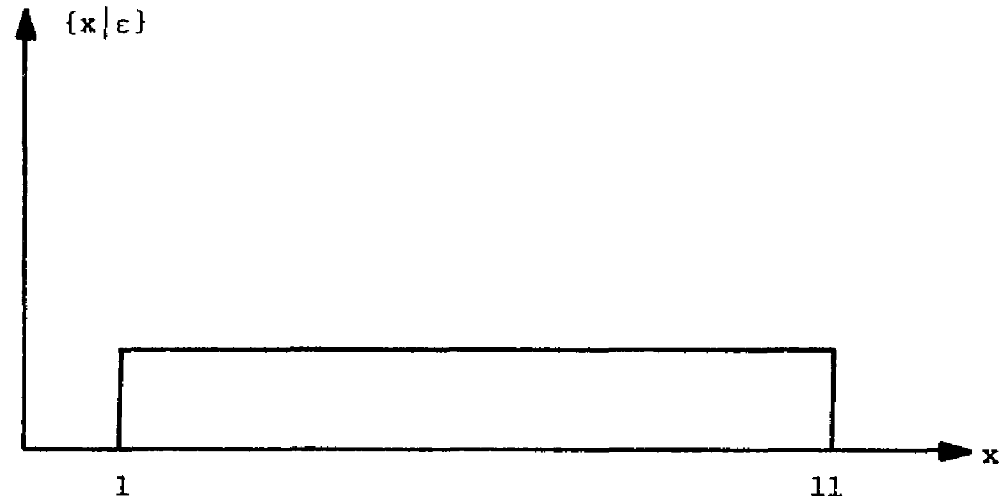


Figure 2.7 Decision Maker's Prior

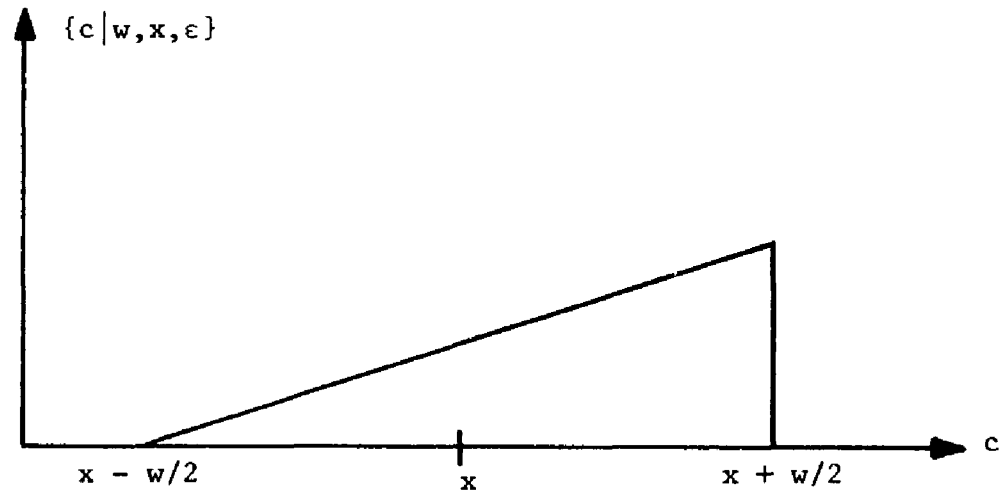


Figure 2.8 Likelihood Function

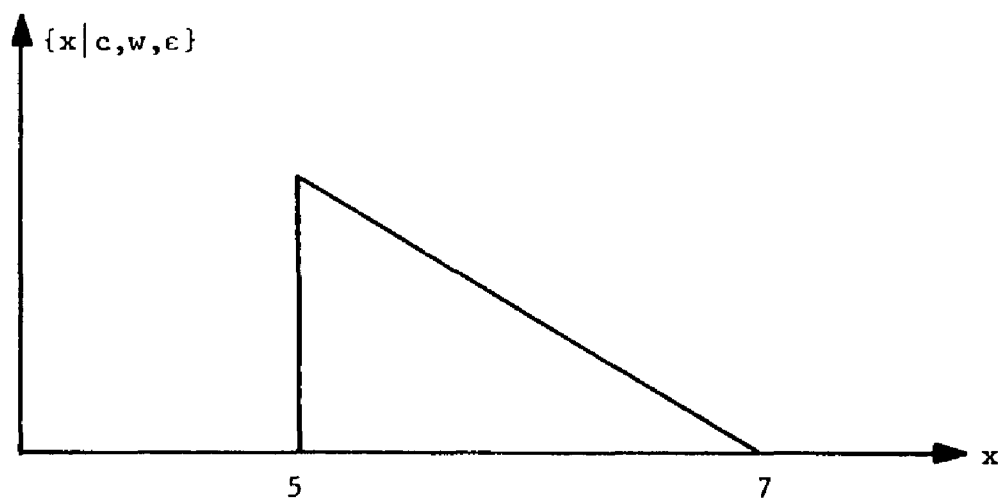


Figure 2.9 Posterior

We are now prepared to update the decision maker's prior using the formula

$$\{x|c,w,\varepsilon\} = k\ \{c|w,x,\varepsilon\}\{x|\varepsilon\} \tag{2.18}$$

Suppose that the expert responds with a uniform prior having width 2 and centered at 6. Therefore

$$\{c|w,x,\varepsilon\} = \begin{cases} 1/2\ (7-x) & 5 \leq x \leq 7 \\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

so that equation 2.18 yields the decision maker's posterior (Figure 2.9):

$$\{x|c,w,\varepsilon\} = \begin{cases} 1/2\ (7-x) & 5 \leq x \leq 7 \\ 0 & \text{otherwise} \end{cases} \tag{2.20}$$

This example shows more clearly how the decision maker's appraisal of the expert affects how the expert's advice is used. We see that, as in Example 2.1, the decision maker is now more certain about x than the expert. As we shall later demonstrate, this type of counter intuitive result can be predicted from certain special properties of the problem. For now it is important to realize that such a case is completely consistent with the theory.

2.4 A Specific Model of an Expert: a Preview of Problems to Come

Recent empirical studies concerning the ability of individuals to make probability assessments have revealed some surprising and fascinating results [1,23]. For example, H. Raiffa found that if a subject encodes a prior on a large number of random variables the true values will fall outside the midrange* of well over fifty percent of the distributions [1]. Even more remarkable, when a group of subjects are asked to evaluate a random variable, Raiffa found consistently that in somewhere between thirty to fifty percent of the assessments, individuals were surprised to find that the true value fell below the .001 fractile and above the .999 fractile.

It is clear that we cannot evaluate an expert's assessment performance from one probability distribution. It is Raiffa's position, however, that we can gauge an individual by many assessment performances. Thus, for instance, if the true value of a random variable consistently falls in an expert's .001 fractile, it is evident that the expert is not "externally calibrated" in Raiffa's terminology. In other words we wouldn't want to use the expert's distribution, yet we might wish to use the expert.

In order to "calibrate" an expert in certain situations, Raiffa has proposed an interesting technique. Study of this technique will introduce a topic that will be discussed extensively later in the dissertation; the phenomenon of a person thinking he knows more or less than he in fact knows. We will defer a careful look at the concept of calibration until Section 2.6.

---

*The midrange of a distribution is defined to be the interval between the .25 and the .75 fractiles. Therefore the area of the distribution over the midrange is .5.

Raiffa's model

Raiffa has introduced a procedure of altering subject probability assessments by modifying Beta distributions fitted to the assessments. He applies his techniques to variables that are defined to be proportions or fractions. For instance one variable assessed by subjects in his study was, "the percentage of first-year students responding on a questionnaire who prefer bourbon to scotch (excluding those who never drink)." In order to calibrate a subject's distribution, Raiffa determines the number of successes and the sample size of an experiment on a Bernoulli process that would infer the fitted Beta distribution. Thus, "If an assessor starts from a flat prior distribution on an uncertain proportion p and takes a random sample of size (n-2) and observes (r-1) so-called 'successes', then his posterior distribution on p is Beta with parameters [r,n]." Since experimentally the subjects' assessed distributions tended to be too tight, Raiffa explored the use of Betas with statistics indicating a smaller experiment. Specifically he tried Betas with parameters $[\alpha(r-1) + 1, \alpha(n-2) + 2]$ where α is some fraction between 0 and 1. He found that with α's equal to roughly .5 the true value fell in the modified distributions at about the expected frequency (using the modified distributions). Raiffa's conclusion is that,

> "Speaking in a rough way, we could say that these . . . subjects think that they know twice as much as they actually know! But let us caution how crude this statement is, since there is, of course, great question-by-question variability [1]."

Before we interpret these results in terms of our theory, let us explicitly review the assumptions behind Raiffa's technique:

(1) The expert starts out with a uniform prior on the variable of interest.

(2) The variable must be a fraction between 0 and 1.

(3) All the expert's experience relevant to the variable of interest must be the result of observations of independent, two outcome (Bernoulli) trials. That is, he must replicate a Bernoulli sampling process.

(4) If the expert is assessing proportions with respect to subsets of some population that he has experimented upon, he must believe the proportions in the subset of the population to be equal to the proportions in the entire population.

Although we could imagine fitting a Beta distribution to any probability function and carrying out the same procedure, the results obtained would have no meaningful interpretation without the above assumptions. We point out two more implicit assumptions made in using such a "calibrated" expert's prior below. We should first clarify, however, that our intention is not to critique Raiffa's model of the expert. Instead we wish to use it as a simple means of transmitting some of the immense complexity associated with formulating a meaningful analytic model of an expert. In doing so we will be underlining the importance of the ability of the present approach to deal with an expert in one-of-a-kind situations that are impossible to model other than subjectively.

<u>The Bernoulli model interpreted</u>

Let us capsulize Raiffa's "expert model" in the format of our theory. It will be helpful to assume at first that the expert's assessment of the random variable x, characterized by a Beta distribution of parameters [r,n] reflects his true experience. That is, he actually has observed r successes in n trials. We can then write in our notation,

$$\{x|\{x|\rho\},\varepsilon\} = \{x|r,n,\varepsilon\} \tag{2.21}$$

where $\{x|\rho\}$ includes the expert's experimental observations. Expanding we get

$$\{x|r,n,\varepsilon\} = k\{r,n|x,\varepsilon\}\ \{x|\varepsilon\} \tag{2.22}$$

But the likelihood function $\{r,n|x,\varepsilon\}$ is the expert's prior scaled by a constant. If we also assume that the decision maker's prior is uniform, we obtain

$$\{x|r,n,\varepsilon\} = \{x|\rho\} \tag{2.23}$$

There are two important observations to make. First, we notice that in order to use directly the expert's prior with Raiffa's model, we must assume that the decision maker has a uniform prior. Secondly, and perhaps even more fundamental, we see that it was crucial that both the expert's and the decision maker's likelihood function were identical. We assumed that the decision maker and the expert updated their distributions identically in the light of the same experimental observations. More precisely, we made the assumption that the decision maker presumed that the expert used the same fundamental model to characterize the production of the data. In this simple case the assumption seems justified.

We could easily suppose, however, that the decision maker was uncertain about whether the expert assumed a Bernoulli or a Markov model producing the data. In more complex situations such as the case where an expert is assessing the probability of nuclear war in the next ten years, the supposition that the decision maker would use the same subjective model, even given all the expert's relevant experience, is intolerable. We will return to this interesting topic in Chapter VI. For now we simply observe that in all but very simple cases it is <u>not</u> appropriate to implement the expert's prior directly, even if he could specify exactly the experimental evidence that formed its basis.

<u>Uncertainty on the expert's state of information</u>

Let us, however, adhere for the moment to all the implicit and explicit assumptions behind the Bernoulli model of the expert. It is interesting to relate our approach to the situation where we relax the condition, as did Raiffa, that the expert in fact knows as much as he thinks he knows. We can be slightly more general and assume that if the expert reports a Beta distribution indicating (r-1) successes in (n-2) trials, he really observed α(r-1) successes in β(n-2) trials. We need not even assume that α and β are less than 1.

Instead of a likelihood function characterized by [r,n] we should then use a likelihood function characterized by [α(r-1) + 1,β(n-2) + 2]. If the decision maker knew the values of α and β, he could calculate

$$\{x|r,n,\alpha,\beta,\varepsilon\} = k\{\alpha(r-1) + 1,\ \beta(n-2) + 2|x,\varepsilon\}\{x|\varepsilon\} \qquad (2.24)$$

Let us presume, however, that α and β are unknown parameters to the decision maker. He then must compute

$$\{x|r,n,\varepsilon\} = \iint_{\alpha\beta} \{x|r,n,\alpha,\beta,\varepsilon\}\ \{\alpha,\beta|\ r,n,\varepsilon\} \tag{2.25}$$

Substituting equation 2.24 into equation 2.25 and expanding the conditional assessment of α and β allows us to write:

$$\{x|r,n,\varepsilon\} = k \iint_{\alpha\beta} \{\alpha(r-1) + 1,\ \beta(n-2) + 2|x,\varepsilon\}\{\beta|n,\varepsilon\}\{\alpha|\beta,r,n,\varepsilon\}\{x|\varepsilon\} \tag{2.26}$$

where it is assumed that $\{\beta|r,n,\varepsilon\} = \{\beta|n,\varepsilon\}$. $\{\beta|n,\varepsilon\}$ is the decision maker's assessment of the scope of the expert's experience given that the expert reports an experience equivalent to (n-2) sample observations. $\{\alpha|\beta,r,n,\varepsilon\}$ is the decision maker's assessment of the results of the expert's observations given that the expert reports an experience equivalent to (r-1) observed successes in (n-2) trials. Equation 2.26 shows explicitly how to incorporate a slightly generalized form of Raiffa's model into our methodology.

In summary we have related our work to some of Raiffa's and, in so doing, highlighted the generality of our development. We posed Raiffa's model as a particular model of an expert in simple situations and used it to point out some subtle yet important aspects of the process of modeling an expert. Also emphasized was the practical importance of having a framework within which the expert may be subjectively modeled. In fact, we incorporated some subjective modeling aspects into Raiffa's basic model to slightly generalize it. We will build on many of the insights gained from this simple treatment in forthcoming chapters.

2.5 Value of an Expert

We are now in a position to calculate the value of an expert. This will be important for situations in which the employment of an expert is a decision in itself.

Since we view the expert essentially as a stochastic process whose output is a functional random variable, we may calculate the worth of obtaining the true value of the variable in the same way as the value of perfect information is computed in standard decision analysis problems [10]. Thus the expert's revelation of his prior is equivalent to the clairvoyant's revelation of the true value of a random variable.

Let $\langle u|d,\varepsilon\rangle$ connote the expected utility of a given decision situation if the decision maker makes decision d. We suppose that x is the only state variable relevant to the decision. That is, x is all we need to know to specify the utility of a given outcome. Without the expert's advice we can calculate

$$\langle u|d,\varepsilon\rangle = \int_x \langle u|x,d,\varepsilon\rangle \; \{x|\varepsilon\} \tag{2.27}$$

The best decision d* can then be selected by maximizing this quantity:

$$d^* \equiv \max_d{}^{-1} \langle u|d,\varepsilon\rangle \tag{2.28}$$

If, however, the expert gives the decision maker his prior, the decision maker will re-evaluate his decision. Specifically we can calculate an optimal decision based on revelation of the expert's prior as

$$d^* \left(\{x|\rho\}\right) = \max_{d}^{-1}\left[\int_{x} \langle u|x,d,\varepsilon\rangle \ \{x|\{x|\rho\},\varepsilon\}\right] \tag{2.29}$$

However, prior to receiving the expert's prior, the decision maker is uncertain as to what it is, so that the expected utility given that the expert will give advice (denoted by $\langle u|a,\varepsilon\rangle$) is

$$\langle u|a,\varepsilon\rangle = \int_{X} \langle u|d^*(\{x|\rho\}),\varepsilon\rangle\{\{x|\rho\}|\varepsilon\} \tag{2.30}$$

Here we are summing over X, the set of all possible expert priors. The value of the expert is that amount of money which, when deducted from all the outcomes, will make the expected utility conditional on using the expert equal to the expected utility without using him.

An example

For example, suppose that the decision maker is faced with the following situation: If he guesses correctly the outcome of a thumbtack flip he receives ten dollars. If he guesses incorrectly he loses ten dollars. We will assume for simplicity that the decision maker rates lotteries by their expected value. We also posit the same thumbtack and expert as in Example 2.1. For notational purposes let

$\langle v|\varepsilon\rangle \equiv$ the expected value of the decision without consulting the expert,

$\langle v|a,\varepsilon\rangle \equiv$ the expected value of the decision given that the expert will be consulted,

$d(p) \equiv$ the optimal decision given that the expert assesses a prior probability p of heads.

First we notice that $\langle v|\varepsilon\rangle = 0$. Secondly, we observe from Figure 2.6 that if p is greater than 1/2 the decision maker's posterior probability of heads will be greater than 1/2. Similarly if p is less than 1/2 the decision maker's posterior probability of heads is less than 1/2. Therefore we may write

$$d(p) = \begin{cases} \text{guess heads} & p \geq 1/2 \\ \text{guess tails} & p < 1/2 \end{cases} \tag{2.31}$$

We can now calculate

$$\langle v|a,\varepsilon\rangle = \int [\max_{d(p)} \langle v|d(p),p,\varepsilon\rangle]\{p|\varepsilon\} \tag{2.32}$$

Since d(p) depends only on whether p is greater or less than 1/2 we may write equation 2.32 as

$$\begin{aligned} \langle v|a,\varepsilon\rangle = & \int_0^{1/2} [\{T|p,\varepsilon\}\ (10) + \{H|p,\varepsilon\}\ (-10)]\{p|\varepsilon\} \\ & + \int_{1/2}^{1} [\{H|p,\varepsilon\}(10) + \{T|p,\varepsilon\}(-10)]\{p|\varepsilon\} \end{aligned} \tag{2.33}$$

Substituting in the appropriate expressions we obtain

$$\langle v | a,\varepsilon \rangle = \int_0^{1/2} \frac{(10(1-p)^2 - 10p^2)}{2p^2 - 2p + 1} (3p^2 - 3p + 3/2)\, dp$$

$$+ \int_{1/2}^{1} \frac{(10p^2 - 100 - p)^2)}{2p^2 - 2p + 1} (3p^2 - 3p + 3/2)\, dp \qquad (2.34)$$

$$= 7.5$$

We now calculate the expected value of the expert to be

$$\langle v | a,\varepsilon \rangle - \langle v | \varepsilon \rangle = 7.5 \qquad (2.35)$$

In this case the expert is worth three-fourths the value of the maximum prize the decision maker could obtain. We will explore the more interesting case of the value of many experts in Chapter III.

2.6 One Expert--Many Variables

In this section we analyze the situation where a decision maker is confronted with an expert's advice on more than one variable. The extension is straightforward, but is important in order to study conditions under which we can assume independence. In particular we shall discuss formally the interesting and fundamental concept of calibration.

Independence

It is extremely important when undertaking a complex decision analysis to build a model with as little probabilistic dependence in it as possible. Dependent random variables greatly magnify the computational cost in both the probabilistic phase and the information gathering phase. In addition, the assessment of probability distributions is greatly simplified by independence. Perhaps most important of all, when two or more variables are dependent, the human mind is, in most cases, taxed beyond capacity to understand the workings of the model; the crucial aspect of intuitive feedback in good model building is lost.

The typical decision analysis procedure is to assume independence if that is what the expert assesses. Indeed, it seems intuitive that if both the decision maker and his expert assess two variables to be independent the variables should be independent in the analysis. In fact, Raiffa presents a seeming paradox wherein one must choose between making the above independence assumption and agreeing with his experts on what act to take [18].

<u>Many variables</u>

In the most general case suppose that the decision maker is confronted with a joint distribution on n random variables from his expert. Let

$$\underline{x} = (x_1,\ldots,x_n) \equiv \text{a vector of n random variables.}$$

All we need do is substitute $\underline{x}$ for x in the one variable case and the basic result follows:

$$\{\underline{x}|\{\underline{x}|\rho\},\varepsilon\} = k\ \{\{\underline{x}|\rho\}|\underline{x},\varepsilon\}\{\underline{x}|\varepsilon\} \tag{2.36}$$

The crucial element is again the likelihood function. In order to study independence we will, for ease of presentation, concentrate on the two variable case. The result is trivially generalized.

Conditions for independence

Suppose that the two random variables under consideration are x and y. We will first state three basic conditions and then determine whether they imply that the decision maker's posterior on x and y should be independent. The three conditions are as follows:

(1) $\{x,y|\varepsilon\} = \{x|\varepsilon\}\{y|\varepsilon\}$ - the decision maker assesses the variables to be independent prior to receiving the expert's opinion.

(2) $\{x,y|\rho\} = \{x|\rho\}\{y|\rho\}$ - the expert assesses the variables to be independent.

(3) The decision maker knows that the expert assesses the variables to be independent.

We wish to determine whether or not the decision maker's new state of knowledge conditional on obtaining the expert's prior implies an independent posterior; that is, does

$$\{x,y|\{x,y|\rho\},\varepsilon\} = \{x|\{x,y|\rho\},\varepsilon\}\ \{y|\{x,y|\rho\},\varepsilon\}? \qquad (2.37)$$

To start the investigation notice that due to condition 3 the decision maker receives no additional information from knowing the expert's joint prior distribution than from knowing the expert's two prior marginal distributions. Specifically, the following relationship should hold:

$$\{x,y|\{x,y|\rho\},\varepsilon\} = \{x,y|\{x|\rho\}, \ \{y|\rho\},\varepsilon\} \tag{2.38}$$

We may expand the right-hand term of the above equation and then apply condition 1 to obtain:

$$\{x,y|\{x,y|\rho\},\varepsilon\} = k\{\{x|\rho\},\{y|\rho\}|x,y,\varepsilon\}\{x|\varepsilon\}\{y|\varepsilon\} \tag{2.39}$$

Next we expand the large righthand term using the definition of conditional probability to write:

$$\{x,y|\{x,y|\rho\},\varepsilon\} = k\{\{x|\rho\}|x,y,\varepsilon\}\{\{y|\rho\}|\{x|\rho\},x,y,\varepsilon\}\{x|\varepsilon\}\{y|\varepsilon\} \tag{2.40}$$

Since we know from condition 3 that the expert assesses x independently of y, we assume that the decision maker's assessment of the expert's prior on x is independent of y; that is

$$\{\{x|\rho\}|x,y,\varepsilon\} = \{\{x|\rho\}|x,\varepsilon\} \tag{2.41}$$

Using this relation in equation 2.40 we have finally that

$$\{x,y|\{x,y|\rho\},\varepsilon\} = k\{\{x|\rho\}|x,\varepsilon\}\{\{y|\rho\}|\{x|\rho\},x,y,\varepsilon\}\{x|\varepsilon\}\{y|\varepsilon\} \tag{2.42}$$

Now we see that the crucial term is that involving $\{y|\rho\}$. If we assume that

$$\{\{y|\rho\}|\{x|\rho\},x,y,\varepsilon\} = \{\{y|\rho\}|y,\varepsilon\} \quad , \tag{2.43}$$

then we have proved that

$$\{x,y|\{x,y|\rho\},\varepsilon\} = \{x|\{x,y|\rho\},\varepsilon\}\{y|\{x,y|\rho\},\varepsilon\} \tag{2.44}$$

which is the desired relation.

It is clear that independence rests on $\{\{y|\rho\}|\{x|\rho\},x,y,\varepsilon\}$, the assessment of the expert's prior on y given his prior on x, and x and y. Although it is true that the assessment of $\{y|\rho\}$ is independent of $\{x|\rho\}$ alone and x alone it is not necessarily independent of both together.

The joint information $\{x|\rho\}$ and x, in fact, gives an indication of the expert's performance in assessing the random variable x. Although this doesn't affect the decision maker's view of the expert's assessment of the random variable y directly, it may well affect the decision maker's evaluation of him as an expert.

Suppose for instance that we tell the decision maker that the true value of x is x_o and that the expert's prior on x is that shown in Figure 2.10. In words, the expert reported that he was sure x_o fell in a certain narrow interval when, in fact, it didn't. It is easy to conceive how this type of information would affect the decision maker's evaluation of $\{y|\rho\}$ given that true value of y is revealed to be y_o. For instance, he might be moved to attach a significantly higher probability to the expert's prior not containing y_o, since the expert has proven himself capable of such an assessment.

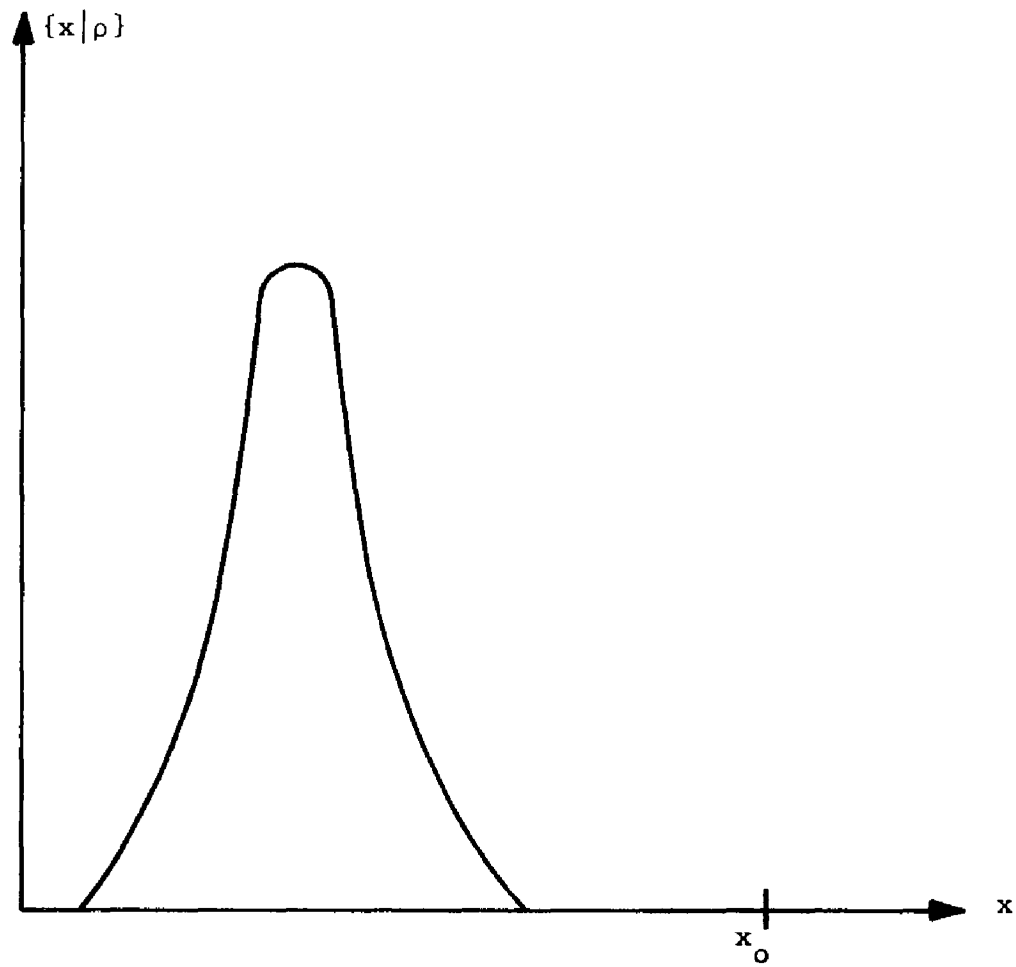


Figure 2.10 The Expert's Assessment Performance

Calibration

It is useful to formalize Raiffa's concept of calibration [1]. We will do so, however, in a somewhat different spirit. Raiffa's definition of calibration has to do with what he calls "external validation." Thus if an expert is asked to give midrange estimates for a large number of variables and, in fact, the true value of the variable falls in the midrange of half the assessments and an equal number of times in the upper and lower quartiles, Raiffa would say that the expert's assessments are externally validated, and he is calibrated. Depending on the number of fractiles an expert assesses to specify each distribution, his calibration (in Raiffa's usage) may vary. Thus one must play off "discriminability" versus external validation.

Unfortunately, this concept of calibration is of no use to the decision maker for one of a kind assessments. We would like a Bayesian notion of calibration that does not depend on a large number of trials. Our definition of calibration will be a subjective one. It will bear approximately the same relation to Raiffa's definition of calibration that the subjectivist definition of probability does to the long run frequency definition of probability. It will not depend on external validation. In particular it will be interesting to observe under what conditions the two definitions are equivalent.

We define the expert to be calibrated on the variable y with respect to the variable x, if the decision maker learns nothing about the expert's assessment of y by observing his performance on x. Note that similar to independence, calibration is in the mind of the decision maker. Additionally, calibration is relative to a particular expert.

For notational purposes let:

$yCx \equiv$ y is calibrated with respect to x.

Formally we may write: yCx if and only if

$$\{\{y|\rho\}|x,y,\{x|\rho\},\varepsilon\} = \{\{y|\rho\}|y,\varepsilon\} \tag{2.45}$$

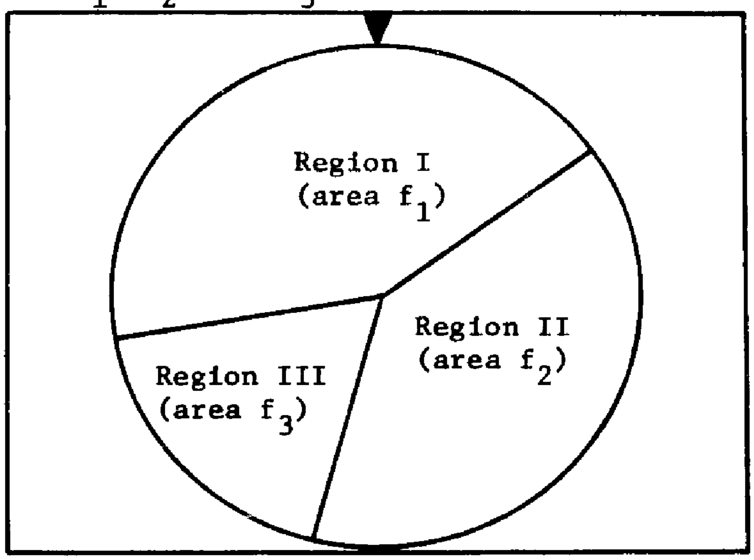
Thus we see that calibration is an additional condition to the three already stated that must hold to guarantee posterior independence. Observe that calibration has other properties that make it similar to independence. For instance, that yCx implies xCy can be easily shown. Also, the calibration between two variables may change as the decision maker gains new knowledge.

We can speak of calibration being weak or strong just as dependence is weak or strong. In general we can roughly say that as the expert performs more and more assessments for the decision maker, his calibration will go up. Thus, if the decision maker has observed the expert many times on <u>similar</u> variables, the observation of one additional variable shouldn't affect much the decision maker's view of the expert's assessment performance. Just as when a tool is calibrated we no longer have to check it against a standard measurement, so when an expert is calibrated we no longer must assess him relative to his previous assessment performances. Chapter V will solidify this point.

We shall now examine in depth the relationship between Raiffa's calibration concept and our own. In so doing we will develop insights into the whole problem of using an expert.

The expert as a stochastic process

Suppose that we view the expert as a stochastic estimating process, invariant to the type of random variable he is assessing. For example, take the case of an expert who assesses only midranges, not entire probability distributions. We could hypothesize that the expert acts like a "wheel of fortune" whose total area of one is divided into three slices whose areas are f_1, f_2 and f_3. Such a wheel is drawn below.



Next we could assert that regardless of what random variable the expert is assessing, the wheel can be spun to determine whether the true value of the variable falls in or out of the midrange. Specifically if the pointer lands in Region I the true value will fall in the lower quartile, if it lands in Region II the true value will fall in the midrange, and if it lands in Region III the true value will fall in the upper quartile.

It is clear that if we know f_1, f_2 and f_3, and nothing beyond what our expert says about the variable of interest, we would be led to assign probabilities f_1, f_2 and f_3 to the variable falling in each of the three regions the expert provides. This would be true for whatever variable he assessed since by assuming that the process is invariant to the type of random variable we are merely assuming that he always spins the same partitioned wheel. The actual numbers specifying the midrange may vary from variable to variable but the process for determining whether the variable falls in the midrange remains the same. In this case, where f_1, f_2 and f_3 are known, the two definitions of calibration are the same. Subjectively, if we know how the expert performs on one variable it doesn't influence our assessment of him on the other. In the subjective sense he is fully calibrated. If we were to spin the wheel repeatedly, corresponding to the expert making many assessments, the fraction of times the pointer falls in region II would be f_2. Since this equals the probability assessment we make each time, in Raiffa's sense we have calibrated the expert. Notice however that the expert has not calibrated himself. In fact, in this case we would not want him to calibrate himself, since by doing so we are gaining nothing, but requiring of ourselves a more sophisticated model--that of the expert's own calibration procedure.

Of course, even if we model the expert as a process of this sort, the above discussion is unrealistic. A more realistic treatment is to allow f_1, f_2 and f_3 to be random variables; that is, we know that the expert acts as a stationary stochastic process but we aren't sure which one. In this case we may view the situation as an inference problem. With each assessment the expert makes (each spin of the wheel) we get information as to the parameters (f_1,f_2,f_3) of the process. We can get arbitrarily close to knowing exactly what the parameters are as the expert performs many assessments. Thus in both definitions of calibration the expert becomes fully calibrated. However, if the expert has made few or no assessments, the long run frequency view of calibration aids us little.

In general, the above sort of model is the only one in which the classical definition makes sense. The crucial assumption is that the expert is a sort of static process who is an independent evaluator of random variables. If we model him as a dynamic process, then clearly we need more than the relative frequencies of his assessment performance to specify how to use him. Similarly if the expert himself assesses the random variables of interest to be dependent, then the long run frequency approach becomes invalid.

In most decisions worthy of a formal analysis we are not fortunate enough to be able to have many independent assessments by the expert. As we shall see in Chapter IV on prior assessment, the expert actually <u>should</u> behave normatively as a dynamic process and, in fact, probably does descriptively. Intuitively we can expect the expert to try to calibrate himself. We shall later capitalize on a fundamentally different frequency interpretation to derive some basic results in this area.

The subjective definition of calibration is a useful tool in practical situations and will prove to be an important concept in our future developments. It has already proven to be of merit in analyzing the property of independence.

2.7 Exchangeability

We present now a review of the concept of exchangeability: an idea important to the remainder of the dissertation.

Exchangeability is a concept whose beauty lies in its simplicity. Exchangeability is a subjective assessment pertaining to a set of random variables which is useful in the same sense as the concept of probabilistic independence: when applicable it facilitates the assessment of probabilities and the solution of inference problems.

De Finetti's original definition of the concept is as follows:

> "We shall say that $x_1, x_2, \ldots, x_n, \ldots$ are exchangeable random quantities if they play a symmetrical role in relation to all problems of probability, or in other words, if the probability that $x_{k_1}, x_{k_2}, \ldots, x_{k_n}$ satisfy a given condition is always the same however the distinct indices $k_1 \ldots k_n$ are chosen." [7]

The assessment of exchangeability over a set of variables at a given point in time means that at that time the variables are indistinguishable quantities to the assessor. The assessor's state of information about the set must be invariant to a relabeling of the members in the set. Clearly a necessary condition for exchangeability is that all variables in an exchangeable set must have identical assigned prior distributions.

Four related results in the study of exchangeability (all provided by de Finetti [7]) are particularly basic. Let $(x_1, x_2, \ldots)$ be a set of exchangeable variables. We define H_m as the revealed frequency histogram of an <u>arbitrary</u> collection of m of these variables. The four results are:

1. The assessment $\{H_m|\varepsilon\}$ tends to a limit $\{H|\varepsilon\}$ as $m \to \infty$. This allows us to speak unambiguously of assessing a limiting histogram, or frequency distribution, H.
2. The probability distribution of x_i given H is just the histogram itself. Writing H as a function $H(\cdot)$ we have $\{x|H(\cdot),\varepsilon\} = H(x)$.
3. The variables are conditionally independent given H; that is

$\{x_i | \underline{x}, H, \varepsilon\} = \{x_i | H, \varepsilon\}$ for an arbitrary finite set of variables $\underline{x}$ not containing x_i.

4. The observation of H_m for large m implies that, with practical certainty, all future histograms will be, within arbitrarily fine specification, identical. Thus the limiting histogram H may for all intents and purposes be treated as a measurable, observable quantity.

The results become clearer in the Bernoulli trial context. In this case H is a long run frequency. Result 1 allows us to assess the long run frequency of successes. Result 2 is the intuitive notion that the probability of success given the long run frequency is the long run frequency. Result 3 makes the inference problem mathematically identical to that where data is being produced by a Bernoulli model of unknown parameter. Result 4 forms a theoretical basis for being "sure" that a fair coin will, in the long run, come up heads fifty percent of the time.

Second order analysis

We may derive an immediate application of exchangeability. Conceptually we can envision a large set of variables exchangeable to the variable of interest x. If we denote by H, the long run histogram of this set, we may expand:

$$\{x | \{x|\rho\}, \varepsilon\} = \int_H \{x | H, \{x|\rho\}, \varepsilon\}\{H | \{x|\rho\}, \varepsilon\} \tag{2.46}$$

However, if H is known, by exchangeability the distribution on x is specified independent of $\{x|\rho\}$. The second term under the integral may be expanded, allowing equation 2.46 to be rewritten as

$$\{x|\{x|\rho\},\varepsilon\} = k \int_H H(x)\{\{x|\rho\}|H,\varepsilon\}\{H|\varepsilon\} \tag{2.47}$$

when again, H(x) is the histogram depicted as a function evaluated at the point x.

The above equation forms a different sort of decomposition than was available previously. We may speak very roughly of H(x) being the "true" distribution of x. Then $\{H|\varepsilon\}$ is the decision maker's appraisal of this distribution. The term $\{\{x|\rho\}|H,\varepsilon\}$ may be viewed as an assessment of how closely the expert's own assessment matches the "true" probability distribution. This assessment may in many cases be easier than the direct assessment $\{\{x|\rho\}|x,\varepsilon\}$. In Chapters IV and V the concept of exchangeability will be further exercised to provide more fundamental insights and assessment techniques.

2.8 Summary

In this chapter we have provided the basic framework upon which we will build in the remainder of the thesis. We have addressed the problem of how to use one expert in a very general yet precise context. In particular we have already included many of the components of the intuitive model presented in Chapter I.

Basically we concluded that the expert should be used as any experimental data should be used. There are, however, some conceptual differences. In a typical experiment the data comes in the form of a number, whereas the expert produces a function. Also, the expert provides a challenging modeling opportunity. He is not a well-defined experiment. We found in our formulation that the likelihood function provides us with the structural basis on which to model the expert. As one example of how to model the expert in simple, well-defined situations we borrowed an idea of Raiffa's, and interpreted it in terms of our treatment. We were able to specify exactly what assumptions were necessary in order to use an expert via Raiffa's modeling. In so doing we were also able to indicate the theoretical generality of the present approach.

The calculation of the value of an expert fell out naturally from the theory. We simply viewed the expert as a clairvoyant giving the decision maker perfect information on a random variable (the expert's prior).

We then generalized our development to the one expert-many variable case. In doing so we showed that the assumption of independence of two or more random variables after the decision maker consults an expert rests on several more basic assumptions. Specifically, we determined that in order to assume posterior independence for the decision maker we must assume that the expert is subjectively calibrated relative to the decision maker. This result shed some light in an area often confused in the literature. Finally, the concept of exchangeability was discussed and illustrated.

In summary, we have created a solid foundation upon which to build but have really just begun. We shall find in the remainder of this dissertation that regardless of how complex the material gets, we will continually draw on the basic concepts of this introductory chapter.

Chapter III

MULTI-EXPERT USE

3.0 Introduction

The purpose of this chapter is to provide a basic framework within which problems concerning the relationship of a decision maker to a group of experts may be viewed. It will be recalled that the proper use of a panel of experts was one of the prime motivating problems of this dissertation. We will solve this problem in its simplest form and postpone more complex variations until Chapter V.

The basic solution for many experts giving advice on one variable is given in Section 3.1. In Section 3.2 we extend the discussion to the new concept of cross-calibration. Section 3.3 will further develop the intuitive implications of the theory through the use of three different examples. Finally, the specific problem of how to select a panel of experts is addressed in Section 3.4.

The termination of this chapter will mark the end of our first pass at the problem of using experts. Starting in Chapter IV we will use the theory developed to present and solve more complex and interesting problems, and to relate our treatment to others.

3.1 The Basic Problem: Many Experts, One Variable

Suppose our decision maker is confronted with the advice of not one, but m experts. Let

$\rho_i \equiv$ the i^{th} expert's state of information prior to to giving the decision maker advice

$\{x|\rho_i\} \equiv$ the i^{th} expert's prior on x.

The formal problem we address ourselves to in this section is:

Given 1) $\{x|\varepsilon\}$, the decision maker's prior on the random variable x

2) $\{x|\rho_i\}$ $i = 1,\ldots,m$, the priors of m experts on the random variable x

How should the decision maker's prior be altered upon reception of the set of expert priors?

The logically correct way to solve this problem is, as in Chapter II, an application of Bayes' Theorem. In this case the expression to be derived may be written as:

$$\{x|\{x|\rho_1\},\{x|\rho_2\},\ldots,\{x|\rho_m\},\varepsilon\} \tag{3.1}$$

Applying Bayes' Theorem once gives:

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_m\},\varepsilon\} =$$

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_{m-1}\},\varepsilon\}\ \{\{x|\rho_m\}|x,\{x|\rho_1\},\ldots,\{x|\rho_{m-1}\},\varepsilon\} \tag{3.2}$$

By induction, we may repeatedly apply Bayes' Theorem to obtain the basic result:

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_m\},\varepsilon\} =$$

$$k\{x|\varepsilon\}\{\{x|\rho_1\}|x,\varepsilon\}\{\{x|\rho_2\}|x,\{x|\rho_1\},\varepsilon\}\ldots\{\{x|\rho_m\}|x,\{x|\rho_1\},\ldots,\{x|\rho_{m-1}\},\varepsilon\} \tag{3.3}$$

We see that to correctly use the advice of many experts, the decision maker must not only specify his feelings about what each expert's distribution will be, but must also make assessments about the relationships between the experts. Thus in evaluating

$$\{\{x|\rho_2\}|x,\{x|\rho_1\},\varepsilon\} \tag{3.4}$$

the decision maker, with no further modeling, must answer questions as:

> "If I know Expert 1's distribution on x and, in addition, I know where the true value of x will lie in this distribution, what does this information lead me to believe about what Expert 2's distribution will be?"

Notice that the above question is similar to the question asked of the decision maker in the one expert case to ascertain whether the expert was calibrated. In this case, however, we are not calibrating relative to an expert's past assessment, but rather to another expert's performance. We are thus led by a natural extension of the definition of calibration to a definition of what shall be named cross-calibration.

3.2 Cross-calibration

We define Expert 1 to be cross-calibrated relative to Expert 2 if the decision maker's assessment of $\{x|\rho_1\}$ given x is independent of $\{x|\rho_2\}$. That is, if

$$\{\{x|\rho_1\}|x,\{x|\rho_2\},\varepsilon\} = \{\{x|\rho_1\}|x,\varepsilon\} \tag{3.5}$$

This definition can be extended to many experts by defining Expert 1 to be cross-calibrated relative to experts 2 through m if and only if

$$\{\{x|\rho_1\}|x,\{x|\rho_2\},\ldots,\{x|\rho_m\},\varepsilon\} = \{\{x|\rho_1\}|x,\varepsilon\} \tag{3.6}$$

Again we emphasize that cross-calibration is similar to calibration in that both concepts explicitly use the concept of performance, but cross-calibration is with respect to two or more expert performances and calibration is with respect to one expert's past performance. Also note that in the definition of cross-calibration the x is not dropped out of the conditioning side of the expression. The decision maker's assessment of the expert still depends on the random variable.

As in the single expert case, each of the factors in the numerator of equation 3.3 requires, most generally, a subjective assessment. In many applications we expect that the decision maker's feelings about each of his experts will be invariant to the information gained by knowing the other experts' priors in addition to the true value of the random variable; we expect the experts to all be cross-calibrated. In this case equation 3.3 simplifies to

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_m\},\varepsilon\} = k\{x|\varepsilon\}\{\{x|\rho_1\}|x,\varepsilon\}\ldots\{\{x|\rho_m\}|x,\varepsilon\} \tag{3.7}$$

Therefore, when the experts are all cross-calibrated, the decision maker has only to specify m independent likelihood functions. As a caveat it should be noticed that, although cross-calibration may seem an easy condition to satisfy, the experts need not even be aware of each other's existence for cross-calibration not to obtain. The simplest example of this would be two experts in the same field whom the decision maker feels are extremely competent. If one expert assesses a narrow distribution about some interval, the decision maker might be led to affix a high probability to the second expert assessing a narrow distribution about the same interval.

The extension of the theory to the many variable case is direct. In general there may be dependencies between the experts due to dependencies among the random variables. Thus it may be displayed that, for guaranteed posterior independence, the experts must not only be each calibrated relative to all variables, but must also be cross-calibrated relative to all possible sets of variables.

Dependence of the assessment of experts on advice received

A further interesting fact is that regardless of whether the experts are cross-calibrated or not, the reception of the advice of one expert changes the decision maker's assessment of the other experts. To see this we return to the two expert case. The decision maker's prior assessment of Expert 2 is $\{\{x|\rho_2\}|\varepsilon\}$ which can be calculated by integration:

$$\{\{x|\rho_2\}|\varepsilon\} = \int_x \{\{x|\rho_2\}|x,\varepsilon\}\{x|\varepsilon\} \tag{3.8}$$

However, after obtaining Expert 1's advice we can determine:

$$\{\{x|\rho_2\}|\{x|\rho_1\},\varepsilon\} = \int_x \{\{x|\rho_2\}|x,\{x|\rho_1\},\varepsilon\}\{x|\{x|\rho_1\},\varepsilon\} \tag{3.9}$$

If the experts are cross-calibrated

$$\{\{x|\rho_2\}|\{x|\rho_1\},\varepsilon\} = \int_x \{\{x|\rho_2\}|x,\varepsilon\}\{x|\{x|\rho_1\},\varepsilon\} \tag{3.10}$$

Since, in general, $\{x|\varepsilon\}$ is not equal to $\{x|\{x|\rho_1\},\varepsilon\}$ we have displayed the assertion. The crucial element is that even though the decision maker's assignment of $\{x|\rho_1\}$ conditional on x remains the same, the assessment of x itself changes. Thus integrating over x produces a different posterior on $\{x|\rho_1\}$. This fact has important ramifications in determining the proper selection of a "panel of experts" (Section 3.4).

We will study other conditions that imply simplification in Chapter V. For now we shall present three illustrative examples; the first an extension to the thumbtack flipping example in Chapter II with two cross-calibrated experts, the second a similar extension for the case of two experts who are not cross-calibrated, and the third an example of a current situation involving experts of highly differing opinions in a slightly more complex situation.

3.3 Examples

Example 3.1 - Thumbtack Flipping: Two Cross-calibrated Experts

Assume that the decision maker employs the use of a second expert in addition to the expert he used in Example 2.1. We let the subscript "1" denote the original expert and subscript "2" denote the additional expert. The original expert's prior is again characterized by p, the probability he assigns to heads; the second expert's prior is characterized by q, the probability he assigns to heads. Thus

$$\{H|p,q,\varepsilon\} = \frac{\{H|\varepsilon\}\{p|H,\varepsilon\}\{q|H,p,\varepsilon\}}{\{p,q|\varepsilon\}} \tag{3.11}$$

from equation 3.3. Next assume that the decision maker's prior opinion about both experts is identical, i.e.,

$$\begin{aligned} \{p|H,\varepsilon\} &= \{q|H,\varepsilon\} \\ \{p|T,\varepsilon\} &= \{q|T,\varepsilon\} \end{aligned} \tag{3.12}$$

and that knowing the performance of Expert 1 doesn't influence the decision maker's assessment of the performance of Expert 2; the experts are cross-calibrated so that

$$\{q|H,p,\varepsilon\} = \{q|H,\varepsilon\},\ \{q|T,p,\varepsilon\} = \{q|T,\varepsilon\} \tag{3.13}$$

We may rewrite equation 3.11 as

$$\{H|p,q,\varepsilon\} = \frac{\{H|\varepsilon\}\{p|H,\varepsilon\}\{q|H,\varepsilon\}}{\{p,q|\varepsilon\}} \tag{3.14}$$

Similarly, we can calculate $\{T|p,q,\varepsilon\}$.

The joint prior on p and q is determined by summing over the numerator of equation 3.14:

$$\{p,q|\varepsilon\} = \begin{cases} 9/2[p^2q^2 + (1-p)^2(1-q)^2] & 0 \le p \le 1,\ 0 \le q \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.15}$$

We are now prepared to compute

$$\{H|p,q,\varepsilon\} = \begin{cases} \dfrac{p^2q^2}{p^2q^2 + (1-p)^2(1-q)^2} & 0 \le p \le 1,\ 0 \le q \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

Equation 3.16 is displayed in Figure 3.1. The curves represent the posterior as a function of p for different values of q. We see that if the second expert responds with a uniform prior ($q = 1/2$), the decision maker's posterior is the same as in the one expert case irrespective of how the first expert responds. Also notice that the points at which the decision maker's posterior equals Expert 1's prior depends on Expert 2's prior. In fact, as the line marked $p = q$ in the diagram shows, even if the two experts respond identically, the decision maker's posterior is the same as their priors only when

(a) $p = q = 1$ or

(b) $p = q = 0$ or

(c) $p = q = 1/2$.

In the typical case, if the experts respond unanimously the decision maker gains additional confidence in the outcome.

It is interesting to calculate the decision maker's posterior if the two experts have complementary priors. By this we mean that Expert 1's prior probability of heads equals Expert 2's prior probability of tails: $q = 1 - p$. When this is the case equation 3.16 reduces to

$$\{H|p,\ 1-p,\varepsilon\} = \frac{p^2(1-p)^2}{p^2(1-p)^2 + (1-p)^2p^2} = 1/2 \tag{3.18}$$

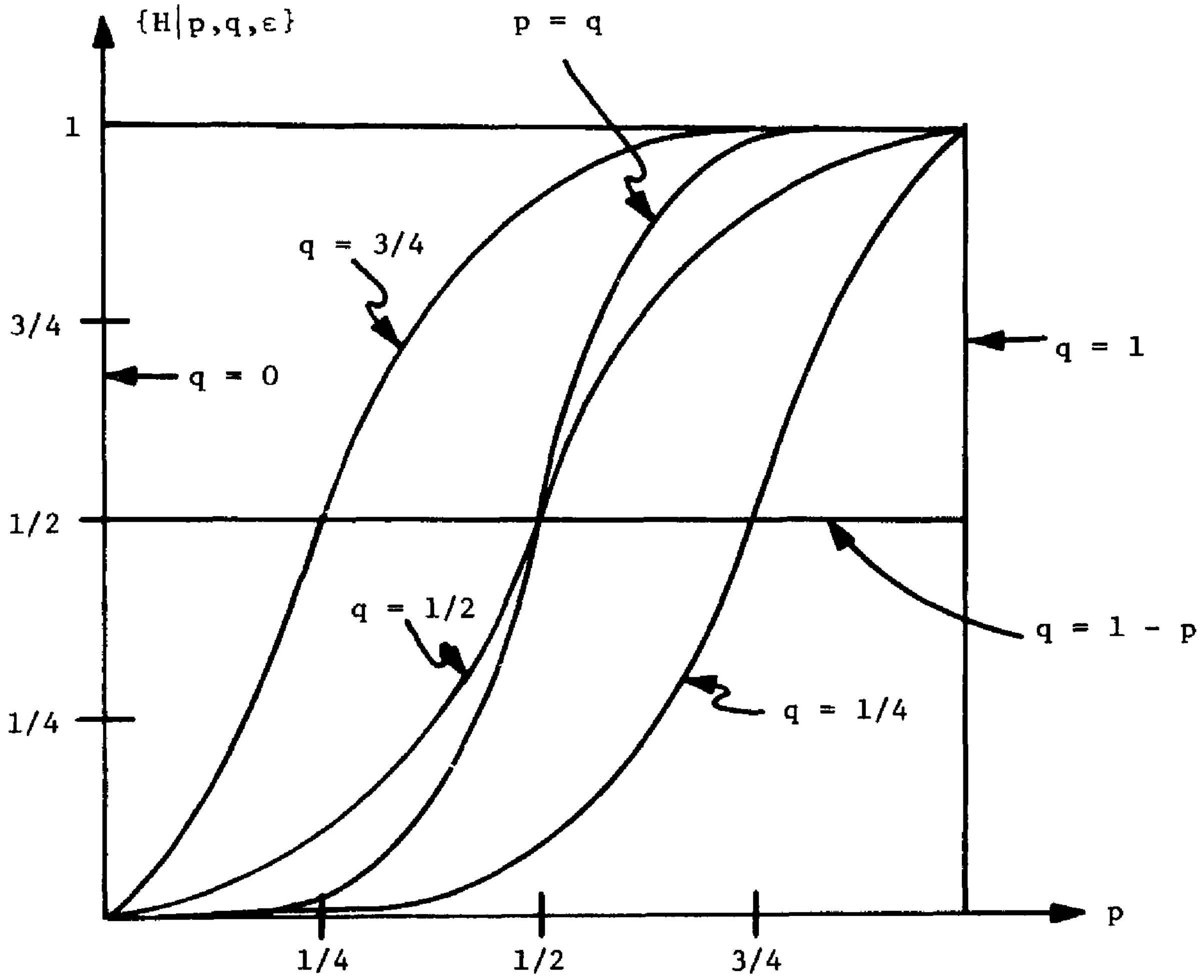


Figure 3.1 Decision Maker's Posterior as a Function of Both Expert Priors

Therefore the decision maker learns nothing about the thumbtack. We should be careful to distinguish this from the case where the experts always answer with complementary distributions for one reason or another. In this instance the additional value of the second expert will be zero.

Finally we can calculate the decision maker's assessment of Expert 2 given that he has received Expert 1's prior:

$$\{q|p,\varepsilon\} = \{q|H,p,\varepsilon\}\{H|p,\varepsilon\} + \{q|T,p,\varepsilon\}\{T|p,\varepsilon\}$$

$$= \frac{3(p^2q^2 + (1-p)^2(1-q)^2)}{2p^2 - 2p + 1} \tag{3.19}$$

We plot this in Figure 3.2 for different values of p, noting that for p = 1/2 it reduces to the decision maker's prior on q. As p approaches 1 the decision maker's assessment of q approaches the likelihood function on q given heads, as expected. Similarly as p approaches 0 the assessment of q approaches the likelihood function on q given tails.

The closer p gets to 0 or 1 the more likely q will be close to 0 or 1. Therefore, the surer Expert 1 is, the surer the decision maker expects Expert 2 to be. Also, the decision maker himself becomes more confident in predicting the outcome of the lottery. It will be interesting in Section 3.4 to calculate the cumulative effect of these two phenomena on the dependence of the value of Expert 2 on Expert 1's prior assessment.

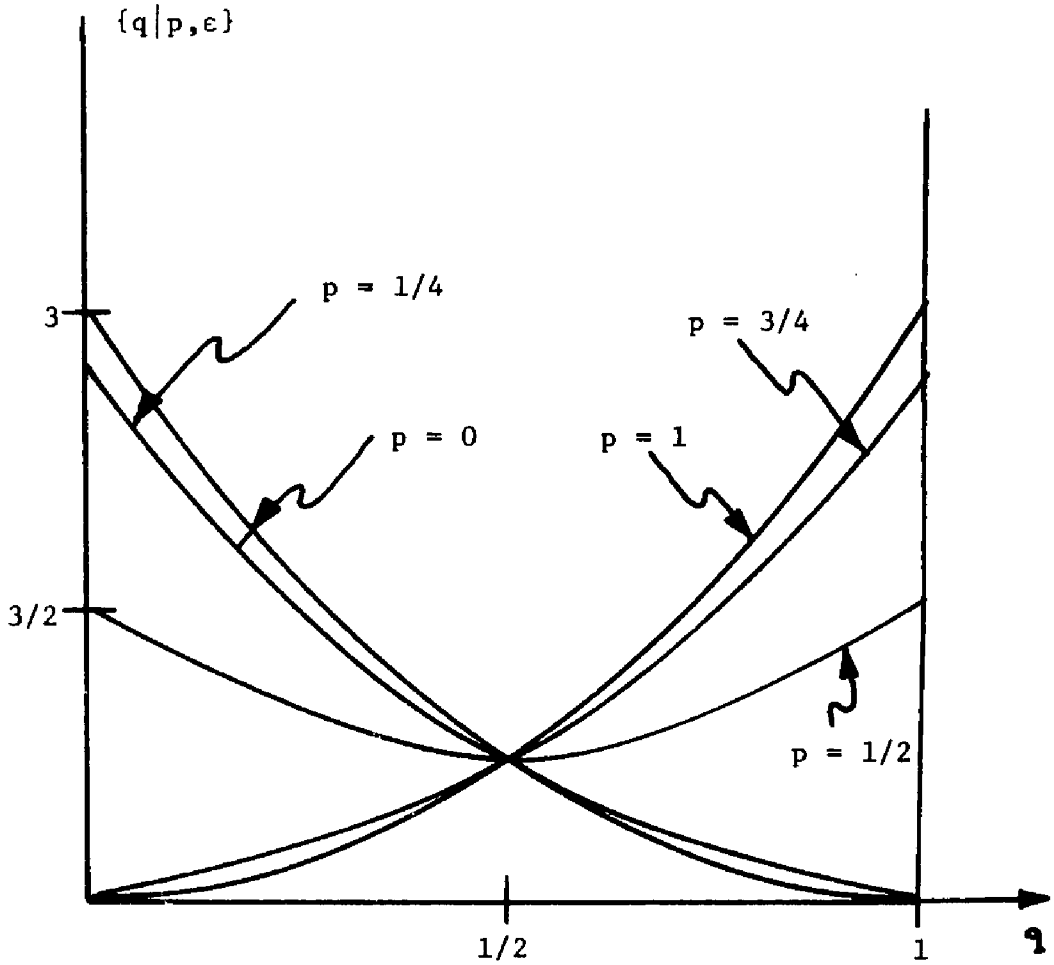


Figure 3.2 Decision Makers Conditional Assessment of Second Expert

Example 3.2 - Thumbtack Flipping, No Cross-calibration

We now approach the same problem as that in Example 3.1 with the variation that the decision maker feels that Expert 2 will probably do better in judging the correct outcome than Expert 1. For example, if the actual outcome is heads, the decision maker feels that Expert 2 is more likely to assign a higher probability to heads than Expert 1. Specifically, suppose (Figure 3.3):

$$\{q|H,p,\varepsilon\} = \begin{cases} 10/4 & p - .1 \leq q < p,\ .1 \leq p \leq .9 \\ 30/4 & p \leq q \leq p + .1,\ .1 \leq p \leq .9 \end{cases} \tag{3.20}$$

and

$$\{q|T,p,\varepsilon\} = \begin{cases} 30/4 & p - .1 \leq q < p,\ .1 \leq p \leq .9 \\ 10/4 & p \leq q \leq p + .1,\ .1 \leq p \leq .9 \end{cases} \tag{3.21}$$

We only analyze the case where $.1 \leq p \leq .9$ since it would introduce unnecessary complication to include the whole domain of p, and will not affect our general conclusions. We can calculate, as in Example 3.1:

$$\{p,q|\varepsilon\} = \begin{cases} 30/8[p^2 + 3(1-p)^2] & p - .1 \leq q < p,\ .1 \leq p \leq .9 \\ 30/8[3p^2 + (1-p)^2] & p \leq q \leq p + .1,\ .1 \leq p \leq .9 \\ 0 & \text{otherwise} \end{cases} \tag{3.22}$$

and from equation 3.11:

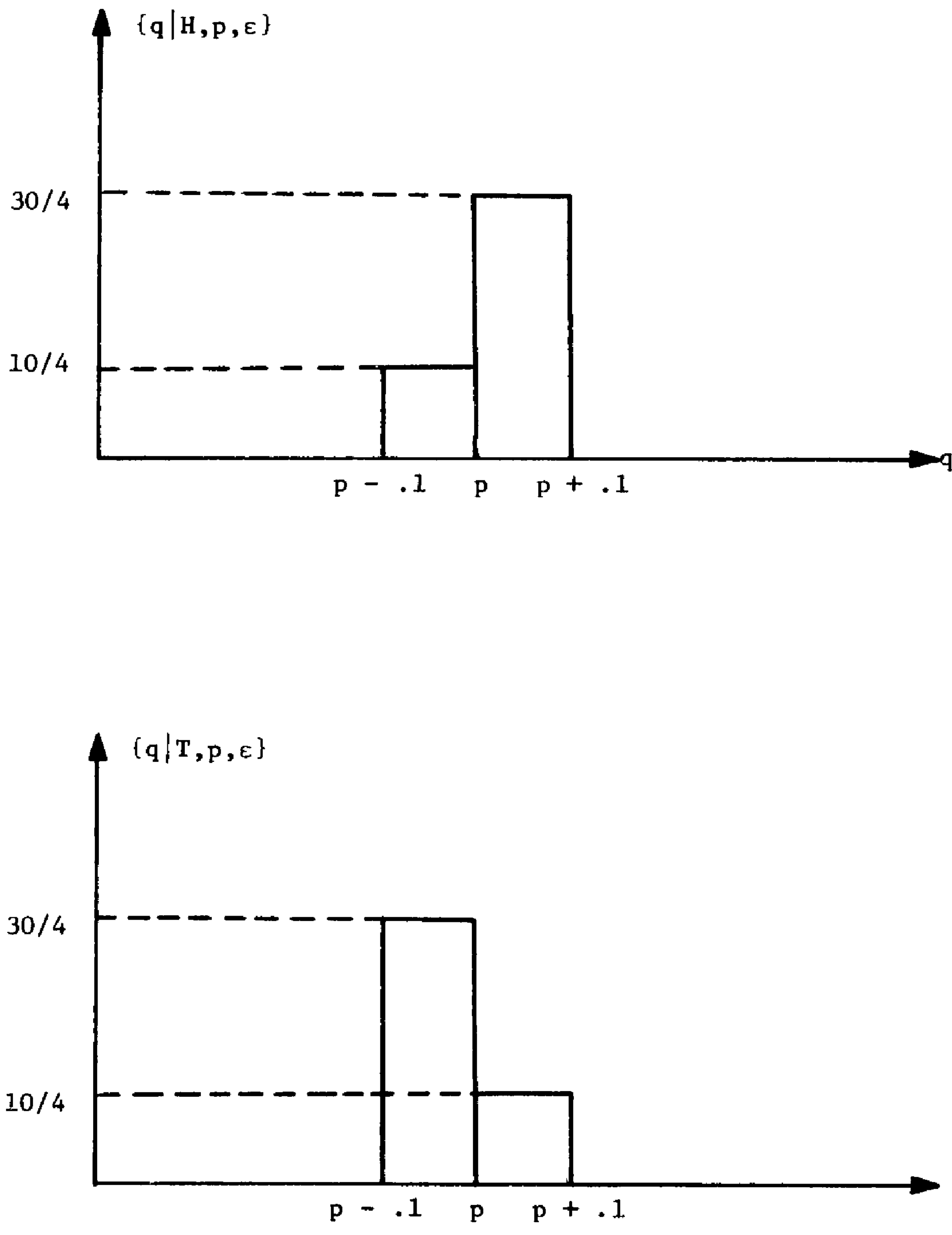


Figure 3.3 Decision Maker's Conditional Distributions on q

$$\{H|p,q,\varepsilon\} = \begin{cases} \dfrac{p^2}{p^2 + 3(1-p)^2} & p - .1 \leq q < p, \ .1 \leq p \leq .9 \\ \\ \dfrac{3p^2}{3p^2 + (1-p)^2} & p \leq q \leq p + .1, \ .1 \leq p \leq .9 \end{cases} \tag{3.23}$$

In Figure 3.4 is plotted the decision maker's posterior, derived from equation 3.23, as a function of both experts' responses. We see that, depending on whether the second expert states the probability of heads as greater or less than the first expert's probability of heads, the correct posterior is determined from the upper or lower line in the diagram. The middle curve determines the correct posterior based only on Expert 1's prior. We examine, for example, the situation where the first and second expert assign .75 and .70 respectively for the probability of heads. In this case, using only Expert 1, the decision maker updates his probability of heads from .5 to .9. However, after hearing Expert 2, the decision maker is forced to lower his assessed probability of heads to .75. This is intuitive since the second expert's response is more likely given tails than heads.

The decision maker is more certain of the outcome in this case with only one expert than with two. In general, if the following conditions obtain:

(a) $p \leq .5$, $q \geq p$ or

(b) $p \geq .5$, $q \leq p$

the decision maker will feel more confident as to one outcome or the other before hearing Expert 2, than after. This does not, however, imply that Expert 2 has a negative value to the decision maker.

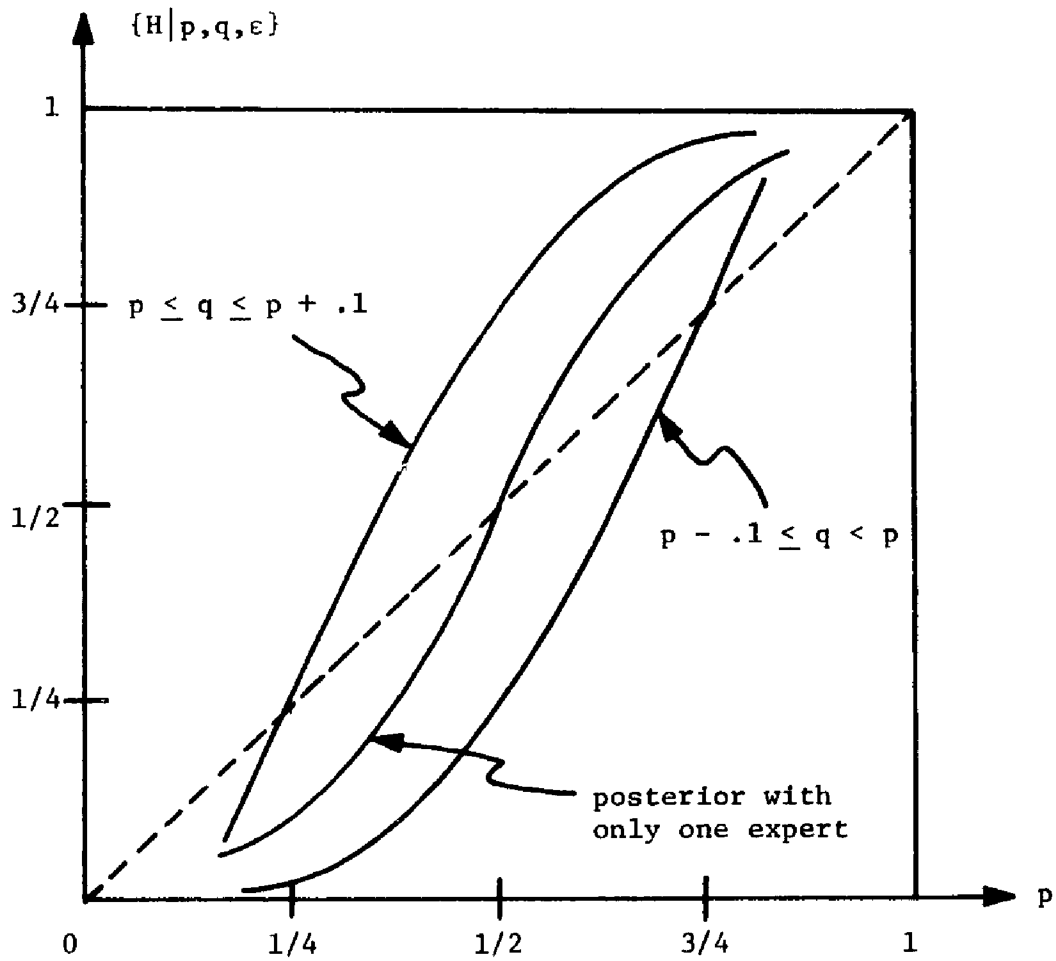


Figure 3.4 Posterior Based on Two Dependent Experts

Example 3.3 - Two Disparate Experts

An interesting case is one in which the decision maker is faced with contradictory opinions from two experts, each of whom he trusts. A recent example of just such a situation arose over debate on the Vietnam War. Two well studied scholars addressed the question of how many deaths would be directly attributable to a Communist purge if the United States immediately withdrew all U.S. troops from Vietnam. The date of actual occurrence of this example was May 1970. Their answers to this question were widely divergent and might be approximated as in Figure 3.5, where the experts are labeled by their political leanings. The left wing expert was "positive" that the true number of purge deaths would be between 1 percent and 2 percent. The right wing expert was just as sure that the true numbers of purge deaths would fall between 9 percent and 11 percent. Furthermore, each was very aware of the other's position and had already built this information into his own assessment.

For sake of exposition, suppose that the decision maker is the President of the United States. Further suppose that upon questioning the President we find that he assesses the experts independently (they are cross-calibrated), and his prior feelings are the same about each expert. His own assessment on purge deaths is uniform from 1 percent to 19 percent (Figure 3.6).

Assume for simplicity, that the experts' responses can be characterized by uniform distributions and that the decision maker assesses:

$$\{w_1|x,c_1,\varepsilon\} = \{w_2|x,c_2,\varepsilon\} = \begin{cases} 1/3 & 0 \leq w_1(w_2) \leq 3 \\ 0 & \text{otherwise} \end{cases} \qquad (3.24)$$

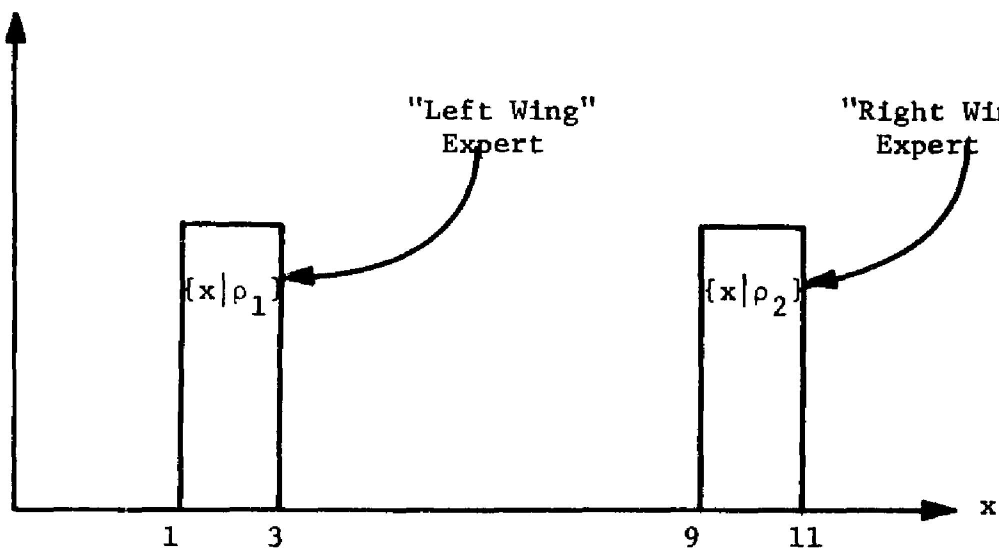


Figure 3.5 Two Disparate Experts

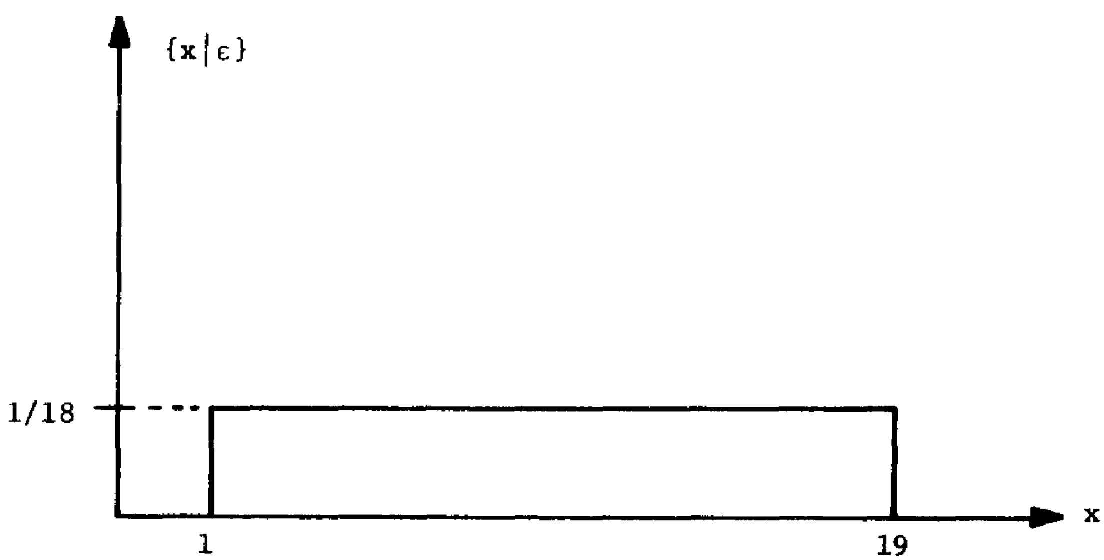


Figure 3.6 Decision Maker's Prior

$$\{c_1|x,\varepsilon\} = \{c_2|x,\varepsilon\} = \begin{cases} 7/16 & x - 1 \leq c_1(c_2) \leq x+1, 1 \leq x \leq 19 \\ 1/256 & c_1(c_2) \leq x-1 \text{ or } c_1(c_2) \leq x+1, 1 \leq x \leq 19 \\ 0 & \text{otherwise} \end{cases} \tag{3.25}$$

where c_1 and w_1 are the center and width of the left wing expert's prior and c_2 and w_2 the center and width of the right wing expert's prior. These assessments are depicted in Figures 3.7 and 3.8.

Based on the revelation of c_1, c_2, w_1 and w_2 the President can update his distribution as follows:

$$\{x|c_1,c_2,w_1,w_2,\varepsilon\} = k\{c_1,c_2,w_1,w_2|x,\varepsilon\}\{x|\varepsilon\} \tag{3.26}$$

By cross-calibration and our assumptions about w_1 and w_2 we can write

$$\{x|c_1,c_2,w_1,w_2,\varepsilon\} = x\{c_1|x,\varepsilon\}\{c_2|x,\varepsilon\}\{x|\varepsilon\} \tag{3.27}$$

Calculation of the constant is straightforward but tedious. We first derive

$$\{c_1,c_2|x\ \varepsilon\} = \begin{cases} \dfrac{49}{(16)^2} & c_1-1\leq x\leq c_1+1,\ c_2-1\leq x\leq c_2+1,\ 1\leq x\leq 19 \\ \dfrac{17}{(16)^3} & c_1-1\leq x\leq c_1+1,\ c_2+1\leq x\leq \text{or } c_2-1\leq x,\ 1\leq x\leq 19 \\ & \text{OR} \\ & c_2-1\leq x\leq c_2+1,\ c_1+1\leq x \text{ or } c_1-1\leq x,\ 1\leq x\leq 19 \\ \dfrac{1}{(16)^4} & x\geq c_1+1 \text{ or } x \leq c_1-1,\ x\geq c_2+1 \text{ or } x\leq c_2-1, 1\leq x\leq 19 \end{cases} \tag{3.28}$$

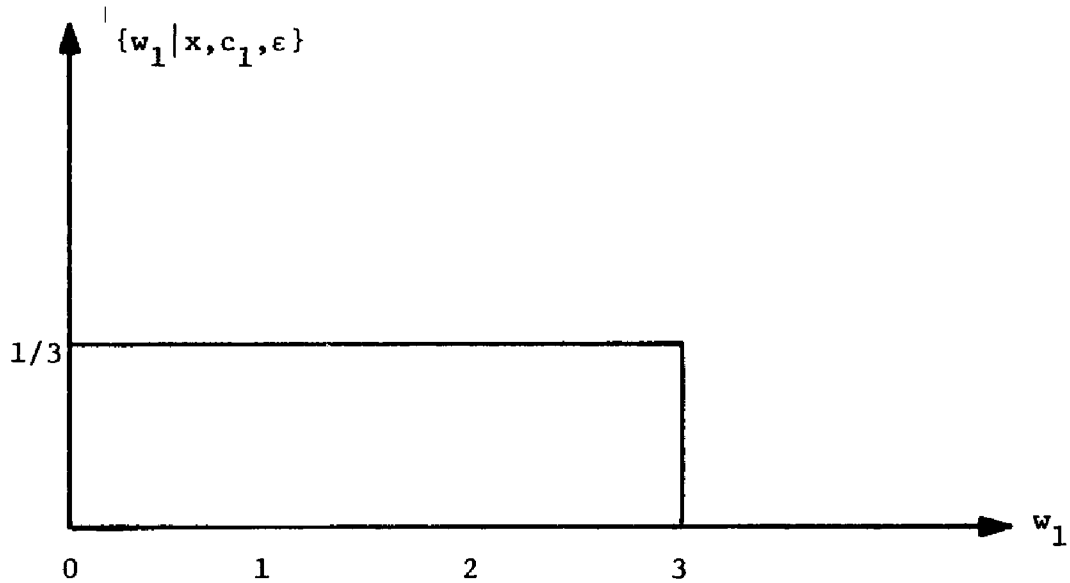


Figure 3.7 Decision Maker's Assessment of the Width of the Experts' Priors

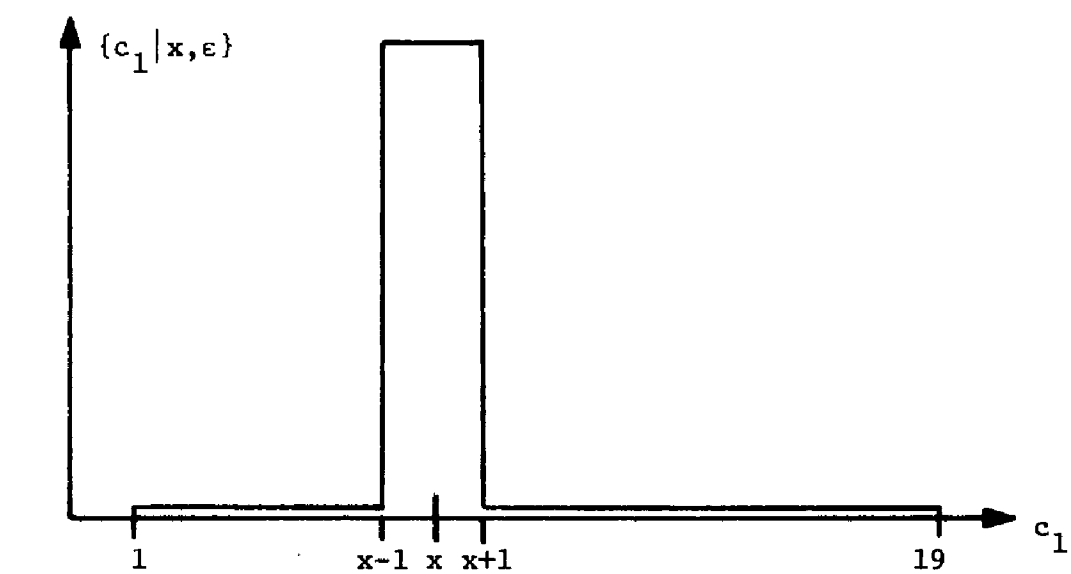


Figure 3.8 Decision Maker's Assessment of the Center of the Experts' Priors

Substituting $c_1 = 2$, $c_2 = 11$, $w_1 = 2$, $w_2 = 2$ corresponding to the given expert priors gives

$$\{c_1,c_2|x,\varepsilon\} = \begin{cases} 7/(16)^3 & 1 \le x \le 3 \text{ or } 10 \le x \le 12 \\ 1/(16)^4 & 3 \le x \le 10 \text{ or } 12 \le x \le 19 \end{cases} \tag{3.29}$$

This is integrated over $\{x|\varepsilon\}$ to provide $\{c_1,c_2|\varepsilon\}$. We are now prepared to calculate (Figure 3.9):

$$\{x|\{x|\rho_1\}, \{x|\rho_2\},\varepsilon\} \approx \begin{cases} .242 & 1 \le x \le 3 \text{ or } 10 \le x \le 12 \\ .002 & 3 \le x \le 10 \text{ or } 12 \le x \le 19 \end{cases} \tag{3.30}$$

We notice immediately that the bimodal posterior could have been obtained approximately by weighting each expert's distribution equally and normalizing. This confirms intuition about the symmetry of the situation. It can easily be shown that if the decision maker had only consulted with one expert his posterior would be very nearly that expert's prior. Thus, in this case, the consulting of two experts rather than one drastically altered the decision maker's posterior.

3.4 Selection of a Panel of Experts

In many decision situations the decision maker is faced with the problem of selecting the set of individuals upon whom he will rely for advice. The previous theory developed provides an immediate conceptual solution to this problem.

We first point out an immediate and interesting consequence of our

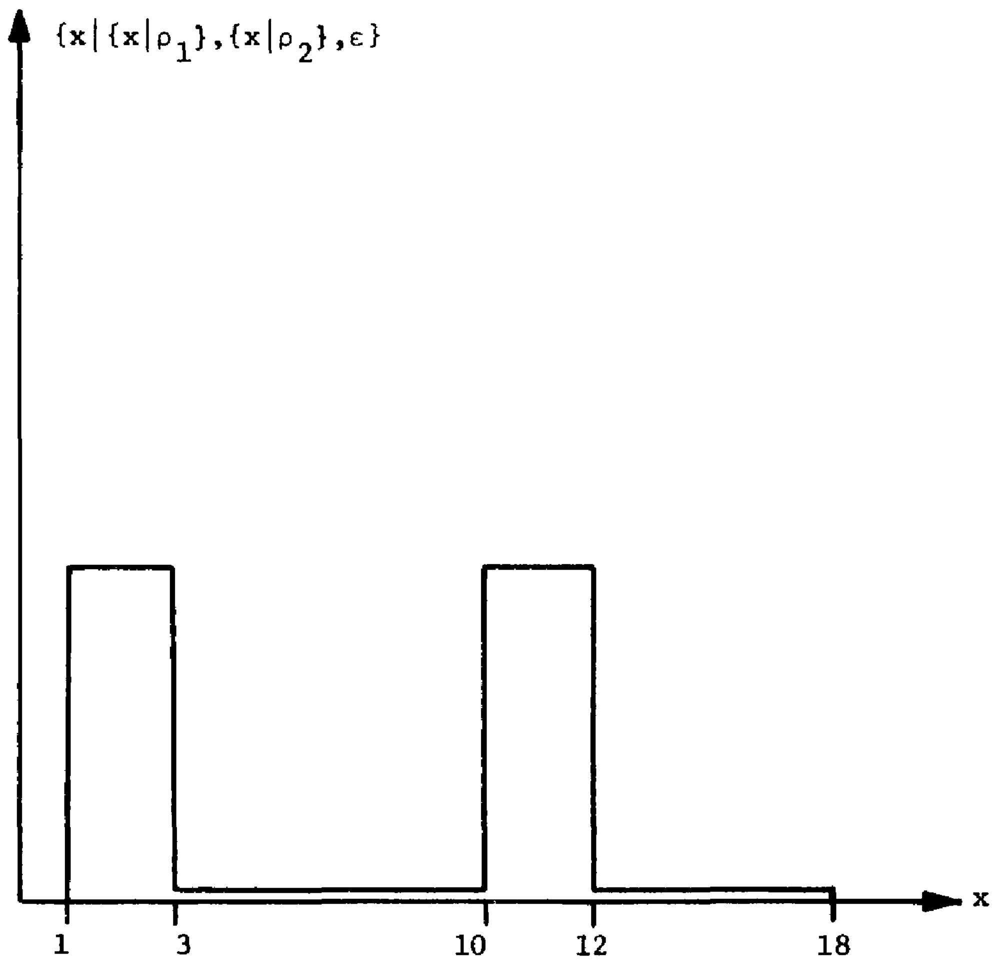


Figure 3.9 President's Posterior Assessment

prior work: there is no such thing as "bad advice." The value of an expert is always greater than or equal to zero. The common misunderstanding probably exists due to the false premise that an expert's advice should be totally accepted or rejected. In fact, we have shown this to be the rare case.

If the experts as scarce resources are free, the decision maker should use them. It is always a mistake to ignore completely the advice of anyone, regardless of his expertise. Thus, in this section we will address ourselves only to the problem of selecting experts whose use involves the expenditure of some of the decision maker's resources.

The value of a panel of experts

The calculation of the value of a specific panel of experts can be done completely analogously to the calculation of the value of a single expert. In the discrete case we determine the expected utility given that each expert gives certain advice and multiply it by the prior probability of obtaining this advice. We then sum over all possible combinations of advice to obtain the expected utility given that the panel will be consulted. When the cost of the panel is adjusted so that the expected utility with the panel is equal to the expected utility without, this cost is the maximum the decision maker should be willing to pay for the panel of experts.

Panel selection

Conceptually all we need do to choose the optimum panel of experts for the decision maker to use is select that panel whose use provides the decision maker with the highest expected utility. Unfortunately, the only general way to perform such a calculation is to try all combinations of experts.

We note, however, that in many situations the decision maker is not constrained to hire all his experts simultaneously. In such situations it is prudent for the decision maker to obtain his experts sequentially. Thus, the use of one expert may preclude the economic viability of the use of another. The following example will illustrate this point.

Example 3.4 The value of two experts; sequential hiring

Let us again consider the use of the two cross-calibrated experts described in Example 3.1 in the decision situation outlined in the example in Section 2.5. We first wish to calculate the value of both experts. Toward this end we define

$\langle v|\varepsilon\rangle \equiv$ the expected value of the decision without consulting either expert

$\langle v|a_1,a_2,\varepsilon\rangle \equiv$ the expected value of the decision given that both experts (1 and 2) will be consulted

$d(p,q) \equiv$ the optimal decision given the prior assessment of heads by both experts

We first notice that

$$d(p,q) = \begin{cases} \text{guess heads} & \{H|p,q,\varepsilon\} \geq \{T|p,q,\varepsilon\} \\ \text{guess tails} & \{H|p,q,\varepsilon\} < \{T|p,q,\varepsilon\} \end{cases} \tag{3.31}$$

which we may calculate from equation 3.16 or note visually from Figure 3.1 as

$$d(p,q) = \begin{cases} \text{guess heads} & p + q \geq 1 \\ \text{guess tails} & p + q < 1 \end{cases} \tag{3.32}$$

We may also calculate the expected value contingent upon both experts' responses as:

$$\langle v|p,q,\varepsilon\rangle = \begin{cases} \dfrac{10p^2q^2 - 10(1-p)^2(1-q)^2}{p^2q^2 + (1-p)^2(1-q)^2} & p + q \geq 1 \\[2ex] \dfrac{10(1-p)^2(1-q)^2 - 10p^2q^2}{p^2q^2 + (1-p)^2(1-q)^2} & p + q \leq 1 \end{cases} \tag{3.33}$$

Next, by integrating over the sample space (shown in Figure 3.10), we determine:

$$\begin{aligned} \langle v|a_1,a_2,\varepsilon\rangle &= \iint_{pq} \langle v|p,q,\varepsilon\rangle\{p,q|\varepsilon\} \\ &= \iint_{p+q\geq 1} 9/2[10p^2q^2 - 10(1-p)^2(1-q)^2]dpdq \\ &\quad + \iint_{p+q<1} 9/2[10(1-p)^2(1-q)^2 - 10p^2q^2]dpdq \\ &= 45\int_0^1 [\frac{q^2}{3} - \frac{(1-q)^2}{3} - \frac{2q^2(1-q)^3}{3} \\ &\quad + \frac{2(1-q)^5}{3} - 2(1-q)^4 + 2(1-q)^3]dq \end{aligned} \tag{3.34}$$

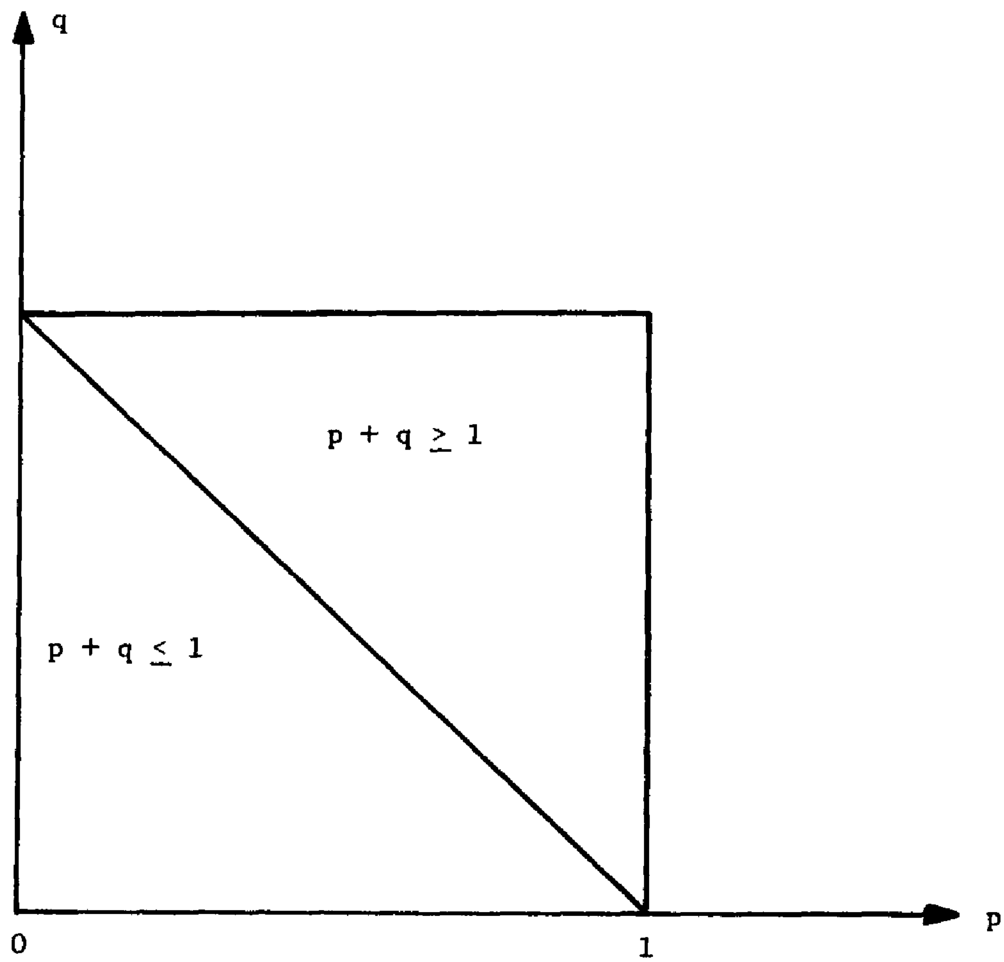


Figure 3.10 Sample Space

The above integral is a summation of Beta functions, the value of which may be found in standard integral tables to give:

$$\langle v|a_1,a_2,\varepsilon\rangle = 9 \tag{3.35}$$

The value of one of these experts was calculated to be 7.5 dollars in the example in Section 2.5. Therefore, we see that the second expert is worth only an additional 1.5 dollars if bought in a package deal with the first expert. We will now investigate the dependence of the value of the second expert on the response of the first in a sequential buying situation.

We wish to calculate the incremental worth of the lottery given that the first expert assesses the probability of heads to be p and the second expert will be consulted. Specifically, what we want to determine is:

$$\langle v|p,a_2,\varepsilon\rangle - \langle v|p,\varepsilon\rangle \tag{3.36}$$

This will give us the additional worth of the second expert. We calculated $\langle v|p,\varepsilon\rangle$ in Chapter II. To calculate the first term we write

$$\begin{aligned}\langle v|p,a_2,\varepsilon\rangle &= \int_q \langle v|p,q,\varepsilon\rangle \{q|p,\varepsilon\} \\ &= \int_{1-p}^{1} \frac{3[10p^2q^2 - 10(1-p)^2(1-q)^2]}{2p^2 - 2p + 1} \\ &\quad + \int_0^{1-p} \frac{3[10(1-p)^2(1-q)^2 - 10p^2q^2]}{2p^2 - 2p + 1}\end{aligned} \tag{3.37}$$

Evaluating the above integrals and collecting terms we obtain the lengthy formula:

$$\langle v|p,a_2,\varepsilon\rangle = \frac{30}{2p^2 - 2p+1}\left[\frac{p^2}{3} - \frac{(1-p)^2}{3} + 2(1-p)^3 - 2(1-p)^4 - \frac{2p^2(1-p)^3}{3} + \frac{2(1-p)^5}{3}\right] \tag{3.38}$$

We first check equation 3.38 by noting for p = 1 or p = 0 that

$$\langle v|p,a_2,\varepsilon\rangle - \langle v|p,\varepsilon\rangle = 0$$

corresponding to the case where the first expert is a clairvoyant rendering the second expert valueless. For p = 1/2 we obtain

$$\langle v|p,a_2,\varepsilon\rangle - \langle v|p,\varepsilon\rangle = 7.5$$

as expected, since the first expert gave the decision maker no new information on the thumbtack. Table 3.1 includes more detailed calculations.

The information from Table 3.1 is displayed graphically in Figure 3.11. We first observe that our intuition in Example 3.2 is confirmed. The second expert becomes more valuable, the less confident Expert 1 is in the outcome. Restating the reasons for this phenomenon we cite the simultaneous occurrence of two effects:

(1) The closer p is to 0 or 1, the higher the decision maker's confidence in the outcome becomes, thereby reducing the value of the second expert.

Table 3.1

	$\langle v\|p,a_2,\varepsilon\rangle$	$\langle v\|p,\varepsilon\rangle$	Incremental Value of Expert 2
p = 1	10	10	0
p = 0	10	10	0
p = 1/2	7.5	0	7.5
p = 1/4	8.875	8	.875
p = 3/4	8.875	8	.875
p = 3/8	7.9	4.7	3.2
p = 5/8	7.9	4.7	3.2

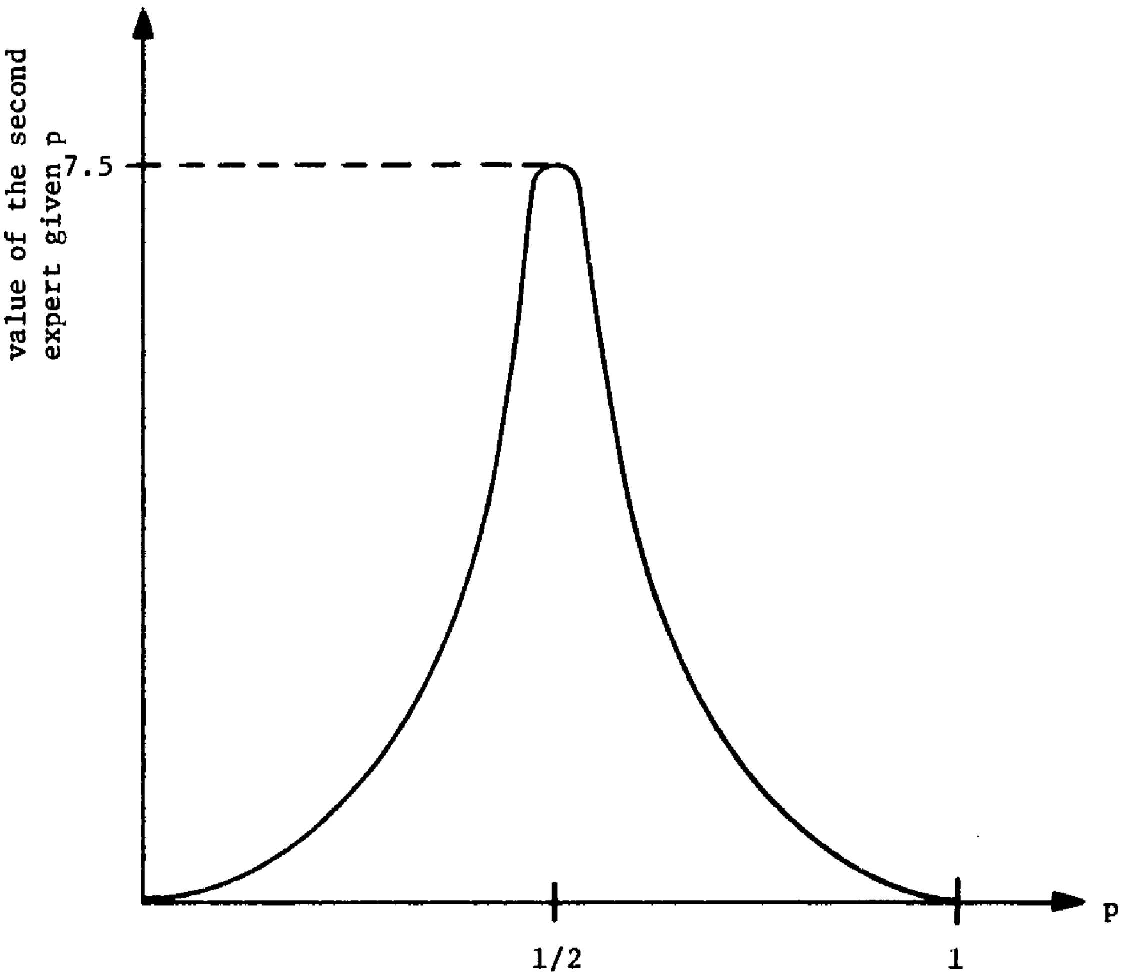


Figure 3.11 Value of the Second Expert as a Function of the Response of the First

(2) The closer p is to 0 or 1, the higher the probability that the second expert will simply justify the optimal decision based on the first expert's advice only (see again Figure 3.2).

We also observe that in the sequential hiring situation the incremental value of the second expert is greater than 1.5 for p roughly between 19/64 and 45/64. Integrating $\{p|\varepsilon\}$ over this interval we obtain approximately 1/3. This means that prior to consulting any experts there is a 2/3 probability that the second expert will be worth less than 1.5, the amount the decision maker would pay for the second expert if he had to make an immediate decision. Posterior to receiving the first expert's comment the second expert is, with prior probability 2/3, worth less than before the first expert's comment.

3.5 Summary

We have presented in this chapter a conceptual solution to the basic problem of how to use the advice of many experts. Mathematically we found the solution to be a straightforward extension of the one expert case. However, the additional implications and practical complexities inherent in using two or more experts are many.

The idea of cross-calibration was introduced--a concept that will prove valuable in the remainder of the dissertation. The implications of the theory for the selection of an optimal panel of experts was then studied. We showed that when a decision maker observes the advice of one expert it changes the value of all his other experts due to two effects. The first effect is that the decision maker's probability distribution on the random variable changes, and the second effect is that the decision maker's assessment of the other experts is altered.

The development thus far is theoretically satisfying; however the results are not yet in a form convenient for practical application. Chapter V will develop an operational methodology for using the concepts here developed.

Chapter IV

RESULTS IN SUBJECTIVE PROBABILITY THEORY

4.0 Introduction

This chapter will put forward new concepts and insights in the general area of subjective probability assessment. The topic of inquiry is naturally related to the study of the use of experts, yet is important for its own sake. Thus the chapter is designed to be read as a separate logical entity although the theory of expert use is used motivationally and philosophically throughout. The direct connection of the concepts presented here to the central topic will be made in Chapter V.

The following section provides a structure and conceptual framework for the methodological results to follow in Sections 4.2 and 4.3. Section 4.4 presents a new technique for probability assessment called "The Method of Equivalent Intervals." This is followed by a simple experiment providing an example of its application.

4.1 Background and Motivation

In general, the revelation of a random variable y provides information which changes our state of knowledge about another random variable x. We can conceptually decompose this information into two types:

(1) Information concerning assessment performance;

(2) Information about the physical environment.

Put in another way, when the value of a random variable is revealed, we receive information about the real world as well as information about ourselves as processors of information about the real world.

That information concerning assessment performance should affect future assessments is eminently reasonable. Experiments done by Edwards [8] show that humans are not consistent information processors. Further experiments by Raiffa and others (such as those quoted in Section 2.5) provide striking data pertaining to subjective probability assessment performance. We make probability assessments on what the real world is perceived to be, not on what it actually is. It can therefore be argued that information about ourselves as perceivers of the world is just as potentially relevant as information about the world itself.

The explicit recognition of assessment performance knowledge as an important ingredient in the formation of probability assessments has fundamental implications. Perhaps the foremost of these implications is the immediate availability of zero cost (other than time spent assessing) information. The sole effect of an information gathering experiment is the altering of prior probability assignments. The information provided by an individual's assessment performance can also drastically affect these assignments. The interesting difference between the two types of information is that one kind is derived solely from observations on the external environment, while the other kind is internally generated and is a measure of the relationship between an individual's conceptualization of the environment and the real thing.

Suppose, for example, we determine from an assessor that one hundred random variables are probabilistically independent (disregarding assessment performance). Now consider the case where the assessor is told that every one of the first ninety-nine revealed values of the random variables falls to the right of the median of each associated distribution. It seems fair to assert that most persons in a similar situation will drastically alter the remaining assessment.

In the above example it is reasonably clear what gross effect the information would have in updating the remaining assessment for most persons. The original distribution would be squeezed and shifted so that most, or all, of it lies to the right of the original median. In general it is not clear how to use such information consistently. It is evident that some sort of transformation must be applied to the original prior but it is not immediate how to make this transformation consistent with prior knowledge about assessment expertise. For instance, suppose that we wish to relate assessment performance on y to our assessment of the variable x. Technically all we must do is use Bayes' Theorem to write

$$\{x|y,\varepsilon\} = k\{y|x,\varepsilon\}\{x|\varepsilon\} \tag{4.1}$$

However, this equation offers no insight relative to the question of how to take assessment performance into account explicitly. It would be fruitful to have a formal means of encoding the dependence of our assessment of a random variable on our assessment performance of another. To illustrate the point further, consider the analogous development in the theory on expert use.

Suppose that a decision maker is concerned with determining the effect an expert's revealed assessment performance on a variable y (given by the information set $[y,\{y|\rho\}]$) should have on the assessment of x based only on the expert's prior on x. Assume that the decision maker knows y and has computed

$$\{x|y,\{x|\rho\},\varepsilon\} \tag{4.2}$$

The information provided by the expert's prior on y then provides his assessment performance, which the decision maker can process and decompose as follows:

$$\{x|y,\{x|\rho\},\{y|\rho\},\varepsilon\} = k\underbrace{\{x|y,\{x|\rho\},\varepsilon\}}_{\substack{\text{independent of}\\ \text{assessment}\\ \text{performance}}}\underbrace{\{\{y|\rho\}|y,x,\{x|\rho\},\varepsilon\}}_{\substack{\text{dependent on}\\ \text{assessment per-}\\ \text{formance}}} \tag{4.3}$$

The term $\{\{y|\rho\}|y,x,\{x|\rho\},\varepsilon\}$ provides a direct means for encoding information about dependencies between assessment performance on x and y. It is important to note that if this term is independent of x and $\{x|\rho\}$, then it has no effect on the final result.

Equation 4.3 provides motivation for trying to achieve the same sort of decomposition relative to individual prior assessment. We should like a general way to decouple operationally the information decoupled conceptually at the beginning of this section. It would also be theoretically and intuitively expedient to provide means whereby an individual could make an assessment which completely encapsulated all his relevant knowledge about his own assessment performance. This would provide a formal mechanism to circumvent the common practice of assessing a prior independent of past performance, and then narrowing or widening it heuristically to appease uneasiness about the empirical results quoted earlier. More precisely, we should like to reshape this unconditional prior in a structured way that is philosophically and logically consistent. We shall find this to be an interesting task, but less straightforward than the ease of the derivation of equation 4.3 indicates.

4.2 Explicit Recognition of Assessment Performance Information

The essential difficulty that prevents the direct extension of the theory on experts to a theory helpful in individual assessment is that an individual cannot directly treat his own assessments as random variables. To the decision maker the expert's prior $\{y|\rho\}$ is an uncertain quantity which, when combined with y, forms an information set providing the expert's assessment performance. As shown in the previous section, decomposition was attained essentially by treating $\{y|\rho\}$ as a random variable. The decision maker's own assessment performance on y is, however, completely provided by the revelation of y alone. That is to say, by definition:

$$\{x|y,\varepsilon\} \quad = \quad \{x|y,\{y|\varepsilon\},\varepsilon\} \tag{4.4}$$

It is instructive to pursue the point further. For example, consider an assessment of a random variable of current interest. Let ε_1 denote our present state of information and ε_2 our state of information five hours from now after intensive thinking about y (and no additional data observation). It is tempting to treat $\{y|\varepsilon_2\}$ as a random variable which, when revealed, will allow us to treat our future self as an expert. For example, we can compute:

$$\{y|\{y|\varepsilon_2\},\varepsilon_1\} = k\{\{y|\varepsilon_2\}|y,\varepsilon_1\}\{y|\varepsilon_1\} \tag{4.5}$$

Unfortunately, although this expression is valid, it is useless for decision purposes because when $\{y|\varepsilon_2\}$ is actually specified we have a new state of information--namely ε_2. Though superficially plausible, there seems to be no direct way to evaluate oneself as an expert.

Dependence on assessment performance only

We shall first restrict the discussion to cases where stochastic dependence between variables is due only to assessment performance relationships. It is also convenient to suppose all variables are continuous. Both restrictions will be later relaxed.

We define the assessment performance of an assessor, relative to a continuous random variable, as the fractile index of the assessor's prior corresponding to the true value of the variable. The fractile index is defined such that the n^{th} fractile has index n. Denoting by p_i, the assessment performance on the variable x_i, we have:

p_i = the index n such that f_n^i equals the revealed value of x_i

where f_n^i is the n^{th} fractile of $\{x_i|\varepsilon\}$. In other words, the revealed value equals $f_{p_i}^i$.

A pair of random variables x_1 and x_2 will be said to be dependent only through assessment performance if

$$\{x_1|x_2,\varepsilon\} = \{x_1|p_2, \varepsilon\}$$

and (4.6)

$$\{x_2|x_1,\varepsilon\} = \{x_2|p_1,\varepsilon\}$$

In this case, the only information about each variable relevant to the assessment of the other is the revealed assessment performance. Our problem is to determine the conditional density function $\{x_1|x_2,\varepsilon\}$ from the assessment $\{p_1|p_2,\varepsilon\}$, which is an explicit appraisal of performance dependencies.

First note that by definition the revealed value of p_1 is just the prior cumulative of x_1 evaluated at the revealed value of x_1. Thus p_1 and x_1 are related functionally through

$$p_1 = \{x_1 \leq x_{1o}|\varepsilon\} \ , \tag{4.7}$$

where x_{1o} is the revealed value of x_1.

Next we perform a change of variables from p_1 to x_1. Denoting the cumulative by the function $g_1(x_1)$ for notational convenience, we may write

$$\{x_1 \leq c|p_2,\varepsilon\} = \int_{-\infty}^{g_1(c)} \{p_1|p_2,\varepsilon\}dp_1 \tag{4.8}$$

since g_1 is monotonically increasing by construction. To calculate the density function of x_1 we differentiate with respect to c:

$$\{x_1 \leftarrow c|p_2,\varepsilon\} = \frac{dg_1(c)}{dc}\{p_1 \leftarrow g_1(c)|p_2,\varepsilon\} \tag{4.9}$$

Notice that the derivative of the prior cumulative is just the prior $\{x_1|\varepsilon\}$ evaluated at the point c. Therefore we may write:

$$\{x_1|p_2,\varepsilon\} = \{x_1|\varepsilon\}\{p_1 \leftarrow g_1(x_1)|p_2,\varepsilon\} \tag{4.10}$$

Since x_2 is given in the expression $\{x_1|x_2,\varepsilon\}$, and $g_2(x_2)$ (the prior cumulative on x_2) is known, we may further obtain

$$\{x_1|x_2,\varepsilon\} = \{x_1|\varepsilon\}\{p_1 \leftarrow g_1(x_1)|p_2 \leftarrow g_2(x_2),\varepsilon\} \tag{4.11}$$

The result is that the prior on x_1 may be updated by multiplying it by the function $\{p_1|p_2,\varepsilon\}$ evaluated at the prior cumulatives of x_1 and x_2. We see that the assessment performance term is exactly analogous to a likelihood function. Note, however, that the variables x_1 and x_2 are on opposite sides of the conditioning bar to the corresponding likelihood function $\{x_2|x_1,\varepsilon\}$.

Remarks on consistency

The three assessments $\{x_1|\varepsilon\}$, $\{x_2|\varepsilon\}$, and $\{p_1|p_2,\varepsilon\}$ have certain relationships that must hold to guarantee prior consistency. By construction the functions $\{p_1|\varepsilon\}$ and $\{p_2|\varepsilon\}$ must be uniform from 0 to 1. Also necessary for mutual consistency is that the expansion rule applies:

$$\int_{p_2} \{p_1|p_2,\varepsilon\}\{p_2|\varepsilon\} = \{p_1|\varepsilon\} \tag{4.12}$$

Given that $\{p_1|\varepsilon\}$ and $\{p_2|\varepsilon\}$ are uniform, the possible consistent assessments of $\{p_1|p_2,\varepsilon\}$ fall in a very restrictive class. Conversely, an explicit assessment of performance dependencies will generally imply an inconsistency in one or both prior assignments.

This type of inconsistency is not surprising, however, as priors are usually assessed without overt recognition of assessment performance. A further aim of this chapter is to develop a direct means whereby an assessor can guarantee consistency between his prior assignments and knowledge about his own assessment ability.

In any event, selection of the set of things over which to require consistency is ultimately a matter of choice. We have displayed a means for directly appraising assessment performance and shown how to infer a posterior consistent with the prior assessment of the variable of interest. An example is outlined in Figure 4.1 relating the stepwise determination of the posterior from a given set of assessments.

Joint dependence

The problem posed by the situation where x_1 is dependent on x_2 for reasons additional to revealed assessment performance is easily solved by direct extension. We assume that the assessor is able to assess the dependence of x_1 on x_2 independently of assessment performance considerations. Labeling this assessment by $\{x_1|x_2,\varepsilon\}$, we denote the relevant performance indicator by $p_1(x_2)$, where the revealed fractile index p_1 is now a function of x_2. By change of variables it is possible to compute the desired quantity $\{x_1|x_2,\varepsilon\}$ from the assessment $\{p_1(x_2)|p_2,x_2,\varepsilon\}$ exactly as in the simpler case.

Figure 4.1 -- An Example Calculation

Step 1: Assess priors

$$\{x_1|\varepsilon\} = \begin{cases} 3x_1^2 & 0 \le x_1 \le 1 \\ 0 & \text{otherwise} \end{cases} \qquad \{x_2|\varepsilon\} = \begin{cases} 4x_2^3 & 0 \le x_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Step 2: Assess performance dependencies

$$\{p_1|p_2\varepsilon\} = \begin{cases} \dfrac{p_1}{p_2} & 0 \le p_1 \le \sqrt{2p_2} \quad , \quad 0 \le p_2 \le 1/2 \\ 1 & 0 \le p_1 \le 1 \quad , \quad 1/2 \le p_2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Step 3: Evaluate at cumulatives

$$\{p_1 \leftarrow g_1(x_1)|p_2 \leftarrow g_2(x_2),\varepsilon\} = \{p_1 \leftarrow x_1^3|p_2 \leftarrow x_2^4,\varepsilon\}$$

where $g_1(x_1) = x_1^3$

$g_2(x_2) = x_2^4$

$$= \begin{cases} \dfrac{x_1^3}{x_2^4} & 0 \le x_1^3 \le \sqrt{2}\, x_2^2 , 0 \le x_2^4 \le 1/2 \\ 1 & 0 \le x_1^3 \le 1 \quad , \quad 1/2 \le x_2^4 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Step 4: Calculate posterior

$$\{x_1|x_2,\varepsilon\} = \begin{cases} \dfrac{3x_1^5}{x_2^4} & 0 \le x_1^3 \le \sqrt{2}\; x_2^2 \le 1 \\ 3x_1^2 & 0 \le x_1 \le 1 , 1/2 \le x_2^4 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

check: posterior is a probability density function.

4.3 Self Calibration

It may be argued that every perception we make reveals the value of a random variable, so that we receive constant information concerning assessment performance. In general it takes a large number of assessment performance observations to provide much relevant information. However, the formal inclusion of all assessment performance knowledge into each probability assignment is a formidable task. It is therefore of extreme practical importance to have a straightforward means whereby an individual can learn enough about himself as an assessor so that additional observations have negligible effect on his probability assessments. This would also eliminate the need for worry about consistency with all past performance data. The results below provide such a means as well as a conceptually satisfying parallel to the theory on expert use.

Continuous random variables

The treatment of the continuous variable case is, surprisingly, more direct than the analogous discrete variable case. We shall consider the variable x_o to be the primary variable of interest and suppose that an assessor has performed a large number N of prior assessments in addition to the one on x_o. Thus he has specified N + 1 priors

$$\{x_i|\varepsilon\} \qquad\qquad i = 0,\ldots,N$$

on the continuous variables x_o through x_N.

Our object is to apply the concept of exchangeability in a meaningful way. Towards this end consider the set of random variables

$(p_o,\ldots,p_N)$ where each p_i is, as in the previous section, the fractile index of $\{x_i|\varepsilon\}$ corresponding to the revealed value of x_i. The key idea is to consider this set of performance indicators as candidates for a subjective appraisal of exchangeability. This is to be sharply contrasted with the possibility of considering $(x_o,\ldots,x_N)$ itself as an exchangeable set. Given an arbitrary x_o it would be extremely difficult to find, in general, a collection of N other exchangeable random variables. Furthermore, such a set would be of no a priori use in learning about assessment performance. However, given a random variable x_o, the specification of a set of random variables which has the property that their performance indicators are exchangeable is, in many cases, a reasonable task.

Recall that exchangeability assessed for a set of random variables means that each variable "plays a symmetrical role in relation to any probability problem" [7]. The assumption that the variables in the set $(p_o,\ldots,p_N)$ are exchangeable is intuitively that the assessor cannot distinguish between his assessment ability on the x_i's.

For example, suppose we define a surprise relative to the variable x_i as the event that the true value falls outside the fractile interval $[f^i_{.01},f^i_{.99}]$, where f^i_j is the j^{th} fractile of the i^{th} prior. Next suppose that we observe the revealed values of x_i and tell the assessor only that exactly one-third of the revealed values will be surprises. To be consistent with exchangeability he must consider any possible set of N/3 variables as equally likely candidates.

The implications of the set of p_i's being exchangeable are simple yet powerful. Consider the frequency histogram P_m made up of revealed values of an arbitrary set of m of the p_i's. Recalling the discussion in Section 2.7, we are entitled to speak of the frequency distribution

$$P = \lim_{m\to\infty} P_m \tag{4.13}$$

as a measurable quantity and assess it as such. We write $\{P|\varepsilon\}$ to denote a prior distribution on possible frequency distributions.

Expanding the assessment of p_o over all possible P's we obtain

$$\{p_o|\varepsilon\} = \int_P \{p_o|P,\varepsilon\}\{P|\varepsilon\} \tag{4.14}$$

Assuming that the p_i's form an exchangeable set we note that the density function on p_o, given the long run histogram P, is just P itself. Therefore writing P as a function, we have

$$\{p_o|\varepsilon\} = \int_P P(p_o)\{P|\varepsilon\} \tag{4.15}$$

We shall use the notation $\langle F|t,\varepsilon\rangle$ to denote the expectation of the function F evaluated at the point t. Thus

$$\langle F|t,\varepsilon\rangle = \int_F F(t)\{F|\varepsilon\} \tag{4.16}$$

where the integral is over the possible functions F. Thus a functional expectation is a pointwise weighting of a set of functions which is itself a function. We may now write

$$\{p_o|\varepsilon\} = \langle P|p_o,\varepsilon\rangle \tag{4.17}$$

Note that the above equation places a consistency requirement on the prior assessment of P. This situation has no parallel in the standard coin flipping examples of exchangeability, since in those examples the analogous probability of heads or tails is derived from the expected frequency. In this case the probability had to be assessed prior to the assessment of exchangeability. Section 4.4 will pursue this point to suggest a new alternative in prior assessment.

Of more immediate interest is an examination of the effect of an additional performance observation p_1. By repeated expansion we obtain:

$$\begin{aligned} \{p_o|p_1,\varepsilon\} &= \int_P \{p_o|p_1,P,\varepsilon\}\{P|p_1,\varepsilon\} \\ &= k\int_P P(p_o)\{p_1|P,\varepsilon\}\{P|\varepsilon\} \qquad (4.18) \\ &= k\int_P P(p_o)P(p_1)\{P|\varepsilon\} \end{aligned}$$

where $\{p_1|P,\varepsilon\}$ equals the distribution P evaluated at the point p_1 by previous logic.

It is clear by induction that the posterior based on N observations of assessment performance is

$$\{p_o|\underline{p},\varepsilon\} = K\int_P P(p_o)P(p_1)\ldots P(p_N)\{P|\varepsilon\} \qquad (4.19)$$

where $\underline{p} = (p_1,\ldots,p_N)$.

Analogous to the previous section we can perform a change of variables to compute the posterior on x_o:

$$\{x_o \leq c|\underline{p},\varepsilon\} = k\int_{-\infty}^{g(c)} dp_o \int_P dP\ P(p_o)P(p_1)\ldots P(p_N)\{P|\varepsilon\} \quad (4.20)$$

where again g is the cumulative of $\{x_o|\varepsilon\}$. By reversing the order of integration we may write:

$$\{x_o \leq c|\underline{p},\varepsilon\} = k\int_P dP\ P(p_1)\ldots P(p_N)\{P|\varepsilon\}\int_{-\infty}^{g(c)} dp_o\ P(p_o) \quad (4.21)$$

Taking the derivative with respect to c we have the final result:

$$\{x_o|\underline{p},\varepsilon\} = k\{x_o|\varepsilon\}\int_P P(g(x_o))P(p_1)P(p_2)\ldots P(p_N)\{P|\varepsilon\} \quad (4.22)$$

This expression explicitly states how to update the prior on x_o in light of the assessment performance $\underline{p}$. If N is large (corresponding to many observations) the histogram P is almost exactly known and the posterior may be written

$$\{x_o|\underline{p},\varepsilon\} = \{x_o|\varepsilon\}P(g(x_o)) \quad (4.23)$$

Important to note is that the posterior becomes insensitive to new performance information on random variables that the assessor considers equivalent in terms of his assessment performance capabilities; it is therefore hard to imagine cases where the observation of a non-exchangeable random variable would further affect the updated distribution.

To summarize we have demonstrated an easy way to incorporate assessment performance knowledge into probability assessments. The essential concept was that of considering exchangeability of the performance indexes of prior distributions rather than exchangeability of the assessed random variables themselves. Additionally we discovered a means for self-calibrating an assessor by the following recipe:

(1) Assess marginal distributions on a set of random variables, including the variable of interest, whose performance indices are exchangeable.

(2) Observe assessment performance on the additional random variables.

(3) Update the prior on the variable of interest using the long run distribution measured in step 2.

The extension to the case where a distribution is specified only by a finite number of fractiles is immediate.

It should be emphasized that locating a set of random variables having the required exchangeability property need not be a difficult job. One advantage is that the necessary condition of equal prior probabilities of each exchangeable event is satisfied by definition. Armed with an almanac, it might in many cases be harder to find a variable without the desired quality than to find one with the desired property.

A final remark is that the assessment of exchangeability in no way rests on the assumption that the assessor is equivalent to some sort of time invariant assessment process like the wheel of fortune process introduced in Chapter II. Exchangeability simply means that the assessor's state of information is such that he can discern no difference between his assessment ability on a set of random variables. That the mathematics of both ways of modeling an assessor are equivalent is a reflection of the power of the concept of exchangeability--it allows a much richer view of the world without engendering additional mechanical complexity.

Discrete random variables

The consideration of the discrete case is not an obvious analog to the continuous case since the measure of performance is not immediate. We will treat the Bernoulli situation for simplicity where the assessor is considering an event E_o to which he attaches prior probability p_o of occurring.

Assume that in order to gauge his assessment accuracy the assessor has assessed the probabilities $(p_1,\ldots,p_N)$ of occurrence of N other possible events $(E_1,\ldots,E_N)$ which he considers independent other than through assessment performance. The occurrence of an event is labeled a success.

Next we construct a collection of sets S(.), where S(p) is the set of all events whose prior probability of occurrence is p. The fundamental condition to be assumed is that each set S(p) is composed of exchangeable events. This allows us to meaningfully define the function F(p) as the long run frequency of successes in the set S(p).

First we expand, for each i, over the possible frequencies $F(p_i)$:

$$\{E_i|p_i,\varepsilon\} = \int_{F(p_i)} \{E_i|p_i,F(p_i),\varepsilon\}\{F(p_i)|p_i,\varepsilon\}$$

$$= \int_{F(p_i)} F(p_i)\{F(p_i)|\varepsilon\} \tag{4.24}$$

$$= \langle F(p_i)|\varepsilon\rangle$$

where it is assumed that the assessment of $F(p_i)$ is independent of the particular assessment p_i. Equation 4.24 gives an immediate consistency condition on the prior assessment of each $F(p_i)$: The expected frequency of successes in $S(p_i)$ must equal p_i.

Suppose that we wish to reassess the probability of E_o after having observed that the event E_1 occurred. We expand over all functions F to obtain:

$$\{E_o|p_o,E_1,p_1,\varepsilon\} = k\int_F \{E_o|F,p_o,E_1,p_1,\varepsilon\}\{E_1|F,p_1,p_o,\varepsilon\}\{F|\varepsilon\}$$

$$= k \int_F F(p_o)F(p_1)\{F|\varepsilon\} \tag{4.25}$$

The expansion above actually need only be done over $F(p_o)$ and $F(p_1)$. Viewed in this way, the assessment $\{F|\varepsilon\}$ is a joint assessment of two frequencies. To generalize, suppose that we observe R successes in N events. By relabeling so that successes are listed first we may write:

$$\{E_o|\underline{E},\underline{p},R,\varepsilon\} = k\int_F \left(\prod_{i=0}^{R} F(p_i) \prod_{i=R+1}^{N} [1-F(p_i)]\right)\{F|\varepsilon\} \tag{4.26}$$

where $\underline{E} = (E_1,\ldots,E_N)$ and $\underline{p} = (p_1,\ldots,p_N)$. When enough events have been observed so that F is practically known the above result may be simplified:

$$\{E_o|\underline{E},\underline{p},R,\varepsilon\} = k\left(\prod_{i=0}^{R} F(p_i) \prod_{i=R+1}^{N} [1 - F(p_i)]\right) \tag{4.27}$$

Since the event E_o must either occur or not we have

$$k = \left(\prod_{i=1}^{R} F(p_i) \prod_{i=R+1}^{N} [1 - F(p_i)]\right)^{-1} \tag{4.28}$$

so that

$$\{E_o|\underline{E},\underline{p},R,\varepsilon\} = F(p_o) \tag{4.29}$$

This is the intuitive result (commonly assumed in the literature [15,17]) that the probability of occurrence of an event E_o is the long run frequency of successes observed on a set of events to which the assessor assigned the same prior probability. When this frequency is not known, however, the above development suggests that information relative to other sets is important. For example, the long run frequency of successes in sets of events with prior probability .7 will in general affect our state of knowledge about the long run frequency of successes in sets of events with prior probability .8.

Relation to expert use

The results for both the continuous and discrete case form an interesting parallel to the theory of the use of experts. In both cases the fundamental quantity to the inference operation was a long run performance measure. In a rough sense, assessing this histogram or frequency is analogous to assessing oneself as an expert. In the next chapter the relationship will become more obvious.

4.4 The Method of Equivalent Intervals: A New Probability Assessment Technique

The theory presented in the preceding section provides an additional practical bonus: an alternate means for probability assessment. The method presented below, which for descriptive purposes is named "the Method of Equivalent Intervals," has a basic advantage over existing techniques: the resulting probability distribution is automatically calibrated relative to new assessment performance information. An additional feature (not unique to the method) is that no quantitative probability numbers need be assigned.

Let x_o be the variable of interest. Suppose that intervals are marked off on the x_o axis such that specification of the probability that x_o will fall in each interval determines to the assessor's satisfaction $\{x_o|\varepsilon\}$. No knowledge, other than a rough specification of the practical range of x_o, need be invoked tc select these intervals; they may be as regular or irregular and as numerous as desired. Assume that there are N marks labeled x_{o1} through x_{oN} as in Figure 4.2. Next suppose that a set of n random variables has been selected to form the set $(x_o,\ldots,x_n)$, which has the exchangeability property we shall demand below. The assessment process consists of choosing N marks $(x_{ij}, j=1,\ldots,N)$ for each x_i other than x_o such that:

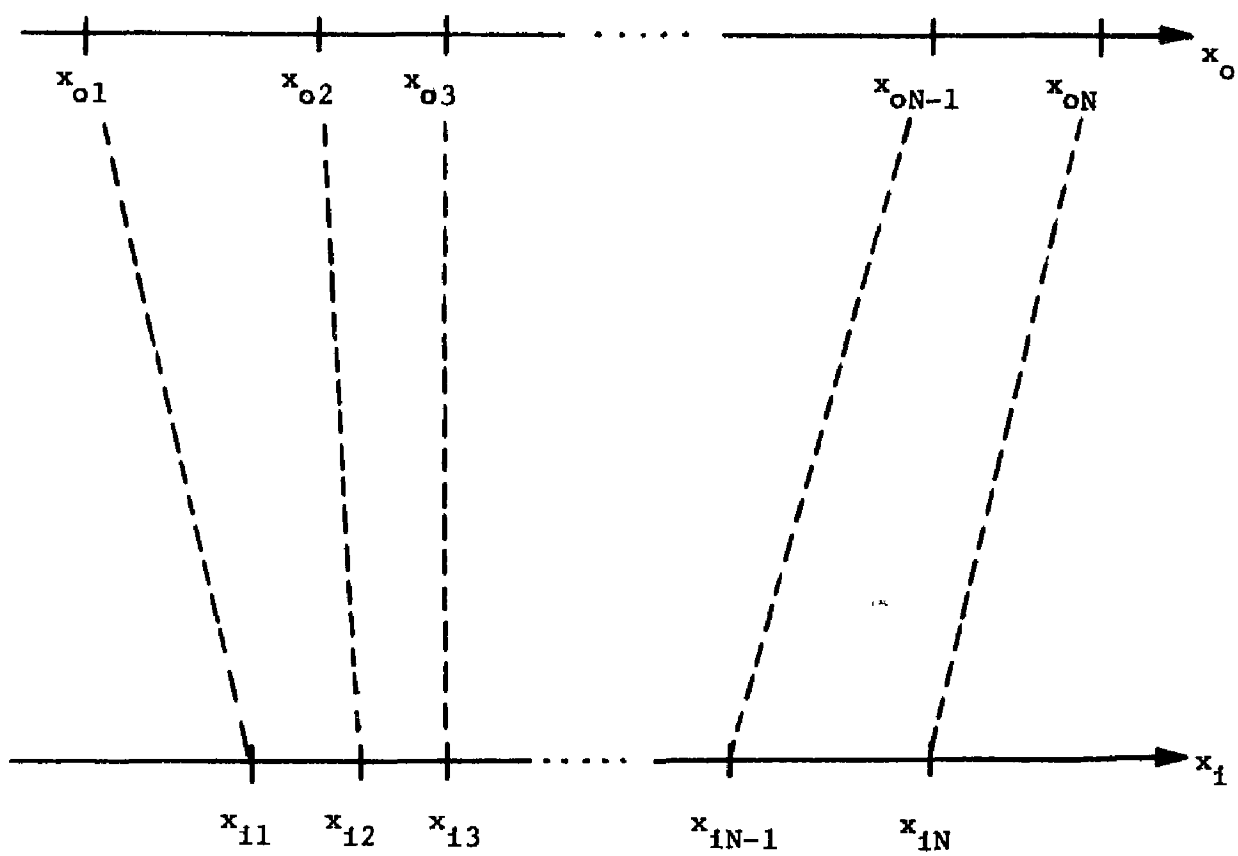


Figure 4.2 Assessment of x_i

$$\{x_i \le x_{ij} | \varepsilon\} = \{x_o \le x_{oj} | \varepsilon\} \qquad \begin{array}{l} i = 1,\ldots,n \\ j = 1,\ldots,N \end{array} \tag{4.30}$$

Define a set of performance indicator variables as:

$$e_{ij} = \begin{cases} 1 & x_i \le x_{ij} \\ 0 & \text{otherwise} \end{cases} \tag{4.31}$$

With i fixed, each variable e_{ij} is a random variable, denoting assessment performance on the variable x_i. The above procedure assures that the performance indices satisfy the necessary condition for exchangeability of having the same prior probability distribution. For example, the probability that each random variable x_i falls to the left of x_{i5} must be equal.

There are many possible ways in which the x_{ij}'s may be specified operationally to satisfy equation 4.30. Perhaps the easiest, and most straightforward of these is the one directly inferred from the formula. Thus, for each variable x_i the numbers x_{ij} are specified by sequentially incrementing the j index. For example, consider the variable x_1. We first determine the point x_{11} such that the assessor would be indifferent between betting that x_1 falls to the right (or left) of x_{11} or that x_o falls to the right (or left) of x_{o1}. Similarly, each successive mark can be established. Figure 4.2 depicts the result of such a process. The probability that each random variable falls to the left of any of the dotted lines is identical.

Of course there are many consistency checks. In particular, the probability that an arbitrary set of x_{ij}'s satisfy a given condition must equal the probability that the corresponding set of x_{oj}'s satisfy the condition. Thus an easy consistency requirement to check is that each interval on the x_i axis must be equally likely to contain a success as the matching interval of the x_o axis.

After the assessment procedure has been followed, if the performance indices are exchangeable by subjective assessment, the material in Section 4.3 may be implemented to provide a probability assignment for the variable of interest. We define the variable

$$F_k = \frac{1}{n} \sum_{i=1}^{N} \sum_{j=1}^{k} e_{ij} \qquad k = 1,\ldots,N \tag{4.32}$$

which is the fraction of times the revealed values of the set $(x_1,\ldots,x_n)$ fall below the k^{th} mark. The distribution on x_o may be inferred by observation of the fractions F_k, for all k, as

$$\{x_o \leq x_{ok} | F_k,\varepsilon\} = F_k \qquad k = 1,\ldots,N \tag{4.33}$$

If the fractions F_k are not observeable the procedure still provides a fundamental link through:

$$\{x_o \leq x_{ok} | \varepsilon\} = \int_{F_k} F_k\{F_k | \varepsilon\} = \langle F_k | \varepsilon \rangle \tag{4.34}$$

which allows an assessor to directly appraise his own assessment performance. When this is an important factor it becomes preferable to expose it explicitly.

The recipe

The entire procedure is summarized below:

(1) Mark off regions for the random variable of interest such that specification of the probability of each region determines satisfactorily the probability distribution for that variable.

(2) Select a set of random variables as candidates for an exchangeable class.

(3) Mark off equiprobable regions on each of these variables corresponding to the initial regions of the variable of interest.

(4) Assess whether the performance indicators as defined by (3) are indeed exchangeable.

(5) Observe the fractions F_k and deduce the probability distribution.

Notice that our earlier claim is justified. Nowhere is a numerical assessment necessary. The only subjective inputs are the assessment of exchangeability and the specification of equivalent (for betting purposes) regions.

In a fundamental sense the theory presented here is to general probability assessment what equally likely events are to simple assessments. In the same way, the general problem is reduced to more accessible, direct components when certain conditions hold. As de Finetti says, "When one accepts the subjectivist point of view, such ought to be the effective meaning and the value of any criterion at all" [7]. An additional feature of the theory just presented is that the assessor has control over the necessary conditions by selection of the proper set.

The next section provides an illustrative example of an actual assessment performed by the above methodology.

4.5 An Experiment in the Use of the Method of Equivalent Intervals

As an experiment in the foregoing method of probability assessment, a subject was asked to perform a series of appraisals. The object of this study was to obtain a probability distribution on the following random variable (hereafter designed by x):

> The daily sale of beer (in pints) of the world's largest beer selling establishment as stated in the 1968 "Guinness Book of World Records" [15].

Toward the desired end the subject was asked to specify two numbers x_ℓ and x_r such that the probability that x falls between them was qualitatively high. The results of this rough assessment are shown in Figure 4.3 for easy reference.

Next the subject was given a collection of twenty-five random variables and asked to specify two points for each ($x_{i\ell}$ and x_{ir}), such that he would be indifferent between owning lottery A or lottery B and indifferent between owning lottery C or lottery D in Figure 4.3. After specification of the appropriate number pair for each random variable, the subject was asked to select from the given collection an exchangeable set of twenty variables. To verify exchangeability operationally, a number of specific questions were presented such as:

> If all you knew was that eight revealed values fall above the rightmost mark, two fall below the leftmost mark, and ten fall in the middle, are all sequences consistent with this information equally likely?

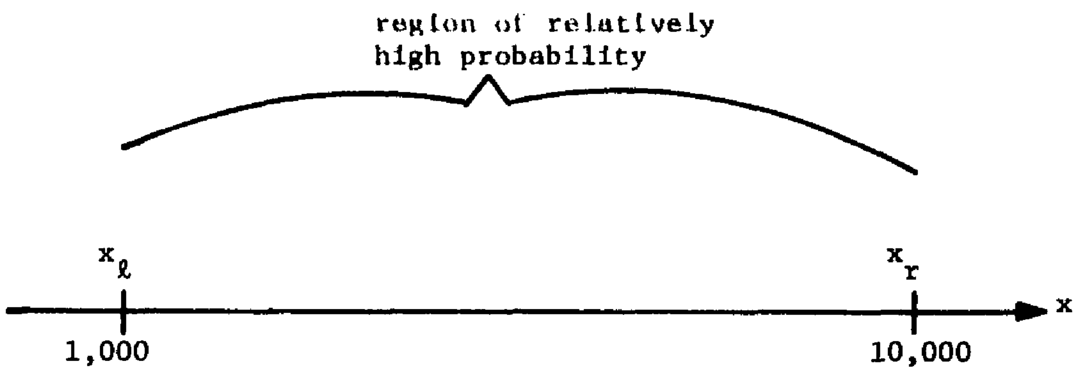


Figure 4.3 Specification of x_ℓ and x_r

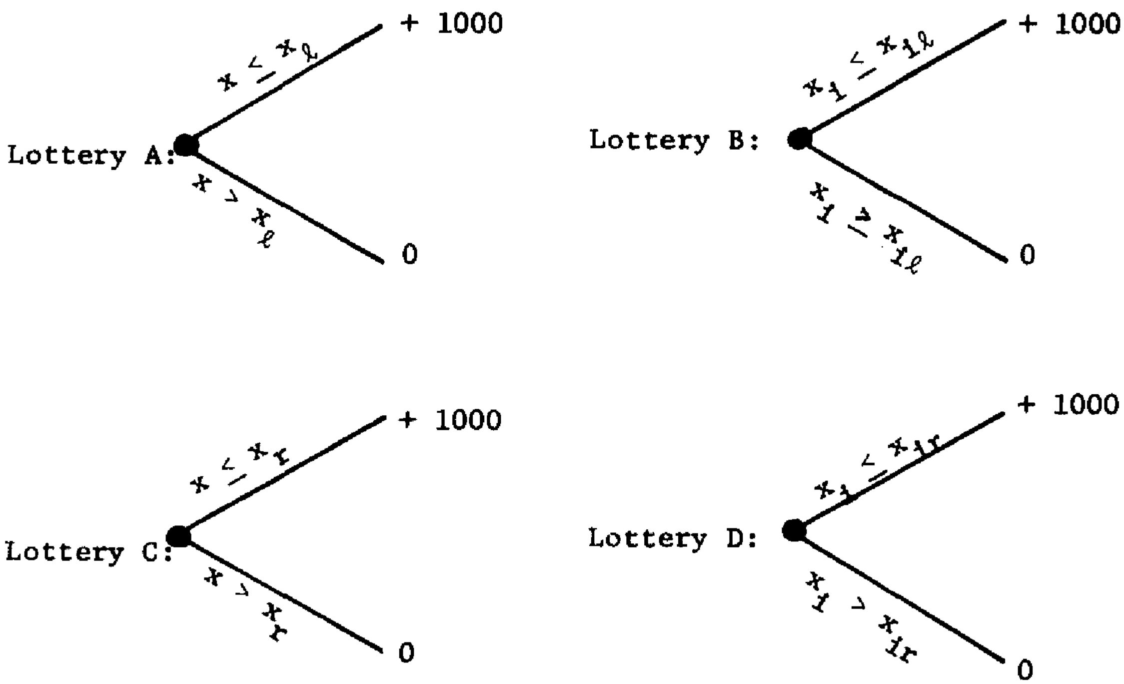


Figure 4.4 Lotteries

As a result, the set of variables and numbers depicted in Table 4.1 was obtained. All variables were defined as stated in reference [15]. For comparative purposes the subject was asked, prior to learning the results, to assess his subjective distribution on x. His assignments were:

$$\begin{aligned} \{x \le x_\ell | \varepsilon\} &= .01 \\ \{x \le x_r | \varepsilon\} &= .9 \end{aligned} \tag{4.35}$$

From the results (denoted R) in Table 4.1 the derived probability assignments are:

$$\begin{aligned} \{x \le x_\ell | R,\varepsilon\} &= .2 \\ \{x \le x_r | R,\varepsilon\} &= .45 \end{aligned} \tag{4.36}$$

The assignments made using the method of equivalent intervals are clearly drastically different than the assignments made using the conventional direct technique. For instance, the probability that x falls in the upper interval changes from .1 to .55. The results are consistent with those discussed in Chapter II in that the number of surprises was far greater than the expected value.

As a final exercise the subject was asked, prior to learning the results, to assess a prior on U, the number of times the true value would fall in the upper region. Figure 4.5 represents a continuous approximation of the subject's response. The interesting thing to note is that the mean of this distribution (represented by the shaded area on the graph) is approximately six. This is, of course, inconsistent with the earlier assessments displayed by equations 4.35. Such an inconsistency

Table 4.1

Symbol		x_ℓ	x_r	Revealed value
x	Daily sale of world's largest beer selling establishment (pints)	1,000	20,000	67,250
x_1	World's tallest totem pole (ft.)	35	100	160
x_2	Earliest recorded successful appendix operation (year)	1,700	1,850	1,736
x_3	Length of world's longest river (mi)	1,000	2,100	4,145
x_4	Height of shortest heavyweight champion (in.)	66	69	67
x_5	Highest price dinner service (thousands of dollars)	20	175	579.6
x_6	Age of youngest professor elected to a chair in a major university (yrs.)	15	19	22
x_7	Length of professorship of most durable professor (yrs.)	65	80	63
x_8	World's longest beard (ft.)	5	9	11.97
x_9	Deepest man-made hole (ft.)	5,500	10,000	11,246
x_{10}	Length of world's longest cigar (ft.)	3	6	5.6
x_{11}	Longest fasting record (days)	42	90	382
x_{12}	Fastest measured speed of a spider (ft/sec)	2	4.5	1.73
x_{13}	Modern pole sitting record (days)	7	21	211
x_{14}	World's longest sermon on record (hrs)	8	23	48.8
x_{15}	Typical length of world's largest ant (in.)	2	5	1.3
x_{16}	World's longest straight length of railroad track (mi.)	70	500	297
x_{17}	Weight of world's largest recorded gorilla (lbs)	600	1,200	716.5
x_{18}	Longest recorded jump of flea (in.)	36	60	13
x_{19}	Longest labor strike (months)	24	40	396
x_{20}	Diameter of world's largest rope (in.)	10	25	47

Results (x_1 through x_{20}):

11 revealed values right of x_r
4 revealed values left of x_ℓ
5 revealed values between x_ℓ and x_r
20 total

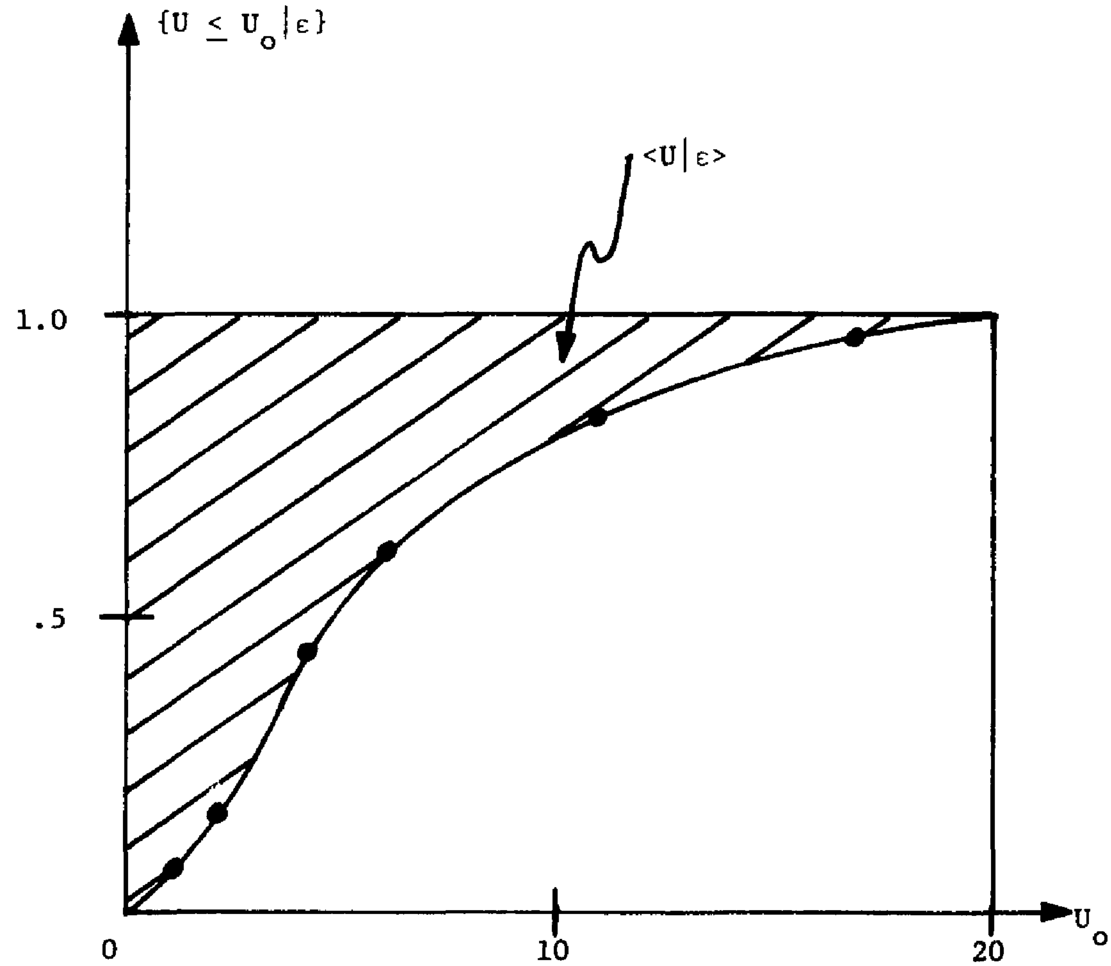


Figure 4.5 Prior Assessment of Number of Upper Interval Successes

is to be expected; it was, in fact, one of the central motivations for this research. The result also displays why many such experiments are misleading. The empirical data of fifty-five percent occurrences in the upper region is seen to be much closer to what the subject would have expected if his knowledge of assessment performance was initially incorporated consistently in his probability assessment.

4.6 Summary

In this chapter the theory of subjective probability was extended by explicit treatment of assessment performance information. A conceptual framework was formed out of which fell essentially three methdological aids:

(1) A means for guaranteeing consistency between prior assessments and information relating to assessment performance.

(2) An operational method for self-calibration.

(3) The Method of Equivalent Intervals; a new probability assessment technique with some useful qualities.

The following chapter will exploit some of the central ideas in this chapter to provide practical insights into the use of experts.

Chapter V

DEVELOPMENT OF A METHODOLOGY: SOME FUNDAMENTAL RESULTS

5.0 Introduction

The theoretical models for expert use developed to this point are philosophically useful but impractical. Direct application of the formulas presents a formidable task. By extension of the conceptual structure presented thus far we shall derive fundamental practical results, both useful and surprising.

The methodology presented here has important side benefits. The first is a more detailed framework in which to view existing theories as limiting cases. Another benefit, perhaps more important, is the physical insight provided by focusing on practical aspects of the theory. Additionally, the results will complement the study of individual prior assessment treated in the previous chapter.

Section 5.1 concentrates on the general problem of assessing functions. In Section 5.2 we present a detailed study of the relationship of exchangeability to expert use. This sets the stage for Section 5.3, in which an important result is derived: the decision maker's posterior can be obtained, in the one expert case, by multiplying his prior by the expert's subjectively calibrated prior and normalizing. The multi-expert case is then treated in Section 5.4 and the concept of surrogate prior is introduced. This is followed by a detailed analysis in Section 5.5 of the calibration function: the device by which experts are subjectively calibrated. The entirety of Section 5.6 is devoted to examples. The examples are especially significant in this chapter since important special cases are analyzed, and the relationship of this dissertation to other theories on expert use is more clearly articulated. Section 5.7 offers a normative approach to the study of conditions under which a group of experts or decision makers should achieve a consensus of opinion.

5.1 Assessment of Functions

The treatment of probability density functions as random variables is not a new concept. The idea arises naturally in Bayesian game theory (Beckman [2]), and in various other situations (de Finetti [7], Howard [13], Matheson [14]). In fact, the problem of encoding uncertainty about functions has long been the subject of intense study in the field of stochastic processes.

It is not our purpose here to undergo a thorough mathematical treatment of the subject. The interested reader may refer to de Finetti [7] for such a presentation. Instead, we shall parallel the development in stochastic processes where, for the purpose of solving actual problems, certain regularities are taken advantage of and fundamental assumptions are made. The goal is to provide modeling tools general enough for practical use, yet not so general as to have no real world application.

Parameterization

The space of all probability distributions is of infinite dimension. In any practical problem involving the assessment of functions the class of possible functions must be parameterized. In this case the likelihood function $\{\{x|\rho\}|x,\varepsilon\}$ becomes an assessment of a probability distribution on a vector of parameters. With regard to expert use this step is justifiable theoretically, as well as practically, for two reasons. First, an expert only makes a finite number of assessments. All assessed continuous distributions are either interpolated or chosen from a previously parameterized set (curve fitting). Conceptually, it is useful to view an expert's probability function as a finite component vector of assessments. Second, even if the expert's distribution did contain an infinite number of assessments, the decision maker surely perceives only a finite data set. For example, a given distribution could be approximated by its value at a large number of points separated by a distance δ. It is safe to assert that, in diminishing δ, there exists a point such that the decision maker's posterior will be insensitive to further changes. The forthcoming results will provide a convenient means for deriving decision sensitivity to the degree with which the expert's prior is specified.

There are many ways to parameterize probability distributions. An obvious yet practical way is to form a catalog of possible expert distributions, thereby reducing the problem to an assessment of the catalog index [14]. An orderly way of forming such a catalog is to select rich families of parameterized curves such as the Beta and Normal families.

An important means of parameterizing distributions is through moments. It may be true, for instance, that the mean and variance of an expert's prior contain as much relevant information (to the decision maker) as the entire distribution. In this case the problem resolves to a joint assessment of these two moments. As a sidelight it is interesting to note that for the class of all distributions expandable in a Taylor series, the likelihood assessment may be done as precisely as desired by specifying a joint distribution on the central moments and the corresponding derivatives of the given expert assessment evaluated at the mean.

Reduction of dimensionality

A further important and interesting parameterization can be made which is especially pertinent to expert use. Consider the two parameters s and ℓ, where s specifies the shape of the expert's distribution and ℓ specifies its location. The shape of an expert's prior is defined by its graph with an unlabeled horizontal axis. Specifying the location means labeling the axis. For example, a typical Beta distribution, as a function of x, may be written as:

$$f_\beta(x|r,n)$$

We can consider the constants r and n as specifying the shape of this distribution while allowing for variable location by forming the function

$$f_\beta(x - \ell|r,n),$$

where ℓ is defined to be the location.

Having yet made no simplifying assumptions we write:

$$\{\{x|\rho\}|x,\varepsilon\} = \{s,\ell|x,\varepsilon\} = \{\ell|s,x,\varepsilon\}\{s|x,\varepsilon\} \qquad (5.1)$$

For reasons to be made clear later $\{\ell|s,x,\varepsilon\}$ is an extremely important quantity. It is a probabilistic specification of the location of the expert's prior, given its shape and the revealed value of the random variable. The quantity $\{s|x,\varepsilon\}$ is the assessment of the shape of the expert's prior given only the revealed value of the random variable. In many cases we expect that this assessment will be independent of x. The main instance where the assumption might not be valid is when the decision maker thinks in terms of precision or percentages. In this case the fundamental component of shape is, roughly, the spread or variance of the expert's prior. In the forthcoming chapter we will introduce (in the section on expert bias) a new, and for our purposes more convenient, way to characterize this parameter. In the remainder of this chapter we shall assume independence.

The consequences of the assessment of s being insensitive to x alone are powerful. In this case we may write:

$$\{x|\{x|\rho\},\varepsilon\} = k\{\ell|s,x,\varepsilon\}\{x|\varepsilon\} \quad , \tag{5.2}$$

the factor $\{s|x,\varepsilon\}$ having been submerged into the constant. Note that the shape has no longer to be assessed. This is crucial since the parameter s has many dimensions in general while the parameter ℓ has but one. In other words, when such an assumption can be made, the problem reduces from a large joint assessment $\{s,\ell|x,\varepsilon\}$ to a single conditional assessment $\{\ell|s,x,\varepsilon\}$. Note that this assumption does not make the shape of the expert's prior irrelevant. In fact, the assessment of ℓ will, as we shall see, be extremely dependent on s in general.

Having made the above assumption, there is no longer any reason to assess $\{\ell|s,x,\varepsilon\}$ for each possible shape to guarantee consistency. Therefore it is expedient to have the expert present his distribution without location prior to any assessment of it. The decision maker, having observed s, can then appraise $\{\ell|s,x,\varepsilon\}$, after which the expert reveals the true value of ℓ.

5.2 Exchangeability and Expert Performance

The powerful applicability of the concept of exchangeability to the theory of individual probability assessment provides strong incentive to search for similar conceptual tools relative to the study of expert use. Superficially, the concept of exchangeability as used to aid an appraiser structure his knowledge about assessment performance seems readily transplantable to the use of experts. In both cases performance is, roughly speaking, the fundamental phenomenon of interest. However, as we shall display, the natural extension of the methodology developed in Chapter IV is fruitless. The key to viewing the problem correctly is the formulation of the conceptual dual of the original problem.

Direct extension

It is instructive to consider the direct extension approach because it will give us insights into basic assumptions underlying other theories. To start, we shall consider an assessment of the variable x_o and presume that the expert has supplied his prior on x_o. Therefore the decision maker is concerned with assessing $\{x_o|\{x_o|\rho\},\varepsilon\}$.

We next proceed analogously to Chapter IV. First the set $(x_o,\ldots,x_N)$ can be formed and the expert's prior on each additional random variable obtained. Let f_n^i equal the n^{th} fractile of the expert's prior on x_i. We define the performance index (or indicator) random variable p_i as the n that satisfies

$$f_n^i = x_{io} , \tag{5.3}$$

where x_{io} is the true, or revealed, value of x_i. Conversely stated, the revealed value of x_i corresponds to the $p_i{}^{th}$ fractile of $\{x_i|\rho\}$.

The performance indicators are defined relative to the expert's priors, not the decision maker's. However, if the decision maker, after obtaining each expert prior, assesses $(p_o,\ldots,p_N)$ to be an exchangeable set, he can proceed just as in the last chapter. Specifically, he can assess the expected frequency histogram of the performance indices, and use the area of this function over a specified interval to compute the posterior probability that x_o falls within the region determined by the corresponding fractiles of $\{x_o|\rho\}$. If the expert's performance is then observed on a large set of variables, exchangeable in performance index, the probability of any x_o interval is determined by the relative frequency of occurrences in the corresponding intervals of the variables in the set.

Important to note is that the above logic underpins current "calibration techniques" as briefly described in Chapter II. The basic necessary assumption is that, to the decision maker, the expert's performance indices are exchangeable. We shall now probe deeper into this assumption, revealing first its inflexibility in the single expert case, and then further demonstrating its inability to capture the essence of the many expert case.

In the single expert case consider the normalized scale depicted in Figure 5.1, where the marks are fractiles of the expert's priors. In other words, each mark represents a different number for each variable x_i, defined by the corresponding fractile of $\{x_i|\rho\}$. Suppose that the decision maker establishes, for each x_i, a posterior distribution $\{x_i|\{x_i|\rho\},\varepsilon\}$ and plots it on the normalized scale. Two such plots are shown in Figure 5.1 where, for example, m is the height of the decision maker's posterior on x_1 at the .1 fractile of $\{x_1|\rho\}$.

A necessary condition for exchangeability is now easily described: the decision maker's posteriors on all variables, x_o through x_N, must be identical when plotted on the normalized scale. However this is a difficult requirement. Since the plot is of posteriors, the necessary condition must rely on the shape of the decision maker's own prior. Due to his own knowledge about the variables the decision maker may not assess exchangeability as above defined, even though he intuitively feels the expert's probability assessment ability to be equivalent on each variable. Stated another way, the above definition relies on the decision maker's joint assessment of the basic random variables and the expert's assessment performance, when assessment performance is the only matter of issue. The example depicted in Figure 5.2 shows one possible violation of exchangeability. Intuitively, it is apparent that if the decision maker considers the expert's knowledge about x_1 and x_2 roughly equivalent, he would assign a higher probability to x_1 falling above the .9 fractile of $\{x_1|\rho\}$, than to x_2 falling above the .9 fractile of $\{x_2|\rho\}$.

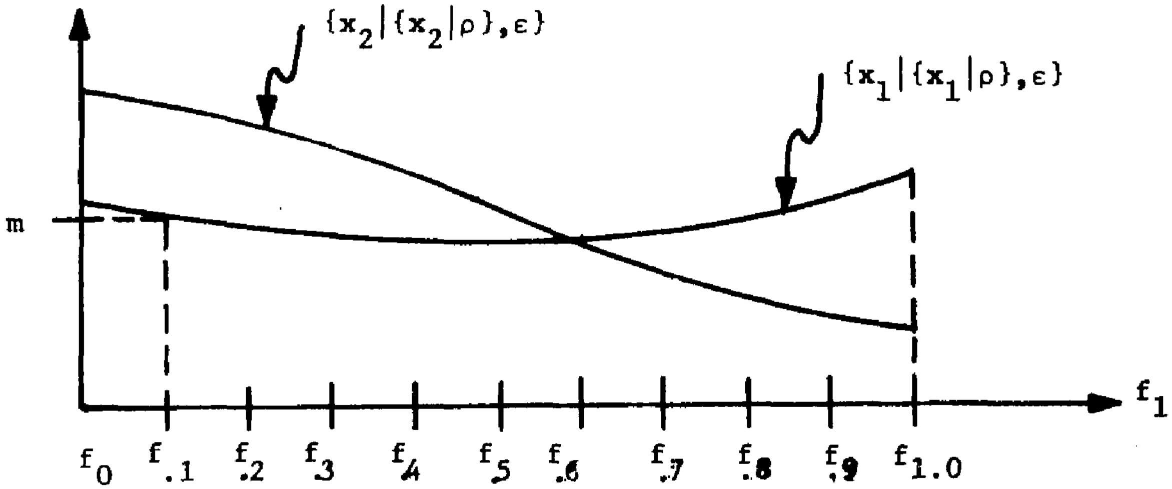


Figure 5.1 A Normalized Scale

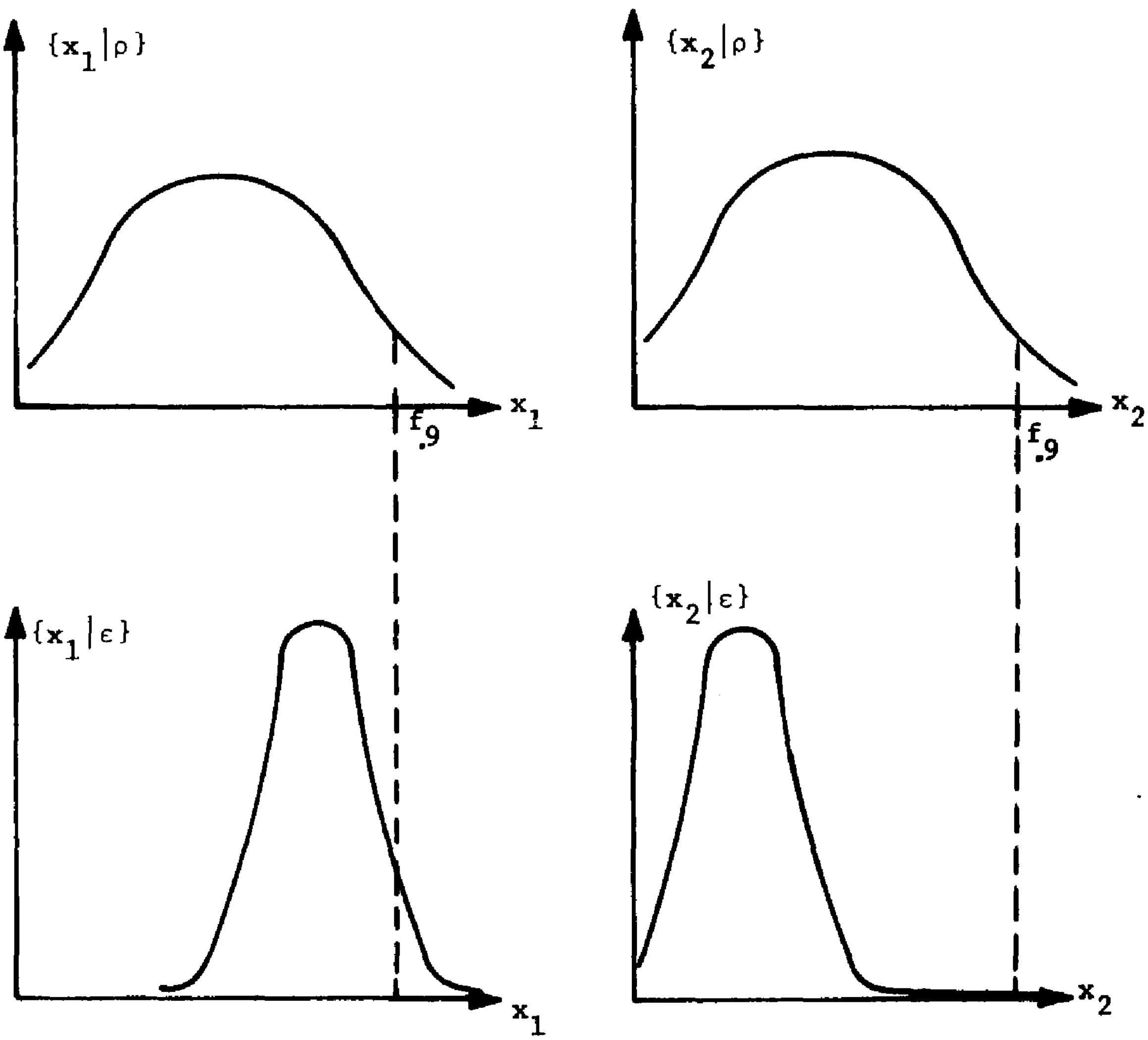


Figure 5.2 Violation of Exchangeability

To make the case stronger note that when the decision maker's posteriors are all equal to the expert's priors, the resultant normalized plots would be identical (all uniform). In this sense the necessary condition is just as stringent as requiring the decision maker to directly use his expert's prior as his own posterior.

When more than one expert is being considered the case against the implicit assumption underlying the conventional technique becomes stronger. It is inconsistent for the indicators to be exchangeable when defined on the same variables for two experts, except in the extremely rare case where both experts have identical priors on all variables and the decision maker appraises each expert equivalently. Even if separate sets of variables are used, the variable of interest must be jointly appraised. This is also a powerful argument against the single expert assumption because it is hard to justify that one set of variables should be exchangeable for one expert only.

Clearly, if exchangeability is to be useful in using experts, it is not in the above fashion. We should like a means whereby the decision maker's prior knowledge about a random variable can be decomposed from his exchangeability assessment of the expert's performance. Furthermore we should benefit from a further decomposition of the decision maker's knowledge about the expert as an expert on the random variable of interest, and his knowledge about the expert as a probability assessor.

The conceptual dual

As previously indicated, the basis for a rich analysis of the decision maker-expert situation is provided by the conceptual dual of the single assessor situation. Specifically, we have been concerned with the assessment of the performance of an expert as measured by the fractile of his given prior in which the random variable of interest is revealed to fall. A more relevant and interesting assessment situation to consider is that where the random variable itself is known, but the location of the expert's prior is unknown. In this case the performance index is revealed only when the expert's prior is specified.

By viewing the situation from this new vantage point the assessment of the expert's performance (as measured by the p_i's) is decoupled from the assessment of the basic random variables. Exchangeability is now a reasonable and intuitive condition. It simply means that the decision maker cannot discern between the expert's probability assessing ability on variables forming the basis for the exchangeable set. As in the previous chapter, it seems harder to find a non-member of such a set than a member.

Fortunately we have already formulated a structure in which the appraisal of an unknown prior, given the true value of the assessed random variable, plays a central role. In the next section the above concepts will be exercised within this framework to provide important new results in expert theory.

5.3 A Fundamental Result

We come now to what is probably the most interesting development of the dissertation: a result not only quite surprising theoretically, but also fundamental in providing a link between theory and practice. Restating the basic expression under study, we have:

$$\{x|\{x|\rho\},\varepsilon\} = \{x|\ell,s,\varepsilon\} = k\ \{\ell|x,s,\varepsilon\}\{x|\varepsilon\} \quad (5.4)$$

The basic item of interest is $\{\ell|x,s,\varepsilon\}$, the assessment of the location of the expert's prior, given its shape and x. The definition of location can be arbitrary; however the reader might find it helpful to consider ℓ to be the mean of the expert's prior.

Suppose that the decision maker is considering the variable x. We presume that he has available a large (possibly conceptual) set of other variables, exchangeable in performance index (defined in equation 5.3). Recall that this means that the fractile indices of each expert prior are indistinguishable uncertain quantities to the decision maker. We define the performance distribution $\mathfrak{X}$ as the frequency distribution with the following property:

$$\int_{n_1}^{n_2} \mathfrak{X}(f)df = \text{The long run relative frequency of times that the expert priors in the exchangeable set contain the true value between the } n_1 \text{ and } n_2 \text{ fractiles}$$

Conceptually it is easier to think about the cumulative of $\mathfrak{X}$, denoted $\mathfrak{X}_{\leq}$. Therefore, $\mathfrak{X}_{\leq}(n)$ is the relative frequency of "successes" occurring below the n^{th} fractile.

In the most general case the exchangeable set, and hence the distribution $\mathfrak{X}$, will depend on the revealed value of x. The discussion of this circumstance will be postponed, however, as the essential results are unchanged by assuming $\mathfrak{X}$ to be independent of x.

It is further useful to construct the function g as follows:

$g(c) =$ the index of the fractile of $\{x|\rho\}$ corresponding to the point $\ell+c$. (Recall that ℓ is the location of $\{x|\rho\}$.)

For example, if the point two units to the right of the location of $\{x|\rho\}$ specifies the .8 fractile, then $g(2) = .8$. One last term of interest is L, the distance from the location of $\{x|\rho\}$ to the revealed value of X. Figure 5.3 depicts the relationship of the newly defined variables. Notice that

$$\ell = x - L \tag{5.5}$$

For clarity we shall use the symbol x_o for a particular value of x in the ensuing development. Therefore, while $\{x|\rho\}$ is thought of as a "picture" of the function, $\{x_o|\rho\}$ is thought of as the function evaluated at the point x_o.

The reason for the preceding definitions will now become clear. First, we expand the term of interest over all possible observable performance distributions:

$$\{\ell|\ x,s,\varepsilon\} = \int_{\mathscr{X}} \{\ell|x,s,\mathscr{X},\varepsilon\}\{\mathscr{X}|\varepsilon\} \quad , \tag{5.6}$$

where probabilistic independence of $\mathscr{X}$ on x and s has been assumed. In order to obtain $\{\ell|\ x,s\mathscr{X},\varepsilon\}$ by change of variables we note that, by the exchangeability assumption, the probability that the variable L is less than some arbitrary number a, is the long run frequency of occurrences of the corresponding events in the exchangeable set: that is, by inspection of Figure 5.3, the long run frequency of times that the variables fall below the $g(a)^{th}$ fractile. Therefore

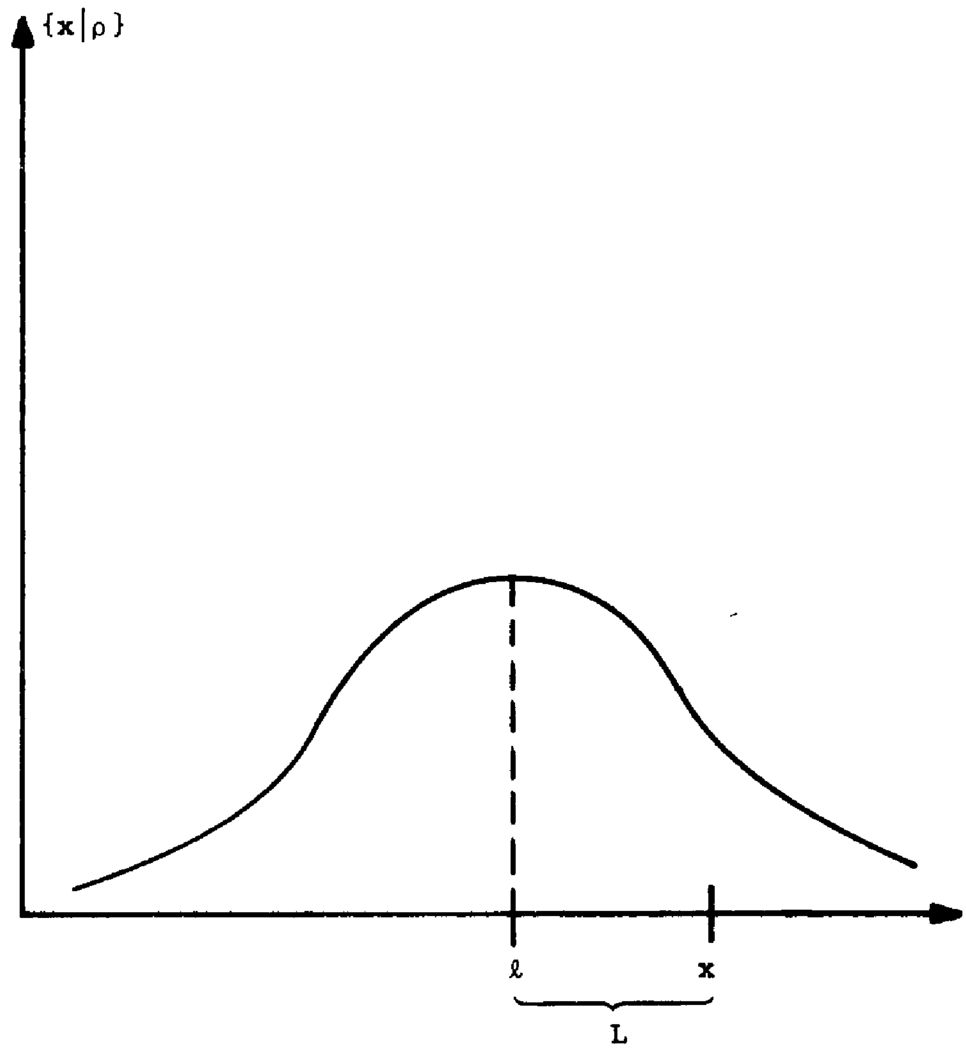


Figure 5.3 Relationship of Variables

$$\{L \leq a|x,s,\mathscr{F}\varepsilon\} = \mathscr{F}_{\leq}[g(a)] \tag{5.7}$$

The complementary cumulative probability that ℓ is less than a particular value ℓ_o is calculated via equation 5.5 as

$$\{\ell \geq \ell_o|x,s,\mathscr{F}\varepsilon\} = \{L \leq x - \ell_o|x,s,\mathscr{F}\varepsilon\} \quad , \tag{5.8}$$

which from equation 5.7 is $\mathscr{F}_{\leq}[g(x - \ell_o)]$. By taking the negative derivative of equation 5.8 we obtain the desired density function on ℓ:

$$\{\ell|x,s,\mathscr{F}\varepsilon\} = - \frac{d}{d\ell}\mathscr{F}_{\leq}[g(x - \ell)] = \mathscr{F}'_{\leq}[g(x - \ell)]g'(x - \ell) \tag{5.9}$$

where $\mathscr{F}'_{\leq}$ is the derivative of $\mathscr{F}_{\leq}$, which of course is the density $\mathscr{F}$, and g' is the derivative of the function g.

Recalling again the definition of g, we note that when the location ℓ is known we may write:

$$g(c) = \int_{-\infty}^{\ell+c} \{x|\rho\} \tag{5.10}$$

so that

$$g(x_o - \ell) = \int_{-\infty}^{x_o} \{x|\rho\} \tag{5.11}$$

$$= \{x \leq x_o|\rho\}$$

which is the expert's prior cumulative distribution. Thus we obtain the startling result:

$$\{\ell|x_o,s,\mathcal{H},\varepsilon\} = \{x_o|\rho\}\bar{\mathcal{H}}[\{x \leq x_o|\rho\}] \tag{5.12}$$

Furthermore, returning to equation 5.6, we see that since $\{x|\rho\}$ is not a function of a particular performance distribution, it factors out of the integral leaving

$$\{\ell|x,s,\varepsilon\} = \{x|\rho\}\bar{\mathcal{H}}[g(x)] \ , \tag{5.13}$$

where $\bar{\mathcal{H}}$ is a compact notation for the (functional) expectation of the distribution $\mathcal{H}$ (refer to chapter IV), and is a single valued function of $g(x)$. The result is that the likelihood function is the product of two terms, each of which is a distribution.

Before proceeding it should be noted for completeness that the above result is not entirely correct if $\{x|\rho\}$ is zero on an interval of finite length in the range of possible x's (defined by $\{x|\varepsilon\}$). In this case there are non-unique fractiles and equation 5.7 is invalid. The theory can be extended to such special cases but would not add insight. Additionally, an expert trained in Bayesian techniques would not attach zero probability to an interval of finite width.

Reassembling the original posterior of interest we have that:

$$\{x|\{x|\rho\},\varepsilon\} = k\bar{\mathcal{H}}(x)\{x|\rho\}\{x|\varepsilon\} \tag{5.14}$$

Before further discussion of $\bar{\mathfrak{F}}(x)$ note the operational significance implied by the above equation. To use an expert we simply multiply his prior times the decision maker's prior times another term and normalize. Regardless of what the expert's prior is, independent of how he is viewed as an expert, it should always multiply the decision maker's prior to obtain the posterior. Further notice that nowhere does a distribution on distributions need be specified--only the expected value of such a distribution. Even this assessment will be further simplified; a development we postpone until the next section.

The calibration function

For convenience let us relabel the composite function $\bar{\mathfrak{F}}[\{x \leq x_o|\rho\}]$ as $C(x_o)$ and call it the calibration function. The reason for this name will soon become apparent. Figure 5.4 provides a graphical calculation of the calibration function. It will be useful to understand the figure thoroughly. The specific point $C(a)$ is obtained by calculating $\{x \leq a|\rho\}$ and then determining the value of the function $\bar{\mathfrak{F}}$ at this point.

The calibration function and self-calibration

An interesting and theoretically useful way to view the calibration function is as the transformation, which when applied to the expert's prior, converts it to the assessment he would submit after self calibrating himself with the performance distribution $\bar{\mathfrak{F}}$. To see this, recall the development in Chapter IV, where it was displayed that if $\{x|\varepsilon\}$ is any prior, with $g(x)$ its cumulative, and P is any probability distribution

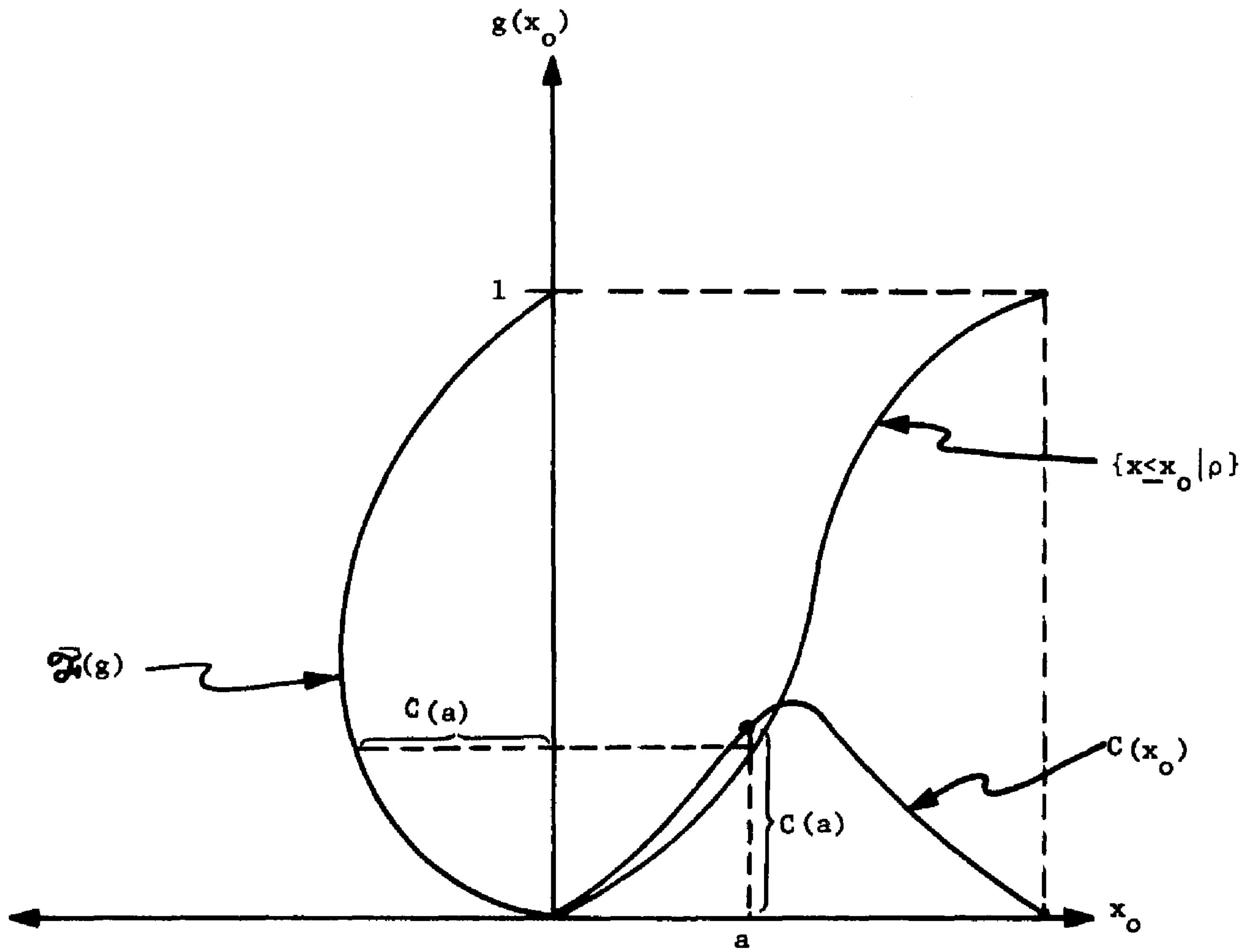


Figure 5.4 Construction of the Calibration Function

on the interval [0,1], then $\{x|\varepsilon\}P[g(x)]$ is the new probability distribution on x inferred by observation of P as the frequency distribution of performance indices. Thus, if the expert observes $\mathcal{P}$, his updated distribution will be $\{x|\rho\}\mathcal{P}[g(x)]$ which is just his original prior times the calibration function. Viewing from another angle, if $\mathcal{P}$ is known deterministically to both the expert and the decision maker then, assuming they each assess exchangeability, they will each agree on how the expert should update his prior. With respect to the expert's self-calibrated prior the calibration function becomes uniform.

We have obtained a basic result: when the expert is self-calibrated relative to a class of priors the decision maker himself considers exchangeable, the resulting posterior given the expert's prior is:

$$\{x|\{x|\rho\},\varepsilon\} = k\{x|\rho\}\{x|\varepsilon\} \tag{5.15}$$

Note that the above expression is symmetric in the expert's and decision maker's priors. This mathematical property provides intuitive insight: a decision maker should weight a calibrated expert exactly like himself. Thus, in a way, the calibration function translates the expert's probability language into the decision maker's. It provides a basis by which the decision maker can compare his expert's probabilistic assertions with his own.

In a real sense, C(x) calibrates the expert, which justifies the name "calibration function." In the case where $\mathcal{P}$ is not known, the function C(x) is the expected function which will calibrate the expert. It is fundamental to observe that the calibration procedure is done independently of the decision maker's own prior. Thus, if there is general agreement about the form of the calibration function, there will be no controversy over how to use an expert. Section 5.7 will further develop this point.

Diffuse prior

An additional special case of interest is that in which the decision maker's own prior is diffuse relative to the expert's and roughly uniform. We then have immediately that:

$$\{x|\{x|\rho\},\varepsilon\} = C(x)\{x|\rho\} \tag{5.16}$$

The decision maker's posterior is just the expert's prior times the calibration function.

Another interesting point is that whenever a self-calibrated expert submits a uniform prior, the decision maker's posterior is equal to his own prior weighted by a calibration function equal to [illegible] on a renormalized interval equal to the width of the expert's prior. Before proceeding to the many expert situation we make one more extension.

Dependence of performance assessment on x

A case can be made that in general the assessment of the performance distribution should be a function of x. Before discussing this notion intuitively we shall quickly derive its mathematical implications.

The assertion is technically that $\{$[illegible]$|x,\varepsilon\}$ is not independent of x as was assumed in equation 5.6. Supposing this to be the case, it is easy to demonstrate that the development up to equation 5.13 is identical. The only difference is that the expectation $\bar{\mathcal{P}}_1$ is now a function of x explicitly:

$$\bar{\mathcal{P}}_1 = \bar{\mathcal{P}}_1[g(x),x] \tag{5.17}$$

Thus we may still define the calibration function as a real function of x only and the basic result remains intact. Figure 5.5 presents the extension of Figure 5.4 to the present case.

Two immediate cases come to mind in which it might seem reasonable to assume dependence on the revealed value. The first is where the decision maker feels that he and the expert lack cross-calibration. If the decision maker is surprised by the revealed value it might increase the probability to him that the expert will be so surprised. The second case is due to border effects. If it is known, for instance, that x lies between 0 and 1, the additional knowledge that x is very close to either extreme will certainly affect the probability that it is contained in a tail of $\{x|\rho\}$.

However, by the very nature of this circumstance, the effect will be subdued by multiplication of the tail of the expert's distribution. If the expert's prior was narrow enough to be large in this region it would no longer be a border effect case. At any rate, the specification of $\bar{\mathcal{P}}_1$ as a function of x is not prohibitively demanding.

5.4 Many Experts

In the case of N cross-calibrated experts we have, recalling equation 3.7, the immediate result:

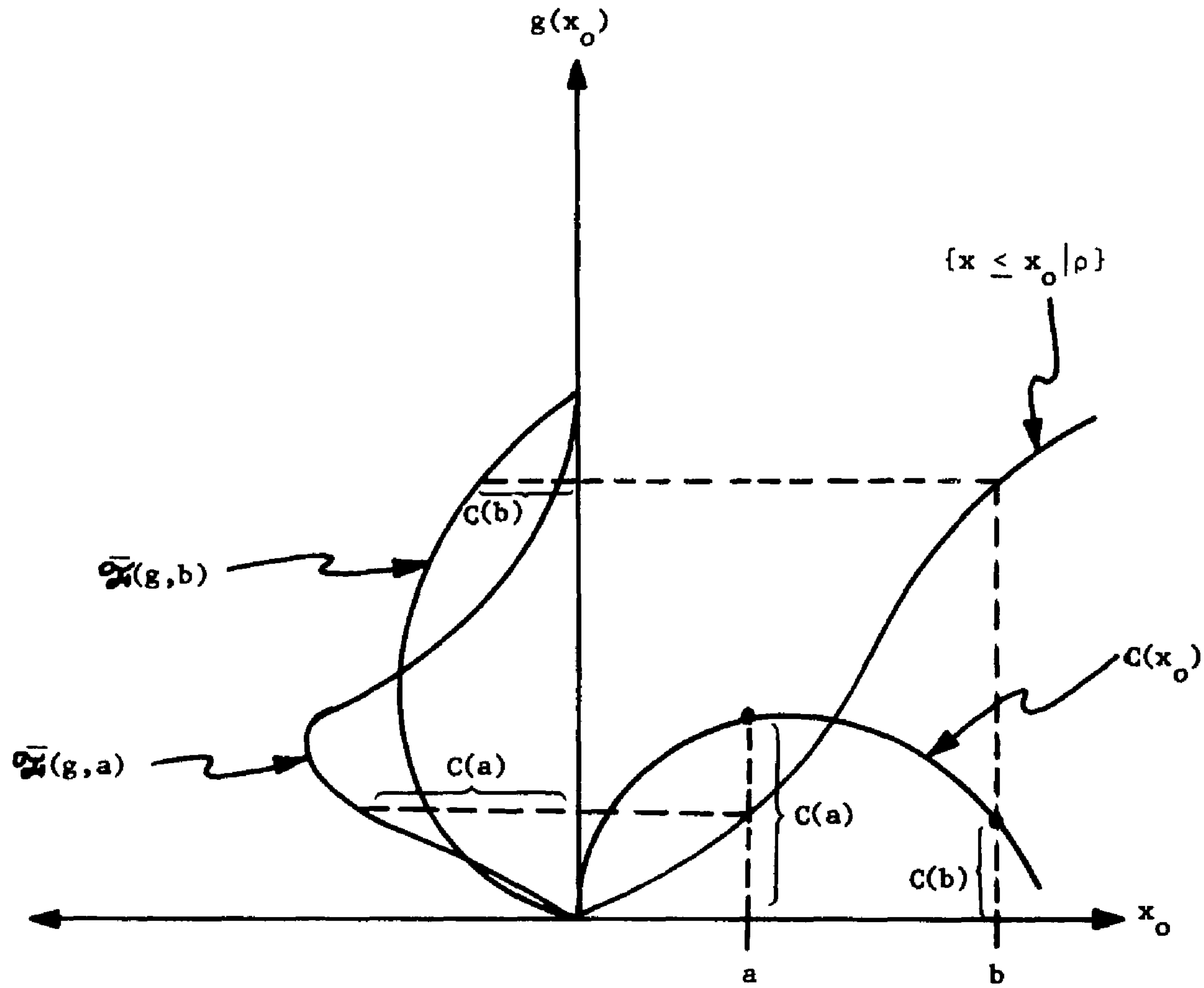


Figure 5.5 Generalized Calibration Function

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_N\},\varepsilon\} = k\prod_{i=1}^{N} C_i(x)\{x|\rho_i\}\{x|\varepsilon\} \tag{5.18}$$

where $C_i(x)$ is the calibration function pertaining to the i^{th} expert.

Dependence

The situation is more complex and challenging when the experts are not independently assessed. Consider the multi-expert case where ℓ_i and s_i are, respectively, the location and shape of Expert i's prior. We can write:

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_N\},\varepsilon\} = k\{\underline{\ell}|\underline{s},x,\varepsilon\}\{x|\varepsilon\} \tag{5.19}$$

where $\underline{\ell} = (\ell_1,\ldots,\ell_N)$, $\underline{s} = (s_1,\ldots,s_N)$, and where it is assumed again that the shape of the expert priors are independent of x.

Next consider the scenario where each expert has assessed a prior on each of a large set of random variables. Let f_n^{ij} be the n^{th} fractile of the j^{th} expert's prior on x_i. The performance index p_i^j is then defined such that

$$f^{ij}_{p_i^j} = x_{io} \quad , \tag{5.20}$$

where, again (recall equation 5.3), x_{io} is the revealed value of x_i. In words, the revealed value of x_i corresponds to the $p_i^{j\,th}$ fractile of $\{x_i|\rho_j\}$.

We now define a performance matrix P whose ij^{th} element is p_i^j. Thus

$$P = \begin{bmatrix} p_1^1 & p_1^2 & \cdots & p_1^N \\ p_2^1 & p_2^2 & \cdots & p_2^N \\ \vdots & \vdots & & \vdots \\ p_M^1 & p_M^2 & \cdots & p_M^N \end{bmatrix} , \tag{5.21}$$

where M is the number of variables under consideration. Of particular interest are the rows of P, each of which is a vector

$$\underline{p}_i = (p_i^1, \ldots, p_i^N) \tag{5.22}$$

called the i^{th} performance vector. Each $\underline{p}_i$ is regarded as a random quantity containing all assessment performance information on the variable x_i.

Our assumption will be that the set of all performance vectors is exchangeable to the decision maker. This means that his state of information about the variable $\underline{p}_i$ is invariant to the particular value of the subscript i. Clearly, a necessary condition for exchangeability is that the elements of the column vectors

$$\underline{p}^j = (p_1^j, \ldots, p_M^j) \tag{5.23}$$

form an exchangeable set. This merely means that the $\underline{p}_i$'s are component wise exchangeable and corresponds to the assumption made in the previous section: the performance indices corresponding to the individual experts must form an internally exchangeable set.

The additional intuitive meaning of vector exchangeability is that the experts retain, in the decision maker's mind, the same relative probability assessment ability from variable to variable. For example in the two expert-two variable case, if the decision maker is told only that the true value of one of the variables is the .3 fractile of Expert 1's prior and the .7 fractile of expert 2's prior, he must assess it equally likely that the assessments pertained to either variable.

Given exchangeability, the joint distribution

$$\underline{\mathscr{F}}(f_1,\ldots,f_N) = \underline{\mathscr{F}}(\underline{f}) \tag{5.24}$$

can be defined whose integral over any region is the long run frequency of the true values contained in that region. For example, the frequency of times that all priors contain the revealed value of each variable to the right of the .9 fractile is

$$\int_{f_1=.9}^{f_1=1} \cdots \int_{f_N=.9}^{f_N=1} \underline{\mathscr{F}}(\underline{f})\,d\underline{f} \tag{5.25}$$

Parallel to the single expert case we may, in order to obtain the joint density on $\underline{\ell}$, calculate

$$\{\underline{\ell} > \ell_o | \underline{s}, x, \varepsilon\} \tag{5.26}$$

where the inequality between vectors means the relation for each component. Next we define

$$g_i(c) \equiv \text{the index of the fractile of } \{x|\rho_i\} \text{ corresponding to the point } \ell_i + c \tag{5.27}$$

Thus $g_i(c)$ equals $\{x \leq \ell_i + c|\rho_i\}$. Using precisely the same reasoning as in the single variable case it may readily be established that:

$$\{\underline{\ell}|\underline{s}, x, \ , \varepsilon\} = -\frac{\partial}{\partial \ell_1} \cdots \frac{\partial}{\partial \ell_N} \int_0^{g_1(x-\ell_1)} \cdots \int_0^{g_N(x-\ell_N)} \mathcal{F}(\underline{f}) d\underline{f}$$

$$= g_1'(x-\ell_1) \ldots g_N'(x-\ell_N) \mathcal{F}[\underline{g}(x)], \ \underline{g} = (g_1, \ldots, g_N)$$

$$= \prod_{i=1}^{N} \{x|\rho_i\} \mathcal{F}[\underline{g}(x)] \tag{5.28}$$

Joint calibration function

The entire result may be written as:

$$\{x|\{x|\rho_1\}, \ldots, \{x|\rho_N\}, \varepsilon\} = k \prod_{i=1}^{N} \{x|\rho_i\} \underline{C}(x)\{x|\varepsilon\} \quad , \tag{5.29}$$

where $\underline{C}(x)$ is defined as the expectation of $\mathcal{F}[\underline{g}(x)]$ and is called the joint calibration function. The result is identical to equation 5.18 with the exception that in the independent case the priors are all multiplied by a number of independent calibration functions. We note that, indeed, in the limit, when the experts are independent:

$$\underline{C}(x) = \prod_{i=1}^{N} C_i(x) \tag{5.30}$$

This is due to the fact that individual expert exchangeability is a necessary condition for joint exchangeability.

The surrogate prior

It is useful to define the density function

$$\{x|S\} = \frac{\underline{C}(x) \prod_{i=1}^{N} \{x|\rho_i\}}{\int_x \underline{C}(x) \prod_{i=1}^{N} \{x|\rho_i\}} \tag{5.31}$$

which is the calibrated product of expert priors normalized to have area one. We call the distribution $\{x|S\}$ the surrogate prior because it represents to the decision maker, the composite knowledge of all his experts. Notice that

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_N\},\varepsilon\} = k\{x|S\}\{x|\varepsilon\} \tag{5.32}$$

The surrogate prior is processed exactly as if it had been submitted by one calibrated expert. Thus we may conceptualize a surrogate expert whose state of information pertaining to a random variable is the union of the individual experts' states of information.

The concept of a surrogate expert, whose knowledge is represented by a surrogate prior, is useful, as well as interesting, because it allows us to speak specifically of the total body of knowledge of a group. For instance, we could theoretically calculate for a policy maker a representation of the total amount of knowledge in the United States concerning the possibility of discovery of a new source of power in the next twenty-five years.

In the next section we shall derive additional results simplifying to a great extent the assessment of the joint calibration function and demonstrating further its intuitive meaning. This will allow us to specify the surrogate prior of a group of experts in a practical way.

5.5 Analysis of the Calibration Function

The previous two sections have demonstrated the fundamental nature of the calibration function concept to the study of expert use. We shall now develop tools for the assessment of this function which will cement the link between the theory just presented and its application to everyday practice.

One expert

Recall in the one expert case that the calibration function was derived from the functional expectation:

$$\bar{\mathcal{F}} = \int_{\mathcal{F}} \mathcal{F}\{\mathcal{F}|\varepsilon\} \tag{5.33}$$

The assessment of $\bar{\mathcal{F}}$ is of course much easier than the complete assessment $\{\mathcal{F}|\varepsilon\}$. However, functional expectations over the entire space of probability distributions are hard to conceptualize.

Given, as before, a large set of variables exchangeable in performance index, let us define

$\langle f_a|\varepsilon\rangle \equiv$ the expected fraction of revealed values falling below the a^{th} fractile of the corresponding expert priors

Thus, $\langle f_a|\varepsilon\rangle$ is the expected fraction of performance indices having value less than a. We next expand over $\mathcal{F}$ to obtain:

$$\langle f_a|\varepsilon\rangle = \int_{\mathcal{F}} \langle f_a|\mathcal{F},\varepsilon\rangle \{\mathcal{F}|\varepsilon\} \tag{5.34}$$

However, the quantity f_a is specified by $\mathcal{F}$, so that

$$\langle f_a|\mathcal{F},\varepsilon\rangle = \int_o^a df\mathcal{F}(f) \tag{5.35}$$

Substituting in equation 5.34 and reversing the order of integration gives

$$\langle f_a|\varepsilon\rangle = \int_o^a df \int_{\mathcal{F}} \mathcal{F}\{\mathcal{F}|\varepsilon\} \tag{5.36}$$

$$= \int_o^a df\, \bar{\mathcal{F}}(f)$$

This fundamental result states that the cumulative of $\bar{\mathcal{F}}$ evaluated at the point a can be determined by specifying the expected fraction of times the expert's priors will contain the true value below the a^{th} fractile. The quantity $\langle f_a | \varepsilon \rangle$ is conceptually straightforward to assess--it does not involve probability assignments over the space of functions.

Many experts

In the many expert case the analogous result is similarly derived. The cumulative of the joint calibration function evaluated at the point $\underline{a} = (a_1, \ldots, a_N)$ is equal to $\langle f_{\underline{a}} | \varepsilon \rangle$ where

$\langle f_{\underline{a}} | \varepsilon \rangle \equiv$ the expected fraction of performance vectors whose components are jointly less than the corresponding components of $\underline{a}$

The assessment of $\langle f_{\underline{a}} | \varepsilon \rangle$, although a great simplification, still involves the specification of a function over a multi-dimensional space. We alleviate this difficulty below.

Trajectory analysis

The two expert case will be analyzed for descriptive convenience. It is fruitful to view $\underline{C}(x)$ as the height of $\underline{\mathfrak{F}}$ along a trajectory which is the projection of $\underline{C}(x)$ in the $f_1 - f_2$ plane. Each point on the trajectory is defined by the fractiles of $\{x|\rho_1\}$ and $\{x|\rho_2\}$ corresponding to a given value of x. Thus, as x increases, the trajectory moves up and to the right. Figure 5.6 illustrates a sample trajectory and its calculation.

The operational significance of the trajectory is derived by the fact that it is defined only by the experts' given priors and is independent of the assessed expected distribution $\underline{\mathfrak{F}}$. For obtaining the poster-

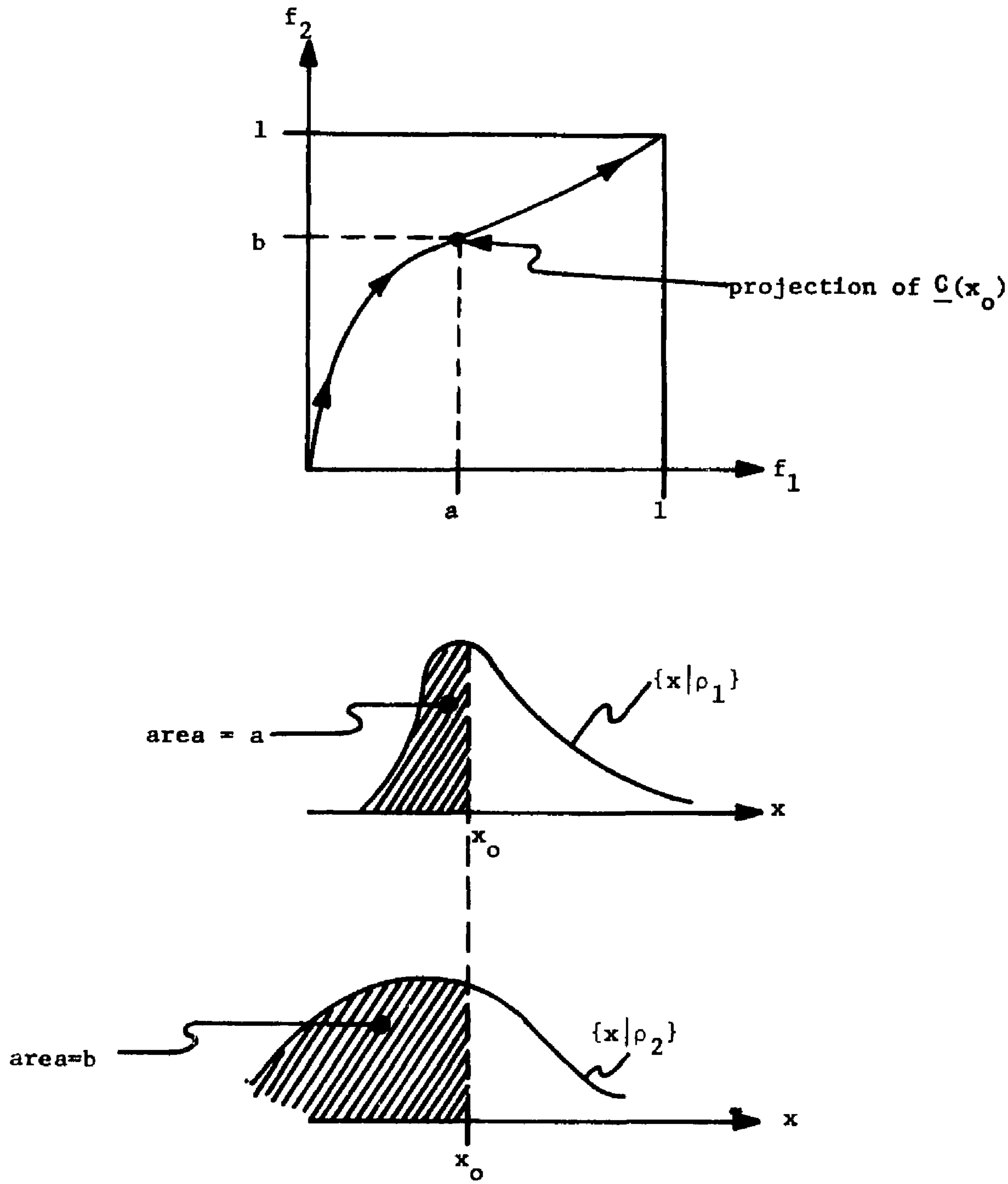


Figure 5.6 Specification of Trajectory Diagram

terior, $\underline{\tilde{\mathfrak{F}}}$ need only be specified along the trajectory.

Unfortunately, specification of the cumulative of $\underline{\tilde{\mathfrak{F}}}$ along a single dimensional path does not determine the value of $\underline{\tilde{\mathfrak{F}}}$ on that path uniquely. Theoretically, to compute density fraction at a point, the cumulative must be specified for some open set about that point. However, the trajectory determines a narrow band over which the cumulative must be specified.

For practical purposes the joint density $\underline{\tilde{\mathfrak{F}}}$ on the trajectory is well approximated by a series of steps. The height of any step is given by the volume of the density function over a rectangular area, divided by that area. The volume is specified over any area, by the decision maker, as the expected fraction of times that the performance vectors in an exchangeable set will fall in the rectangle. For example, consider the sequence of rectangles in Figure 5.7. Suppose that the decision maker assesses an expected frequency of $\bar{f}$ relative to the shaded rectangle whose area is A. As shown in Figure 5.8, the resulting calibration function has uniform height $\bar{f}/A$ over the interval $[x_1,x_2]$.

A further look at special cases brings intuitional rewards. Consider the two expert case where the experts are assessed symmetrically dependent. Thus, $\underline{\tilde{\mathfrak{F}}}$ is roughly shaped like a mountain range whose peak runs along the line $f_1 = f_2$. The height of the calibration function varies as the trajectory deviates from this line. In Figure 5.9 three special cases are shown. Case a is where the experts' priors are identical. As the figure shows, trajectory t_a is a straight diagonal following the peak of $\underline{\tilde{\mathfrak{F}}}$. In Case b the priors have equal shape but far different location. Trajectory t_b traces a path along the upper border of the diagram which implies a U-shaped calibration function. Finally, Case c

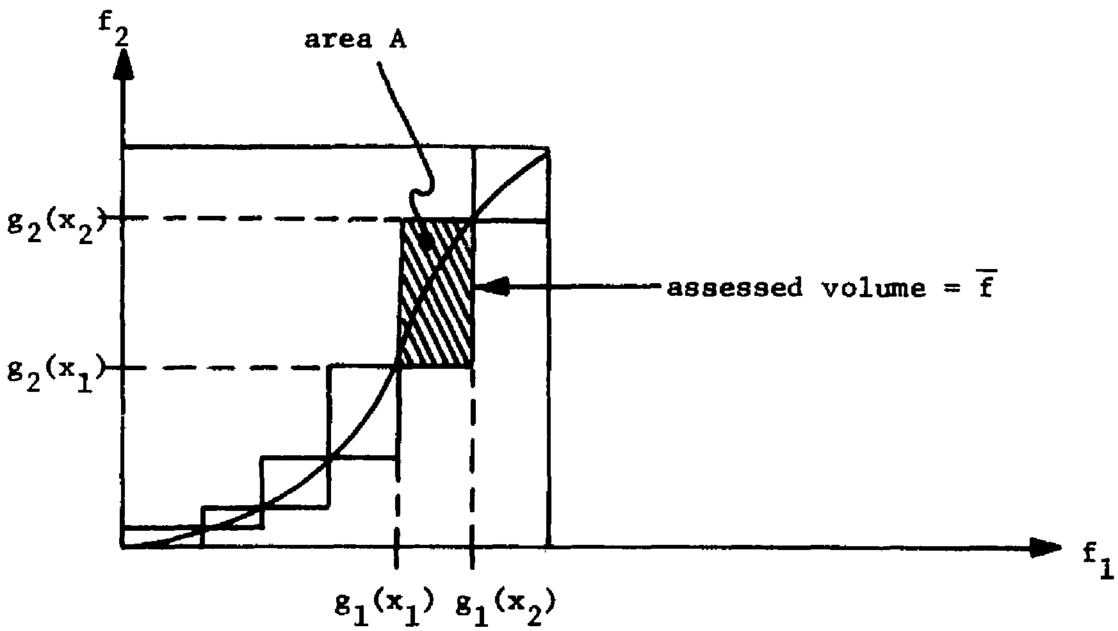


Figure 5.7 Approximate Trajectory Analysis

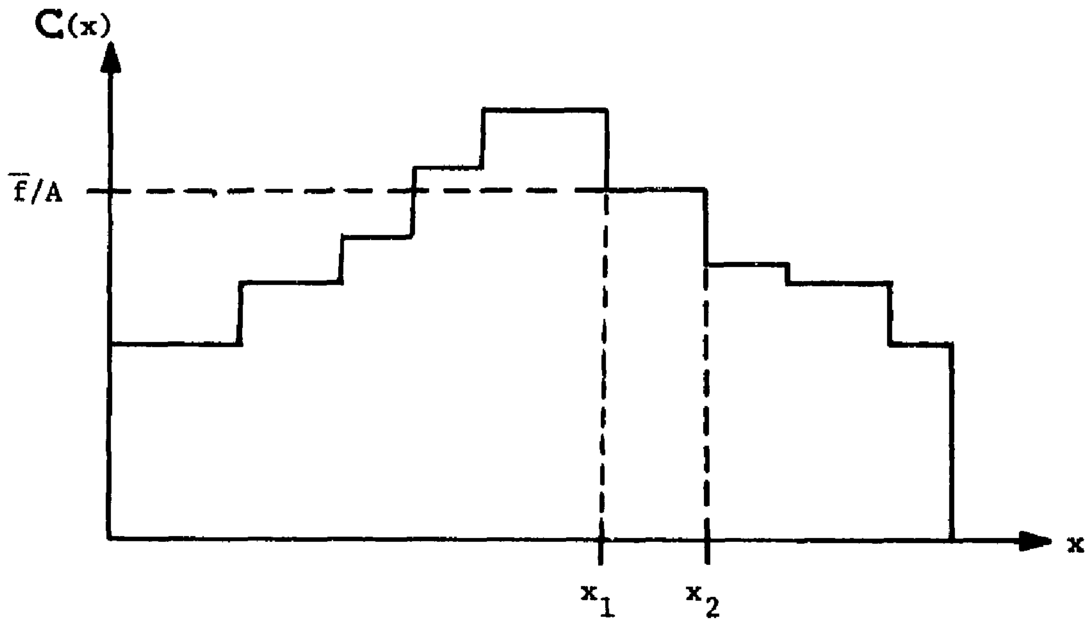


Figure 5.8 Approximate Calibration Function

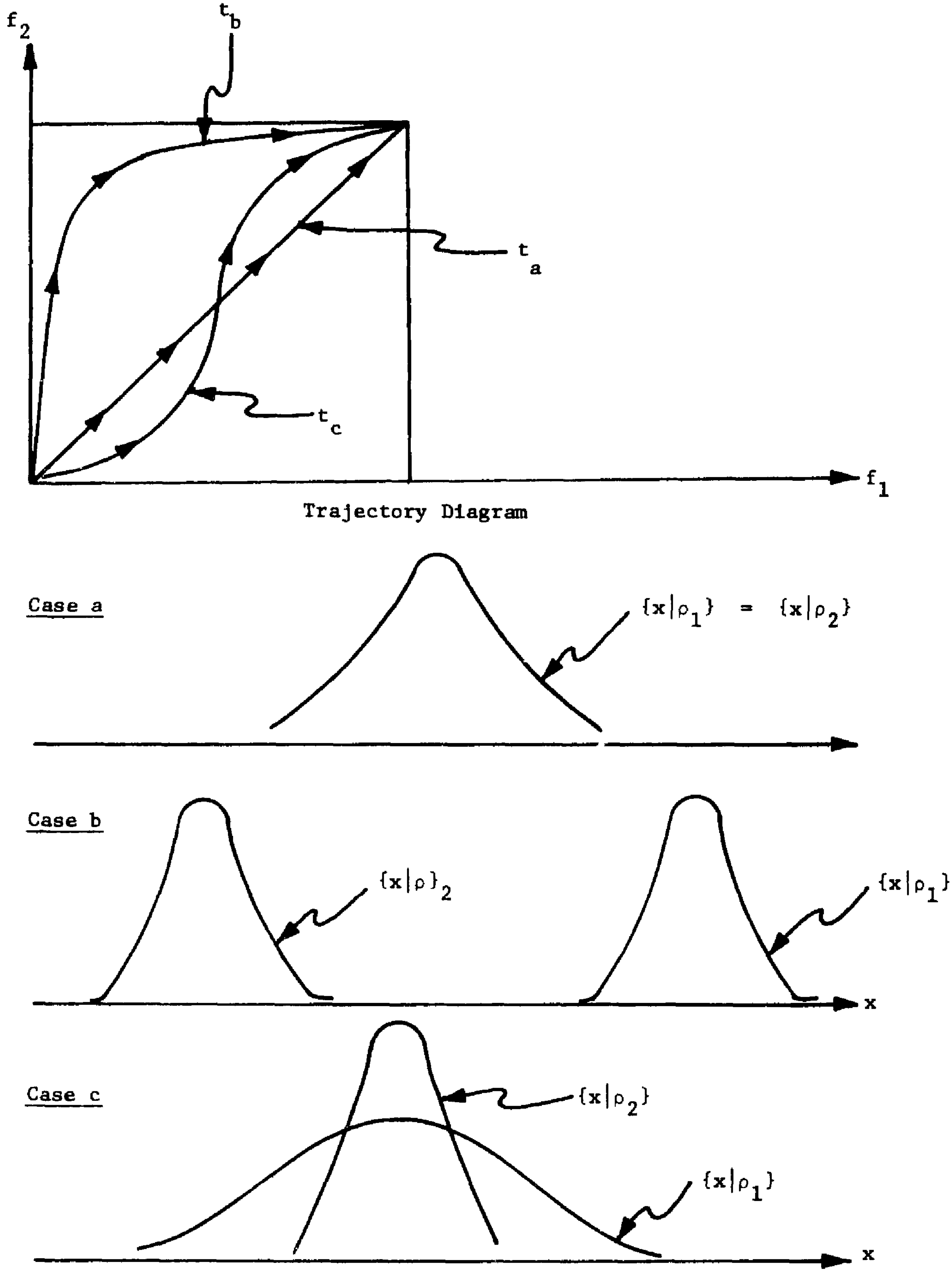


Figure 5.9 Trajectory Analysis

depicts the situation where the priors have the same location but different shape. In this case the trajectory t_c traverses both halves of the diagram, causing the calibration function to be rather sharply peaked. This type of analysis allows us to obtain a great deal of physical insight into an expert use problem without having to perform intricate mathematical calculations.

In summary we have displayed a practical methodology for specifying both the single expert and multi-expert calibration functions and thus, the surrogate prior. The next section presents examples to further amplify understanding of the material presented.

5.6 Examples

Example 5.1--Use of the Calibration Function

Suppose the decision maker feels that his two experts on the variable t are equivalent in terms of probability assigning ability. Consistent with recent empirical experiments he has assigned the expected performance distribution shown on the left of Figure 5.10. This distribution indicates that the decision maker feels that both experts are likely to be surprised a large fraction of times.

Now for comparative purpose suppose that Expert 1 is very knowledgeable about t and that Expert 2 is not so well informed. Their two priors are depicted in Figure 5.11 with cumulatives depicted to the right in Figure 5.10. Figure 5.10 also shows the calibration function for each expert scaled down by a factor of two for ease of presentation. Finally, referring back to Figure 5.11, each calibrated prior is drawn by multiplying by the calibration function and normalizing.

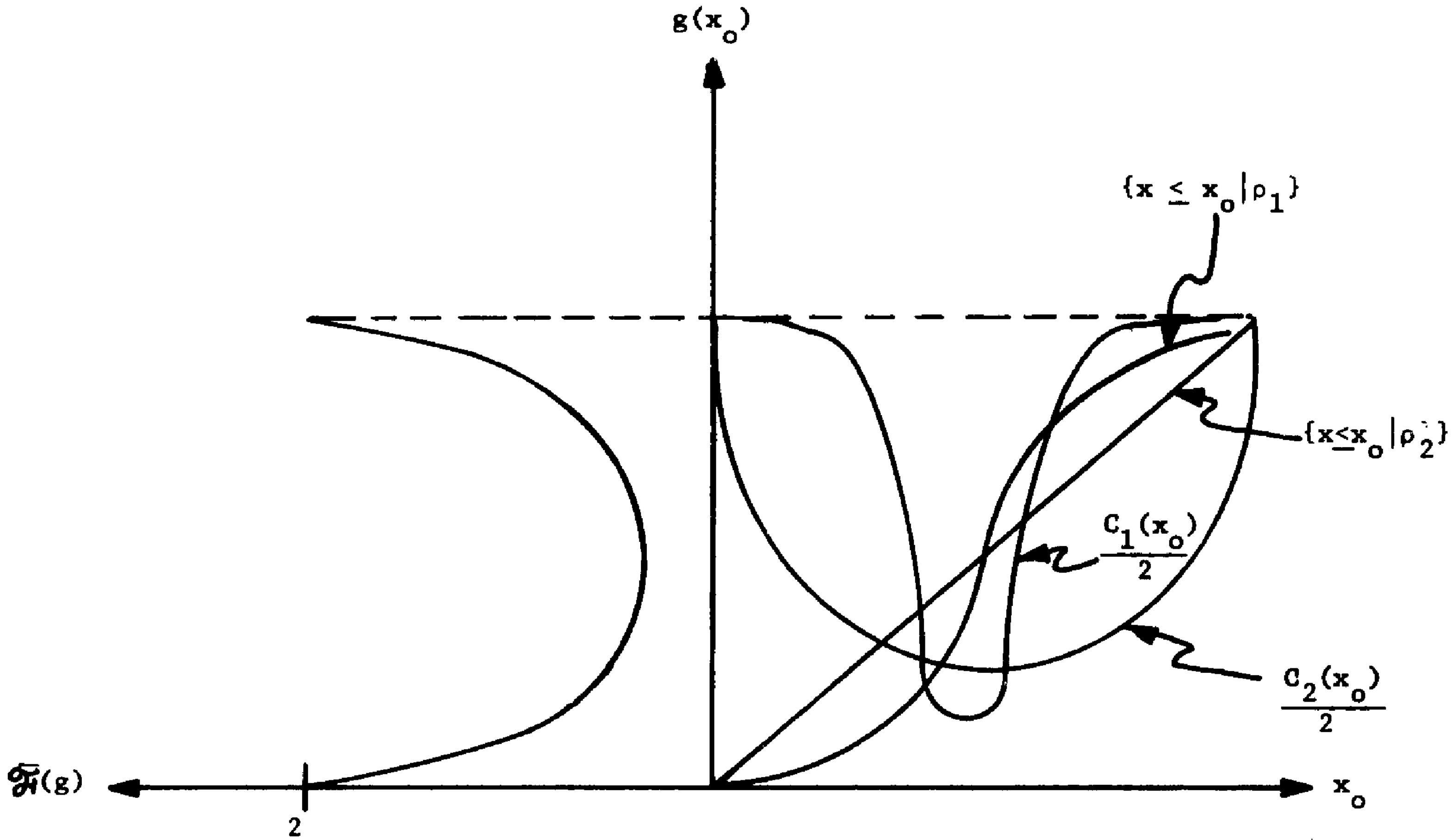


Figure 5.10 Calibration Calculation

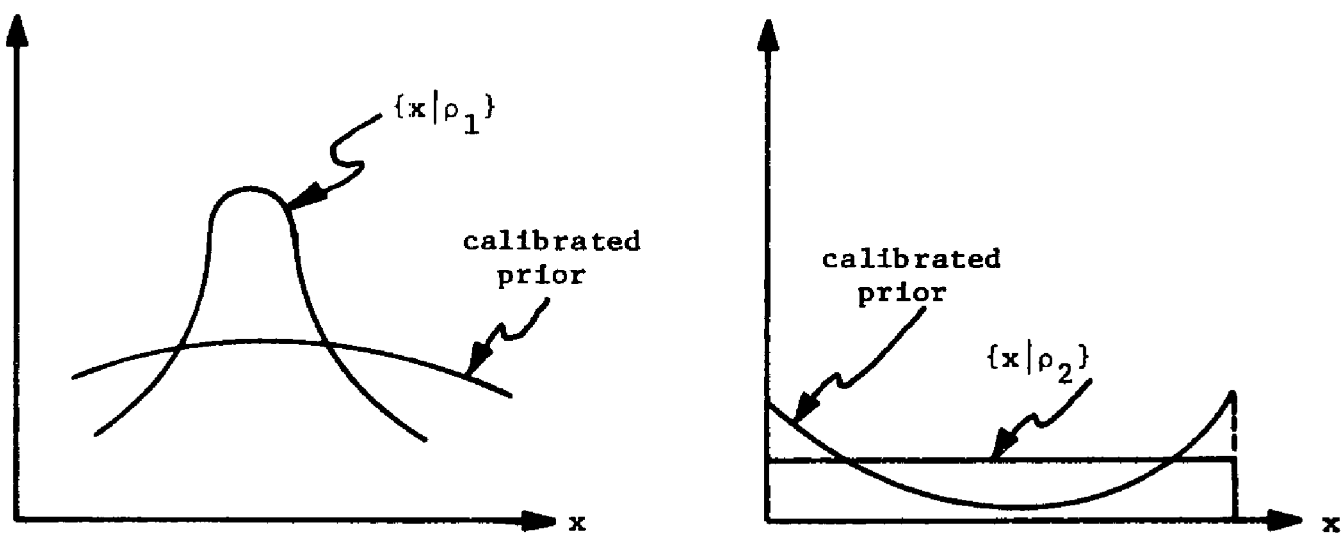


Figure 5.11 Experts' Original and Calibrated Priors

It is seen that the effect of the calibration function is to make the narrow prior more diffuse and the diffuse prior more narrow. Also interesting is that if the decision maker treats the experts as independent, his resulting posterior is approximately equal to his own prior. The experts, in essence, cancel each other out.

Example 5.2--The Forecasting Problem

A well studied case in the literature concerns the use of experts in weather forecasting [17,19]. Basically, the problem is to determine the probability of rain given that the weatherman says the probability is p.

Since the random variable is discrete (rain or not rain), application of the preceding techniques is not direct. However, we can redefine a performance distribution on the relative frequency of the number of times, given rain, that the weatherman will give probabilities in each interval. Thus

$$\int_{p_1}^{p_2} \mathcal{H}_R(p)dP \tag{5.37}$$

is defined as the long run fraction of times that the weatherman forecasts a probability of rain between p_1 and p_2 on days when it actually does rain. Similarly a function $\mathcal{H}_{\bar{R}}$, can be defined given that it did not rain.

The fundamental exchangeability assumption now may be stated: given a sequence of rainy days and the fraction of those days on which the weatherman assessed p in the interval $[p_i, p_j]$ (where p_i and p_j are arbitrary probabilities such that $0 \leq p_i \leq p_j \leq 1$), each possible sequence is equally likely. Given exchangeability we may now write:

$$\{p_o | R, \mathfrak{X}_R, \varepsilon\} = \lim_{\Delta \to 0} \frac{\{p_o \leq p \leq p_o + \Delta | R, \mathfrak{X}_R, \varepsilon\}}{\Delta} \tag{5.38}$$

$$= \mathfrak{X}_R(p_o) \, ,$$

where p_o is the weatherman's given probability of rain. We may now compute the decision maker's posterior probability of rain given the weatherman's forecast:

$$\{R | p_o, \varepsilon\} = \frac{\{p_o | R, \varepsilon\}\{R | \varepsilon\}}{\{p_o | R, \varepsilon\}\{R | \varepsilon\} + \{p_o | R', \varepsilon\}\{R' | \varepsilon\}}$$

$$= \frac{\bar{\mathfrak{X}}_R(p_o)\{R | \varepsilon\}}{\bar{\mathfrak{X}}_R(p_o)\{R | \varepsilon\} + \bar{\mathfrak{X}}_{R'}(p_o)(1 - \{R | \varepsilon\})} \tag{5.39}$$

$\bar{\mathfrak{X}}_R$ is the functional expectation of $\mathfrak{X}_R$. As a special case suppose that the expert's expected performance on rainy days or non-rainy days is assessed to be equivalent; that is

$$\bar{\mathfrak{X}}_R(p_o) = \bar{\mathfrak{X}}_{R'}(1 - p_o) \quad , \tag{5.40}$$

and that the fraction of rainy days on which he says p_o is proportional to p_o. We shall call such an expert calibrated. In this case we may write

$$\{R|p_o,\varepsilon\} = \frac{p_o\{R|\varepsilon\}}{p_o\{R|\varepsilon\} + (1 - p_o)(1 - \{R|\varepsilon\})} \tag{5.41}$$

Contrary to previous developments, the posterior doesn't equal the calibrated expert's prior unless the decision maker's prior probability of rain is 0, 1, or 1/2. It is interesting to see that equation 5.41 is symmetric in p_o and $\{R|\varepsilon\}$. This simply affirms again in the discrete case that the decision maker should weight a calibrated expert's opinion exactly as heavily as his own.

Example 5.3--The Calibrated Expert

This example will display how the calibration function operates in converting the expert's prior to that which he would provide if self-calibrated. Assume the expert's prior is

$$\{x|\rho\} = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.42}$$

and the measured long run frequency distribution is

$$\mathfrak{X}(f) = \begin{cases} 12(f - 1/2)^2 & 0 \le f \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.43}$$

Both distributions are shown in Figure 5.12. It is most convenient to think in terms of cumulatives. First we shall consider how the expert would calibrate himself.

The point x_o corresponds to the x_o^2 fractile of the expert's prior. After observing $\mathfrak{X}$, the expert will assign a probability

$$\int^{x_o^2} \mathscr{F}(f)\,df = 4x_o^6 - 6x_o^4 + 3x_o^2 \tag{5.44}$$

to the event that x falls below x_o. Therefore, the experts self calibrated prior, denoted $\{x|\rho\}_C$, is, taking the derivative of the above expression (Figure 5.13),

$$\{x|\rho\}_C = \begin{cases} 6(4x^5 - 4x^3 + x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.45}$$

Next we pursue the problem from the decision maker's point of view and calculate the calibration function:

$$\begin{aligned} C(x_o) &= \mathscr{F}[\{x \le x_o|\rho\}] \\ &= \mathscr{F}[x_o^2] \\ &= 12(x_o^2 - 1/2)^2 \end{aligned} \tag{5.46}$$

Multiplying this by the expert's prior we have:

$$\begin{aligned} C(x)\{x|\rho\} &= 2x(12(x^2 - 1/2)^2) \\ &= 24x^5 - 24x^3 + 6x \end{aligned} \tag{5.47}$$

which is the expert's self calibrated prior.

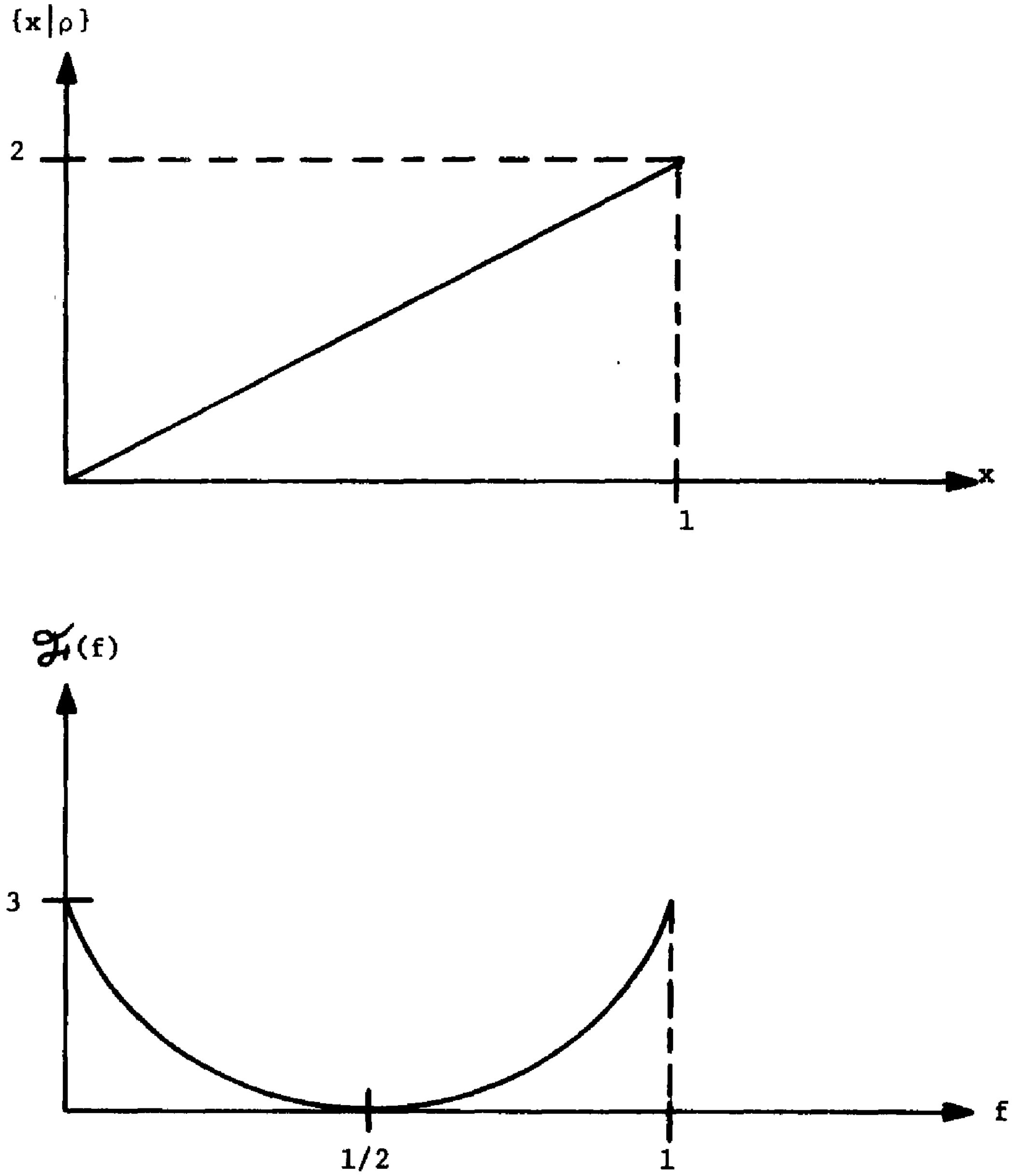


Figure 5.12 Expert's Prior and Performance Distributions

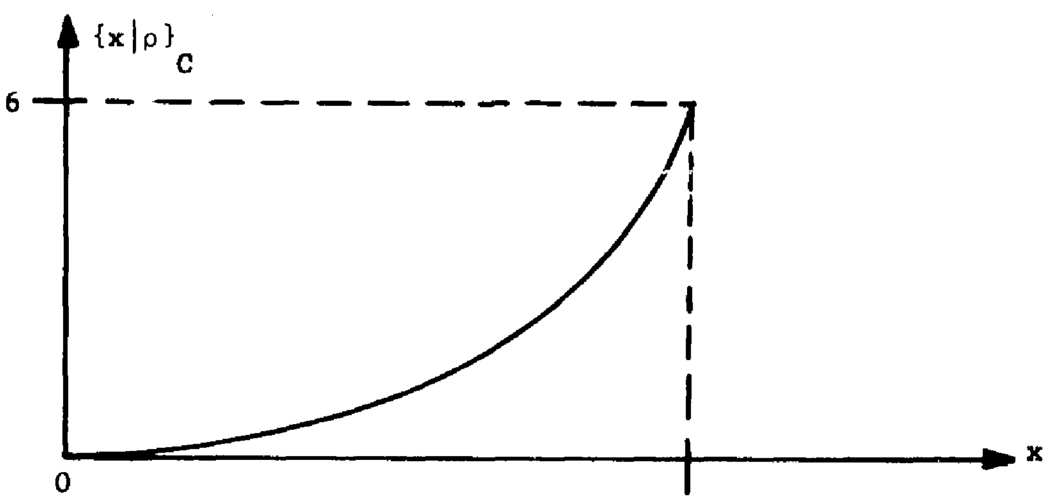


Figure 5.13 Self-calibrated Prior

Example 5.4--The Weighted Average and Natural-Conjugate Methods Revisited

The theory here presented sheds light on the two methods named above. As discussed in Chapter I, the Weighted Average Method takes as its basic premise, that the decision maker's posterior should be a linear combination of expert priors:

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_N\},\varepsilon\} = \sum_{i=1}^{N} \alpha_i\{x|\rho_i\} \tag{5.48}$$

However, even given cross calibration, equation 5.18 implies that not only are the weights not constant in general ($C_i(x)$ is a function of x), but the posterior is a weighted product of distributions including the decision maker's own prior. The two cases converge only when the decision maker's prior is diffuse and almost uniform, each expert has been calibrated, and each expert prior has its bulk in mutually exclusive regions. Even in this extreme case the framework here developed has the advantage of defining precisely what the constant weights are: a function of the decision maker's prior and the relative heights of the tails of each expert's prior.

The Natural Conjugates Method is more readily interpretable. Recall that the method fits each expert prior with a named conjugate distribution and forms the posterior from the same class by calculating the posterior parameters as a function (usually linear) of the parameters of each expert prior. Consider the case where each expert prior can be approximated by a Beta distribution as follows:

$$\{x|\rho_i\} = f_\beta(x|r_i,n_i) \tag{5.49}$$

Defining

$$R_i = r_i - 1$$
$$N_i = n_i - 2 \tag{5.50}$$

we have from equation 5.29:

$$\{x|\{x|\rho_1\},\ldots,\{x|\rho_N\},\varepsilon\} \quad = \quad k\underline{C}(x)f_\beta(x|\sum_{i=1}^{N} R_i-1, \sum_{i=1}^{N} N_i)\{x|\varepsilon\} \tag{5.51}$$

If the decision maker's prior is also approximately a Beta distribution with parameters r_o and n_o then $\{x|\varepsilon\}$ drops out of the above expression and the sums go from 0 to N. Notice that if the joint calibration function is constant, or a Beta distribution itself, then the expression reduces to a Beta distribution parameterized by a functional combination of the parameters of each expert prior. This result is similar to that of a natural conjugate analysis, except that the appropriate functions are well defined. However, no assumption need be made that each expert's experience is equivalent to observations of a Bernoulli process. Also, the decision maker's own prior information on the random variable contributes to the result as, in general, it must. Most important, the theory here developed provides the underlying assumptions that must be made in order to use the Natural Conjugate method or any other operational technique.

Example 5.5--A Sample Trajectory Analysis

Consider the case of a decision maker with a uniform prior on x,

$$\{x|\varepsilon\} = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} , \tag{5.52}$$

who consults two experts whose priors are (Figure 5.14):

$$\{x|\rho_1\} = \begin{cases} 4x^3 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.53}$$

$$\{x|\rho_2\} = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{5.54}$$

The trajectory in the $f_1 - f_2$ plane over which the performance distribution is defined is specified by the implicit equations:

$$\begin{aligned} f_1 &= \{x \le x_o|\rho_1\} = x_o^4 \\ f_2 &= \{x \le x_o|\rho_2\} = x_o^2 \end{aligned} \tag{5.55}$$

Thus, the trajectory lies on the line:

$$f_2 = f_1^2 \tag{5.56}$$

Proceeding as outlined in Section 5.5, the expected performance distribution $\mathcal{F}$ is elicited from the decision maker. We find that it can be specified on the trajectory as:

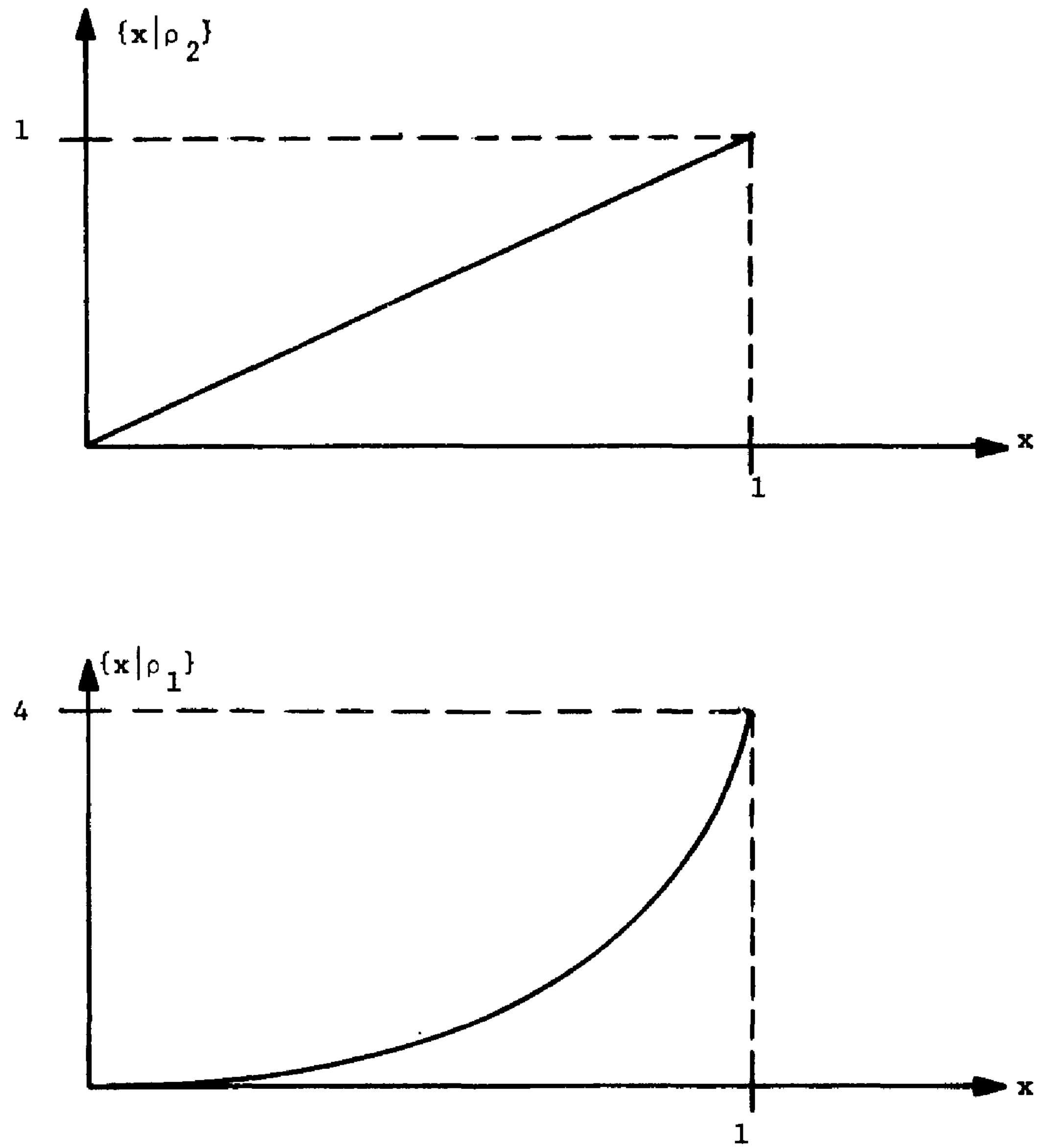


Figure 5.14 Experts' Priors

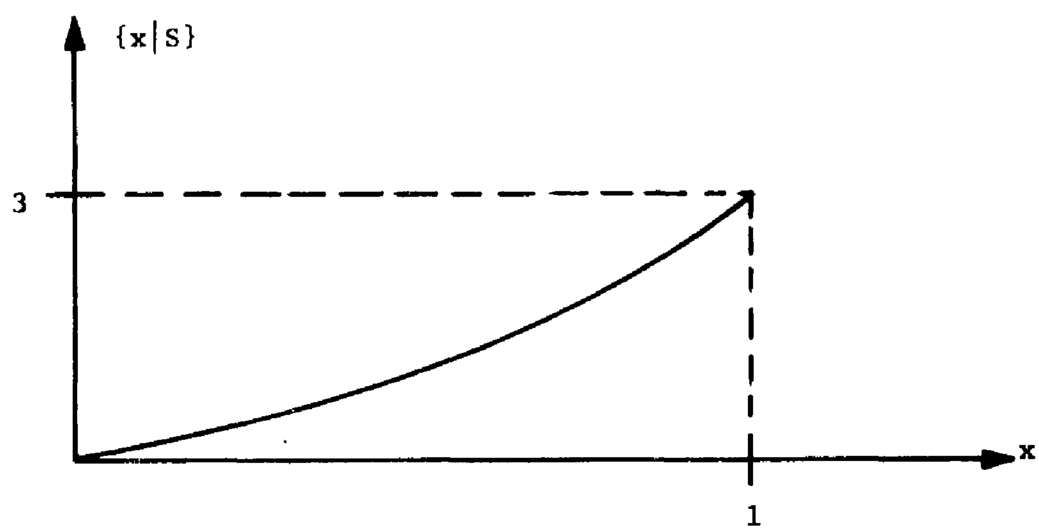


Figure 5.15 Surrogate Prior

$$\mathfrak{F}(f_1,f_2) = \frac{f_2}{f_1} \;;\; f_2 = f_1^2 \tag{5.57}$$

We can now calculate the calibration function:

$$\begin{aligned} C(x) &= \mathfrak{F}(x^4,x^2) \\ &= \frac{1}{x^2} \qquad\qquad 0 \le x \le 1 \end{aligned} \tag{5.58}$$

The surrogate prior is then readily computed as (Figure 5.15):

$$\{x|S\} = \frac{C(x)\{x|\rho_1\}\{x|\rho_2\}}{\int_x \text{numerator}} \tag{5.59}$$

$$= \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Notice that the surrogate prior implies that the surrogate expert has an intermediate measure of confidence in the variable as compared with the two experts he represents. Also notice that since the decision maker's prior is uniform, he effectively becomes the surrogate expert: his posterior state of knowledge leads to the same probability assessment.

5.7 Implications for Group Decision Making: Conditions for Consensus

The preceding developments provide interesting implications for the normative behavior of a group of decision makers or experts. We first study conditions under which a group of experts should reach a consensus.

Consensus

Let us postulate a different sort of environment than the decision maker-expert structure. The new environment is one in which a group of N experts are interacting in order to each derive information from each other. We assume that the experts have interacted to the point at which they each have a prior assessment on the variable of interest, invariant to further interaction other than an exchange of priors.

Denoting by $\underline{C}^i(x)$, the i^{th} expert's joint calibration function on the other experts, we write the i^{th} expert's posterior, conditional on prior exchange, as:

$$k\, \underline{C}^i(x) \prod_{i=1}^{N} \{x|\rho_i\} \tag{5.60}$$

A necessary and sufficient condition for consensus is clearly that the function $\underline{C}^i(x)$ is independent of the superscript.

In the case where each expert assesses cross-calibration, $\underline{C}^i$ may be written, for every i, as:

$$\underline{C}^i(x) = \prod_{j \neq i} C^i_j(x) \quad , \tag{5.61}$$

where C^i_j is the i^{th} expert's appraised calibration function for the j^{th} expert. This creates a natural condition for consensus. If each expert performs a number of assessments, exchangeable to all, C^i_j reduces to a deterministic function. Since each expert becomes self-calibrated in the process, C^i_j times the j^{th} expert's old prior equals his new prior. In other words, if everyone agrees that everyone else is calibrated, then the posterior of each expert must be the normalized product:

$$k\{x|\rho_1\}\{x|\rho_2\} \ldots \{x|\rho_N\} \tag{5.62}$$

Thus, when the experts are independent by mutual assessment of each other, calibration provides the normative condition for consensus: a condition under which expert disagreement is inconsistent with the basic axioms of probability theory.

When the experts are not cross-calibrated by joint assessment the conditions for consensus are much more demanding. First, even if the joint performance of all experts is empirically observed, there is no guarantee that the functional form of $\underline{C}^i$ is independent of i. For this to be the case each expert must have the same assessment performance relative to the rest of the group. Second, even if the form of the performance distribution is invariant to the subgroup measured, a difference in expert priors implies different trajectories. Therefore $\underline{C}^i(x)$ might not be independent of i, even if the corresponding performance distribution was.

In general, when experts are dependent by mutual assessment, conditions under which consensus should be reached are extremely restrictive. In the typical case, when exchange of data does not create agreement, methods concerned with bringing a group of experts to unanimity are aimed at an inconsistent goal.

Implementation of experts: conditions for agreement

A more important result from our point of view relates to the situation in which a group of decision makers are considering the use of one or more experts. In this situation agreement may easily be reached on how to use the experts.

Corresponding to the i^{th} decision maker is a surrogate prior $\{x|S_i\}$ representing his view of the group of experts as a whole. If a group of decision makers agree on the expected performance distribution $\tilde{\mathfrak{F}}$ (or if it has been measured) then each decision maker's surrogate prior will be identical; i.e., his surrogate expert will be the same as the other surrogate experts.

Furthermore, in many decision situations each decision maker's prior will be roughly uniform relative to the surrogate prior. When this is the case, not only will the decision makers agree on how to use a panel of experts, but will agree on posterior probability assessments.

The above ideas have powerful implications for group policy making under uncertainty. Corresponding to any uncertain variable, many experts can be questioned. If these experts are then mutually calibrated, a non-controversial surrogate prior can be formed. The decision makers need have no interaction with the assessment on calibration process, yet the surrogate prior forms a theoretically consistent way to summarize the composite of all relevant opinion on a given topic.

5.8 Summary

The contribution of this chapter is a practical methodology for the use of experts. A group of experts' prior probability distributions are processed in two steps:

1) A surrogate expert prior is formed by multiplying the product of the expert priors by the assessed joint calibration function and normalizing.

2) The decision maker's posterior is calculated as the normalized product of his prior and the surrogate prior.

In the single expert case the surrogate prior is the distribution that the decision maker would expect the expert to submit if the expert was self-calibrated. In the multi-expert case the surrogate prior represents the calibrated prior of a surrogate expert, whose state of information represents the combined knowledge of the group of experts.

The assessment of experts was reduced to an assessment of calibration functions. The specification of a calibration function was then facilitated by demonstrating that they can be derived from the result of a series of assessed expectations, whose meaning is intuitive and well defined. An important further simplification was demonstrated in the many expert case. We showed that the joint calibration function specifies a trajectory defining all the necessary information needed about the expected performance distribution. Thus, the assessment of this distribution over a many dimensional space is reduced to an assessment along a path.

The results have important implications for group decision making problems. We demonstrated by use of the surrogate prior concept that, in many cases, the use of experts can and should produce agreement among a group of decision makers concerning probabilities of uncertain outcomes. Additionally, we displayed conditions under which a group of experts should achieve a consensus.

Chapter VI

MODELING EXPERTS

6.0 Introduction

Our research concludes with the detailed study of three important special topics in expert resolution. The first analysis is of expert bias. We demonstrate an approach that isolates important characteristics and provides physical insight.

The next topic is the use of experts in experimental situations. Through the use of exchangeability, an expert's likelihood function is viewed as providing information about which model accurately describes the production of data in the world. An interesting special case is the subject of prior exchange (expert-expert or decision maker-expert).

Finally, we study the relationship between models and experts. By treating a model conceptually like an expert, insights are gained into aspects of model use important in their own right.

6.1 Expert Bias

In this section we address the problem of how to handle expert biases. Historically this has been a matter of great interest to those who have studied the use of experts. In fact, a considerable amount of research has been done on how to remove expert bias [17,20]. This focus can be explained in large part by the classical practice of using the expert's prior as the decision maker's posterior. We shall view the problem in a Bayesian framework and, in so doing, develop methods for using experts in spite of their biases.

Definition of bias

Although we shall use a specific definition of expert bias it will become clear that the model can be generalized to many other situations. We shall define an expert to be biased if, when submitting his prior to the decision maker, he misrepresents his current state of knowledge; he does not submit his true prior. The only difficulty in the above definition rests with the meaning of "true prior." For our purposes we should like to define a true prior as one invariant to any reward structure surrounding the expert. Thus, we can define the expert's true prior as the one he would use for his own decision making.

The above definition allows an expert to be biased in many ways. A thorough study of the reasons for an expert to be biased would divert us into the realm of behavioral science; however, a few short remarks are in order. First it is clear from the definition that the reason for an expert to be biased is the perceived reward structure attached to the form of the expert's probability assignment. This reward structure may take many forms. For instance, the expert may be economically affected by the decision maker's decision. In this case, aside from the moral costs of lying, it is clearly of benefit to the expert to alter his true prior before submitting it to the decision maker. Other reward structures may have nothing to do with the physical allocation of resources. The expert may only be concerned with his own prestige. For example, the expert might feel that a broad prior reflects unfavorably upon his degree of expertise and therefore may submit a prior overstating his true knowledge. An expert operating in a group situation may feel group pressure and alter his prior due to fear of being different. There are many other examples of situations when it is reasonable to suppose that an expert may be biased.

Uncertainty in Expert Bias

The crucial concept relevant to the study of expert bias is that an expert's bias is of concern only when there is some uncertainty about what the bias is. If the decision maker knows with certainty the relationship of the expert's stated prior to his true prior, he can derive the true prior directly and there is no problem. Therefore we cannot say, in general, that the degree of an expert's bias determines the weight with which his opinions should be implemented. It is the decision maker's own state of knowledge about the expert's bias that is decisive.

For notational purposes the following two quantities are defined:

$\{x|\rho\}_t$ = the representation of the expert's true state of knowledge about x--his true prior.

$\{x|\rho\}_s$ = the expert's stated prior--that which he submits to the decision maker.

The decision maker's problem, given the expert's advice, is to calculate:

$$\{x|\{x|\rho\}_s,\varepsilon\} = k\{\{x|\rho\}_s|x,\varepsilon\}\{x|\varepsilon\} \tag{6.1}$$

Notice that the critical term $\{\{x|\rho\}_s|x,\varepsilon\}$ could theoretically be assessed directly by the decision maker. Our purpose is to structure the problem so that the expert's bias can be explicitly appraised and taken account of in an operational way. Toward this end we define a bias vector Δ which contains all the information needed to calculate the expert's true prior from his stated prior. In general, Δ may require a very complex specification. For example, Δ may be defined as the difference:

$$\Delta = \{x|\rho\}_s - \{x|\rho\}_t \quad , \tag{6.2}$$

in which case Δ is a function. A major task will be to characterize Δ in a straightforward, meaningful way.

To begin, we expand the posterior over Δ to obtain:

$$\{x|\{x|\rho\}_s,\varepsilon\} = \int_\Delta \{x|\{x|\rho\}_s,\Delta,\varepsilon\}\{\Delta|\{x|\rho\}_s,\varepsilon\} \tag{6.3}$$

The factor $\{\Delta|\{x|\rho\}_s,\varepsilon\}$ is the assessment of the expert's bias. For notational convenience we shall assume that this assessment is independent of $\{x|\rho\}_s$. If it is not, the later results will be structurally unaffected by reintroduction of $\{x|\rho\}_s$ behind the conditioning bar.

The information $\{x|\rho\}_s$ and Δ provide the expert's actual prior. We define $F_\Delta(x)$ as the expert's true prior implied by the bias vector Δ, where it is assumed that $\{x|\rho\}_s$ is known. Equation 6.3 may now be written:

$$\begin{aligned}\{x|\{x|\rho\}_s,\varepsilon\} &= \int_\Delta \{x|F_\Delta(x),\Delta,\varepsilon\}\{\Delta|\varepsilon\} \\ &= \{x|\varepsilon\}\int_\Delta \frac{\{F_\Delta(x)|x,\Delta,\varepsilon\}\{\Delta|\varepsilon\}}{k(\Delta)}\end{aligned} \tag{6.4}$$

where the assessment of x is assumed independent of Δ alone, and $k(\Delta)$ is a function of Δ but not x. Specifically

$$k(\Delta) = \int_x \{F_\Delta(x)|\Delta,x,\varepsilon\}\{x|\varepsilon\} \tag{6.5}$$

so that $k(\Delta)$ can always be determined from other assessments by simple integration. $\{F_\Delta(x)|x,\Delta,\varepsilon\}$ is the assessment of the expert's true prior; exactly the same assessment that has been discussed in previous chapters.

Conceptually we are finished; however, as we shall see, fundamental practical problems remain. Recall that the calibration function $C(x_o)$ is the composite function $\mathfrak{F}(\{x \leq x_o|\rho\}_t)$, whose form depends on the shape of the expert's prior. We therefore define $C_\Delta(x)$ as the calibration function implied by $F_\Delta(x)$. This allows us to write, using the results in Chapter V,

$$\{x|\{x|\rho\}_s,\varepsilon\} = \{x|\varepsilon\} \int_\Delta \frac{C_\Delta(x)F_\Delta(x)}{k(\Delta)} \{\Delta|\varepsilon\} \tag{6.6}$$

The above equation implies that for every Δ a new calibration function must be computed. Aside from the obvious computational complication involved, the equation denies the application of straightforward approximation techniques. Suppose, for example, that the decision maker's prior is relatively uniform over a wide range, so that

$$\{x|\{x|\rho\}_s,\varepsilon\} = \int_\Delta C_\Delta(x)F_\Delta(x)\{\Delta|\varepsilon\} \tag{6.7}$$

where $k(\Delta)$ equals one since $C_\Delta(x)F_\Delta(x)$ is a density function integrating to one over x independently of Δ. We compute the posterior mean by integrating over x:

$$\langle x | \{x|\rho\}_s, \varepsilon \rangle = \int_\Delta \left[\int_x x \; C_\Delta(x) F_\Delta(x) \right] \{\Delta|\varepsilon\} \tag{6.8}$$

Notice that the first term in the Δ integrand is the expectation of $C_\Delta(x)F_\Delta(x)$. However this expectation is generally a very complex function of Δ, whose integral (over Δ) is a nonspecific quantity varying from problem to problem. Therefore, no physical insight can be gained without a detailed mathematical investigation. The following analysis will decompose elements of the problem and in so doing suggest practical ways in which Δ should be defined.

Change of coordinates

It is fruitful to examine the cumulative distributions of the expert's true prior and his stated prior. An interesting way to view their relationship is through a function $h(x_o)$ defined as follows:

$$\{x \leq x_o|\rho\}_t = \{x \leq h(x_o)|\rho\}_s \tag{6.9}$$

It is clear that for any two cumulatives a function h may be found satisfying the above equation. For example, consider the calculation of h(a) depicted in Figure 6.1. In this case h(a) equals b since

$$\{x \leq b|\rho\}_s = \{x \leq a|\rho\}_t \tag{6.10}$$

For each Δ we define $h_\Delta(x_o)$ as the function relating the stated cumulative to the true cumulative. The true prior is then determined by differentiation as

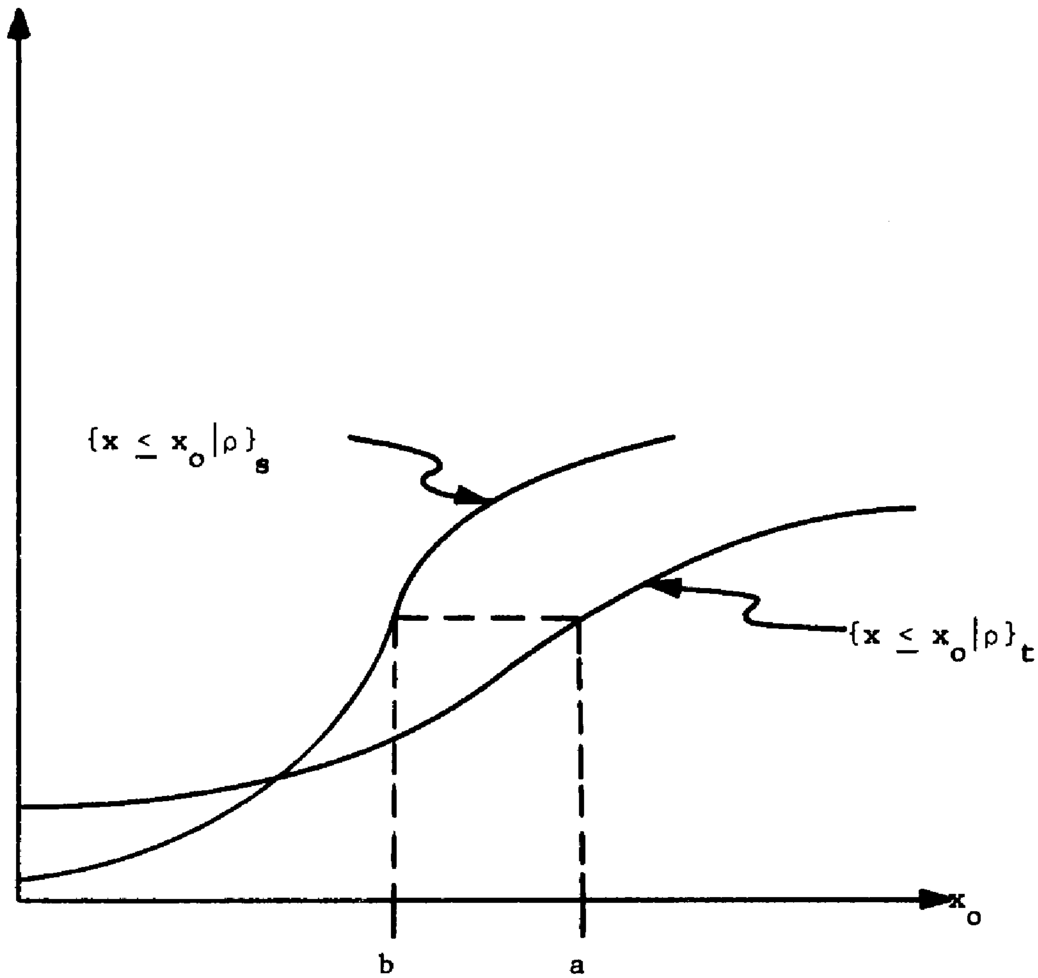


Figure 6.1 Calculation of h

$$\frac{d}{dx_o}\{x \leq h_\Delta(x_o)|\rho\}_s = \frac{d}{dx_o}\int_{-\infty}^{h_\Delta(x_o)} \{x|\rho\}_s$$

$$= h'_\Delta(x_o)\ F_s(h_\Delta(x_o)) \tag{6.11}$$

where F_s is, for convenience, another symbol for the expert's stated prior.

The above manipulations are useful because now we may write:

$$\begin{aligned} C_\Delta(x)F_\Delta(x) &= \mathfrak{F}[g_\Delta(x)]h'_\Delta(x)F_s(h_\Delta(x)) \\ &= h'_\Delta(x)\mathfrak{F}[g_s(h_\Delta(x))]F_s(h_\Delta(x)) \end{aligned} \tag{6.12}$$

where g_Δ and g_s are the expert's true and stated prior cumulatives and where the functional forms of both $\mathfrak{F}$ and F_s are, on the right-hand side of the equation, independent of Δ. We define a function G as follows:

$$G(x) = \mathfrak{F}[g_s(x)]F_s(x) \tag{6.13}$$

The major advance may now be articulated: G(x) is the result of calibrating $\{x|\rho\}_s$ with $\mathfrak{F}$ as if it were the true prior. The calibration procedure need only be done once, whereas before (recall equation 6.6) it had to be applied continuously as a function of Δ. Other benefits of this approach will soon become apparent.

Summarizing our results thus far we have

$$\{x|\{x|\rho\}_s,\varepsilon\} = \{x|\varepsilon\}\int_\Delta \frac{G(h_\Delta(x))h'_\Delta(x)}{k(\Delta)}\{\Delta|\varepsilon\} \tag{6.14}$$

The remaining matter of interest is the function h_Δ. Depending on how the bias Δ is defined, h will be more or less amenable to analysis. Below we develop an operationally powerful way to characterize h and Δ.

Characterization of bias

It is useful to differentiate between bias in location and bias in shape. For convenience, the location of a distribution will be specified by its mean. Denoting the mean of the stated prior by $\overline{x}$, the parameter α is defined such that $\alpha\overline{x}$ is the mean of the true prior.

The hard quantity to specify is the shape of a distribution. For our purposes variance is not a good measure for two reasons:

1. The expert's stated prior may be impossible to parameterize by mean and variance only.
2. The functional form of h generally depends in an extremely complex way on the variance.

The key is to investigate different forms of h_Δ directly. An especially interesting and simple characterization of h_Δ is the linear form βx. Note in Figure 6.2 the various cumulatives $\{x \le \beta x_o | \rho\}$ as a function of β. It is seen that as β decreases the resulting cumulatives represent smoothly widening priors with location shifts.

With this partial motivation we define a new function

$$h_\Delta(x) = \beta x + (1 - \alpha\beta)\overline{x} \tag{6.15}$$

Rather than trace through the derivation of the above equation we state its basic properties. Suppose we are given the distribution $F_s(x)$. The distribution F_Δ implied by the definition of h (see again equation 6.11) is

$$\begin{aligned} F_\Delta(x) &= F_s(h(x))h'(x) \\ &= \beta F_s(\beta x + (1 - \alpha\beta)\overline{x}) \end{aligned} \tag{6.16}$$

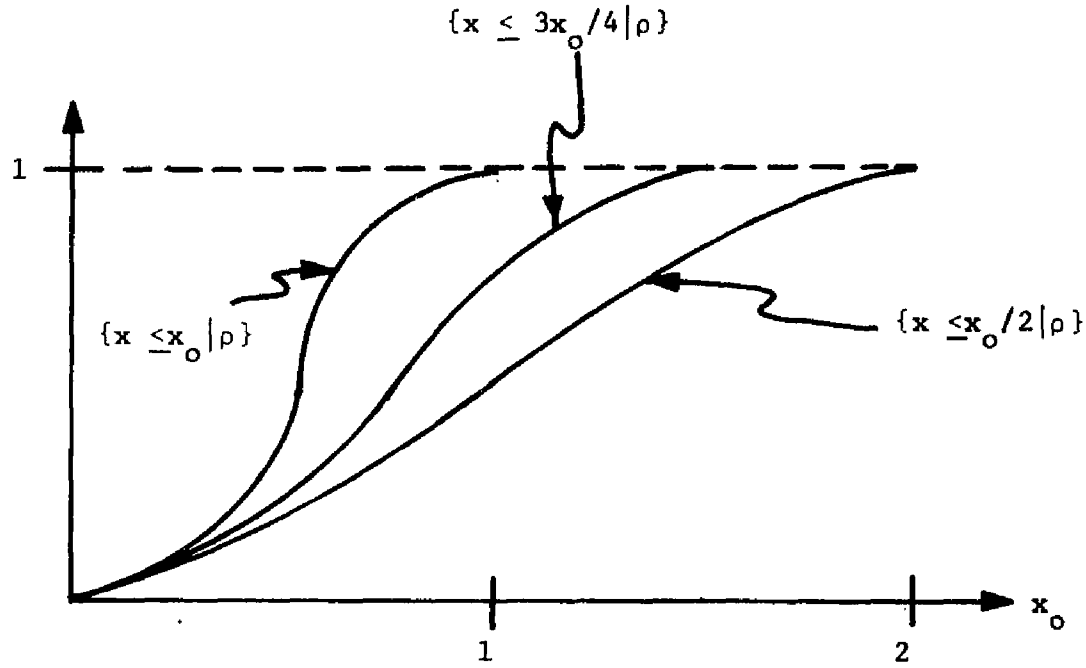


Figure 6.2 Cumulatives

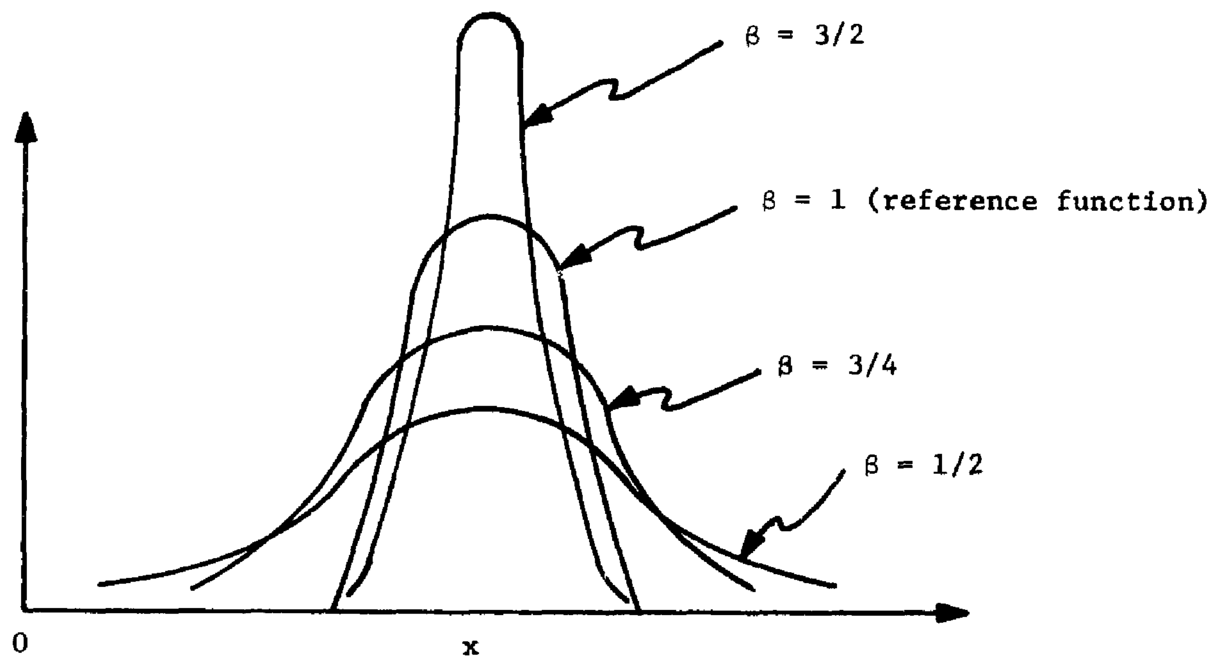


Figure 6.3 Scaling Effect of β

The mean of this new distribution can easily be determined to be $\alpha\bar{x}$. Also, if $F_s(x)$ is symmetric, so that

$$F_s(\bar{x} + x) = F_s(\bar{x} - x) \tag{6.17}$$

we have

$$\begin{aligned} F_\Delta(\alpha\bar{x} + x) &= \beta F_s(\bar{x} + \beta x) \\ &= \beta F_s(\bar{x} - \beta x) \\ &= F_\Delta(\alpha\bar{x} - x) \end{aligned} \tag{6.18}$$

proving that F_o is also symmetric about its mean $\alpha\bar{x}$. Therefore, we may view α as a location parameter and β as a scale parameter.

The transformation defined in equation 6.15 provides a rich family of different shapes and locations from a given reference distribution. For α equal to one (corresponding to no translational bias), a family of curves is **partially** sketched in Figure 6.3 for different β's.

It is natural to assume that, however the expert biases his true prior, it may be fitted by proper choice of α and β. With this assumption we may write

$$\begin{aligned} \{x|\{x|\rho\}_s,\varepsilon\} &= \{x|\varepsilon\}\int_{\alpha,\beta} \beta\,\frac{G(h(x))}{k(\alpha,\beta)}\{\alpha,\beta|\varepsilon\} \\ h(x) &= \beta x + (1 - \alpha\beta)\bar{x} \end{aligned} \tag{6.19}$$

where

$$k(\alpha,\beta) = \int_x \beta G(h(x))\{x|\varepsilon\} \tag{6.20}$$

Approximate techniques

It is worthwhile to calculate the important moments of the posterior since the integration in equation 6.19 may be difficult, even though the integrand is easily specified. Our major assumption will be that the decision maker's prior is approximately uniform relative to the expert's prior (however biased). Such a case is typical in expert use situations. This allows us to calculate:

$$\langle x|\{x|\rho\}_s,\varepsilon\rangle = \int_{\alpha,\beta x}\int \beta x \; G(h(x))\{\alpha,\beta|\varepsilon\} \tag{6.21}$$

where again $k(\alpha,\beta)$ equals one, since G is a density function for all α and β.

By changing variables via the formula

$$y = h(x) \tag{6.22}$$

the integral with respect to x may be written

$$\int_y \left(\frac{y - (1 - \alpha\beta)\bar{x}}{\beta}\right) G(y) = \frac{\langle x|c\rangle - \bar{x}}{\beta} + \alpha\bar{x} \tag{6.23}$$

where $\langle x|c\rangle$ is the mean of the expert's calibrated stated prior. In the usual case, the calibration function and the expert's stated prior are both symmetric so that $\langle x|c\rangle$ equals $\bar{x}$. We shall for convenience (and no real loss of generality) assume this to be the case, and further expand equation 6.21:

$$<x|\{x|\rho\}_s,\varepsilon> = \int_{\alpha,\beta} \alpha <x|c>\{\alpha,\beta|\varepsilon\}$$

$$= <\alpha|\varepsilon><x|c> \tag{6.24}$$

This important practical result states that the posterior mean of x, given a (possibly) biased expert prior, is the product of the expected locational bias and the mean of the distribution obtained by calibrating the expert's stated prior as if it were not biased.

We may also calculate the posterior variance as

$$^{v}<x|\{x|\rho\}_s,\varepsilon> = \int_{\alpha,\beta}\int_{x} \beta\Big(x - <\alpha|\varepsilon><x|c>\Big)^2 G(h(x))\{\alpha,\beta|\varepsilon\} \tag{6.25}$$

Again, through change of variables, the inside integral may be written

$$\int_y \left(\frac{y - (1 - \alpha\beta)\overline{x} - \beta\overline{\alpha}\ \overline{x}}{\beta}\right)^2 G(y) \quad \overline{\alpha} = <\alpha|\varepsilon>,\ \overline{x} = <x|c>$$

$$= \int_y \left(\frac{(y - \overline{x}) + \beta\overline{x}(\alpha - \overline{\alpha})}{\beta}\right)^2 G(y) \tag{6.26}$$

$$= \int_y \left(\frac{(y - \overline{x})^2}{\beta^2} + \overline{x}^2(\alpha - \overline{\alpha})^2 + 2\beta\overline{x}(y - \overline{x})(\alpha - \overline{\alpha})\right) G(y)$$

$$= \frac{^{v}<x|c>}{\beta^2} + <x|c>^{v}<\alpha|\varepsilon>$$

Substituting back into equation 6.25 we have:

$$ {}^{v}\langle x|\{x|\rho\}_s,\varepsilon\rangle = \langle\beta^{-2}|\varepsilon\rangle {}^{v}\langle x|c\rangle + \langle x|c\rangle^2 \, {}^{v}\langle\alpha|\varepsilon\rangle \ , \tag{6.27} $$

which provides a convenient and practical formula. The term ${}^{v}\langle x|c\rangle$ is of course the posterior variance width assuming no uncertainty about the expert's bias. Viewed as a function of ${}^{v}\langle x|c\rangle$, the above expression shows explicitly the way in which uncertainty in expert bias affects posterior uncertainty about the random variable of interest.

Summarizing, we have displayed a convenient way to handle expert bias. The crucial step was to restructure the problem, focusing attention on the cumulatives of the expert's stated and true priors. This enabled us to characterize bias in a computationally convenient way, while providing a framework where immediate physical insight can be gained by observing certain characteristics of the calibration function and the expert's stated prior.

6.2 Experimentation and the Use of Experts

It is common practice in decision making situations for the decision maker to relegate the responsibility for experimentation relative to crucial state variables to his experts on those variables. Typically, the experts are not only asked to design and conduct the experiments, but are also totally depended upon to make inferences based on the experimental results. For the one expert situation where the decision maker is presumed to use the expert's probability assessment directly the procedure seems valid. However, the results thus far deem it important to treat the use of experiments in a more basic way. The one expert case will be developed since the corresponding many expert treatment is analogous.

Suppose that the decision maker obtains some experimental data y in addition to the advice of his expert. We determine his conditional probability function as follows:

$$\{x|y,\{x|\rho\},\varepsilon\} \;=\; k\{y|x,\{x|\rho\},\varepsilon\}\{\{x|\rho\}|x,\varepsilon\}\{x|\varepsilon\} \tag{6.28}$$

In general it is possible for the expert to misstate or misinterpret the experimental results. Here we shall assume a well defined experiment with impeccable data, in no way affected by the expert, so that

$$\{y|x,\{x|\rho\},\varepsilon\} \;=\; \{y|x,\varepsilon\} \tag{6.29}$$

Thus, with no further interaction with his expert the decision maker can ignore him in making inferences about experiments.

Next consider the situation where the expert is also allowed to update his own prior conditional on the experimental data. If the expert's likelihood function $\{y|x,\rho\}$ is known, the decision maker can easily obtain the expert's posterior himself. However, if the likelihood function is unknown, it can be shown that the expert's given posterior is not sufficient to calculate it. Thus, the expert's likelihood function is always at least as much value to the decision maker, relative to an experiment, as the expert's posterior. It is therefore clearly prudent to use an expert only as an assessor rather than a processor in relation to a given experiment.

We shall henceforth assume that the expert gives the decision maker his complete likelihood function on the data y. In addition, it will be assumed that the expert's prior $\{x|\rho\}$ in no way affects, in the decision maker's mind, how the expert models the experiment generating y. Thus, with no loss of generality, the object of our further analysis is

$$\{x_o|y_o,\{y|x,\rho\},\varepsilon\} \quad , \tag{6.30}$$

where subscripts have been introduced to highlight the functional character of this expression. The above function may be expanded and manipulated to give

$$\{x_o|y_o,\{y|x,\rho\},\varepsilon\} = k\{\{y|x,\rho\}|x_o,y_o,\varepsilon\}\{y_o|x_o,\varepsilon\}\{x_o|\varepsilon\} \tag{6.31}$$

The term $\{\{y|x,\rho\}|x_o,y_o,\varepsilon\}$ is, intuitively, an assessment of how well the expert's experiment model can duplicate the functioning of the real world which produced the data y_o from a state defined by x_o. The decision maker must assess how the expert's likelihood function depends on the actual physical characteristics of the system he is modeling; he must assess the expert's modeling expertise. Unfortunately the assessment implied by equation 6.31 is very complex and hard to conceptualize. The decision maker must not only assess how well the expert's model will predict y_o, given the actual revealed value of x, but must assess what the model will predict given other arbitrary values of x. To understand the situation in a more explicit way we must turn to a deeper modeling. The most general case will first be presented followed by an intuitive interpretation of the results.

A more detailed modeling

Our approach will be to conceptualize a conditionally exchangeable set of random variables. More precisely we define a set $S(x)$ of random variables which are exchangeable with the data variable y, given that x is known. This entitles us to further define the function $w(y|x)$ as the long run frequency distribution of the variables in the set $S(x)$. Although the set $S(x)$ depends on the assessor's state of information, the function $w(y|x)$ is theoretically well defined and conceptually can be measured. A rough way of viewing $w(y|x)$, legitimitized by de Finetti's results [7], is as the specification of an underlying process parameterized by x. This will become clearer shortly.

Next we visualize $w(y|x_o)$ as a function of y given the single value x_o, and expand over all such functions:

$$\{\{y|x,\rho\}|x_o,y_o,\varepsilon\} = \int_{w(y|x_o)} \{\{y|x,\rho\}|x_o,y_o,w(y|x_o),\varepsilon\}\{w(y|x_o)|x_o,y_o \tag{6.32}$$

The first term in the integrand may be rewritten as

$$\frac{\{y_o|w(y|x_o),\{y|x,\rho\},x_o,\varepsilon\}}{\{y_o|w(y|x_o),x_o,\varepsilon\}} \{\{y|x,\rho\}|x_o,w(y|x_o),\varepsilon\} \tag{6.33}$$

from which it may be deduced that the original term is independent of y_o, since by exchangeability $w(y|x_o)$ completely specifies the distribution of y_o and reduces the fraction in the expression to unity. The second term in the integrand may, in standard fashion, be written

$$\{w(y|x_o)|x_o,y_o,\varepsilon\} = \frac{\{y_o|x_o,w(y|x_o),\varepsilon\}\{w(y|x_o)|x_o,\varepsilon\}}{\{y_o|x_o,\varepsilon\}} \tag{6.34}$$

Combining our results thus far into equation 6.31 we obtain:

$$\{x_o|y_o,\{y|x,\varepsilon\},\varepsilon\} = k\int_{w(y|x_o)} \{y_o|w(y|x_o),\varepsilon\}\{\{y|x,\rho\}|w(y|x_o),\varepsilon\}\{w(y|x_o)|x_o,\varepsilon\}\{x_o|\varepsilon\} \tag{6.35}$$

For brevity, x_o is omitted from the conditioning information in the first two terms since it is specified in $w(y|x_o)$.

We have, by exchangeability, that

$$\{y_o|w(y|x_o),\varepsilon\} = w(y_o|x_o) \tag{6.36}$$

Interpreting the discrete case, the probability of the value y_o is the long run frequency of occurrences of this value in an exchangeable set, which is just $w(y_o|x_o)$. To understand the other terms it is useful to view the situation in a more explicit modeling context.

Metamodel approach

The distribution specified by the function $w(y|x_o)$ may, in many cases, be regarded as that implied by a given stochastic model with parameter x_o. (In general x_o may be a vector.) The set of all stochastic models generated by each possible value of w is, following Smallwood [22], termed the metamodel. The data may then be conceptually treated as having been produced by one of the models in the metamodel. Therefore we may write, in the discrete case,

$$\{w(y|x_o)|x_o,\varepsilon\} = \{M_k|x_o,\varepsilon\} \quad k = 1,\ldots,N \tag{6.37}$$

where $\{M_k|x_o,\varepsilon\}$ is the probability that model k actually produced the data, or, more accurately, that model k implies the same probability distribution as $w(y|x_o)$. Analogously we may write

$$\{y|w(y|x_o),\varepsilon\} = \{y|M_k,x_o,\varepsilon\} \ , \tag{6.38}$$

where $\{y|M_k,x_o,\varepsilon\}$ is the likelihood function defined by model k. Important to recognize is that equation 6.38 is only valid given the exchangeability assumption.

Next we suppose that the likelihood function specified by the expert can be characterized by M_e, one of the models in the metamodel. It is then clear that the assessment $\{\{y|x,\rho\}|w(y|x_o),\varepsilon\}$ is equivalent to $\{M_e|M_k,\varepsilon\}$, which is the assessment of the expert's model, given that model k is actually producing the data. In this context, we have an explicit formulation of the notion of modeling expertise: the expert is a good model builder insofar as his model (likelihood function) is, in fact, the one producing the data.

The complete posterior on x may be written in the model notation as

$$\{\mathbf{x}|y,M_e,\varepsilon\} = k \sum_{i=1}^{N} \{y|M_i,x,\varepsilon\}\{M_e|M_i,\varepsilon\}\{M_i|\varepsilon\}\{x|\varepsilon\} \tag{6.39}$$

The above equation aids our intuition. Through it may be seen that the reason the expert's model is of value to the decision maker is that it allows the decision maker to update his own prior on the actual model producing the data. The equation clearly demonstrates that if the decision maker is certain of the data producing model, or if his assessment of M_e is independent of the "true" model, then the revelation of the expert's model is of no value.

Exchange of priors--the metaexpert

A specific application of interest concerns the exchange of priors between a decision maker and his experts, or between the experts themselves. The analysis of such situations is straightforward and analogous to the above treatment. The exchange of priors may be viewed as each expert receiving experimental data. The results are, as previously, that the relevant quantities are the experts' likelihood functions. The decision maker must gauge how well each expert can assess either the other experts or the decision maker himself. This allows the decision maker to learn about his or his other experts' abilities as assessors.

We call an expert who provides the decision maker with information about other assessors (possibly the decision maker himself) a metaexpert. Chapter V developed the fact that a likelihood assessment on experts can be summarized by a calibration function. Thus, the metaexpert's assessment of an expert provides the decision maker with additional knowledge about the expert's performance distribution. Similarly, the metaexpert's assessment of the decision maker provides information about the decision maker's own performance distribution. A metaexpert's value in a given situation depends on the extent to which the decision maker has calibrated his experts or himself.

6.3 Modeling and the Use of Experts

The use of experts and the use of formal models in decision making are intimately related. In this section we relate how the theory on the use of experts pertains to the difficult problem of modeling real world situations. We will not only display the proper relationship between experts and models, but will in addition use our results to provide insight into the modeling process in general.

The subject of model building has long been the object of intensive thought by theorists in many disciplines. This considerable attention is certainly justified by the importance of modeling in the solution of practical problems. In the science of decision making the focus on the problem of modeling the real world is particularly keen. In most real decision problems worthy of intensive study the applicability of the analysis rests heavily on the ability of the model maker to encapsulate reality with his model.

The use of formal numerical models for decision making, or for any other purpose, has in recent years been widely maligned. Much of the antagonism toward analytic models, particularly in the public policy domain, has been rightfully directed at the lack of realism displayed in such models. It is correctly asserted that the results obtained from invalid models are themselves invalid. Unfortunately, it is typically presumed that models must either be completely relied upon or discarded entirely in favor of expert advice. Here it will be shown that the less understood a situation, the more appropriate it is to build numerical models. We shall also display that experts and models should not be used to each other's exclusion, but should be used side by side in a mutually supportive fashion.

Uncertainty in the representation of the real world

Basically, the fundamental difficulty that arises relative to the use of a model concerns uncertainty about the extent to which the model portrays the actual physical system of interest. If everyone agrees that a model is an accurate representation of reality there should be no controversy over the results derived using the model.

In the general sense, a model is conceptually exactly like an expert. Just as an expert gives the decision maker one or more probability assessments on random variables, so does a stochastic model. For example, suppose that the decision maker is concerned with the value v of a given policy. If we denote the decision maker's formal model by M we can display the model's output to the decision maker by $\{v|M\}$. Consistent with the philosophy of subjective probability the decision maker should alter his prior on v (prior to using the model) as follows:

$$\{v|\{v|M\},\varepsilon\} \quad = \quad k\{\{v|M\}|v,\varepsilon\}\{v|\varepsilon\} \qquad (6.40)$$

The term $\{v|\{v|M\},\varepsilon\}$ reduces to $\{v|M\}$ only under very special conditions. Also obviously true is that $\{v|\{v|M\},\varepsilon\}$ reduces to $\{v|\varepsilon\}$ under equally stringent conditions. Thus it has been shown with little effort that, in general, models should not be used without qualification--but they should be used!

Although equation 6.40 is a conceptual solution to how to use models, there are additional insights and practical benefits that can be derived by taking advantage of further structuring. The relationship of experts to model building is also best studied in a less broad context.

It will be useful to view the modeling situation as if the decision maker receives a deterministic model from a model maker and receives all probabilistic information concerning the model from one or more experts, possibly including the model maker. For clarity we shall always suppose that the purpose of the model is to relate a set of state variables $\underline{x} = (x_1,\ldots,x_n)$ to one outcome variable v for a given decision policy. Thus, in general, a model is a set of relationships defining a functional f as follows:

$$v = f(\underline{x}) \quad . \tag{6.41}$$

Our attention will be focused on that information derived by observation of the functional f. Other possible information relating to the logic by which the model maker computed f will be submerged into the decision maker's state of information ε, and will not be specifically analyzed. Thus, for our purposes, we shall speak of the functional and the model interchangeably.

A deterministic metamodel approach

Let us suppose that we have a collection of models, each with a set of parameters. Further suppose that one of these models with a given set of parameters is known to depict accurately the relationship of the input variable $\underline{x}$ to the output variable v. For notational simplicity we can subscript each model with a vector n which specifies a model form and a specific value for each model parameter. Let

$f_n \equiv$ the event that the n^{th} model accurately depicts the relation between $\underline{x}$ and v.

If the decision maker receives a model f_e from his model builder he will be faced with the following assessment:

$$\{v|f_e,\varepsilon\} \tag{6.42}$$

This may be expanded to give:

$$\{v|f_e,\varepsilon\} = \int_n \{v|f_e,f_n,\varepsilon\}\{f_n|f_e,\varepsilon\} \tag{6.43}$$

We can then apply Bayes' Theorem to obtain:

$$\{v|f_e,\varepsilon\} = k \int_n \{v|f_n,\varepsilon\}\{f_e|f_n,\varepsilon\}\{f_n|\varepsilon\} \quad , \tag{6.44}$$

where it is assumed that v is independent of f_e, if the "true" model is known. The term $\{f_e|f_n,\varepsilon\}$ is the probability that the model maker will give the decision maker the model f_e, given that model f_n represents physical reality. The term $\{f_n|\varepsilon\}$ is the decision maker's prior on which model actually formalizes reality. To obtain the remaining term we can further expand to write:

$$
\begin{aligned}
\{v|f_n,\varepsilon\} &= \int_{\underline{x}} \{v|\underline{x},f_n,\varepsilon\}\{\underline{x}|f_n,\varepsilon\} \\
&= \int_{\underline{x}} \{v|\underline{x},f_n,\varepsilon\}\{\underline{x}|\varepsilon\}
\end{aligned}
\tag{6.45}
$$

where we have assumed that the state variables are independent of the model used to relate them to v. Equation 6.45 describes a change of variables from $\underline{x}$ to v since v is completely determined from $\underline{x}$ and f_n. For completeness we rewrite equation 6.44 to display it in final form:

$$
\{v|f_e,\varepsilon\} = k \int_{n} \{f_e|v,f_n,\varepsilon\}\{f_n|\varepsilon\} \int_{\underline{x}} \{v|\underline{x},f_n,\varepsilon\}\{\underline{x}|\varepsilon\} \tag{6.46}
$$

The expression $\{v|\underline{x},f_n,\varepsilon\}$ displays most explicitly the possible operational shortcoming of the metamodel concept in this application. The decision maker must, in essence, exercise each possible model to evaluate this term. Depending on the complexity of the situation, this may or may not be desirable. Of course it may be appropriate to specify $\{v|\underline{x},f_n,\varepsilon\}$ as a direct subjective evaluation. If the set of possible n's is large this becomes a formidable task.

However, equation 6.46 provides a framework in which the use of experts is clear. If the decision maker receives the advice of an expert on the state variables $\underline{x}$, two factors would possibly be affected. The term $\{\underline{x}|\varepsilon\}$ can be updated to $\{\underline{x}|\{\underline{x}|\rho\},\varepsilon\}$. The term $\{f_e|v,f_n,\varepsilon\}$ would be altered by further conditioning it on $\{\underline{x}|\rho\}$, only if the decision maker feels there is a relationship between the model maker's modeling expertise and the expert's expertise in assessing $\underline{x}$. Even in the typical modeling situation where the model maker is the state variable expert we expect this to be a rare case.

Another approach

If the decision maker is unable to formulate a deterministic meta-model of practical value we can apply Bayes' theorem directly to give

$$\{v|f_e,\varepsilon\} = k\{f_e|v,\varepsilon\}\{v|\varepsilon\} \tag{6.47}$$

The term $\{v|\varepsilon\}$ is immediately recognizable as the decision maker's prior on v. Consistent with the Bayesian philosophy, this assessment encodes all the information the decision maker has relevant to v before obtaining the model. To understand better the other numerator term we can further expand to obtain:

$$\begin{aligned}\{f_e|v,\varepsilon\} &= \int_{\underline{x}} \{f_e|v,\underline{x},\varepsilon\}\{\underline{x}|v,\varepsilon\} \\ &= \int_{\underline{x}} \{f_e|v,\underline{x},\varepsilon\} \frac{\{v|\underline{x},\varepsilon\}\{\underline{x}|\varepsilon\}}{\{v|\varepsilon\}}\end{aligned} \tag{6.48}$$

Combining equations 6.47 and 6.48 we can write:

$$\{v|f_e,\varepsilon\} = k \int_{\underline{x}} \{f_e|v,\underline{x},\varepsilon\}\{v|\underline{x},\varepsilon\}\{\underline{x}|\varepsilon\} \tag{6.49}$$

The two terms $\{v|\underline{x},\varepsilon\}$ and $\{\underline{x}|\varepsilon\}$ allow the decision maker to encode his prior information on v in a detailed way. The interesting term, which we shall examine in detail, is the assessment of f_e given v and $\underline{x}$. This is an assessment of a function, just as in the general expert use formulation where probability functions must be assessed. In the general case the decision maker must not only assess how accurately the model will predict v from the actual state variable inputs, but must also assess the performance of the model for all other inputs. Figure 6.4 depicts a one dimensional example where $\underline{x}$ is scalar and can take on three possible values: $\underline{x}_1$, $\underline{x}_2$, and $\underline{x}_3$. Therefore, f_e is specified by its value at three points. The figure pictorially displays the assessment of f_e given that $\underline{x}_2$ is the true value ($\underline{x}_t$) of $\underline{x}$. The point ($\underline{x}_t$,v) gives a reference point, since a perfect model must produce v from $\underline{x}_t$. To combine the three assessments into a complete assessment of f_e we have to make the simplifying assumption that the three assessments shown are independent. Thus even in this extremely simplified case the assessment requirements are vast.

A fundamental simplification

In many situations, however, such a demanding exercise need not be performed. To see this we rewrite the assessment $\{f_e|v,\underline{x},\varepsilon\}$ as follows:

$$\{f_e|v,\underline{x},\varepsilon\} = \frac{\{v|f_e,\underline{x},\varepsilon\}\{f_e|\underline{x},\varepsilon\}}{\{v|\underline{x},\varepsilon\}} \tag{6.50}$$

A logical assumption to make is that $\{v|f_e,\underline{x},\varepsilon\}$ is equivalent to $\{v|f_e(\underline{x}),\underline{x},\varepsilon\}$. That is, the only pertinent information contained in a model, given that the true values of the state variables are known, is the output of the model with the true values as the inputs. For example,

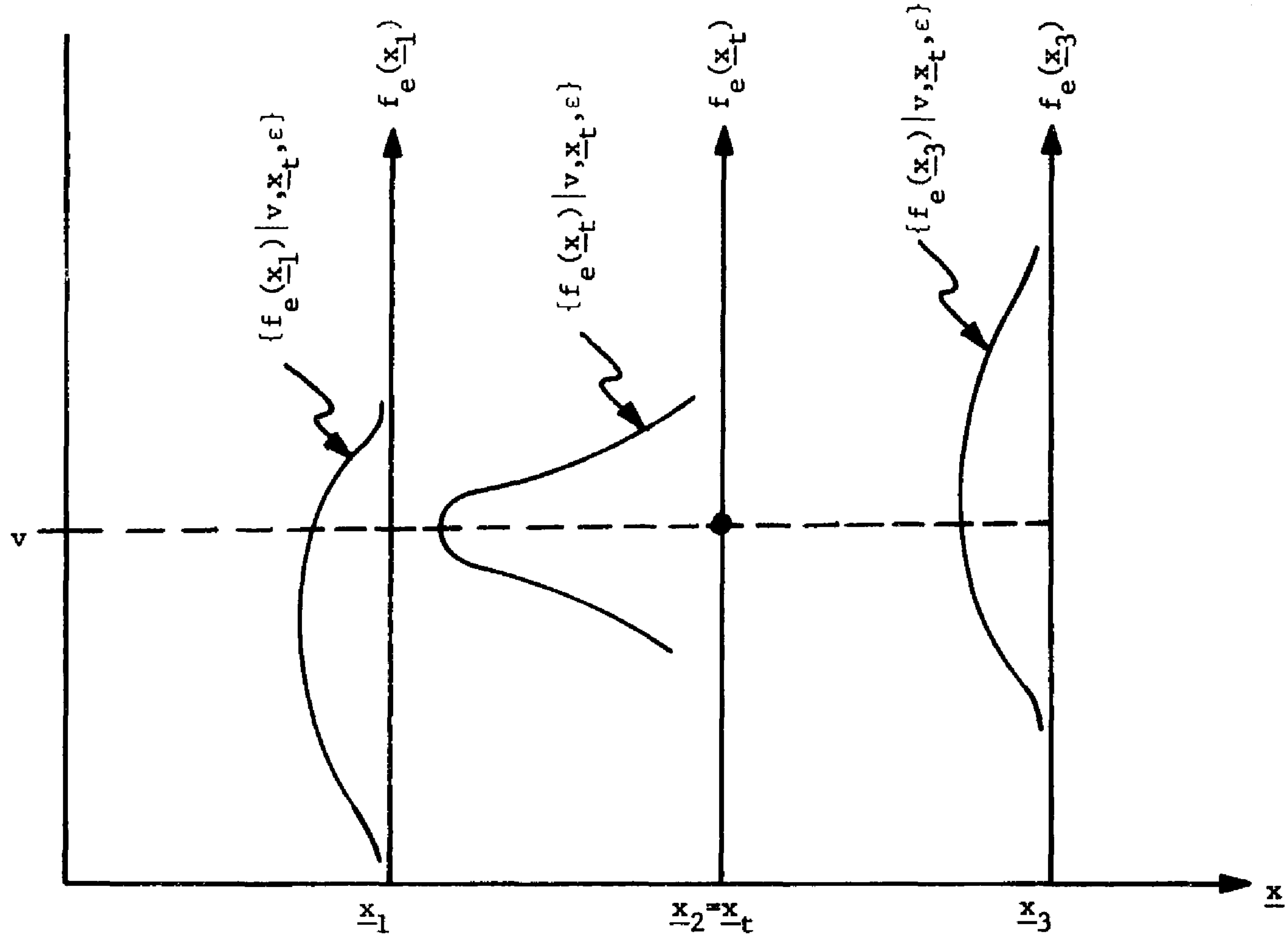


Figure 6.4 A Pictorial Representation of the Assessment $\{f_e|v,\underline{x},\varepsilon\}$

suppose that a computer model is built relating profit π to cost c in an industrial situation: $\pi = \pi(c)$. Next suppose that the decision maker performs a study and determines with certainty that the cost will be a fixed value c_o. What data will the decision maker require from the computer? The assumption, seemingly reasonable, that all he cares about is the profit the model computes from the actual cost; namely, $\pi(c_o)$.

We shall also assume that the revealed values of the state variables alone give no information by themselves about the model form. Thus

$$\{f_e|\underline{x},\varepsilon\} = \{f_e|\varepsilon\} \tag{6.51}$$

Equations 6.50 and 6.51 allow us to write:

$$\{f_e|v,\underline{x},\varepsilon\} = \{f_e(\underline{x})|v,\underline{x},\varepsilon\} \frac{\{f_e|,\varepsilon\}}{\{f_e(\underline{x})|\underline{x},\varepsilon\}} \tag{6.52}$$

This may be verified by applying Bayes' Theorem to $\{f_e(\underline{x})|v,\underline{x},\varepsilon\}$. This equation relates the density function on functions to a density function on the single parameter $f_e(\underline{x})$. Inserting equation 6.52 back into equation 6.49 we have

$$\{v|f_e,\varepsilon\} = k \int_{\underline{x}} \frac{\{f_e(\underline{x})|v,\underline{x},\varepsilon\}\{v|\underline{x},\varepsilon\}\{\underline{x}|\varepsilon\}}{\{f_e(\underline{x})|\underline{x},\varepsilon\}} \tag{6.53}$$

where the term $\{f_e|\varepsilon\}$ is now part of the constant. The useful result is that the problem is reduced to an assessment of a scalar $f_e(\underline{x})$, which is the value given by the model at the true value of the state variable vector. This is a reasonable and intuitive assessment to make.

A further natural assumption can simplify matters even further. In many cases we expect the decision maker's assessment of model performance to be independent of the actual revealed value of $\underline{x}$. In this case the term

$$\{f_e(\underline{x})|v,\underline{x},\varepsilon\} \tag{6.54}$$

is a function of $\underline{x}$ only through the model's dependence on $\underline{x}$. Now define an error

$$e = f_e(\underline{x}) - v \quad , \tag{6.55}$$

where e is the difference between the model's predicted value and the true value. The assumption rephrased is that the decision maker assesses e independent of $\underline{x}$. Therefore the term $\{f_e(\underline{x})|v,\underline{x},\varepsilon\}$ may be computed by change of variables from $\{e|v,\varepsilon\}$. The important thing to notice is that the decision maker need have no formal interaction with the model. The change of variables may be done by experts from the specified model and the decision maker's simple appraisal. The following example will illustrate this point.

Example 6.1--A Campaign Decision

Politician X is investigating the potential effects of a campaign tour. He is a Democrat and knows that his total number of votes v will be a function of d, the number of Democrats who hear him speak. His prior assessment of v given d is (Figure 6.5):

$$\{v|d,\varepsilon\} = \begin{cases} 1/5 & d+1 \leq v \leq d + 6 \\ 0 & \text{otherwise} \end{cases} \tag{6.56}$$

where all variables are in units of millions. His prior on d is (Figure 6.6):

$$\{d|\varepsilon\} = \begin{cases} 1/10 & 0 \leq d \leq 10 \\ 0 & \text{otherwise} \end{cases} \tag{6.57}$$

Thus we may calculate (Figure 6.7):

$$\{v|\varepsilon\} = \begin{cases} \frac{v-1}{50} & 1 \leq v \leq 6 \\ 1/10 & 6 \leq v \leq 11 \\ \frac{16-v}{50} & 11 \leq v \leq 16 \\ 0 & \text{otherwise} \end{cases} \tag{6.58}$$

The politician then hires a consultant whose job it is to build a model relating d to v. The politician's prior estimate of the error of this model, as defined by equation 6.55, is:

$$\{e|v,\varepsilon\} = \begin{cases} 1 & -1/2 \leq c \leq 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{6.59}$$

In other words, he feels that the consultant's model is equally likely to be off by no more than a half unit in either direction. Now suppose that the consultant comes up with the simple linear model:

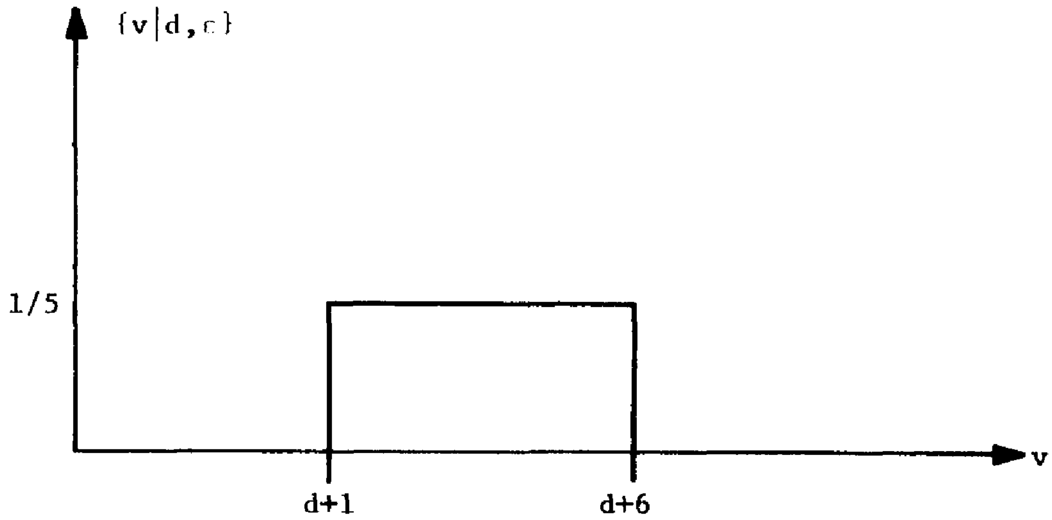


Figure 6.5 Conditional Prior on v

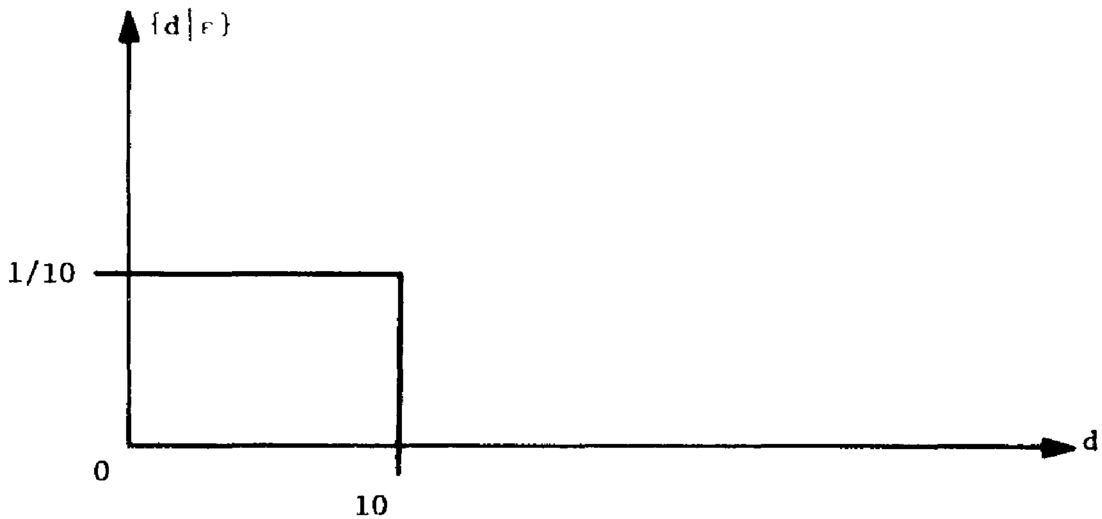


Figure 6.6 Prior on d

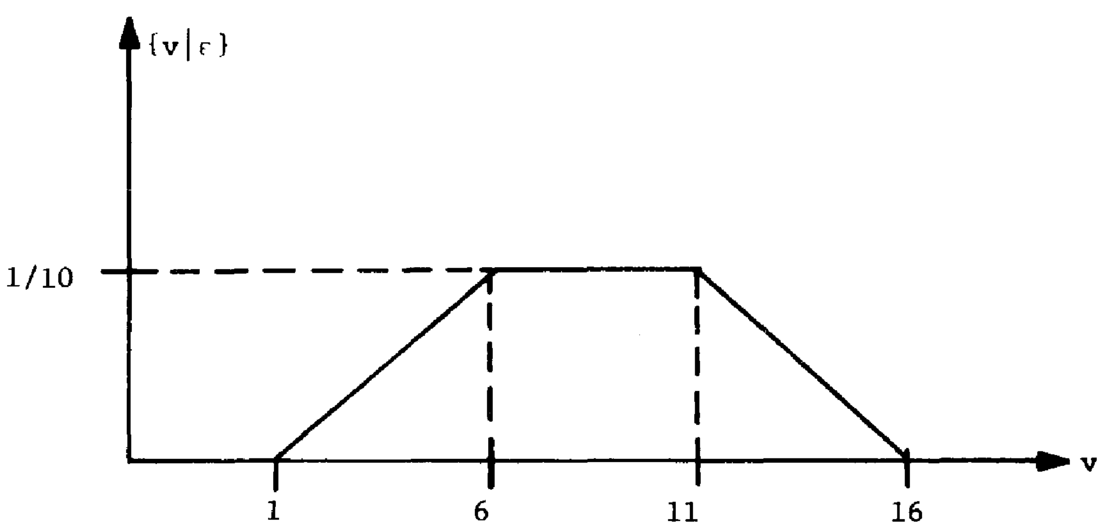


Figure 6.7 Prior on v

$$f_e(d) = 2d + 3/2, \quad d \geq 0 \tag{6.60}$$

The problem is to calculate the decision maker's posterior $\{v|f_e,\varepsilon\}$.

First we calculate from equations 6.59 and 6.55:

$$\{f_e(d)|v,d,\varepsilon\} = \begin{cases} 1 & v - 1/2 \leq f_e(d) \leq v + 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{6.61}$$

$$= \begin{cases} 1 & v - 2 \leq 2d \leq v - 1 \\ 0 & \text{otherwise} \end{cases} \tag{6.62}$$

From equation 6.53 we may calculate:

$$\{v|f_e,\varepsilon\} = \int_d \frac{\{f_e(d)|v,d,\varepsilon\}\{v|d,\varepsilon\}\{d|\varepsilon\}}{\{f_e(d)|d,\varepsilon\}} \tag{6.63}$$

from the given quantities. The computation is conceptually straightforward but tedious, and therefore will not be given.

The study of model use is an interesting topic. The preceding development shows much promise for additional benefits from further research. Such analysis, however, would divert us far afield from the main topic.

6.4 Summary

This chapter presented a number of diverse, yet related, developments. First we studied expert bias and formulated a particular characterization of bias which allowed us to present a practical methodology with emphasis on structural insights. The use of experts to aid the decision maker solve problems of inference was then investigated. We found by analysis that the problems could be reformulated into one where the basic issue is uncertainty surrounding the decision maker's model of an experiment. In this instance, the expert is roughly viewed as an expert on which model actually produced the given data. The development was shown to have interesting implications for the study of prior exchange. Finally, we addressed the general problem of how to use subjective inputs (experts) in relation to objective inputs (mathematical models). This led to a brief investigation of the proper use of models--an important topic in its own right.

Chapter VII

SUMMARY

7.0 Conclusions

The main objective of this dissertation was to create a theory of expert use and derive from it practical tools for the solution of real world problems. In achieving this goal we have ventured into many different areas of research and application.

We began by developing a theoretical framework resting philosophically on the same foundations as the theory of decision analysis. The single expert case was treated first where the basic idea was to consider the expert's prior distribution as an addition to the decision maker's own state of information. This immediately led to the concept of the expert's probability distribution as a random variable. The theory was then naturally extended to include in its domain the multi-expert environment. Our preliminary analysis considered such topics as independence, subjective calibration, the value of an expert, and the selection of a panel of experts.

We devoted an entire chapter to a detailed investigation of an interesting aspect of subjective probability theory. A methodology was developed to allow an assessor to encode and use his knowledge about himself as an assessor. The concept of self-calibration was introduced and a new probability assessment technique, the Method of Equivalent Intervals, created to enable an assessor to self-calibrate himself. Aside from important practical implications, the developments also provided useful concepts pertaining to expert theory.

Our major effort was directed towards the development of a practical methodology. We derived a process where a decision maker can transform a set of expert priors into a single surrogate prior. The surrogate prior is used symmetrically with the decision maker's own prior to produce a consistent posterior probability assessment.

The surrogate prior is the result of the specification of a subjective calibration function. By use of an invariance principle we demonstrated an operational way to assess the calibration function. Important for multi-expert applications was the development of the tool of trajectory analysis. This proved a useful way to reduce the amount of assessment necessary to specify the calibration function.

We further gained insight into one aspect of the group decision making problem. Conditions were specified under which a group of decision makers should agree on how expert advice is implemented. Also presented was an analysis of normative conditions under which a group of assessors should achieve a consensus.

Our final results were primarily oriented towards a detailed modeling of three basic aspects of expert use. A practical way of handling expert bias was developed which provides immediate insight by identifying several fundamental characteristics of a given expert resolution problem. We developed a framework for understanding the use of experts relative to experimentation and then considered the relationship of the use of experts to the use of formal models. In doing so we investigated in some detail the way in which certain types of models should be used.

In summary we have accomplished our objectives. We have created a practical theory fulfilling all the requirements of the conceptual model presented in Chapter I. The results suggest that application of the ideas here presented should be an immediate task for those with important decisions to make.

7.1 Topics for Further Research

The study of the use of experts encompasses many broad and diverse topics. It is hoped that the ideas presented in this paper will help form the foundation for various future research efforts. The list of promising topics is certainly a rich one.

Clearly there is much use for more detailed modelings of how experts make assessments. The basic structure provided should pave the way for further analysis and new experiments aimed at a deeper understanding of the way in which experts behave.

The material on expert bias provides a convenient base from which to study reward/penalty systems designed to optimally control an experts bias. Present work on this topic is directed towards rewarding experts to remove all bias. However, as we have seen, it is uncertainty about bias, rather than bias itself, that creates a problem for the decision maker. Also, current techniques ignore the factors that cause an expert to be biased.

Another area ripe with many possibilities is the subject of modeling theory. It appears that the brief conceptualization presented in this paper may, when further exercised, provide answers to questions currently conceptually intractable. For instance, the theory goes a long way towards answering the question, "what do we do about things we haven't modeled"?

The treatment of the use of experts in this dissertation is certainly not an exhaustive one. The horizons for future research are only a function of the skill and interest of those who endeavor to study the topic.

BIBLIOGRAPHY

1. Alpert, Marc, and Raiffa, Howard, A Progress Report on the Training of Probability Assessors, Unpublished manuscript, Harvard University (1969).

2. Beckman, Otto R., A Decision Analysis Approach to Two Person Games, Ph.D. Dissertation, Department of Engineering-Economic Systems, Stanford University, March (1970).

3. Brown, Bernice, and Helmer, Olaf, Improving the Reliability of Estimates Obtained from a Consensus of Experts, The RAND Corporation, P-2986, September (1964).

4. Dalkey, Norman C., Delphi, The RAND Corporation, P-3704, October (1967).

5. Dalkey, Norman C., The Delphi Method: An Experimental Study of Group Opinion, The RAND Corporation, RM-5888-PR, June (1969).

6. deFinetti, Bruno, "Does it Make Sense to Speak of 'Good Probability Appraisers'?," The Scientist Speculates--An Anthology of Partly-Baked Ideas, Irving S. Good, ed., London: Heineman, 357-364 (1962).

7. deFinetti, Bruno, "La prévision: ses lois logiques, ses sources subjectives," Annals de l'Institut Henri Poincaré, Vol. 7, 1-68 (1937). Trans. by Henry E. Kyburg, Jr. as "Foresight: Its Logical Laws, Its Subjective Sources," Studies in Subjective Probability, Kyburg and Smakler, ed., Wiley, New York, 97-158 (1963).

8. Edwards, Ward, "Conservatism is Human Information Processing," Formal Representation of Human Judgment, B. Kleinmuntz, ed., New York: Wiley, 17-52.

9. Helmer, Olaf, Convergence of Expert Consensus Through Feedback, The RAND Corporation, P-2973, September (1964).

10. Howard, Ronald A., "Decision Analysis: Perspectives on Inference, Decision and Experimentation," Proceedings of the IEEE, Vol. 58, 632-643, May (1970).

11. Howard, Ronald A., "Information Value Theory," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, 211-219 (1968).

12. Howard, Ronald A., "The Foundations of Decision Analysis," IEEE Transactions on System Science and Cybernetics, Vol. SSC-4, No. 3, 211-219 (1968).

13. Howard, Ronald A. and Murray, George R., Jr., An Example of Bayesian Statistical Decision Theory, Unpublished manuscript, Stanford University.

14. Matheson, James E., "The Economic Value of Analysis and Computation," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, 325-332 (1968).

15. McWhirter, Norris and Ross, The Guinness Book of World Records, Bantam Books, New York (1968).

16. Phillips, L. D., and Edwards, W., "Conservatism in a Simple Probability Inference Task," Journal of Experimental Psychology, 346-354, Vol. 72 (1966).

17. Raiffa, Howard, Assessments of Probabilities, Unpublished manuscript, Harvard University (1969).

18. Raiffa, Howard, Decision Analysis: Introductory Lectures on Choices under Uncertainty, Addison-Wesley, Reading, Massachusetts (1968).

19. Roberts, Harry V., "On the Meaning of the Probability of Rain," Proceedings of the First National Conference on Statistical Meteorology, Boston: American Meteorological Society, 133-141 (1968).

20. Roberts, Harry V., "Probabilistic Prediction," Journal of the American Statistical Association, 50-62, Vol. 60 (1965).

21. Savage, L. J., The Foundations of Statistics, Wiley, New York (1954).

22. Smallwood, Richard D., "A Decision Analysis of Model Selection," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, 333-342 (1968).

23. Staël von Holstein, C.-A.S., Assessment and Evaluation of Subjective Probability Distributions, The Economic Research Institute at the Stockholm School of Economics (1970).

24. Winkler, Robert L., "Scoring Rules and the Evaluation of Probability Assessors," Journal of the American Statistical Association, 1073-1078, Vol. 64 (1969).

25. Winkler, Robert L., "The Consensus of Subjective Probability Distributions," Management Science, Vol. 15, No. 2, October (1968).

26. Winkler, Robert L., and Murphy, Allan H., "Evaluation of Subjective Precipitation Probability Forecasts," Proceedings of the First National Conference on Statistical Meteorology, Boston: American Meteorological Society, 148-157 (1968).

27. Winkler, Robert L., and Murphy, Allan H., "Nonlinear Utility and the Probability Score," Journal of Applied Meteorology, 143-148, Vol. 9, No. 1, February (1970).