TRANSACTIONS OF SOCIETY OF ACTUARIES
1980 VOL. 32

THE QUANTUM INTERPRETATION OF PROBABILITY

MYRON H. MARGOLIN

ABSTRACT

This paper proposes a new interpretation of probability, the quantum interpretation of probability (QIP), so named because it adapts certain of the principles of quantum mechanics. According to this interpretation, probability is an objective, measurable quantity whose value is subject to an inherent uncertainty or indeterminacy. In any random experiment the amount of uncertainty is complementary to the quantity (mass) of data.

The rationale for QIP follows from an analysis of the frequentist interpretation, according to which probability is objective and precise. The frequentist interpretation is defined, however, only for time-homogeneous sequences of random experiments—that is, experiments occurring under "similar" conditions and for which the probability is taken to be constant. But many random processes, such as mortality and morbidity, are time-heterogeneous; that is, they occur under changing conditions. If a probability were both objective and precise in the time-heterogeneous case, then changes in its value should be distinguishable from the random fluctuations of the observed data. The analysis indicates, however, that such distinctions are conceptually untenable. On the other hand, it is clear that there are identifiable, objective factors that cause changes in observed frequencies. The conclusion is, then, that probability is objective yet imprecise—the basic thesis of QIP. Analysis of the subjectivist-Bayesian view of probability shows that this, too, is defective relative to time-heterogeneous processes.

Credibility theory furnishes a case history of a practical problem in actuarial science for which different interpretations of probability may lead to different solutions. QIP suggests a practical, realistic solution to this problem, while the frequentist and subjectivist-Bayesian views lead only to formal solutions with no practical significance for time-heterogeneous claim processes.

I. INTRODUCTION

PROBABILITY is the cornerstone of actuarial science. The size and strength of the life insurance industry evidence the successful application of probability theory in this domain. Of course, probability is also widely used in such diverse fields as statistical mechanics, genetics, experimental psychology, and gambling.

Yet—paradoxically, in view of its widespread and successful uses—the very concept of probability is the subject of much controversy. Scholars, including philosophers, statisticians, and economists, have proposed a variety of interpretations of probability, all aiming to explain its basic meaning, but no single interpretation has received general acceptance. (For surveys of most of the leading interpretations see [8], [12], and [22].)

Most of these interpretations fall into either of two categories, frequentist and subjectivist. In brief, frequentists assert that probability is an objective quantity closely associated with observed frequencies, while subjectivists interpret probability as a personal or subjective degree of belief. Thus, a frequentist would interpret the statement "The probability of rolling an ace with this die is one-sixth" to mean that the long-term frequency of an ace is very nearly one-sixth. If it turns out that the long-term frequency is not nearly one-sixth, the statement is false.

The subjectivist would say that the probability is one-sixth for *you*, if *you* would accept betting odds of 5 to 1 on an ace or 1 to 5 on a non-ace. The probability could be different for someone else who, for whatever reason, put different odds on the occurrence of an ace. Your degree of belief in the occurrence of an ace is neither true nor false, according to the subjectivist interpretation, and you are free to revise your opinion as you acquire additional information.

The frequentist version is the one most commonly taught in textbooks on probability. For example, the text currently cited in the Society's Course of Reading characterizes probability theory as that branch of mathematics concerned with random phenomena. Experience shows that many such phenomena exhibit a "statistical regularity" when repeated frequently. Accordingly, the statement that a certain experimental outcome has probability p is interpreted to mean that if the experiment is repeated a large number of times, that specific outcome would be observed "about" $100p$ percent of the time [14].

This paper argues that both frequentist and subjectivist interpretations are defective. It offers a new interpretation, the quantum interpretation of probability (QIP). According to this interpretation, probabilities are objective quantities that are *measured* by frequencies, but their values are inherently imprecise—they are subject to an objective uncertainty analogous to the typical uncertainties of quantum mechanics. If the mass of data is relatively large, the uncertainty in the corresponding probability is relatively small, but the probability is a characteristic of the entire mass and is not well localized in space-time. Conversely, if the mass is small, the probability is relatively well localized but is subject to a large uncertainty.

The rationale for QIP emerges upon close analysis of the frequentist interpretation. The frequentist interpretation explicitly presupposes that the repetitions of a random experiment occur under similar conditions. However, in actuarial science and elsewhere, we frequently encounter phenomena for which conditions change in time. Section II, after formalizing the notion of a random experiment, considers how the frequentist interpretation might be extended to time-heterogeneous (nonstationary) conditions. It also considers the conceptual relationship between random variation and variation induced by changing conditions. Using mortality rates as paradigmatic, it concludes that an objective probability cannot be precise and that random variation cannot be distinguished unambiguously from causal variation—conclusions that contradict the frequentist interpretation.

QIP itself is discussed in detail in Section III. Its principles, some of which are adapted from quantum mechanics, preserve the concept of objective probability while remedying inadequacies of the frequentist position.

Section IV analyzes the subjectivist interpretation and its offspring, Bayesian statistics, from the standpoint of QIP. Subjectivism is not entirely incompatible with QIP, since QIP explicitly recognizes subjective estimates of probability. The analysis does, however, disclose certain inadequacies of subjectivism and Bayesianism, especially in the case of time-heterogeneity.

Section V outlines the case history of a sector of actuarial science: credibility theory. The frequentist and Bayesian approaches to the problem of credibility apparently have failed to produce useful solutions. QIP suggests why these approaches have failed and furnishes the rationale for a workable, empirical solution to this long-standing problem, thus showing that the dispute over probability is not entirely academic or theoretical but may have practical implications for certain applications.

II. RANDOM EXPERIMENTS; FREQUENTIST INTERPRETATION

As previously noted, frequentists often try to express the relationship between probability and frequency by using the notion of a repeatable random experiment. Without prejudging the nature of the relationship, let us try to formalize this notion.

A *random experiment*, denoted by ϵ, refers to any kind of occurrence that is characterized by

Π: An identifiable process that generates specified types of observable data.

S and T: Parameters that delineate or specify spacelike and timelike regions of observations, in order to distinguish ϵ from other similar experiments.

R: The set of all possible observations for ϵ. (Let r denote the actual outcome of ϵ, where, of course, $r \in R$.)

Consider two illustrations:

1. The dollar amount of group major medical claims for the employees of the ABC Corporation in 1978 is equivalent to the observed outcome of an experiment ϵ. Here $\pi \in \Pi$ is the process of major medical claim generation, $s \in S$ specifies the ABC Corporation, $t \in T$ specifies the year 1978, and R is the set of real numbers corresponding to possible claim amounts. If we were interested in the number of claims, then the values of Π, S, and T would be the same, but R would be the set of nonnegative integers corresponding to numbers of claims; this would be a distinct experiment, ϵ'.
2. In this example, π is the process of coin tossing; s specifies the coin I now hold in my hand; t means the next ten tosses from "now"; and R is $\{0, 1, \ldots, 10\}$, the possible number of heads. Alternatively R might be all 2^{10} possible outcomes where orders are considered distinct—defining a different experiment, ϵ'.

A *set of repetitions* ϵ^* is a set of experiments $\{\epsilon_1, \ldots, \epsilon_i, \ldots, \epsilon_n\}$ for which the values of Π, S, and R are identical for all ϵ_i's while the values of T are disjoint: $t_{i+1} > t_i$, where $t_i \in T$. The set ϵ^* is itself an experiment; its π^* and s^* are those of the ϵ_i's; $t^* = \{t_i\}$; and if the r_i's are numerical, r^* may be the sum or an appropriate average of the r_i's. At this point, there is *no* assumption that the repetitions occur under similar or identical conditions. They are repetitions because of the identity of Π, S, and R—the same coin is being tossed, or the same group contract is exposed to risk. Note that ϵ^* may itself be a member of a set of repetitions.

(The requirement that the R_i's be identical in a set of repetitions may be relaxed to permit certain types of nonessential differences. For example, let ϵ_1 refer to the experiment that measures the mortality rate among United States males aged 35 in 1977, and let ϵ_2 refer to the corresponding measurement for 1978. The number of exposures is not the same for both years, consequently the respective sets of possible results are not precisely identical, but we should still consider ϵ_1 and ϵ_2 to be repetitions.)

Other sets of experiments $\{\epsilon_1, \ldots, \epsilon_n\}$ may be identical with respect to Π, T, and R but disjoint with respect to S. Such a set might be termed a *set of repetitions in space.* Again, there is no assumption that the ϵ_i's occur under similar conditions. (As above, the requirement that the R_i's be identical may be relaxed.)

The term *binary experiment* will refer to an experiment ϵ with only two possible outcomes: $R = \{r, \text{Not } r\}$.

The *frequency* of r in a set ϵ^* of binary experiments is $r^* = \Sigma\, r/n$, where r has the value 1 and Not r has the value 0. Here the set may be one of timelike repetitions, spacelike repetitions, or a combination of both.

Finally, let us suppose that associated with any random experiment ϵ there is a quantity θ. Unlike the other characteristics—π, S, r, and so on—θ has no direct interpretation in terms of empirical observation. It is not an identifiable process, nor is it measurable as are space and time, nor is it directly observable as is the result or outcome of ϵ.

When ϵ is binary, we will call θ the probability of r. Otherwise, θ is a "statistical parameter"—often the "expected value" of r. This paper is concerned primarily with the interpretation of probability per se, but much of the discussion is relevant also to the more general case of non-binary experiments and statistical parameters.

The earliest definition of probability, the so-called classical definition, long preceded the frequentist and subjectivist interpretations. Laplace [18] expressed it along the following lines: If a trial can result in N equally likely outcomes, of which M are deemed favorable, then the probability of obtaining a favorable result is M/N. Much later it was realized that this definition has two serious flaws. First, there are relatively few kinds of trials or experiments in which the possible results are in any sense symmetrical and can be judged equally likely. Second, the definition turns out to be a vacuous tautology. It presupposes an understanding of "equally likely," a phrase which can only mean "equiprobable," otherwise the definition is internally inconsistent—but "equiprobable" presupposes an understanding of probability.[1]

The tautology would be equally apparent if we were to translate Laplace's definition into the formalism for ϵ, which we have just introduced. The term "equally likely" would have to be translated into a common θ for each of the N possible outcomes; but θ cannot be used as an argument, since it is what we are seeking to define.

We shall see that these two flaws, limited scope and tautology, also afflict the frequentist interpretation.

Frequentist Interpretation (FI)

The frequentist interpretation refers to any binary experiment that, ostensibly, can be repeated a large number of times under similar condi-

[1] Of course, we can use Laplace's definition to advantage in *estimating* probabilities in cases where symmetries do seem to be present. QIP provides explicitly for estimating probabilities.

tions. Then the probability θ_i associated with ϵ_i is deemed to have the following characteristics:

1. θ_i is objective. Its numerical value is independent of our personal or subjective opinions.
2. The value of θ_i is precise.
3. $\theta_i = \theta$, constant throughout the set of repetitions $\epsilon^* = \{\epsilon_i\}$.
4. In principle, as the repetitions are continued indefinitely, the observed frequency r^* approaches θ as a limiting value. In practice, if n is very large, the observed r^* can be taken as a close approximation to the true value of θ. In effect, such an r^* serves as an approximate *measure* of θ.
5. It is generally understood that the individual r_i's are "random." Precisely what this means has proved difficult to define. Let us, however, set aside any such difficulties for the time being and assume that the r_i's satisfy our usual notion of randomness.

A common objection to FI is that it is not possible in fact to observe an experiment repeated without limit. Also, it may be difficult or impossible to repeat an experiment finitely many times—perhaps even once—under similar conditions. The FI response might be to acknowledge that the imputed set of repetitions is an idealization or abstraction, but to point to many other instances in science in which idealization, abstraction, or so-called thought-experiments are used to define or clarify concepts. Furthermore, frequentists can point to certain random processes whose actual behavior does in fact conform closely to this idealization (e.g., radioactive decay). Accordingly, it is not clear that we ought to reject FI solely because it has recourse to some sort of idealization.

It does seem clear, however, that FI limits its scope to only certain types of repetitions: those for which the conditions of the experiments are ostensibly uniform in time, or *time-homogeneous*. But for many types of experiments, and especially those which we encounter in actuarial science, the data give unmistakable evidence of time-heterogeneity. The time variation of the results is much too great to be construed solely as random fluctuation. Mortality rates are an exemplary case in point. Take the year-by-year sequence of observed mortality rates for United States males age x, where almost any value of x will serve the purpose. Wherever data are available, history has shown striking long-term declines in mortality rates in the United States and in many other nations. These changes are much too large to be compatible with a constant θ under any of the usual notions of random fluctuation. Mortality rates are also subject to moderate seasonal variation and to occasional significant nonrandom fluctuation due to wars and epidemics. Similarly, the probability of snow is evidently not constant in time. Here, in addition to the obvious seasonal variations, long-term climatic changes seem to play a role.

Let us then assume that sets of repetitions are either time-homogeneous or time-heterogeneous. (Later we shall reconsider whether this simple dichotomy is tenable). Let us also assume, provisionally, that FI fits those sets that are time-homogeneous. How might FI be extended to cover the time-heterogeneous case?

Recall the five characteristics of the FI conception of θ: objective, precise, constant, limiting value, and randomness of results. Obviously, the characteristics of constancy and limiting value must be set aside or greatly modified in the case of time-heterogeneity. It is clear also that, as a *practical* matter, it will not be possible to ascertain the precise value of θ_i. Is the assumption still tenable that precise θ_i's exist *in principle* and that the individual r_i's are "randomly" distributed about the θ_i's? Our analysis will show that the answer is no.

The simplest way to extend FI would be to define θ_i for a time-heterogeneous process to be the limiting value that *would* result if the value of θ did not change after t_i. This would be tautological, however, since it presupposes that we already have a basic idea of what the probability θ means in the first place.

A more ingenious approach hinges on the evident connection between gross changes in observed frequency and changes in the conditions of the experiment. For example, in the case of mortality rates, it is clear that the long-term decreases in mortality rates are related in some way to changes in the conditions of society, especially to improvements in nutrition, sanitation, and medical care. Such connections are evidence in support of the objective character of θ. They seem similar to the cause-and-effect relationships of science in general—for example, if the water supply in Indian villages is purified, the probability of death at age x will decrease. From such connections, one may also infer that the value of θ is a function of the conditions of the experiments. The final step in the argument is to define θ at a specific time as the limiting value that would result if all conditions were to remain unchanged thereafter. Like the original definition of θ, this would be a kind of definition *in principle*, or a *thought experiment*, since the argument would concede that it may be impossible in fact to maintain constant conditions for factually heterogeneous processes.

The flaw in this argument lies with the problem of distinguishing those conditions that are relevant to θ from those that are relevant to r. In order to produce an idealized sequence of experiments $\{\epsilon_i\}$ in which θ is constant but the r_i's vary "randomly" about θ, one must relax the constraint that the experiment be performed under uniform conditions. If the experiments were performed under precisely *identical* conditions, the results necessarily would be identical. Conceptually, we must demand variations in just those conditions that affect r but not θ, while permitting absolutely no variation in those conditions that do affect θ. Yet the notion that there is a clear distinction between two such types of conditions is quite unsubstantiated. Indeed, merely to assert that there exists a class of conditions that affect only r, and not θ, presupposes a clear conception of θ and is therefore tautological.

This point can be developed more clearly in the specific context of mortality rates. Let $\{\epsilon_t\}$ represent the exposures to mortality in years t for a specified population, and let $\{r_t\}$ denote the corresponding mortality rates. The argument for extending FI presupposes that (1) there exists a set $\{\theta_t\}$ such that each r_t is randomly distributed about θ_t, (2) the changes $\theta_{t+1} - \theta_t$ correspond to certain "macroscopic" changes in conditions, and (3) these are precisely distinguishable from other "microscopic" changes corresponding to $r_{t+1} - \theta_{t+1}$ and $r_t - \theta_t$. Examples of macroscopic changes might be severe epidemics or natural disasters that give rise to abrupt, nonrecurring changes in θ, and also the numerous specific improvements in nutrition, sanitation, and medical care to which we attribute the long-term decline in θ. That which we call an epidemic, however, differs only in degree from any occurrence of contagion. Likewise, a serious disaster differs only in degree from any accident that kills just one person. Conversely, an occurrence in which one life is saved by application of medical technology is an instance of the general application of such technology. In general, the changes in conditions that are said to cause significant, nonrandom changes in mortality rates, that is, changes in θ, are *not* essentially different from those that cause the individual deaths or save the individual lives.

It follows that it is impossible to define a nonarbitrary procedure for precisely dividing $(r_{t+1} - r_t)$ into a nonrandom component $(\theta_{t+1} - \theta_t)$ and random components $(r_{t+1} - \theta_{t+1})$ and $(r_t - \theta_t)$.

Note, too, the danger implicit in the notion that deaths *caused* by nonrandom, macroscopic changes can be distinguished from those caused by microscopic, random variation. Carried to its logical conclusion, this notion would exclude *all* deaths, for all deaths have causative antecedents. There is a significant sense in which no death is purely random. Accordingly, any precise distinction between randomly occurring deaths and those caused by changes in conditions must be fundamentally arbitrary.

Problem of Empirical Validation

Another serious objection to FI arises when we consider how one might attempt to validate empirically a statistical model employing a precise θ for a time-heterogeneous process.

Let $\{r_t\}$ be the results of a time-heterogeneous set of repetitions $\{\epsilon_t: t = 1, \ldots, n\}$. It is asserted that the results are explained by a certain mathematical model containing a parameter $\theta_t = \theta(t)$ and a conditional density function $f(r|\theta)$ characterizing the random fluctuation of r about θ_t. Let σ_f denote the standard deviation of f.

This model is consistent with the set $\{r_t\}$ if, for the specified set $\{\theta_t: t = 1, \ldots, n\}$, it is almost always the case that $|r_t - \theta_t| < 3\sigma_f$. On the other hand, it is obvious that this criterion can be satisfied by indefinitely many different sets $\{\theta_t\}$. Any model for which most of the θ_t's lie within the intervals $r_t \pm 3\sigma_f$ will be consistent with the set $\{r_t\}$. No amount of data can dispel the uncertainty as to which is the correct model, since the experiment at $T = t$ can occur only once. Subsequent repetitions, for $T > t$, do not help to distinguish which value for θ_t is the objectively correct one.

This line of reasoning is not unrelated to the previous discussion of changes in conditions.[2] In both cases, it initially seems plausible to consider *gross* changes—either in conditions or in the value of θ—to be factual characteristics of the set $\{\epsilon^*\}$. The trouble arises when we try to be precise. We cannot precisely distinguish a change in condition from the changes prerequisite to random variation, nor can we precisely distinguish that the true value of θ_T at $T = t$ is θ_t, as opposed to a θ'_t having nearly the same value.

The arguments have shown thus far that, for time-heterogeneous processes, an objective probability cannot be precise. They do not, however, exclude the possibility that probabilities are somehow objective yet imprecise. Indeed, there remain several good reasons that tend to support the notion that probabilities are objective. First, there are cases (e.g., rates of radioactive decay) where probabilities can be measured precisely. This leaves open the possibility that probability is both objective and precise, at least with respect to time-homogeneous repetitions involving large masses of data. Second, gross changes in observed frequencies often seem connected with observable changes in conditions. There is evidentiary support for such statements as "The influenza epidemic of 1918–19 substantially increased the probability of death for United States males age 35" or "The reduction of speed limits in the

[2] Note that this argument, as well as the argument of the preceding section, can be turned against the notion of a precise θ in the time-homogeneous case. A model of the form $\theta_T = \theta =$ constant cannot be conclusively validated to the exclusion of any alternative, regardless of how many data were available. An alternative model according to which θ_T differed slightly from θ at finitely many places could not be falsified. Subtle but undetectable changes in relevant conditions might have occurred at these places. These arguments will surface later.

United States to 55 m.p.h. reduced the probability of death from motor vehicle accident." A third reason, and perhaps the most compelling, is the argument presented in the next section, dealing with spacelike variation.

Spacelike Variation

From about age 10, mortality rates are an increasing function of age. This generalization is almost universally true, wherever adequate data are available. It holds in each state or region of the United States and in most, if not all, of the countries of the world. For sufficiently large populations, the variation by age is so regular as almost to follow a precise mathematical function such as the Gompertz and Makeham "laws" of mortality.

There are numerous other types of regularities associated with mortality rates. The mortality differential for females is similar in different regions of the country, and it changes only gradually with respect to age. The added mortality associated with high blood pressure is undoubtedly similar in San Francisco to that in New York. The mortality advantage for teachers as compared with laborers undoubtedly holds in both Houston and Minneapolis and at both age 45 and age 55. A recent study suggested that the mortality associated with skin cancer is a monotonic function of geographical latitude, undoubtedly the result of different exposures to sunlight.

No one can reasonably dispute the factual and objective nature of such regularities. They evidently have causal explanations, like the undoubted relationship of the age slope of mortality to changes in cellular physiology. Also, they may serve as the basis for sound predictions. For example, a prediction that in a very large population, the mortality rate at age $46\frac{1}{2}$ lies between the rates for ages 46 and 47 would almost certainly prove correct. The facts that such regularities are so prevalent, have causal antecedents, and lead to useful predictions suggest that we take seriously the notion that probability is objective.

To carry the argument another step, consider how life insurance underwriters utilize these kinds of regularities in evaluating individual insurance applications. In essence, the probability of death ascribed to the applicant is based initially on his or her age and sex, but is then modified by a series of adjustment factors (often called debits and credits [23]). These adjustments correspond to various indexes of the physical condition, medical history, occupation, and other objective characteristics of the applicant. The amount of each adjustment factor is not a mortality rate as such but a quantity abstracted from the relationships among the mortality rates of various subsets of the total population. For example, let us take the case of a male fireman age 45 with blood pressure of 151/86. There are no reliable statistics for applicants having precisely these characteristics. Instead, one begins with a rate for males age 45. (This rate probably would have been derived by interpolation or graduation from recent mortality studies by quinquennial age.) Then one adds a debit based on the observed excess mortality for a class such as males age 35–55 with blood pressure in the range 145–154/83–87. Such excess is calculated in relation to males age 35–55 with standard blood pressures. Similarly, a debit is added for firemen in relation to a standard underwriting class, within some broad age span. Note that the adjustments are based on objective measurements, but that the classes of population from which these measurements are derived correspond only loosely to the age-sex class of the applicant, namely, male age 45.

This methodology of underwriting seems to presuppose the following conceptualization of the "spacelike" behavior of probabilities of death. Each characteristic of the applicant, such as age, sex, blood pressure, and occupation, corresponds to a dimension in a multidimensional nonmetric space, a space $S = \langle S_1, S_2, \ldots, S_n \rangle$, where each S_i corresponds to one such dimension. The domain of some S_i's is the set of positive real numbers, as in the case of age or systolic blood pressure, while for other S_i's the domain is equivalent to a set of discrete integers corresponding, for example, to occupation.

Probability is, in some sense, a function of the dimension or parameter S_i. One can measure and compare the values of the probability function in various subspaces, such as firemen age 35–55 in relation to white-collar workers age 35–55. Such measurements generally tend to support the notion that the probability value changes smoothly with respect to a particular S_i (at least when the values of the S_i's are naturally ordered as in the case of age). These measures enable one to quantify, at least approximately, the variation in the probability with respect to the S_i's, and to assign values to the probability function at individual points in space.

We must be careful, however, in how we interpret this conceptualization. Although the observed data exhibit many regularities and their variations can be quantified, this does not prove the existence of a precise probability function. As in the case of θ_T, it remains impossible to confirm that θ_S has a specific value at $S = s$. In assigning a precise value to the probability at $S = s$, the underwriter is making an estimate of the value of θ_s. Although this estimate derives largely from objective data, a review of its derivation from the original observations would disclose many subjective aspects as well. Subjectivity enters into the graduation of the initial age-sex mortality rates, into any *precise* evaluation of the effects of the elevated blood pressure, into a determination of which parameters are to be considered, and perhaps elsewhere as well. No two underwriters or insurance companies will arrive at precisely the same conclusion.

Why, then, should we suppose that this conceptualization points to an objective character for probability? The answer begins with the intuitive notion that a purely subjective approach to probability, which could ignore the existence of these regularities, would be wrong. We can carry this notion a step further. A methodology for making predictions or estimates that does take these kinds of regularities into account will lead to demonstrably better predictions than a methodology that does not. An insurance company that presupposes that probabilities have a regular spacelike variation will have better results than one that does not. To take an extreme case, a company whose rates did not vary by age would soon go bankrupt, the victim of antiselection. Even if antiselection is set aside and two companies underwrite equivalent cross-sections of the population, the results of the company that discriminates by age will be demonstrably superior according to a simple root-mean-square test. (For each insured, i, define the "error of estimation" as $e_i = r_i - q_i$, where r_i is 1 if the individual dies and 0 if he survives, and q_i is the imputed probability estimate. Then, although $\Sigma\, e_i$ would be the same for both companies, $\Sigma\, e_i^2$ will almost certainly be substantially less for the company that considers age. Analogous results will apply for other parameters, such as occupation and medical characteristics.)

The discussion suggests that one can objectively discriminate, to some extent, between alternative sets of estimates and between alternative methodologies of estimating. Although each method would still be partly subjective, the one that gave better predictions could be construed as being more nearly correct and somehow nearer to objective validity.

A further point is needed to clarify the underwriter's methodology, namely, his presupposition that the value of the probability function will remain approximately constant in time. The underwriter's data are based solely on past results, yet he is setting a rate or assigning a probability with respect to the applicant's future exposure to death. Obviously, his assertion that the past observations are indicative of future probabilities presupposes that general environmental conditions and the applicant's own circumstances do not change significantly. This basic presupposition, that past results are a guide to future expectations provided that relevant conditions do not change, tacitly underlies virtually all probability estimates.

The principle of spacelike conceptualization can be extended to other kinds of processes. To cite a nonactuarial illustration, the probability of snow at a specified time and place surely depends on such dimensions or parameters as geographical latitude, altitude, and season of the year. Note, however, the emergence of a rather interesting paradox. The more parameters we take into account, the more nearly we may come to a deterministic prediction of the actual results. Knowledge of current temperature, humidity, and wind conditions in the surrounding atmosphere may enable us to predict almost with certainty whether or not it is now or will soon be snowing. To some extent the same paradox applies in the case of underwriting, where detailed medical information might disclose the presence of a fatal illness. This paradox appears to be related to the earlier discussion of how to distinguish those conditions that alter R from those that alter θ.

Preface to the Quantum Interpretation

The discussion of FI has brought out several reasons for supposing that probability is objective: the possibility of measuring its value in cases of ostensibly strict time-homogeneity; the evident connection between changes in θ and objective changes in external conditions; and the presence of more or less measurable regularities in the spacelike variation of θ. The last reason, in turn, implies that methodologies for estimating probabilities can be objectively compared as to predictive ability.

At the same time, the discussion disclosed at least two anomalies with which an objective interpretation of probability must contend. One is that an objective probability cannot also be precise. This is shown both by the impossibility, in principle as well as in practice, of confirming the exact value of a probability for a time-heterogeneous process, and by the impossibility of distinguishing unambiguously those changes in conditions that may alter the value of θ from those that account for the very randomness of r about θ. Second, the notion of relevant conditions, carried to its logical conclusion, undermines the very notion of randomness. It implies that what we commonly take to be random outcomes are in reality fully determined results of antecedent causes.

The branch of modern physics known as quantum mechanics (QM) may hold the keys to unlocking these puzzles. Its uncertainty principle makes plausible the notion that an objective quantity may yet be intrinsically imprecise. Its principle of complementarity makes plausible the notion that a phenomenon may show two seemingly incompatible faces—in this case, random and caused. QIP will borrow these and several other features of QM in elaborating an interpretation of probability. Accordingly, let us review briefly some of the relevant aspects of QM before developing QIP in detail.

Quantum Mechanics

The essentials of quantum mechanics were developed in the 1920s after experimental research in atomic physics had disclosed phenomena that conflicted irreconcilably with classical mechanics. Quantum mechanics does not replace classical mechanics (especially Newton's laws of motion), but it does introduce several radically new concepts in order to make the atomic phenomena intelligible. Chief among the new concepts are the following:

1. *Quantization.*—The energies of particles and light waves cannot vary continuously but are restricted to integral multiples of a discrete "quantum" of energy. Changes in energy level occur in "quantum jumps."

2. *Uncertainty.*—There are theoretical limits to the precision with which certain physical measurements can be performed. These are not the practical limits of experimental error; they will apply to the most refined measuring devices that will or can ever be built. The limits apply to the simultaneous measurement of certain pairs of variables called conjugate pairs. The irreducible amount of inaccuracy is called uncertainty or indeterminacy, and the relationship between the uncertainties of paired variables is expressed by Heisenberg's uncertainty principle. For one such conjugate pair, momentum and position, the relationship is

$$\Delta p \Delta x \cong h/4\pi ,$$

where Δp is uncertainty in momentum, Δx is uncertainty in position, and h is Planck's constant. Replacing Δp by $m\Delta v$, since momentum is mass times velocity, we obtain

$$\Delta v \Delta x \cong \frac{h/4\pi}{m} .$$

For masses of macroscopic size, Δv and Δx are imperceptible, and the laws of classical physics seem to fit perfectly. However, for measurements of atomic motion, the uncertainties are appreciable. This means that, when one attempts precise measurements of velocity and position, the results are inconsistent with the basic laws of motion. The laws can be saved only by ascribing the appropriate degrees of uncertainty to the measurements.

Notice that the relationship is reciprocal, or "complementary" as physicists call it. This means that, depending on how one chooses to arrange the experiment, one can obtain arbitrarily small uncertainty in one variable, but only at the expense of increased uncertainty in its conjugate. Another important conjugate pair is time and energy.

To this day, the meaning of the uncertainty principle remains a subject of controversy. One minority of scientists holds that it expresses a transitory limitation on human knowledge of atomic phenomena, a limitation that will be removed by future advances in physics. Another minority holds that it expresses a permanent limitation on our knowledge but is not a characteristic of nature itself. However, a majority seems inclined to follow the so-called Copenhagen interpretation of Bohr and Heisenberg, according to which the relationship expresses an essential indeterminacy in the quantities themselves.

3. *Complementarity.*—Bohr subscribed to a yet more extensive principle of complementarity, which stems from the complementary relationship of the uncertainties. He argued that our direct experience of atomic events consists only of reading the results of a measurement from a dial. To interpret these measurements, we form pictures or representations of what we imagine to be the underlying physical reality, choosing that picture which is most consistent with the measurements. For example, one such picture was the early representation of the atom as a solar system, with electrons revolving about the nucleus. (This particular representation has been superseded by modern quantum mechanics.)

According to the principle of complementarity, no single picture suffices to represent all aspects of one phenomenon. For different experiments, conflicting pictures are required to represent the same phenomenon. A well-known example is the phenomenon of light. Some experimental results can be interpreted only by representing light as continuous waves, others by thinking of it as particles. The two pictures are mutually exclusive—a continuous wave cannot consist of discrete particles—yet the totality of what physicists know about light requires both pictures.

Bohr believed that the principle of complementarity might apply to other domains of science, where we often find mutually irreconcilable pictures. He cited the phenomenon of human consciousness. Psychologists study mental activity and behavior in certain ways; physiologists study the brain in other ways. Brain and mind are conflicting representations of the phenomenon, yet perhaps we need both representations.

III. THE QUANTUM INTERPRETATION OF PROBABILITY

Let ϵ be any set of n repetitions,[3] timelike or spacelike, of random binary experiments ϵ_i, and let $r = \Sigma\, r_i/n$.

1. *Quantization.*—The set R of the possible results of ϵ consists of $n + 1$ discrete members. R is "quantized," much as certain quantum-mechanical quantities are restricted to discrete states.

2. *Probability.*—The probability θ associated with ϵ is *measured* by r. In effect, θ is defined by a correspondence rule, which relates it to the frequency r. Note that θ is associated only with ϵ and not with any of the constituent ϵ_i's individually.

3. *Uncertainty principle* (*UP*).—The relationship between θ and its measurement r is subject to an inherent degree of indeterminacy $\Delta\theta$ as follows:

$$\Delta\theta = k\sqrt{[\theta(1 - \theta)/n]}\,,$$

where k is an arbitrary constant. $\Delta\theta$ is also called the "uncertainty," where uncertainty is understood in an objective sense.

UP expresses the idea, which arose repeatedly in the discussion of FI, that θ is objective but does not have a precise value. The formula for $\Delta\theta$ quantifies the degree of imprecision, in such a way that $\Delta\theta$ is complementary to $\sqrt{n}$. When n is relatively large, that is, the "mass" of data is great, the value of θ is relatively precise. Conversely, when n is small, θ becomes relatively imprecise. Indeed, as n approaches 1, θ becomes altogether indeterminate.

The quantity $\sqrt{n}$ plays a role roughly equivalent to "uncertainty in space-time." With $n = 1$ the result is localized to a single ϵ_i, while for large n the result is smeared over many ϵ_i's.

Although the value of k is arbitrary, it is generally useful to choose a value between 2.0 and 3.0, as will be explained. The fact that k is arbitrary parallels the UP of QM, in which the stipulated amounts of indeterminacy are typically expressed as the "root-mean-square error" of the measurement.

UP implies that the θ's associated with different experiments may be equal even though their measured results are not precisely equal. Thus, if ϵ_1 and ϵ_2 are repetitions of each other or are subsets of the same set of repetitions, then $\theta_1 \doteq \theta_2$ if $|r_1 - r_2| < \Delta\theta_1 + \Delta\theta_2$, meaning that θ_1 is equal to θ_2 within the limits of uncertainty.

4. *Inertia.*—The statements that $\theta_1 \doteq \theta_2$ if and only if ϵ_1 and ϵ_2 occur under similar conditions (inertia principle), and $\theta_1 \not\doteq \theta_2$ if and only if

[3] For convenience we will drop the * superscripts in the notation for repetitions, except where necessary for clarity.

ϵ_1 and ϵ_2 occur under dissimilar conditions, express relationships between probabilities and relevant conditions that are much used in estimating future results and in evaluating past results. As indicated in the previous discussion of life insurance underwriting, one usually assumes that future results will be consistent with past results, that is, that the probability will remain constant, *provided* that relevant conditions do not change. If one anticipates a change, however, one modifies one's assessments accordingly.

These statements also suggest a criterion for distinguishing homogeneous from heterogeneous sets of repetitions. Let $\{\epsilon_j^*\}$ be a set of experiments, *each* of which comprises a set of repetitions, and let $\Delta\theta_j$ be the uncertainty of ϵ_j^*. We shall define the set $\{\epsilon_j^*\}$ to be *homogeneous* if there is a quantity $\tilde{\theta}$ such that, for all or nearly all j, $|r_j - \tilde{\theta}| < \Delta\theta_j$; otherwise the set is defined to be *heterogeneous*. With $k = 2.0$–3.0, we will find that the resultant distinctions generally accord with our preconceptions.

Note that in some cases it may be possible to quantify the relationship between changes in conditions and changes in θ. For example, bending a coin will generally alter the probability of heads. It seems plausible to assume that the change in probability can be quantitatively related to parameters that describe the physical changes of the coin. (A suitable "randomizing" coin-tossing apparatus is assumed.)

5. *Complementarity.*—The verbal descriptions "random" and "caused" refer to complementary pictures of phenomena, that is, of the relationships among observable data. The identical data may appear in two alternative sets of experiments. The results of the respective sets are explained by alternative models or theories, one corresponding to our notion of randomness, the other to causality. Each picture and its corresponding theory can be internally consistent, but neither the random nor the causal picture/theory furnishes a complete explanation of the phenomenon.

The datum representing the death of an individual can appear among a set of mortality rates that, by our usual notions of randomness, seem to fit the picture of random fluctuation about an underlying θ. Alternatively, the same death may be the object of a medical examination that proves, on the basis of a theoretical model confirmed with respect to many similar deaths, that this death resulted from specific causal factors.

Accordingly, given a sequence of mortality rates $\{r_j\}$ of which one, say r_m, falls outside the interval $\tilde{\theta} + \Delta\theta_j$, there is no certain answer to the question whether r_m was a purely random fluctuation or was caused by a change in relevant conditions. Any criterion for distinguishing random from caused is arbitrary, in that it must willfully and one-sidedly exclude one or the other equally valid mode of description of the same event. To specify a value for k is tantamount to drawing a line, an inevitably arbitrary line, between the two sides.

Similarly, consider a coin tossed repeatedly by an automatic tossing device. The set of results may conform closely to the usual statistical model with probability p at each toss, yet in principle the outcomes of the tosses are predictable far in advance using the theoretical apparatus of physics.

So-called pseudorandom numbers are another case in point. The digits in the decimal expansion of an irrational number generally conform to any reasonable test of randomness, yet each digit is absolutely predetermined by algorithm. The digits are respectively random or caused, relative to two distinct sets of theoretical apparatus, mathematical probability and arithmetic.

6. *Estimation.*—An *estimate* of the probability associated with an experiment is conceptually distinct from the probability itself. The estimate is a subjective opinion as to the unknown outcome. The probability is θ, an objective quantity measured by r once the outcome is known.

When the experiment is a set of repetitions for which N is very large and $\Delta\theta$ negligible, the distinction accords well with commonsense notions of the relationship between estimate and true value. The result, once known, furnishes the true value of θ and either confirms or denies the correctness of the estimate. Furthermore, we can quantify the error of the estimate and rank different estimates according to their proximity to the result.

When ϵ is a single binary experiment, or in general when N is small and $\Delta\theta$ is not negligible, the situation is more complex. Now we have not two but three quantities to consider: the estimate, θ, and r. The estimate is equally an estimate of θ and of r; the possibility of different estimates for r and θ is excluded, since r measures θ. Now the value of r, once known, can validate the accuracy of the estimate in only an approximate way, since r is subject to the effects of indeterminacy. In this case, estimates cannot be unequivocally ranked.

Methods of estimation typically rely on the inertia principle and on observed regularities in the spacelike behavior of θ. No estimate is purely objective; consequently, no rigid standards of legitimacy versus illegitimacy can properly be imposed on a single estimate.

On the other hand, the accuracy of different sets of estimates can be objectively compared with respect to large aggregates of data. Let $\hat{E}(r_i)$ and $\hat{E}'(r_i)$ refer to distinct sets of estimates based on the same data. If $\Sigma_i\ [\hat{E}(r_i) - r_i]^2 < \Sigma_i\ [\hat{E}'(r_i) - r_i]^2$, and if the difference is large compared with $\Sigma\ (\Delta\theta_i)^2$, then the set $\hat{E}(r_i)$ is clearly superior to the set $\hat{E}'(r_i)$. If these sets can be characterized by distinct methodologies of estimation, then one methodology is clearly superior to the other.

7. *Spatial representation.*—A set of experiments ϵ may be characterized by a set of spacelike parameters $S = \langle S_1, \ldots, S_N \rangle$ in addition to a timelike parameter T such that each ϵ_i is located at a point in $S \times T$. Then θ may be represented as a function of space-time, $\theta(S, T)$, and the value of θ with respect to any region of $S \times T$ can be measured. Such measurements are of course subject to the uncertainty $\Delta\theta$, the value of which depends on the number of binary experiments located in the region.

In Section II we discussed at length the regular behavior of θ with respect to the S_j's, and there is no need to repeat ourselves here. It is important, however, to note that the results of any experiment or measurement are a characteristic of that experiment—that is, of the experimental arrangement chosen by the experimenter. The experimenter selects the particular set $\langle S_j \rangle$ by which he structures S and selects the specific partitions or regions in which he measures θ. Seemingly conflicting inferences may be drawn from the same or similar data under different experimental arrangements. Thus, the gross mortality rate in population A may be higher than that in population B, yet the age-specific rates may be equal. Although this general principle is well known, it has an especially striking expression in QIP, namely, that r (hence θ) is a characteristic of a specific ϵ and not necessarily of any other experiment.

8. *Mathematical probability.*—Mathematical probability is a strictly formal, axiomatic system. As such, it has no relation to events in the real world until an appropriate interpretation is placed on its primitive symbols.

The first rigorous statement of its axioms is due to Kolmogorov [17]. The following, slightly less technical version is from Savage [22]:

> A ***probability measure*** on a set S is a function $P(B)$ that attaches to each $B \subset S$ a real number such that
>
> 1. $P(B) \geq 0$ for every B;
> 2. If $B \cap C = 0$, $P(B \cup C) = P(B) + P(C)$;
> 3. $P(S) = 1$.

As Savage states, "This definition, or something very like it, is at the root of all ordinary mathematical work in probability." $B, C, \ldots$ are generally understood as observable events, which are members of a set S of all possible events, the so-called universal event. Nevertheless, the interpretation of the measure $P(\cdot)$ is left open.

As interpreted by QIP, the probability θ does not satisfy these axioms. θ does not, in general, have a precise value, and therefore cannot be equivalent to the axiomatic probability measure P, which is a real number, and hence precise.

QIP holds that the system of mathematical probability is an idealization that, relative to real events, is precisely valid only if $\Delta\theta = 0$, is approximately valid if $\Delta\theta$ is quite small, and is demonstrably invalid if $\Delta\theta$ is large. If a mathematical theory is empirically valid under a suitable interpretation of its primitive terms, then its implications should be empirically verifiable, as they are, for example, in geometry and classical mechanics. However, when an objective precise value is used for P, the predictions of mathematical probability are generally inaccurate.

Consider the frequency $f(m)$ of m heads in n tosses of a coin. If a precise value p is assigned to the probability of heads at each toss, and if the tosses are taken to be independent, then according to the mathematical theory,

$$f(m) = \binom{n}{m} p^m(1 - p)^{n-m}.$$

But the observed frequency $r(m)$ of m heads in N sets of n tosses each, will not, in general, agree with $f(m)$. Thus the observed results do not confirm the predictions of the theory. As $N \rightarrow \infty$, one supposes that $r(m) \rightarrow f(m)$, but this is only because the uncertainty associated with $f(m)$ goes to zero.

It may be contended that $r(m)$ is itself a sample from a random distribution $g_{mN}(d)$ that predicts the frequency of $d = r(m) - f(m)$, and hence that disagreement between $r(m)$ and $f(m)$ is in accordance with the theory. But the one set of observations $\{r(m); m = 0, \ldots, n\}$ in one experiment of N sets of n tosses will not confirm the accuracy of $g_{mN}(d)$. We could, of course, test g by N' repetitions of the N sets of m tosses. But a test of this sort is clearly the first step in an infinite regression. At each step, there is a discrepancy between result and prediction, and a next step is needed to "explain" the previous discrepancy. Thus, any very large but finite set of data will confirm only the approximate validity of the theory, and then only if the process is homogeneous. In the case of heterogeneity, the discrepancy between prediction and result generally will not diminish with increasing data.

Illustration: Radioactive Decay

The implications of QIP for *time-heterogeneous* experiments should now be clear. In the analysis of FI, we concluded that θ has no precise value for such experiments, and that changes in θ could not be unambiguously distinguished from the random variation of r about θ. The QIP principles of uncertainty and complementarity accord well with these findings.

Here we return to the time-homogeneous case, to consider the process of radioactive decay. This is a seemingly perfect, naturally occurring realization of the laws of mathematical probability and of FI; accordingly, it serves as a good test case for QIP in relation to time-homogeneous experiments.

The standard mathematical model for this process assumes that each atom of a species of radioactive isotope has a definite probability of decaying in a given time interval. This probability is constant and uniform for all atoms of that isotope. Thus, for a sample of N atoms, the number assumed to decay in time dt is $-dN = N\lambda dt$, where λ is the "disintegration constant" for that isotope. By mathematical integration, $N(t) = N(0)e^{-\lambda t}$. Constants related to λ are the *half-life*, $(\ln 2)/\lambda$, and the *mean life*, $1/\lambda$.

Let us assume that the model is well confirmed by a number of sets of repetitions, each involving samples of one isotope. In the case of carbon 14, for example, the published value of the half-life is 5,730 years, giving $\lambda = 1.2097 \times 10^{-4}$. According to the model, the fraction of carbon 14 atoms expected to decay in one year is also 1.2079×10^{-4}, since $1 - N(1)/N(0) = 1 - e^{-\lambda} = 1 - (1 - \lambda + \lambda^2/2 - \ldots) \doteq \lambda$. For a set of 1-gram samples (each of which contains about 4.3×10^{22} atoms), it is reasonable to expect that the observed fractions r_i will all equal 1.2097×10^{-4} to at least five significant digits; indeed, they should agree to about eight or nine digits. If such results were actually obtained for a large number of samples, and also for repetitions with the same samples, they would strongly corroborate the mathematical model and would indicate that the ϵ_i's are time-homogeneous and space-homogeneous, as presupposed by a constant value of λ.

On the other hand, such consistency of results could not normally be obtained for very small samples. If in another set of ϵ_i's each sample involved about 10^6 atoms, we could expect a substantial portion of the r_i's to lie outside the interval 1.10–1.32×10^{-4}, deviations that are certainly not negligible. How should we reconcile these quite dissimilar sets of hypothetical results?

For QIP, the second set of results refutes the supposition that the probability of decay in one year is precisely 1.2097×10^{-4}. Instead, QIP contends that the probability with respect to the first set is indeed 1.2097×10^{-4} with negligible uncertainty, but that for the second set the probability is 1.2097×10^{-4} with uncertainty of about 0.3×10^{-4} (using $k = 3$). The first set demonstrates that the process is both time-homogeneous and space-homogeneous at the *macroscopic* level. For the second set, it is *only* if appropriate allowance is made for uncertainty that the data support the assumption that the θ_i's are all equal and that the process is time- and space-homogeneous at the microscopic level.

The conclusions are sensitive to the value chosen for k. If a much smaller value were chosen for k, say $k = 1$, a substantial portion of the r_i's would contradict the hypothesis that the θ_i's are all equal. There would be significantly many instances for which the model would be contradicted, and we would have to abandon the model at the microscopic level. But the model remains our best tool for predicting $N(t)$ at *any* level. Also, as opposed to the case of manifestly heterogeneous processes, we cannot in this case associate the variation in the r_i's with any identifiable change in conditions. We would like to maintain that this process is homogeneous. This can be achieved by attributing the appropriate uncertainty to θ_i, in which case $r_i \doteq \theta$, except for a very few outliers. On the other hand, suppose that a much larger value were selected for k, such as 10. This would have the slight advantage of eliminating essentially all outliers in the case of homogeneous processes. It would, however, lead to seriously misleading conclusions for manifestly heterogeneous processes by concealing significant deviations in the r_i's. Thus the selection $k = 3$ is an arbitrary distinction, which nonetheless maintains that distinction between homogeneous and heterogeneous results that accords with our general preconceptions.

A proponent of FI, or of another precise, objective interpretation of probability, must deal with the fact that observed data, especially at the microscopic level, do not precisely agree with the model. The fact that agreement is so nearly precise at the macroscopic level in no way proves that either θ or λ is precise at the microscopic level. Science furnishes a number of instances in which facts, theories, or models that are valid at one level of observation are invalid, or only approximately valid, at another level. Newtonian mechanics is, of course, a prime example.

Three other possible lines of defense for a precise probability at the microscopic level are also defective. The first, that the deviations $r - \theta$ conform to the laws of mathematical probability, cannot be confirmed at that level of observation. As we saw earlier, an attempt to demonstrate that the deviations do so conform must entail an infinite regression to ever higher levels of observation. Furthermore, this effort cannot prove that θ did not change value with respect to any single deviation.

A second line of defense might be based on the alleged physical homogeneity of the atoms of the isotope. This fails, however, because, while the atoms are homogeneous in certain respects, such as mass and atomic number, they are heterogeneous in other ways. In particular, their nuclei can occupy different energy levels and can move from one level to another.

Finally, it could be argued that the phenomenon of radioactive decay is explained by the well-confirmed theory of QM, that this explanation utilizes an exact wave function, usually denoted by ψ, and that $|\psi^2|dS$ is interpreted as an exact probability. But confirmation of the precise value of ψ at a point in space-time is no more possible than precise confirmation of λ by observation of one atom. Both ψ and λ are precisely confirmable only with respect to large ensembles of events.

Accordingly, it seems that nature's most nearly homogeneous random process can be successfully interpreted by QIP. The predictive success of the exact model is not itself exact, and the uncertainty principle is upheld.

IV. THE SUBJECTIVIST INTERPRETATION AND BAYESIAN STATISTICS

No brief discussion of the subjectivist interpretation (SI) and Bayesian statistics (BS) can do them justice. Their proponents have written extensively on the philosophical and conceptual foundations of probability and statistics. Here I can but outline the development of BS from its origin in SI, following for the most part the discussion in [9], especially chapters 11 and 12.

As previously noted, SI considers probability to be one's personal degree of belief that a certain event will occur or that a certain statement is true. Probability is not, therefore, a frequency, an objective parameter, or any other objective attribute of an event, or set of trials or random experiments. Different persons are at liberty to assign different personal probabilities to the same event, and an individual is free to revise his opinion as he acquires additional information. The sole constraint imposed is that the individual be rational—that his probabilities for a set of events be *coherent*. For two mutually exclusive events A and B, the conditions for coherency are $0 < P(A) < P(S) = 1$ and $P(A \cup B) = P(A) + P(B)$, where $P(S)$ is the tautological or universal event [11].

Following subjectivist convention and referring to the subject as "You," we can operationally define Your $P(A)$ by asking what odds You will bet for and against A, where *we* then choose which side (for or against) You must take. Note that by the principle of coherence, Your $P(A)$ must equal $1 - P(\text{Not } A)$.

How should You revise Your opinion? DeFinetti introduces the notion of "exchangeable events," which are of particular significance for the orderly revision of opinion. This notion refers to any instance of a set of n trials where, prior to the occurrence of the first, You judge that every combination of exactly m successes and $(n - m)$ failures has the same probability, regardless of order. For example, prior to three tosses of a coin, You may judge that $P(htt) = P(tht) = P(tth)$, regardless of the specific values of $P(h)$ and $P(t)$. Of course, this is equivalent to the subjective judgment, prior to the occurrence of the first event, that $P(h)$ is the same at each trial. (We stress ***prior*** to the first trial, since we shall see how, using Bayes's theorem, You revise Your probability judgments after some of the trials occur.)

DeFinetti next considers the case of an urn containing an unknown mixture of white and black balls. Let $H_1, H_2, \ldots, H_M$ represent a set of mutually exclusive, but exhaustive, hypotheses about the proportion of white balls, where H_i is the hypothesis that the proportion is θ_i. Also, let us assume, as deFinetti seems to, that the probability of drawing white is equivalent to the proportion of white balls. Next, assume that a sequence of N drawings, with replacement, is a set of exchangeable events. Let $f_N(X \mid \theta_i)$ denote the conditional density function, the probability of drawing X white balls in N drawings, given θ_i. Since the drawings are exchangeable,

$$f_N(X \mid \theta_i) = \binom{N}{X} \theta_i^X (1 - \theta_i)^{N-X} .$$

The subjectivist-Bayesian approach explicitly provides for subjective judgments about the composition of the urn, expressed as prior probabilities of H_i. Bayes's theorem serves as the mechanism by which Your prior probabilities may be combined with observations of the data to produce modified or posterior judgments. Let $p(\theta_i)$ denote Your prior probability that the proportion is θ_i, prior to observing any drawings, and let $P(\theta_i \mid X)$ denote the posterior probability. These are related through Bayes's theorem as follows:

$$P(\theta_i \mid X) = \frac{p(\theta_i) f_N(X \mid \theta_i)}{\Sigma_i \, p(\theta_i) f_N(X \mid \theta_i)} .$$

This methodology is quickly extended to cases in which θ is a member of a set of continuous real numbers, instead of being restricted to a set of discrete numbers. Now Your prior opinion about θ is expressed as the prior density $p(\theta)$. Again, the conditional density of X given θ is $f(X \mid \theta)$. After the observations of some data X, the posterior density $P(\theta \mid X)$—Your updated estimate of θ—is

$$P(\theta \mid X) = \frac{p(\theta) f(X \mid \theta)}{\int_\theta p(\theta) f(X \mid \theta) d\theta} ,$$

and Your posterior expected value of θ is

$$E(\theta|x) = \int_\theta \theta P(\theta|X)d\theta .$$

A common objection to this procedure is that the use of a prior subjective probability is vague and variable, and therefore useless for scientific purposes. Bayesians respond that, where extensive observations are available, the form and properties of the subjective prior density have negligible effect on the posterior. Two persons with widely divergent prior opinions but reasonably open minds will be forced into arbitrary close agreement about future observations by a sufficient amount of data [11]. (We will refer to this as the "principle of overwhelming data," or POD.) Even so, a strict subjectivist like deFinetti would consider general agreement that $\lambda = 1.2097 \times 10^{-4}$ for carbon 14 to be coincidence of opinion, not objective fact.

Furthermore, Bayesian statistics seems especially useful when few data are available. For example, suppose that You observe 100 tosses of a coin, of which 65 are heads and 35 tails, and You are asked to bet on the 101st toss. Bayesian statistics permits You to consider both the observed data and Your prior opinion about the coin in deciding how much to bet. You might base Your original opinion of the coin on a physical inspection and, if it seems "fair," Your $p(\theta)$ might be a distribution centered about 0.5. Then, after the data are obtained, Your resultant posterior distribution would be skewed somewhat toward heads, with the amount of skewness dependent on the relative sharpness of $p(\theta)$. Non-Bayesian statistics does not permit such combination of data and prior opinion.

Above all, it is the use of subjective prior judgments $p(\theta)$ that distinguishes Bayesian statistics from classical statistics [19]. However, consistency demands that a subjectivist-Bayesian must also regard the conditional density $f(X|\theta)$ as subjective ([11], p. 199).

Critique

My criticisms of SI and BS should not be construed as an attempt to refute them. QIP provides both for objective probabilities and for estimates of probability, and I am willing to identify the latter with subjective probabilities. The differences between QIP and SI appear to be partly semantic, although a strict subjectivist might not agree.

Moreover, SI offers some definite advantages. Any statement of the form "The probability of H is P," where H is a singular proposition, is easily expressed in SI. For many H's, such statements are at best awkward in QIP and, a fortiori, in FI. Edwards [11] gives as an example the probability that "weightlessness decreases visual acuity."[4] Contingent on a suitable definition of terms and agreement on method, this hypothesis can be proved either true or false. Before the result is known, what is the probability that the hypothesis will be proved true? Some objectivists would consider this question to be nonsense, while others would assert that $P = 1$ or 0, depending on whether H is in fact true or false. For subjectivists, however, it is quite appropriate to think of such a probability, as a measure of one's degree of belief in the truth of the proposition.

QIP is not adept at dealing with the probability of a formally singular proposition. Considered as a random quantity, such a P is viewed as objective, but the amount of uncertainty to which it is subject renders it effectively indeterminate. QIP also permits subjective estimates of this probability, and such estimates may be construed as being equivalent to SI probabilities, though expressed somewhat more awkwardly.

QIP is less concerned with formally singular propositions than with probability as an attribute of sets of repetitions; it is less concerned with estimates than with the objective relationships among the experimental results for various subsets.

Note that even in the case of the proposition "weightlessness decreases visual acuity," some advantage may be gained by considering this H to be a member of a set $\{H_i\}$, where H_i means "weightlessness decreases visual acuity in species i" and i indexes (man, gorilla, chimpanzee, . . .). Alternatively, H_i could mean "G_i decreases visual acuity relative to G_n" where G_i is a set of values for the force of gravity, from 0 to G_n at the earth's surface. If H itself cannot be tested, experimentation with respect to other members of these sets should certainly assist You in fixing Your opinion. The point is that, in this example at least, an ostensibly singular proposition *can* be placed in a set of repetitions, whereupon it becomes amenable to QIP.

Nevertheless, the differences between SI/BS and QIP are most striking when we consider first space-heterogeneity and then time-heterogeneity. The QIP representation of probability as a function of a set of spacelike parameters is altogether lacking in SI. This conception, together with the emphasis on experimentation, compels an insurance company to test whether, for example, probabilities of death vary by age and other factors. But from an SI point of view, the company that does not discriminate by age may be as "coherent" as any other; if exchangeability (of insurance applicants) is purely a subjective judgment,

[4] Note that this paper appeared before the era of space travel.

that company has as much right to ignore age as another company has to consider it. It may apply Bayes's theorem faithfully to its data yet be bankrupted by antiselection.

The subjectivist conception, "which leaves each individual free to evaluate probabilities as he sees fit, provided only that the condition of coherence be satisfied" [7], is simply not good enough. Indeed, it is easy to show that a methodology that is slightly incoherent but takes reasonable account of space-heterogeneity will give much better predictions than one that is quite coherent but ignores space-heterogeneity.

In addition, the QIP spacelike conception provides a rationale for basing probability judgments on data that evidently are *related* though not exchangeable, and it permits one to abstract from the data a quantitative measurement or estimate of the effect of a parameter considered in isolation from specific events.

Similar conclusions apply to SI in the case of time-heterogeneity. If exchangeability is purely a subjective determination, an insurance company can conclude freely that the experience of *all* past years is exchangeable and can set its rates accordingly. The financial results for its annuity business would be disastrous! QIP maintains that past mortality experience is demonstrably time-heterogeneous and that expectations of future results should be based on recent experience, adjusted for expected future changes in conditions. According to SI, either events are exchangeable or they are not, the concept of partial exchangeability notwithstanding ([9], p. 212); QIP holds that we can often measure, at least approximately, the *degree* of time-heterogeneity—that, for example, we can measure the trend of θ—and take account of the gradual effect of changes in conditions. Such notions seem inexpressible in SI.

Similarly, BS runs into trouble when θ (now a parameter, not a simple hypothesis) is time-heterogeneous. The logical derivation of BS from SI presupposes that θ is both objective and time-homogeneous. DeFinetti is explicit on the first point: a subjective probability must refer "exclusively to facts and circumstances which are directly verifiable, and of a completely objective, concrete and restrictive nature" ([9], p. 201). A prior density $p(\cdot)$ is a set of subjective probabilities; accordingly, the θ to which it refers should be *objective*. DeFinetti is not explicit on the second point, but the example of the urn and other illustrations are compatible only with time-homogeneity. Other Bayesians are generally unconcerned about the nature of θ—for example, Lindley [19] calls it simply a parameter—and seem ready to apply the Bayesian paradigm regardless of whether or not θ meets these conditions.

I would like to show that the Bayesian paradigm does not work when

θ is time-heterogeneous, but it will be worthwhile to analyze all four possibilities: θ either objective or subjective, and either homogeneous or heterogeneous.

1. *θ is objective and time-homogeneous.*—This situation presents no difficulty. Let θ correspond to any fixed quantity that, in principle, is precisely measurable. Suppose that θ is measured by an apparatus such that the experimental error of a measurement r is assumed to follow a density $f(r|\theta)$, and the error of the mean $\bar{r}$ of n measurements is assumed to follow $f_n(\bar{r}|\theta)$, derivable from f. In a straightforward manner, using Bayes's theorem, You can modify Your prior opinion $p(\theta)$ about the true value of θ, using the observation $\bar{r}$ to form the subjective posterior $p(\theta|\bar{r})$.

If n is very large, $p(\theta|\bar{r})$ becomes a thin spike centered approximately at θ, thus illustrating the principle of overwhelming data. In principle, as $n \to \infty$, $E(\theta|\bar{r}) \to \theta$, consistent with POD.

According to SI, $f(r|\theta)$ is essentially subjective. In many practical applications, there is general agreement about it; then $f(r|\theta)$ is "public" [11]. QIP has a different interpretation of this function. It considers $f(r|\theta)$ as an objective characteristic of the measuring apparatus used to measure θ. If the apparatus is poorly calibrated, it is *not* the case that $E(r|\theta) = \theta$, and POD does *not* apply. Moreover, the variance of $(r - \theta)$ is clearly a reflection of the *accuracy* of the apparatus. The point is that the apparatus may be calibrated using another quantity, θ', whose true value is known to high precision. Since both θ' and r are objective, we can experimentally determine (a) that the characteristics of the measuring apparatus itself are time-homogeneous; (b) that its measurements are not systematically biased; and, provided that n is large and the uncertainty small, (c) the approximate *objective* character of $f(r|\theta')$. Nevertheless, it is clear that, with enough data, $E(\theta|r) \to \theta$ regardless of the specifics of *either* $p(\theta)$ or $f(r|\theta)$ so long as there is no systematic bias in the measurements. In other words, even if You make poor subjective judgments about either the prior density or the likelihood, large quantities of data will save You—but only if the measurements are *objectively* unbiased.

2. *θ is objective but time-heterogeneous.*—Your body temperature is such a quantity. You may use BS to estimate it, combining prior judgments with measurements taken using a thermometer that is subject to experimental error. In illustrating the application of BS, Edwards et al. [11] used the same example, but they quite overlooked the fact that one's temperature is subject to daily cyclical variations and to other irregular changes.

Suppose that, like Edwards et al., You erroneously overlooked the time-heterogeneity of the temperature process and applied the Bayesian technique as if θ_T were constant. No great harm would result, provided that You took all n measurements within a relatively short time interval —an interval Δt such that changes in the true value of θ were quite small compared with $|r - \theta|$, the experimental error of Your thermometer, and also fairly small compared with $|\bar{r} - \theta|$. Relative to this experiment, θ_T is virtually constant, and this case is approximately equivalent to case 1. If n were sufficiently large, Your $E(\theta|\bar{r})$ would be a very good estimator for the true values of θ_T within time Δt, although it might be a very poor predictor for values of θ_T in the future.

On the other hand, if Δt is relatively long and $\Delta\theta$ is not negligible compared with $|r - \theta|$, various difficulties can ensue that no amount of data can cure. The specific problems will depend on the timing of the measurements relative to the daily cycle, but in general we can say that $E(\theta|\bar{r})$ generally will not give good predictions for θ_T and that different subjects may arrive at posterior distributions that are each relatively sharp yet correspond to quite distinct estimates of θ. POD will not work in this case.

If You do take time-heterogeneity into account, You may still face serious difficulty in making good estimates of θ. For example, You may assume that the true θ_T is of the form $\alpha(T) + \beta$, where α is a specified time-dependent function of T, and β is an unknown constant, whose value You will estimate by applying the Bayesian paradigm. Your results will continue to be poor unless You make a lucky guess for α; that is, unless Your $\alpha(T)$ differs from the true θ only by a constant. You will know whether Your choice was a good one *only* by subsequent empirical testing.

Obviously it would be better to analyze some data first, without any preconceptions about the form of θ_T. The data might indicate that θ_T is subject to both daily and thirty-day cycles. Then You might assume that $\theta(t)$ is of the form $a_0 + a_1 \cos(t - t_0) + a_2 \cos(30t - t_1) + e_t$, where a_0, a_1, a_2, t_0, and t_1 are unknown constants and e_t represents noncyclical irregularity. By means of least-squares techniques, You could estimate values for the unknown constants. The least-squares procedure guarantees that the resultant estimator $\hat{E}(\theta_t) = a_0 + a_1(t - t_0) + a_2(t - t_1)$ is optimal relative to the known data—optimal in the sense that for this $\hat{E}$, $\Sigma_{\text{data}} [\hat{E}(\theta_t) - \text{data}]^2$ is less than for any other estimator of the same form. You can and should test its continued accuracy relative to future data.

Some interesting questions arise concerning the correct interpretation of the "unknown constants" and of the term e_t, but we will set these aside in the interest of brevity. The important point is that, when θ is time-heterogeneous yet objective, the Bayesian paradigm does not guarantee acceptable results. You may be forced to use non-Bayesian techniques, and in any event only empirical testing can confirm the reasonableness of Your estimates.

3. *θ is subjective and time-homogeneous.*—This case poses no practical difficulty, for BS works here much as it does in case 1, but it does raise some theoretical questions. Consider again the process of radioactive decay. Let θ correspond to the one-year probability of decay for a large mass. If, as SI contends, θ is a subjective quantity, then $p(\theta)$ must be translated as Your prior opinion concerning Your opinion or degree of belief in the proposition that an atom will decay in one year.

But the notion of an opinion about an opinion is nonsense. BS works in this case only because θ is objective (or so contends QIP) and is rather accurately measurable. Similarly, $f(r|\theta)$ can be tested empirically, although precise confirmation is limited by the uncertainty principle.

4. *θ is subjective and time-heterogeneous.*—This case combines the difficulties of cases 2 and 3. BS is both ineffectual and illogical. As in case 2, BS will work approximately if θ is nearly homogeneous. "Nearly homogeneous" means that $\Delta\theta$ is negligible compared with $f(r|\theta)$ and small compared with $f(\bar{r}|\theta)$. If, however, the process is severely heterogeneous, You face severe difficulties.

a) As in case 3, $p(\theta)$ is Your subjective opinion about an inherently unobservable, unmeasurable quantity. If this quantity is understood to be subjective, then the interpretation of p makes no sense. QIP considers θ to be objective but imprecise, and it is hard to envision how You fix Your opinion about a quantity that is inherently unobservable and has no precise value.

b) The density $f(r|\theta)$ loses its significance. As understood by SI, this function means the subjective probability of r given the hypothesis $\theta_T = \theta$, but θ has no precise value and it is hard to appreciate on what basis You can make a specific choice for f. When $\Delta\theta$ is not small compared with the putative deviations $|r - \theta|$, the distinction between $r - \theta$ and changes in θ is hopelessly blurred, and f has no objective significance.

c) Undaunted by the conceptual problems entailed by the functions p and f, You may decide to proceed to make assumptions for them and to apply the time-homogeneous Bayesian paradigm. Your results generally will be quite poor by any objective test. This is evident from the discussion of case 2.

It is not clear how You can remedy this situation within the framework of BS. To the best of my knowledge, the problems entailed by time-heterogeneity have been ignored by the Bayesians. Approaches analogous to those suggested for case 2 are considerably more difficult to apply here. Any pattern for the variation in θ will tend to be obscured by the random variations $(r - \theta)$—assuming that one were entitled conceptually to distinguish between these. You might assume that θ has a parametric form in terms of a set of unknown constants and apply the Bayesian paradigm to these constants, but this leaves open many of the same questions: Are these really constants? Are they objective or subjective? On what basis can You formulate an opinion about them? etc.

If, notwithstanding these practical and conceptual difficulties, You forge ahead and somehow derive Your posterior distribution $p(\theta_T|r)$, I think You would be well advised to test it empirically. Of course, if You are a strict subjectivist, You may deny that Your opinion about θ_T is testable. In a sense, QIP agrees, since θ_T is subject to too much uncertainty. Then let us agree to test how well Your posterior distribution predicts future values of r. If You do not agree to submit Your estimates to empirical validation, I have no further interest in them, nor in the mathematical model for θ_T from which You derived them.

The conclusion is that BS has both theoretical and practical shortcomings. When θ is not objective, the density $p(\theta)$ and the likelihood $f(r|\theta)$ cannot be sensibly interpreted. When the process is time-heterogeneous, posterior functions $p(\theta|r)$ and $E(\theta|r)$ may furnish very poor estimates of future results. Empirical testing is required to confirm the reasonableness of the estimates, and it may be easier to dispense with Bayesian methods altogether. These points will be illustrated in the next section.

V. CREDIBILITY THEORY—A CASE STUDY

The subject of credibility theory is an interesting testing ground for the rival interpretations of probability as applied to time-heterogeneous processes. It illustrates how QIP can lead to conclusions and practical approaches radically different from those of either FI or BS. First let us trace the FI and BS development of this subject, and then look at it from the perspective of QIP.

It is impossible even to state the purpose of credibility theory in a noncontroversial way, so let us initially describe the problem as it was seen by A. W. Whitney, the originator of credibility theory [26]. Whitney was interested in drawing certain inferences about the claim experience for a set of workmen's compensation risks. He proposed a statistical model for the claim process, essentially as follows:

1. Each risk within a class of broadly similar risks can be characterized by a unique quantity or parameter, which Whitney called the "true hazard" of that risk. Let us denote that quantity by θ.
2. Within the class of risks, the parameter θ has a probability density $p(\theta)$. Whitney assumed that p is normal with mean μ_θ and variance σ_θ^2.
3. The conditional density of the claims X is $f(x|\theta)$. Whitney assumed this to be binomial with variance n^2, where n is the number of persons covered in the risk.

Using what then was called the method of inverse probabilities (Bayes's rule), Whitney showed that the most probable value of θ can be approximated by an expression of the form $E(\theta|x) \doteq zX + (1 - z)\mu_\theta$, where z is a function of μ_θ, σ_θ, and n. The quantity z is known as the credibility factor. It expresses how much weight or "credibility" should be given to the observation X in relation to μ_θ.

Whitney's approach to credibility seems essentially frequentist. The observed claims X are randomly distributed about a constant "true" and presumably objective parameter similar to the time-homogeneous FI conception of probability. This approach shapes his view of the purpose of credibility theory: to estimate the value of this unobservable parameter, the true hazard of the risk.

Note that Whitney's model includes two untestable distributions, the normal distribution p and the binomial distribution f. His results contain two unobservable parameters, μ_θ and σ_θ. The parameter μ_θ presumably can be estimated as $\bar{X}$, the observed mean claims for the class, but it is not so easy to see how the other parameter, σ_θ, can be estimated.

This model obviously lends itself to other kinds of insurance. Similar approaches, still essentially time-homogeneous and frequentist, were developed by Keffer [16], Shur [24], Hewitt [13], and others for other specific kinds of insurance or for claim processes generally. There are minor differences in the formulas for z depending on the specific assumptions for p and f. The results still include unobservable parameters.

Bailey [1] and Mayerson [21] approached credibility theory from the standpoint of BS. Mathematically, their approach is rather similar to that of the frequentists, but the interpretations placed on some of the symbols are different. The function $p(\theta)$ is now, of course, the subjective prior density of θ, representing Your initial opinion about θ. Presumably $f(X|\theta)$, the likelihood of X given θ, is also subjectively chosen. For the Bayesians, the purpose of credibility theory still is to estimate θ, but it is not altogether clear what they mean by θ and, specifically, whether it represents an objective or a subjective quantity. Their results are also similar to those of Whitney [26], still containing several unobservable parameters, and still taking no account of time-heterogeneity.

Finally, mention should be made of two recent contributions of Hans Bühlmann to the mathematics of credibility theory. First, whereas earlier formulas for z depended on the specific form of $f(X|\theta)$, Bühlmann was able to prove [4] the existence of a best least-squares linear approximation to $E(\theta|X)$, independent of $f(X|\theta)$ when the process is time-homogeneous. The credibility factor z equals $1/(1 + X)$, where $X = \sigma_f^2/\sigma_p^2$, the ratio of the variance of the likelihood function to the variance of the prior distribution. Bühlmann calls θ a "risk parameter" and seems to consider it an objective characteristic of the risk. Unlike subjective Bayesians, he also appears to view $p(\cdot)$ as an objective characteristic of a collection of risks. He calls it a "structure function," describing the idealized frequency of θ in the collective. Later, Bühlmann proposed methods to estimate the values of σ_f^2 and σ_p^2 from actual data [3].

Second, Bühlmann apparently was the first credibility theorist to appreciate the possibility of time-heterogeneity [5]. He considered one special case of time-heterogeneity: an exact, known linear trend (similar to inflation) applying uniformly to all risks, which are otherwise time-homogeneous. Of course, this case is not essentially different from the purely time-homogeneous case, for, by "untrending" the sample claim data, one has transformed the problem back to that of strict time-homogeneity.

The literature on credibility theory is much more extensive than this brief history might indicate. For example, the bibliography published in reference [10] contains 139 entries. Also, some of the mathematical development is much more elaborate than this history might suggest (see, for example, Jewell [15]).

Yet, for all the breadth and depth of the theoretical analyses, a realistic assessment of the practical value of credibility theory to date must find it wanting. For approximately the first fifty years of its history, the credibility formulas derived by the theory were expressed in terms of unobservable parameters and presupposed strict time-homogeneity. Finally, Bühlmann showed how to estimate the values of these parameters (assuming they exist), but only for the strict time-homogeneous case plus one trivial exception. Few of the published papers consider any sort of sample data and, generally, such sample data are hypothetical data, often generated by the preconceived theoretical models (e.g., Bolnick [2] and Cabral and Garcia [6]). One possible exception is Bühlmann himself [3]—the source of his data is not clear—but he fails to show that the data conform to his presupposition of time-homogeneity, and Taylor later concludes that these data indicate some type of time-heterogeneity [25].

The fact is that typical problems in experience rating are inordinately more complex than the theory presupposes. Group major medical expense insurance (GMME) is a case in point. No actuary familiar with GMME can reasonably doubt that the claim amounts X_{st}, expressed as nominal dollars, are heterogeneous with respect to both risk s and time t. Among the many factors contributing to heterogeneity are differences in benefit plans, differences in the age and sex composition of the risks, and inflation.

In order to make the data more tractable, the group actuary has recourse to a manual premium rating system (MPRS) based on his analysis of the timelike and spacelike variation of claim data. The MPRS embodies the actuary's estimate of how X_{st} varies with respect to differences in the plans of benefits, differences in the characteristics of the group, and time. From the MPRS the actuary derives the manual premium M_{st} for a risk characterized by $s = \langle s_1, \ldots, s_n \rangle$ in year t. Then he can express the claims as loss ratios to manual premium, that is $X'_{st} = X_{st}/M_{st}$.

The MPRS is constructed in such a way that the actuary's prior estimate of X'_{st} is a constant c, independent of s and t. Like the individual life underwriter's estimate of a probability of death, this estimate derives largely from objective data, the data that underlie the MPRS. Nevertheless, it is inevitably influenced by subjective judgment as to how to organize the data and how to take account of unusual plans or other characteristics for which there are meager data.

Expressing the claims as loss ratios X'_{st} in a well-chosen MPRS undoubtedly has the effect of reducing some of the elements of timelike and spacelike heterogeneity. Nevertheless, substantial time-heterogeneity remains. In a previous paper [20], I presented sample loss ratios for some 1,000 GMME cases exposed in 1967, 1968, and 1969. These data suggest that the sample correlation coefficients between X'_{t_1} and X'_{t_2} are a decreasing function of $|t_2 - t_1|$. The data do not support the proposition that the correlation coefficients are independent of t_1 and t_2, which would be the case if the X'_t's were time-homogeneous.

It seems quite doubtful that any of the models of credibility theory can be usefully applied to such data. In the all too simple cases that Whitney and others had in mind, one might at least hope to be able to make a suitable guess for $f(X|\theta)$ or for σ_θ. However, I doubt that anyone has even a vague idea of the form of $f(X|\theta)$ when X and θ are expressed as loss ratios in an MPRS. Furthermore, these loss ratios continue to exhibit time-heterogeneity, even through they have been adjusted for all the factors known to cause time-heterogeneity. The alternative, to apply the model to $\{X_{st}\}$ expressed in nominal dollars, seems even worse, since here the data are chaotically time-heterogeneous. Accordingly, I submit that credibility theory can lead only to a dead end, at least with respect to substantially time-heterogeneous risks.

What, then, are the implications of QIP for credibility and experience rating? Above all, QIP argues that it is illegitimate to make assertions about the behavior of individual risks, and, in particular, to assert that the claims X_{st} behave in a specified way with respect to a hidden parameter θ_{st}. In an even moderately time-heterogeneous situation, supposed distinctions between random deviations $X_{st} - \theta_{st}$ and changes in θ itself have no basis in fact. It seems inconceivable that any model purporting to portray such distinctions can be empirically validated. Rejecting the notion of a precise θ, QIP reformulates the problem of credibility as follows: to estimate the future claims of a risk (and not some shadowy parameter), on the basis of its own actual past experience and the known experience of other similar risks.

These principles lead almost directly to the least-squares regression approach that I proposed previously [20]. The regression approach is based on a few rather plausible presuppositions (p. 264) about the aggregate statistical behavior of a set of risks. Plausible or not, these assumptions are empirically testable, and the actual GMME claim data of that paper tend to corroborate them. The resultant formulas for the credibility factors are expressed entirely in terms of observable data: sample variances and covariances of past claim data.

Here, then, is at least one important practical problem in which the matter of interpretation does make a significant difference.

VI. CONCLUDING REMARKS

From a pragmatic point of view, the ultimate test of any interpretation of probability is whether it contributes to the solution of practical problems. A major contention of this paper is that frequentist and Bayesian approaches generally will fail to solve practical problems involving time-heterogeneous random processes, at least where the degree of time-heterogeneity is not negligible. The random processes with which actuarial science is concerned, especially mortality and morbidity, are certainly not time-homogeneous. In some instances the degree of time-heterogeneity may indeed be negligible. In other cases the actuary may be satisfied with rough estimates of the probabilities in question or may rely on margins of conservatism to compensate for faulty estimates. However, the problem of credibility illustrates one case where time-heterogeneity is not negligible and where a more precise solution is desired. There may be, or may arise, other similar problems.

The quantum interpretation per se does not offer an explicit methodology for solving such problems, but it does suggest some useful guidelines. First, one should try to ascertain the severity of time-heterogeneity. Is the uncertainty in the value of the underlying parameter significant, or negligible, in comparison with the variation of the observed data about this parameter? General familiarity with the random process in question may enable one to make a determination. In other cases, some sort of empirical test may be necessary. In the case of group health insurance claim data, severe time-heterogeneity was confirmed by showing that the sample correlation coefficients between X_{st} and $X_{st'}$ are strongly dependent on $|t - t'|$.

Where time-heterogeneity is not negligible, solutions that presuppose a precise underlying parameter generally will not work. Some other type of estimation technique will be necessary. A simple linear regression approach works in the case of group insurance credibility. Other techniques may be useful in other types of situations. Regardless of the estimation technique employed, some sort of empirical test should be used to confirm its validity.

In any case, I hope that this paper will stimulate actuaries to think further about the nature of risk processes and about the relationship between mathematical probability and the actual events generated by these processes.

REFERENCES

1. BAILEY, A. L. "Credibility Procedures," *PCAS*, XXXVII (1950), 7.
2. BOLNICK, H. J. "Experience-rating Group Life Insurance," *TSA*, XXVI (1974), 123.
3. BÜHLMANN, H. "Credibility for Loss Ratios," *ARCH*, 1972.
4. ———. "Experience Rating and Credibility," *ASTIN Bulletin*, IV, Part 3 (1967), 199.
5. ———. "Experience Rating and Credibility," *ASTIN Bulletin*, V, Part 2 (1969), 157.
6. CABRAL, M., and GARCIA, J. "Study of Factors Influencing the Risk and Their Relation to Credibility Theory," *ASTIN Bulletin*, IX, Part 1 (1977), 84.
7. DEFINETTI, B. "Foresight: Its Logical Laws, Its Subjective Sources," in *Studies in Subjective Probability*. New York: John Wiley & Sons, 1964.
8. ———. "Probability: Interpretations," in *International Encyclopedia of the Social Sciences*. New York: Macmillan Co., 1968.
9. ———. *Theory of Probability*. Vol. II. New York: John Wiley & Sons, 1975.
10. DEPRIL, N., et al. "A Bibliography on Credibility Theory and Its Applications." Instituut voor Actuarielle Wetenschappen, Katholieke Universiteit Leuven, 1975.
11. EDWARDS, W.; LINDMAN, H.; and SAVAGE, L. J. "Bayesian Statistical

Inference for Psychological Research," *Psychological Review*, LXX (1963), 193.

12. FINE, T. L. *Theories of Probability: An Examination of Foundations*. New York: Academic Press, 1973.
13. HEWITT, C. C. "Credibility for Severity," *PCAS*, LVII (1970), 148.
14. HOEL, P. G.; PORT, S. C.; and STONE, C. J. *Introduction to Probability Theory*. New York: Houghton Mifflin Co., 1971.
15. JEWELL, W. S. "The Credible Distribution," *ASTIN Bulletin*, VII, Part 3 (1974), 237.
16. KEFFER, R. "An Experience Rating Formula," *TASA*, XXX (1929), 130.
17. KOLMOGOROV, A. N. *Foundations of the Theory of Probability*. English trans. New York: Chelsea Publishing Co., 1956.
18. LAPLACE, S. *A Philosophical Essay on Probability*. English trans. New York: Dover Publications, 1951.
19. LINDLEY, D. V. *Bayesian Statistics: A Review*. Philadelphia: Society for Industrial and Applied Mathematics, 1971.
20. MARGOLIN, M. H. "On the Credibility of Group Insurance Claim Experience," *TSA*, XXIII (1971), 229.
21. MAYERSON, A. L. "A Bayesian View of Credibility," *PCAS*, LI (1964), 85.
22. SAVAGE, L. J. *The Foundations of Statistics*. 2d rev. ed. New York: Dover Publications, 1972.
23. SHEPHERD, P., and WEBSTER, A. C. *Selection of Risks*. Chicago: Society of Actuaries, 1957.
24. SHUR, W. Discussion of R. MAGUIRE, "An Empirical Approach to the Determination of Credibility Factors," *TSA*, XXI (1969), 1.
25. TAYLOR, G. C. "Credibility for Time-heterogeneous Loss Ratios." Research Paper No. 55, School of Economic and Financial Studies, Macquarie University, 1974.
26. WHITNEY, A. W. "The Theory of Experience Rating," *PCAS*, IV (1918), 274.

DISCUSSION OF PRECEDING PAPER

JAMES A. TILLEY:

I have many objections to Mr. Margolin's paper, but I will concentrate my discussion on his description of quantum mechanical principles and their application to his view of probability.

There are numerous inaccuracies in the section entitled "Quantum Mechanics."

1. From a theoretical viewpoint, quantum mechanics *does* replace classical mechanics. Classical mechanics fails to explain many microscopic phenomena. On a macroscopic scale, however, the predictions of the equations of motion of quantum mechanics are the same as those of classical mechanics to an extremely high degree of accuracy, and, from a practical viewpoint, the latter can be used without error.
2. The energies of *free* particles can vary continuously. The energy levels of bound systems are discrete, however, not continuous. An atom of the simplest isotope of hydrogen, for example, is a bound system consisting of an electron and a proton; it has a ground state and excited states with discretely spaced energies.
3. It is *not* true that energies of particles (or bound systems of particles) are restricted to integral multiples of some fundamental quantum of energy. For instance, the levels of a hydrogen atom are not evenly spaced in energy and thus cannot be integral multiples of a fundamental quantum.
4. The uncertainty in the measurement of a variable is *not* the "irreducible amount of inaccuracy" as stated by Mr. Margolin. It is nothing more than the usual statistical standard deviation of the results that would be obtained by measuring the variable when the system is in a given state.

   Consider a simplified system of a single particle in a world with one spatial dimension. If $\psi(x, t)$ and $\varphi(p, t)$ represent the wave functions of the particle in position space and momentum space, respectively, then $\psi^*(x, t)\psi(x, t)$ and $\varphi^*(p, t)\varphi(p, t)$ are the probability density functions of the particle's position and momentum, respectively. (The asterisk denotes complex conjugation.) The functions ψ and φ are *not* independent of each other—φ is the Fourier transform of ψ, and vice versa. Let $\langle A \rangle_t$ denote the expected value at time t of the dynamical variable A (position or momentum, for example) when the system is in a state characterized by $\psi(x, t)$ and $\varphi(p, t)$. Then

$$\langle x\rangle_t = \int_{-\infty}^{\infty} x\psi^*(x, t)\psi(x, t)dx \,, \tag{1a}$$

$$\langle x^2\rangle_t = \int_{-}^{\infty} x^2\psi^*(x, t)\psi(x, t)dx \,, \tag{1b}$$

$$\langle p\rangle_t = \int_{-\infty}^{\infty} p\varphi^*(p, t)\varphi(p, t)dp\,, \tag{1c}$$

$$\langle p^2\rangle_t = \int_{-\infty}^{\infty} p^2\varphi^*(p, t)\varphi(p, t)dp\,, \tag{1d}$$

$$(\Delta x)_t \equiv [\langle x^2\rangle_t - \langle x\rangle_t^2]^{1/2}\,, \tag{1e}$$

$$(\Delta p)_t \equiv [\langle p^2\rangle_t - \langle p\rangle_t^2]^{1/2}\,. \tag{1f}$$

In a particular state of the system, the probability density functions of position and momentum might appear as shown in Figure 1. Quantum

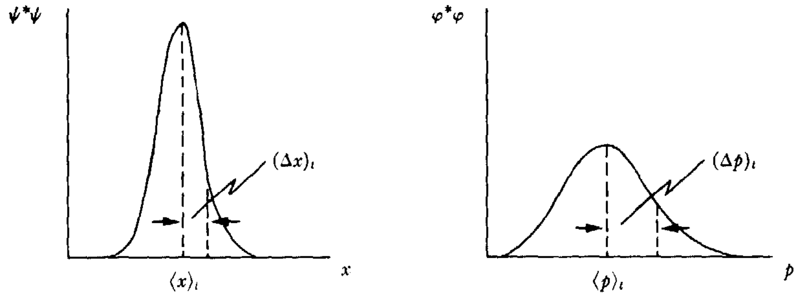

FIG. 1

mechanics does ***not*** preclude the precise measurement of either the position or the momentum of the particle, only the *simultaneous* measurement of both its position and momentum with absolute precision. Theoretically, one could measure the *exact* position of the particle at any instant, but then one would have no idea whatsoever about its momentum. Theoretically, one could measure the *exact* momentum of the particle at a later time but then would lose all information about its position.

5. In the theoretical development of quantum mechanics, the possible states of a system can be represented as vectors in an abstract vector space, and observables such as position, momentum, angular momentum, and energy as linear operators on the space of state vectors. Two operators need not commute, much as the multiplication of square matrices is, in general, noncommutative.

   The possibility of simultaneous precise measurement of two observables is related to whether or not they commute. If A and B are two (Hermitian) operators that do not commute, and

$$AB - BA = iC\,, \tag{2}$$

it can be proved that

$$\Delta A \Delta B \geq \tfrac{1}{2}|\langle C \rangle| , \tag{3}$$

where, for a given state of the system, ΔA and ΔB are the standard deviations or "uncertainties" of the probability distributions of the observables A and B, and where $\langle C \rangle$ is the mean value of the probability distribution of the observable C.

Relation (3) is the general form of Heisenberg's uncertainty relations. The position-momentum uncertainty relation is a special case of the general theorem. The time-energy uncertainty relation, however, is not a special case of this theorem.

It is *not* true that time and energy are conjugate variables. In fact, there is no such thing as a time *operator* in quantum mechanics. Unlike energy, which is a dynamical variable represented by a Hermitian operator called the Hamiltonian, time is only a *parameter*. The time-energy uncertainty relation

$$\tau \Delta E \geq h/4\pi \tag{4}$$

connects the statistical uncertainty ΔE to a time interval of order τ over which an appreciable change in the system could be observed. A quantitative definition of τ is given in chapter viii, section 13, of Albert Messiah's recognized classic *Quantum Mechanics*.[1]

The section in which Mr. Margolin attempts to transfer the concepts of quantum mechanics to his notion of probability is weak. It is not clear to me how his approach contributes to an understanding of the differences between time-homogeneous and time-heterogeneous processes.

1. The notion of quantization in quantum mechanics has highly significant consequences and leads to physical properties of atomic and subatomic systems that cannot be understood in terms of classical mechanics. In the quantum interpretation of probability described by Mr. Margolin, there is nothing profound about quantization in a set of n repetitions of a binary (heads or tails) experiment—it is merely a statement that there are $n + 1$ possible outcomes from (n tails, 0 heads) to (0 tails, n heads) if the order of the outcomes is ignored.
2. The uncertainty relations of quantum mechanics (except for the time-energy one) are direct relationships between the statistical variances of *pairs* of noncommuting observables. There is only a *single* uncertainty, $\Delta\theta$, in Mr. Margolin's uncertainty principle. Moreover, n is not an observable at all: it is the number of repetitions in the experiment ϵ.

[1] Albert Messiah, *Quantum Mechanics* (Amsterdam: North-Holland Publishing Co.; New York: John Wiley & Sons, Inc., 1961). (Original title, *Mécanique Quantique;* translated from the French by G. M. Temmer [New Brunswick, N.J.: Rutgers—The State University].)

If there is any similarity between Mr. Margolin's uncertainty principle and the uncertainty relations of quantum mechanics, it must lie with the time-energy relation. Much as τ is a characteristic of the quantum mechanical system, and E is an observable, it could be argued that n is a characteristic of the experiment and θ is an observable. The analogy has other weaknesses, however. For example, the uncertainty relation in quantum mechanics involves no "arbitrary" constant: Planck's constant h cannot be chosen at will by each physicist.

I would like to conclude my discussion by making a few further comments about Mr. Margolin's presentation. His paper contains the following statements.

If the experiments were performed under precisely *identical* conditions, the results necessarily would be identical.

The more parameters we take into account, the more nearly we may come to a deterministic prediction of the actual results.

. . . yet in principle the outcomes . . . are predictable far in advance using the theoretical apparatus of physics.

Most physicists today would not agree with the above statements as they apply to quantum mechanical systems. Such statements are suggestive of the "hidden variable" theories espoused by a small minority of physicists (including Albert Einstein) who were distressed with the statistical interpretation of quantum mechanics. But there are no experimental results that conflict with the statistical interpretation of quantum mechanics.

It is ironic that Mr. Margolin advances a "quantum" interpretation of probability whereas most physicists appeal to a frequentist view in their description of quantum mechanics. For example, in chapter IV, section 16, of the text quoted previously, it is stated:

Quantum Theory does not generally yield with certainty the result of a given measurement performed on an individually selected system, but the statistical distribution of the results obtained when one repeats the same measurement on a very large number N of independent systems represented by the same wave function.

While Mr. Margolin makes no claim that his interpretation of probability per se offers an explicit methodology for solving actuarial problems, it is important to point out that quantum mechanics does offer an explicit methodology for solving problems in physics and for making predictions about the properties of physical systems that are verifiable through experimentation.

I believe that all actuaries would agree with a statement in the concluding section of Mr. Margolin's paper:

> One should try to ascertain the severity of time-heterogeneity. Is the uncertainty in the value of the underlying parameter significant, or negligible, in comparison with the variation of the observed data about this parameter?

It is common sense and sound actuarial practice, not useful guidelines from quantum mechanics, that require actuaries to scrutinize their models in such fashion.

JAMES C. HICKMAN:

As Mr. Margolin indicates, few would doubt that the foundation of insurance is probability. I contend, however, that the first sentence of his paper could be made more inclusive. Probability is part of the foundations of most of the major intellectual developments of this century.

About 2,300 years ago, Euclid's geometry established a style for mathematical developments that has persisted to the present. Moreover, the influence of the Euclidean model of thought extended far beyond mathematics. Treatises on law, physics, and theology were cast in the Euclidean mold. This model proved somewhat inadequate to describe and predict, with suitable accuracy, the motion of the planets and some other physical phenomena. A new, and originally less formal, set of mathematical ideas was required. The ideas associated with the name Newton filled this need. The extension of these Newtonian ideas and methods throughout physics, the other natural sciences, biology, and the social sciences was a leading item on the agenda of science for about two hundred years. This project was not essentially completed in economics until this century.

The geometric model was displaced by the mechanical model as the cutting edge of science. However, the older geometric modes of thought retain their vitality because they continue to provide insight into many problems and because of the intellectual stimulation and pleasure they provide. Nevertheless, scarcely had the new mechanical models become part of the grammar of science than it became clear that they were inadequate in providing insight into some natural processes. Mr. Margolin's examples of human mortality and the behavior of subatomic particles provide illustrations of the failure of mechanical models to be universally applicable. Probability provided the needed paradigm. Starting almost three hundred years ago in the study of human mortality, continuing on into genetics, physics, and economics, probability became a key element in the basic models of the sciences. Western intellectual history might be summarized by the slogan "geometry, mechanics, and chance."

The acceptance of probability with all its subtleties has not proceeded smoothly. To some, the very idea that the basic models of physics and biology contain random elements comes close to heresy. The definition of probability as the mathematics of uncertainty seems to some to contain an internal contradiction. Yet to be an informed observer of modern science, one must consider the meaning of probability. Mr. Margolin has done the Society of Actuaries a service by directing attention to the basis of the science that actuaries practice.

Our debt to Mr. Margolin starts with his review and criticism of the three main interpretations of probability: the classical interpretation involving a finite outcome space with equally likely elements, the relative frequency interpretation, and the subjective interpretation. It is particularly gratifying that one of his primary sources is the works of the Italian actuary, mathematician, and philosopher Bruno deFinetti. Actuaries should take pride in the fact that a scholar of deFinetti's international reputation also has made significant contributions to actuarial science.

As Mr. Margolin points out, the three interpretations all support a common set of mathematical axioms. Thus the mathematics of probability is not in dispute among those adopting one of these three interpretations. The interpretation of some results and the range of application of probability are at issue, but not the basic axioms and the propositions derived from them. In this reviewer's opinion, one of the major intellectual achievements of this century has been the construction by Ramsey, Savage, deFinetti, and others of a rigorous basis for subjective probability that leads a coherent probability assessor to assign to uncertain events numbers that satisfy the mathematical axioms of probability.

These observations form the background for a question to Mr. Margolin. If, as stated in Section III, 8, the quantum interpretation of probability does not lead to the mathematical axioms for probability, what is left of the mathematics of probability? Since the rather extensive mathematics of probability, including its applications in statistics, physics, and actuarial science, rests on the three axioms for mathematical probability, it appears that one who adopts the quantum interpretation of probability (QIP) must be very cautious in using any existing results. This question should not be taken as a naive acceptance of existing models. Perhaps science can advance only with a new axiom system and the associated system of derived results. However, one needs to be very clear as to the results that remain valid under both the old and the new system of axioms.

Next is a request for clarification. In Section III, 8, both in discussing the coin-tossing experiment and in the radioactive-decay example, the problem of the experimental verification of probability statements under a relative frequency interpretation is probed. Probably I misunderstand Mr. Margolin, but it seems that the very impossibility of verifying probability statements, at what he calls the "microscopic level," is in fact the key element in a definition of random phenomena. Rather than being a criticism of the relative frequency interpretation, the impossibility of successful predictions of a few trials seems to define the phenomenon under study. A consequence of existing probability theory is that statements about expected, not actual, results are more precise, and the probability of large deviations between expected and observed results is smaller, when a large amount of data is used.

Mr. Margolin devotes a great deal of attention to the possibility that a random process under study is not necessarily stationary. He asserts, quite correctly, that most of the models used in science do not make provision for basic shifts in the structure of the system under investigation. This deficiency can be very serious. Mr. Margolin provides examples from business where shifts in the basic process have led observers to incorrect conclusions.

Despite the fact that much of conventional science has at its base some sort of stationarity assumption, I believe it would be fair to say that in the last twenty years considerable progress has been made in developing models that provide for the possibility of events causing shifts in the models.

In reviewing these developments, one must start with time-series analysis. Because of the stress in time-series analysis on differencing or using other transformations to obtain a stationary series before going ahead to identify and estimate a model, Mr. Margolin may not agree that it belongs on a list of models that achieve his goals. However, the specific properties that he sets out in Section VI are possessed by simple autoregressive models. The standard reference is to Box and Jenkins [1], but Miller and Hickman [5] summarized some of the ideas for actuaries.

Intervention analysis is an elaboration of time-series analysis developed by Box and Tiao [2]. It provides for measuring shifts in a time-series model at particular points of time. It has been used to measure the effectiveness of price controls and air pollution control regulations.

The Kalman filter [4] was developed in engineering in connection with spacecraft control problems. To a statistician it is a general linear model with random coefficients generated by a second linear model.

The statistical applications of the Kalman filter and other random coefficient models are outlined by Harrison and Stevenson [3]. Researchers at many universities are now working in this area. At the University of Wisconsin, several doctoral students are working on projects involving random coefficient models. In one such project we gained significant new insight into the behavior of gold prices by adopting a model that permitted random changes in the coefficients of the process generating the prices.

REFERENCES

1. BOX, G. E. P., and JENKINS, G. M. *Time Series Analysis, Forecasting and Control.* San Francisco: Holden-Day, Inc., 1970.
2. BOX, G. E. P., and TIAO, G. C. "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, LXX (1975), 70–79.
3. HARRISON, P. S., and STEVENS, C. F. "Bayesian Forecasting," *Journal of the Royal Statistical Society*, Ser. B, XXXVIII (1976), 205–47.
4. KALMAN, R. E. "A New Approach to Linear Filtering and Prediction Problems," *Transactions of the American Society of Mechanical Engineers, Journal of Basic Engineering*, LXXXII (1960), 35–45.
5. MILLER, R. B., and HICKMAN, J. C. "Time Series Analysis and Forecasting," *TSA*, XXV (1973), 267–330.

S. DAVID PROMISLOW:

I have some hesitation about discussing Mr. Margolin's interesting paper, as I do not feel I have had sufficient time to digest his ideas fully. However, I would like to make some remarks that I hope will be of interest.

Mr. Margolin joins a long list of previous writers on the philosophy of probability theory who challenge the frequentist interpretation. There is no doubt that time-heterogeneity can present difficulties in the *application* of the frequentist interpretation to estimation and prediction. However, this does not necessarily mean that the concept itself is invalid. In any event, Mr. Margolin has made no mention of the key argument proposed in defense of the frequentist interpretation, and I feel that we should at least examine it. This is, of course, the law (or more accurately, laws) of large numbers. I will not give precise statements here, since they can be found in any advanced book on probability, but let me discuss intuitively the simplest version. We begin with a probability measure P defined on the set of events of an experiment. It is true, as Mr. Margolin indicates, that no interpretation is put on P, but this is precisely wherein the beauty and power of the theory lie. It does not matter what P is. It need only satisfy the few simple axioms of Kolmogorov, and we reach the following conclusion. Let A be any event. Then for a sufficiently "large" number n of independent repetitions of the experiment, it is practically certain that the relative frequency r/n of occurrences of A is "close" to $P(A)$. By "practically certain" we mean that the value of P is "close" to 1 when we take a natural extension of the probability measure to the events of the sequence of repetitions. (A precise statement of this theorem of course gives definite meaning to the words "close" and "large" used above.)

To understand the frequentist interpretation fully, it is necessary to note the two distinct qualifying restrictions inherent in the above conclusion. In the first place, we cannot expect the relative frequency to be exactly equal to the probability but only close to it. In the second place, we cannot be absolutely certain of this but only practically certain. It is possible, though unlikely, that in some cases we may obtain relative frequencies that differ greatly from $P(A)$. I do not, therefore, agree with Mr. Margolin when he states in Section III, 8, that "the predictions of mathematical probability are generally inaccurate."

The law of large numbers is, of course, a perfectly valid mathematical theorem. The logical flaw of the classical definition is avoided by the axiomatic treatment. It is true that there is some circularity involved in the *interpretation* of the theorem, since the probabilistic concept of "practically certain" is used to explain the meaning of probability. However, something of this sort is almost always present when we interpret a mathematical theorem in a physical setting.

Some critics claim that this limitation of "almost" rather than absolute certainty prevents meaningful applications of the theorem. As actuaries we must reject this argument. Of course it is *possible* that all the policyholders of an insurance company will die on the same day, immediately bankrupting the company. It is so highly unlikely, however, that we must, and do, ignore the possibility. Accepting the above criticism would mean that the insurance industry would cease to exist.

To summarize, we do not have to begin as believers in the frequentist interpretation. We need only believe in the Kolmogorov axioms and we are inevitably led to accept at least some connection between probability and frequency.

Of course, the axioms of Kolmogorov are not sacred. They can be challenged, although not without some difficulty. Consider the basic additivity axiom, which states that we add the probability of two mutually exclusive events to obtain the probability of their union. This would seem to be more or less inviolable, as it is implied not only by the frequentist interpretation but also, as Mr. Margolin notes, by the coherence requirement of the subjectivist interpretation. It is true, however, that some writers, such as deFinetti, have proposed quite reasonable arguments for rejecting the usual extension of this axiom to sequences of events.

Other writers have questioned the basic mathematical structures involved in the axioms. There are those, for example, who claim that the real number system is not the appropriate object for measuring probabilities and that some other type of ordered set should be used in its place. This theory is closely connected with the comparative school of probability, which has had a long history, going back to the treatise of J. M. Keynes. Mr. Margolin would appear to have somewhat similar ideas, given his belief that probability is objective yet imprecise. I would be happier, however, if he had proposed some definite objects for measuring probability. One of my main difficulties in understanding Mr. Margolin's QIP is that I have no real grasp of what probabilities are supposed to be. I accept as reasonable the postulate that probabilities are imprecise and subject to uncertainty; however, I would like some precision in the mathematical model that one builds to describe such uncertainty. The quantum theorists, for example, model the uncertainty in the physical world by using precise objects (like linear operators in Hilbert space). In the present context one can assign intervals of real numbers as probabilities in order to model uncertainty. A discussion of some work along these lines is given by deFinetti (Mr. Margolin's reference [9], sec. 19 of the Appendix). It may be that Mr. Margolin intends such an interval assignment when he speaks of $\Delta\theta$, but this is not clear to me.

It is interesting that the quantum theorists have challenged the Kolmogorov axioms on a somewhat different ground. It is now the domain rather than the range of the probability measure that is subject to scrutiny. In the usual case, where events are represented as the subsets of a universal set S, we can always speak of the simultaneous occurrence of two events A and B, represented by the event $A \cap B$. However, the uncertainty principle implies that there are pairs of events that are not simultaneously observable. It is argued, therefore, that we must represent the event set by a more general mathematical system. An excellent account of the resulting theory can be found in V. S. Varadarajan, "Probability in Physics," *Communications on Pure and Applied Mathematics*, XV (1962), 189–217.

Finally, when the philosophizing is over, I think that all of us, as actuaries, are really subjectivists whether we admit it or not. A person presents himself for insurance and we must come up with the probability that he will die within the next year. Do we really feel that there is some true objective answer, even an imprecise one? I think not, despite the fact that we may have such a number written in a table somewhere. Indeed, the fact that among different companies, and even within the same company, we have several different numbers for this probability supports our disbelief in objectivity. I think that what we really do is arrive at a subjective opinion, our personal degree of belief in the occurrence of death at the time the insurance is issued. Naturally we are strongly influenced by past experience in arriving at our degree of belief. This is certainly allowed for in the subjectivist interpretation.

In conclusion, I would like to say that Mr. Margolin's provocative paper seems to have achieved the goal that he states at the end, for it has indeed stimulated me to "think further about the nature of risk processes."

FUNG-YEE CHAN:

I would like to point out that the uncertainty principle in quantum mechanics embodies the statement that the experiment disturbs the physical state, and such a perturbation is also one of the basic reasons for the introduction of quantum mechanics. In fact, this principle arises from the noncommutativity of two conjugate operators [2].

To illustrate in Dirac's language [1], let P and Q be the conjugate pair of momentum and position operators. The relationship $PQ \neq QP$ implies that there exists an eigenstate $|p\rangle$ with eigenvalue p with the property $PQ|p\rangle \neq QP|p\rangle$; that is, $PQ|p\rangle \neq pQ|p\rangle$. This means that the operation of Q on the state is such that, as far as P is concerned, the resulting state is no longer an eigenstate of eigenvalue p.

However, in Bayesian statistics, credibility formulas using weighted combinations, or least-squares fitting to experience, it is assumed that the physical state (population) under estimation remains the same before and after the experiment (sampling). The prior distribution, the sample (experiment) result, and the posterior distribution represent the various levels of our knowledge about the *same* physical state.

Accordingly, this basic difference about the presence and absence of the perturbation by experiments makes it unclear to me how quantum mechanics in its real physical sense would be necessary in probability theories.

REFERENCES

1. DIRAC, P. A. M. *The Principles of Quantum Mechanics.* Oxford: Oxford University Press, 1958.
2. VON NEUMANN, J. *Mathematical Foundations of Quantum Mechanics.* Princeton, N.J.: Princeton University Press, 1955.

CHARLES A. HACHEMEISTER:*

It has been my distinct pleasure over the last year to have had the opportunity to discuss with the author the concepts he presents in this paper.

I am particularly honored because he has strongly encouraged me to write this review, knowing full well that we do not reach the same philosophical conclusions regarding the nature of probability. It is remarkable that, in spite of our different philosophical views on the nature of probability, I find myself agreeing wholeheartedly with much of the paper, particularly the author's arguments concerning the defects of the frequentist interpretation and the need for empirical verification. It is only toward the end of the paper that I find myself reaching different conclusions regarding the usefulness of Bayesian methods and credibility. As the author notes in his introduction, many eminent scholars have interpreted the meaning of probability, but no single interpretation has gained general acceptance. If any one interpretation of probability is to prevail, it will require individuals such as the author who are willing to buck the tide of what others perceive to be true.

Actuaries, in my evaluation, should be businessmen. Businessmen make decisions. For the general businessman, the philosophical basis of probability is not important. It is results that count. Within this context, probability should be viewed as one of the tools that help improve business decisions. As long as a situation is sufficiently complex, it is quite likely, in my view, that different businessmen will consider themselves to be facing different problems. It is not just a question of their estimating probabilities differently but is rather one of actually structuring the problem differently. The author cites an example in Section IV, 2, of his paper of the work by Edwards et al. on the measurement of body temperature. According to the author, Edwards and his associates did not take "daily cyclical variations and other irregular changes" into consideration. Mr. Margolin's implicit model of body temperature, which considers cyclical variations, appears to be a "better" model. I would be interested to know whether Edwards and his coworkers might modify

* Mr. Hachemeister, not a member of the Society, is a Fellow of the Casualty Actuarial Society.

their statistical procedures if this omission were pointed out to them. Of course, an important question is whether they and Mr. Margolin are discussing the same thing. It might be that the former are really interested not in the daily cyclical variations in body temperature, but rather in some objective way to evaluate whether an individual "has a temperature" (is sick). On the other hand, the question, "What is the probability that this person's temperature is over 100 degrees Fahrenheit?" is unanswerable without either explicit or implicit assumptions about one's model of the situation.

Because of the impressive usefulness of probability models in aiding real-world decision making, there has been much activity recently in trying to find the "true" probability that an event will occur. As the author himself points out, "the more parameters we take into account, the more nearly we may come to a deterministic prediction." Clearly (to me), probability exists only within the framework of a model. To the extent that there is universal acceptance of a model as a reasonably accurate model of the world, that model and therefore the probabilities associated with it, are deemed objective. But this very process of acceptance is itself subjective. Besides the definitions of "objective" that relate to having to do with known objects, impersonality, and lack of bias, Webster's *New Twentieth Century Dictionary* lists as the second definition: "2. being, *or regarded as being*, independent of the mind; real; actual [emphasis added]." I understand this to mean that our understanding of reality is itself subjective.

Science, in my opinion, is the search for rules of order and regularity. Living beings somehow have the ability to ignore information selectively so as to be able to find relationships in the information retained that allow them to make decisions that are far more effective than random action in dealing with the world. On the more sophisticated level, this filtering of information leads to the formation of a theory. What is remarkable about our theories of the world is not what they explicitly take into account but what they implicitly assume to be irrelevant. The blanket caveat of "all other things being equal" is at least implicit in every theory.

Theories gain acceptance because, when used, they produce better answers in the eyes of users than any of the alternatives. However, most of us accept theories based on the plausibility in our eyes of the theory, as opposed to direct analysis and experimentation. The author discusses the standard mathematical model for radioactive decay, a model that he says, and I agree, "remains our best tool for predicting [radioactive decay]." Nevertheless, I am not a physicist and could not begin to formulate a reasonable alternative model for the process. At least in my case, I am led to the half-life model of radioactive decay because there is no alternative model that I am aware of that produces better results; that is, in my subjective opinion, I do best by accepting the half-life model of radioactive decay.

Another theory discussed in the paper is that of antiselection. My interpretation of this theory is that if you do not take into consideration differences in death or accident rates between different groups of individuals in the underwriting, marketing, and pricing of an insurance product, the resulting mix of insureds who buy your product will exhibit higher death or accident rates than anticipated. Although I have never seen statistics to support this theory, I strongly endorse it. (I have seen statistics showing that death rates vary by age, but I have never seen statistics that show the change in age distribution of plan participants that results when life insurance is sold at the same rate for all ages.) Nevertheless, I do not hesitate to draw the conclusion that antiselection would occur if not guarded against. Moreover, I would not bother quibbling with the author's statements (since I agree with them) regarding the disastrous impact of considering insurance and annuity experience to be comparable. However, it is important when discussing the nature of probability to note that no matter how compelling the reasons for accepting a theory, the acceptance of it remains subjective. The author points out that the only requirement for a subjectivist is to be "coherent." Perhaps, but it is also possible to be coherent but "dumb."

Unfortunately, one can be dumb under any philosophy of probability. As the author points out, the "basic presupposition, that past results are a guide to future expectations provided that relevant conditions do not change, tacitly underlies virtually all probability estimates." Yet different individuals will reach different conclusions as to whether and to what degree relevant conditions have changed. It is interesting that when we make decisions by following past indications without recognizing changes in relevant conditions, we react to those decisions as being either "dumb" or "subject to the changes that could not have been reasonably anticipated." Of course, our choice between these two responses is itself subjective. However, for a fairly broad range of circumstances, individuals can reach a consensus as to what is dumb and what is just bad luck.

The more diverse the backgrounds of those evaluating the acceptance of a theory, the less clear the consensus. A reasonable number of people, I have read, believe in the existence of flying saucers. If pressed, I would say that subjectively there is less than a one-in-a-hundred-thousand chance (whatever that means) that they exist. Therefore, from my point of view, the people that believe in them are either wrong, silly, mistaken, misinformed, or dumb—take your choice. From their point of view, they get to choose an adjective for me from the preceding sentence.

So far in this review, we have discussed two theories that virtually all of us agree with—concerning radioactive decay and antiselection—and one that I think hardly any of us agree with and moreover consider silly. But there are many theories that have been widely and strongly held, and thus were considered objective fact, that are now considered wrong or just plain silly, as, for example, that the earth is flat, that ether fills all space, that the earth is the center of the universe, and that sea serpents exist. Consider estimating the accident rate for ships sailing off the edge of the earth or being attacked by sea serpents, for example. By our very nature, we find it difficult to consider what we believe to be anything but the objective truth. As prudent businessmen and decision makers, however, we must face the fact that we cannot escape subjectivity in our models or theories of the world and, a fortiori, that probability within the framework of those theories is subjective.

The counterargument may be raised that if the theory truly represented reality, then the probabilities would be objective. I do not believe we will ever be able to model reality completely but, if we were, there would no longer be any need for probabilities since everything would be fully determined.

The QIP concept of complementarity as discussed by the author would, in my idea of the perfect model of reality, relegate all observations to the "caused" category. Moreover, θ would be either zero or one. No probability would be left. As soon as we reconcile ourselves to live with models that do not perfectly mirror reality, however, division between r and θ is indeed, as the author contends, subjective and unmeasurable. Yet, as was pointed out above, subjectivity is unavoidable regardless of whether FI, QIP, or SI is the point of view adopted. All three possibilities allow for subjective choice as to which spacelike and timelike variables are relevant. That is, statements such as "the probability that the total major medical claim cost from a group with such and such spacelike and timelike parameters is greater than \$100,000" make sense regardless of one's philosophy. But implicit in the statement is "within the model I believe in." The choice of model is subjective.

All three possibilities also allow for carrying out repeated random experiments. However, the decision as to whether the conditions are such as to properly call an experiment a repetition of another experiment is subjective even if you control for the spacelike and timelike variables.

QIP is unique in that probability is defined only for the aggregate. Both FI and SI define probability relating to elements of the aggregate. For FI, the meaning of repeated experiments becomes foggy, but for SI one can define probabilities at below the aggregate level quickly and subjectively. Of course, this requires the subjective split between r and θ, which is of concern to the author.

QIP requires that the θ of the aggregate be measured *exclusively* by r. No subjective adjustment is allowed. We have already discussed the problem of subjectivity as it concerns the ability to repeat experiments in order to make this measurement. Further, statisticians throw out outliers. How should this be done?

Finally, the author comments that "the accuracy of different sets of estimates can be objectively compared with regard to large aggregates of data." This is accomplished by using the sum of squared-error terms $\Sigma_i [\hat{E}(r_i) - r_i]^2$ as a measure of "goodness." This criterion can also be used for credibility estimates, as has been shown by Bühlmann and Straub [1] and by me [2], and will produce a smallest value over the aggregate of all possible linear estimators that are unbiased in the aggregate.

In summary, in my opinion it is impossible to avoid substantial subjectivity in decision making, and in the setting up of probability models. It is the actuary's job to provide a coherent structure within which to make those subjective judgments. In closing, it should be reemphasized that many of the difficult points of philosophy that I have mentioned in this review are first mentioned by the author himself in his paper. I only wish that, as he concludes, it were possible to find decision models and their associated probabilities that are truly objective. The author has set himself the task of attempting to identify such objective probabilities. I hope I have made reasonable arguments in this review as to why it is impossible to obtain objectivity, unsullied by subjective elements, and I hope I have also made a convincing argument for the value of the subjectivist interpretation and Bayesian statistics.
I want to thank Mike Margolin for encouraging me to write this review, and the Society for publishing it.

REFERENCES

1. BÜHLMANN, H., and STRAUB, E. "Credibility for Loss Ratios," *ARCH*, 1972. (Translated from H. BÜHLMANN and E. STRAUB, "Glaubwürdigkeit für Schadensätze," *Mitteilungen Vereinigung Schweizerischer Versicherungs-mathematiker*, Vol. LXX, No. 1 [1970].)
2. HACHEMEISTER, C. A. "Credibility for Regression Models with Application to Trend." In *Credibility Theory and Applications*, edited by PAUL M. KAHN. New York: Academic Press, 1975, pp. 129–69.

A. P. DAWID:*

While I have yet to be convinced of the potential of a quantum interpretation of probability, I welcome this paper for its clear statement of the difficulties faced by frequentist and subjectivist interpretations of probability models. However, the subjectivist approach does not suffer all the ills attributed to it here. In particular, deFinetti's theory of exchangeability has not been fully appreciated.

A subjectivist, in contemplating a sequence of coin tosses, can assign a value to his subjective probability of any particular finite sequence of heads and tails. He does so directly; he need not necessarily model the tosses as Bernoulli trials with a probability parameter θ, and then give θ a prior distribution. Now his subjective distribution for the tosses will not usually exhibit independence between them, since this would rule out the possibility of using the outcomes of past tosses to improve future predictions. However, it would sometimes seem reasonable to require exchangeability, so that all permutations of a fixed sequence of outcomes are assigned the same probability.

DeFinetti's fundamental theorem on exchangeability asserts that any such exchangeable assignment of probabilities to an infinite sequence of tosses has the mathematical property that it can be imitated by a model based on Bernoulli trials, for a uniquely determined prior distribution for θ. We need not, then, argue whether θ is "subjective" or "objective"—it is merely a mathematical fiction that simplifies calculations. It is possible, though cumbersome, to work all the time with the joint probabilities of sequences of outcomes, avoiding all mention of θ and its distribution.

I find deFinetti's result invaluable in interpreting the time-homogeneous Bernoulli model, and the unique-case probability assignments in it: the model is just a mathematical consequence of the assumptions of symmetry appropriate to this case. Thus does the subjectivist theory shore up the underlying structure on which frequentist interpretations are built. But, at the same time, it suggests extensions not available to the frequentist. For example, we can consider symmetry under other groups than the simple permutation group ([6], [3], [4]); an important application might be to insurance risks classified by two factors (or to events with both spacelike and timelike repetitions), where we can permute the levels of either factor separately [1]. Any such symmetry assumption produces, almost magically, a parametrized statistical model together with a "prior" distribution for its parameter. Again, the model

* Professor Dawid, not a member of the Society, is professor of actuarial science and statistics at the City University, London, England.

probabilities take their meaning from their origin in the symmetry assumptions.

Of course, it is often unrealistic to impose strong symmetries on one's subjective beliefs, and, even when it seems appropriate a priori, the data may turn out to exhibit unsuspected structure. As the author points out, coherence is not enough; some correspondence with external reality is also essential. This may be investigated by means of another subjectivist theorem referring to the behavior of frequencies [5]. Suppose a subjectivist S has an arbitrary (not necessarily exchangeable) distribution for a sequence of events. He observes the events one by one, and at each stage gives his probability (conditional on current information) for the next event. Let us now pick out the subsequence of events to which S assigned a probability between, say, 0.69 and 0.71, and calculate, for this subsequence, the limiting proportion π of times the events in fact occurred. Then S must believe (with probability one) that π will lie in the chosen small interval (0.69, 0.71). In other words, he believes that his subjective probabilities will be verified as frequencies. The snag is that his believing it need not make it happen—and if it doesn't we have empirical evidence that his beliefs, coherent though they might have been, were in error.

Although I strongly favor the subjectivist view of modeling, I see the above difficulty as very real—the more so since it applies to any subjective distribution, however "sophisticated." Perhaps subjectivists will have to accept that their distributions must undergo empirical testing, just like any classical statistical model. If so, they may also have to accept, following Braithwaite [2] and Margolin, that statements of probability can be made only with an uncertainty reflecting their acceptability in the light of data.

REFERENCES

1. Aldous, D. J. "Representations for Partially Exchangeable Arrays of Random Variables." Typescript, University of Cambridge, 1979.
2. Braithwaite, R. B. *Scientific Explanation*. Cambridge: Cambridge University Press, 1953.
3. Dawid, A. P. "Invariant Distributions and Analysis of Variance Models," *Biometrika*, LXIV (1977), 291–97.
4. ———. "Extendibility of Spherical Matrix Distributions," *Journal of Multivariate Analysis*, VIII (1978), 559–66.
5. ———. "The Well-calibrated Bayesian," *Journal of the American Statistical Association*, 1981 (to appear).
6. Kingman, J. F. C. "On Random Sequences with Spherical Symmetry," *Biometrika*, LIX (1972), 492–94.

I. J. GOOD:*

When I first read Mr. Margolin's paper, too hurriedly, I thought he was concerned exclusively with variations in time and in ordinary space, because this was his main emphasis, but in his discussion of spacelike variation in Section II he uses "space" in its more abstract sense. Thus, the variations in probability that he has in mind are those pertaining to variations in any set of attributes. When the attributes are categorical, the probabilities would pertain to the cells of a multidimensional contingency table. The estimation of such probabilities is a topic to which I have tried to apply a partly subjectivistic Bayesian analysis. Extensive calculations have so far been done only for ordinary (two-dimensional) tables ([6], [12], [1]). These examples alone are enough to show that Margolin's comment in Section IV that "the QIP representation of probability as a function of a set of spacelike parameters is altogether lacking in SI [the subjectivist interpretation]" is incorrect.

The emphasis on ordinary space and time is useful in the context of nonstationary stochastic processes such as speech. In this context, the analogy with the formulation of quantum mechanics, with a clear analogue of the uncertainty principle, was developed very effectively by Denis Gabor [3]. Note, however, that his adaptation of the uncertainty principle referred to the complementary variables time and frequency, and he was not concerned with the nonsharpness of probabilities of their estimates. It is even a feature of the ordinary formalisms of quantum mechanics that the probabilities are precise and develop deterministically in accordance with Schrödinger's equation. For this reason I think that the term "quantum interpretation of probability" is somewhat misleading, and that theory of "nonsharp probabilities" or "interval-valued probabilities" or "upper and lower probabilities" or "partially ordered probabilities" might be better, or perhaps "nonsharp parameters," (etc.), if the emphasis is put on parameters rather than on probabilities.

The probabilities in quantum mechanics, when applied to a largish system, would, in principle, be "single-case" physical probabilities. Popper suggested the excellent name "propensities" for them, although he overstated his position by claiming that his nomenclature constituted a "theory."

The author has informed me by telephone that the main point in his paper is as follows: Any statistical procedure for estimating probabilities must contain one or more parameters that are treated as constant, yet

* Professor Good, not a member of the Society, is University Distinguished Professor of Statistics, Virginia Polytechnic Institute and State University.

these parameters are not usually known with certainty. I agree with this position, although, in a hierarchical Bayesian model, I prefer to call a parameter in a prior distribution more specifically a "hyperparameter," and one in a hyperprior distribution a "hyperhyperparameter," and so on. My main philosophical dictum on the hierarchical approach [5] was as follows: "It may be objected that the higher the type [in the hierarchy of probabilities] the woollier the probabilities. It will be found, however, that the higher the type the less the woolliness matters, provided the calculations do not become too complicated." Thus, instead of talking about a "quantum interpretation" I talked about "woolliness." But the question of variations of probability when attributes change is largely a separate issue.

Mr. Margolin argues that the subjective (personal) Bayesian position is unsound because it fails to take into account variations of conditions in space and time. But if "you," as a Bayesian, fail to take *any* clearly relevant facts into account, you can expect to make imperfect judgments of probabilities. This fact has even been proved and discussed by using the Bayesian point of view; that is, it can be shown that, from "your" point of view, there is a positive expected utility in making a free observation [7]. That Bayesians, being human, have made mistakes, cannot be taken as a serious objection to a Bayesian philosophy, especially when it is borne in mind that there are many varieties of Bayesianism [10]. One of the principles that I have emphasized is that apparently non-Bayesian methods often are all right, but one should not ***knowingly*** contradict the usual axioms of subjective probability ([9], principle 21). All statisticians, Bayesian or otherwise, have to judge what information is irrelevant, and this is usually a subjective judgment.

Mr. Margolin emphasizes the problem of estimating probabilities of events that have seldom or never occurred before. This problem has not been ignored by users of Bayesian methods. One example is the estimation of the probabilities corresponding to empty cells in large contingency tables [6]. Even the discussion of the binomial distribution by Bayes involves the estimate $1/(n+2)$ for the probability of an event of frequency zero in a sample of size n. But the relevance to the philosophy of actuarial science is greater when the categorization is more than one-dimensional.

In Section IV, 2, Margolin argues, in effect, that it is frequently necessary to "peek at the data" before forming a model. Although I am often regarded as a Bayesian, I agree with this remark. But the model then selected is taken as a reasonably simple one, largely because a simple model usually has a higher initial probability of being approximately correct than does a more complicated one—again a Bayesian argument!

Also, in Section IV, 2, Mr. Margolin says, "You may be forced to use non-Bayesian techniques," but "forced" is a little too strong. I would often choose to use non-Bayesian techniques when they are easier to apply, because I believe in the "type 2 principle of rationality" ([9], principle 6). This principle advocates maximizing expected utility after allowing for the costs of calculation and thought. Whether this should be called a Bayesian principle is a semantic matter—I prefer to call it Doogian [11].

At the end of Section V, referring to an example of how he attacks a practical problem, Mr. Margolin states that "QIP reformulates the problem of credibility as follows: to estimate the future claims of a risk (and not some shadowy parameter), on the basis of its own actual past experience and the known experience of other similar risks." This agrees with the emphasis of deFinetti [2] and Geisser [4]. I agree that the prediction of observables is of more immediate practical significance than that of parameters, but I take a less extreme view on the matter than does deFinetti [8].

In summary, in my opinion Mr. Margolin's criticisms of the subjectivistic Bayesian approach are untenable when they are applied to the philosophy that I support, which in 1980 can be called a Bayes/non-Bayes compromise or synthesis. Moreover, his main point, according to his telephone conversation, is consistent with the hierarchical Bayesian philosophy. For a recent account of some of the history of this philosophy see [13].

REFERENCES

1. Crook, J. F., and Good, I. J. (1980). Part II of ref. [12]. *Annals of Statistics*, 1980. Scheduled for November.
2. deFinetti, B. "Initial Probabilities, a Prerequisite for Any Valid Induction," *Synthèse*, XX (1969), 2–16; also published in *Induction, Physics and Ethics: Proceedings and Discussion of the 1968 Salzburg Colloquium in the Philosophy of Science*, edited by P. Weingartner and G. Zecha, pp. 3–17. Dordrecht: D. Reidel, 1970.
3. Gabor, D. "Theory of Communication," *Journal of the Institute of Electrical Engineers*, XCIII (1946), 429–41.
4. Geisser, S. "The Inferential Use of Predictive Distributions," In *Foundations of Statistical Inference*, edited by V. P. Godambe and D. A. Sprott, pp. 456–69 (with discussion). Toronto and Montreal: Holt, Rinehart & Winston of Canada, 1971.
5. Good, I. J. "Rational Decisions," *Journal of the Royal Statistical Society*, Ser. B, XIV (1952), 107–14.

6. ———. "On the Estimation of Small Frequencies in Contingency Tables," *Journal of the Royal Society*, Ser. B, XVIII (1956), 113–24.
7. ———. "On the Principle of Total Evidence," *British Journal for the Philosophy of Science*, XVII (1967), 319–21; see also XXV (1974), 340–42.
8. ———. Discussion of deFinetti (ref. [2]), *Synthèse*, XX (1969), 17–24; also published in the *Proceedings* cited in ref. [2].
9. ———. "Twenty-seven Principles of Rationality." In *Foundations of Statistical Inference: Proceedings of a Symposium on the Foundations of Statistical Inference* [1970], edited by V. P. GODAMBE and D. A. SPROTT, pp. 124–27. Toronto and Montreal: Holt, Rinehart, & Winston of Canada, 1971.
10. ———. "46656 Varieties of Bayesians." Letter in *American Statistician*, XXV, No. 5 (December, 1971), 62–63.
11. ———. "The Bayesian Influence, or How to Sweep Subjectivism under the Carpet." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science* (Proceedings of a Conference in May, 1973, at the University of Western Ontario), edited by C. A. HOOKER and W. HARPER, pp. 125–74 (with discussion). Dordrecht: Reidel, 1976.
12. ———. "On the Application of Symmetric Dirichlet Distributions and Their Mixtures to Contingency Tables," *Annals of Statistics*, IV (1976), 1159–89.
13. ———. "Some History of the Hierarchical Bayesian Methodology." Invited paper for the International Meeting on Bayesian Statistics, May 28–June 2, 1979, Valencia, Spain. To be published in *Trabajos de Estadística y de Investigación Operativa*.

(AUTHOR'S REVIEW OF DISCUSSION)

MYRON H. MARGOLIN:

I am delighted that my paper elicited seven such varied and stimulating discussions. Four of the replies are from F.S.A.'s, who need no introduction to this readership. Mr. Hachemeister is a Fellow of the Casualty Actuarial Society who has both executive responsibility in a major reinsurance company and extensive knowledge of the mathematics and philosophy of probability. Professors Dawid and Good are both statisticians who have written widely on the foundations of probability and statistics. How the galley proof of my paper came into their hands I do not know, but I am certainly glad that it did and that they chose to submit comments.

Two themes are recurrent throughout many of the discussions.

1. QIP has nothing really new to offer. The search for time heterogeneity is based on "common sense and sound actuarial practice" (Tilley). Time-series analyses do indeed recognize time heterogeneity (Hickman). The representation of probability as a function of spacelike parameters is not new (Good).

2. Subjectivity is inescapable in actuarial and statistical practice, indicating that probability cannot be objective. It is impossible to avoid substantial subjectivity in decision making and the setting up of probability models (Hachemeister). All actuaries are really subjectivists whether we admit it or not (Promislow).

Let me first respond to these two major themes, then turn to some of the other specific points and criticisms in the seven discussions.

QIP does not pretend to be a new *theory* of probability. The mathematical theory of probability is already quite highly developed, and QIP offers no new extensions of this theory. As an *interpretation*, QIP may instead offer some guidance as to how the mathematical theory should be applied to problems arising in the real world.

I agree with Mr. Tilley that common sense and sound actuarial practice have often led actuaries to consider heterogeneity. Common sense has in fact forced us to break out of the narrow confines of FI and strict Bayesianism. QIP furnishes a conceptual framework that requires due regard for heterogeneity and therefore captures the common sense that is lacking in FI and in strict SI/BS. This lack of common sense is precisely what went wrong with credibility theory, at least as it concerns group insurance. Every practicing group actuary knows that group insurance claim experience is afflicted with the sort of heterogeneity my paper describes, yet the models continually spun out by strict frequentists and subjectivists remained sterile.

Professor Good objects to my assertion that the concept of probability as a function of spacelike parameters is lacking in SI. He cites his own *partly* subjective Bayesian analysis and his Bayesian/non-Bayesian compromise or *synthesis*. Apparently he, too, finds it necessary to depart from strict subjectivity.

I stand by my criticism of strict subjectivity, which incidentally is no straw man. There are theoreticians who are strict subjectivists, such as deFinetti; and it is dogmatic subjectivism that has led credibility theory astray, as I think my paper makes clear.

If we set aside the extremes, we may agree that sound actuarial practice and decision making involve some mixture of objective and subjective elements, and I think it is vital to try to sort them out. This brings us to Mr. Hachemeister's discussion.

Most actuaries are businessmen, as Mr. Hachemeister notes, but, whether businessmen or statisticians, we do not work in isolation. We must communicate and justify our recommendations to our clients or superiors. In these communications and justifications, we are obliged to distinguish as carefully as possible between subjective/personal and objective elements.

A mixed bag of objective and subjective elements enters into almost any human intellectual endeavor. In many contexts the distinction is clear. For example, when a schoolboy solves a problem in plane geometry, the act of finding a solution (the so-called discovery phase) is highly subjective (personal, intuitive). On the other hand, the proof itself (the justification phase) is quite objective. Either a proof is correct or it is not.

Mr. Hachemeister points out that probability "exists" only in the context of a model. I concur, provided that we interpret the word "model" fairly broadly to include any theory or set of assumptions, whether or not explicitly mathematical, about how the world works. But if we intend to communicate with other persons concerning our models, we must describe them as precisely as possible and distinguish between the subjective and the objective elements in our use of them.

Actuarial science is hardly unique in using precise models. Their use is common in the natural sciences, and mathematical models flourish especially in physics. These models are considered theoretical, provisional, or corrigible, but I do not think that anyone considers the models or their constituent parts to be subjective.[1] We should be skeptical of the notion that the mathematical models of one part of science are subjective while those of other parts are not.

QIP treats models and probabilities as objective, but it does not follow that there is a uniquely correct model for each experiment or decision problem. My paper makes this point clearly in the subsection "Problem of Empirical Validation"; it is also a direct consequence of the principles of uncertainty and complementarity. But the bounds of uncertainty are defined, and models whose predictions fall outside these bounds must be regarded as invalidated by the data. In other words, some models are objectively ruled out by the data; others are consistent with the data, and you may select from them, subjectively. Yet there are other aspects to this question of subjectivity, as shown in the following illustration.

The example is fairly typical of decision problems in (life) actuarial practice. An actuary is updating the premium rates and dividend scales of his company's individual life insurance portfolio. A major part of the job is to revise the large array of select and ultimate mortality rates

[1] On the other hand, someone's acceptance of a model is surely a subjective matter. Logicians distinguish between (*a*) a specified factual proposition and (*b*) a person's assertion or belief that a specified proposition is correct. A similar distinction applies to models.

that, exclusive of loadings for conservatism, will represent his estimate of the probabilities of death to be experienced in the future. Assume that mortality studies of the latest four years of the company's experience show the following aggregate results (percentages of "expected" on a basic table): 1976, 100.0 percent; 1977, 98.1 percent; 1978, 96.5 percent; 1979, 93.8 percent. Assume further that *one* standard deviation, based on the number of deaths, is roughly 0.5 percent.

With just these data, the actuary can formulate plausible models that can serve as a basis for the new scale of rates. Let θ_t represent the percentage of actual to expected for t = year − 1976. Then one model that is consistent with the data is $\tilde{\theta}_t$ = 100 percent $\times$ $(1 - 0.02t)$.

Let us now proceed to distinguish subjective and objective elements of the problem.

1. The actuary has chosen to structure the data in a certain way, an aggregate percentage of actual to expected. He is not using any information at a more detailed level. Of course, at a more detailed level (less uncertainty in position) there would be more uncertainty in the value of θ.

   Moreover, he has ignored other possibly relevant data, such as the trend of population mortality data. To use such other data, he would need to elaborate his model to define a relationship between his company's data and the other data, but he chooses not to do so.

   In any case, QIP makes clear (Sec. III, 7) that experimental data and inferences depend on how an experiment is structured (as does QM). How to structure an experiment is a subjective choice.
2. The model $\tilde{\theta}_t$ = 100 percent $\times$ $(1 - 0.02t)$ is consistent with the data. So are $\tilde{\theta}_t$ = 100 percent $\times$ $(1 - 0.021t)$ and $\tilde{\theta}_t$ = 100 percent $\times$ $(1 - 0.02t + 0.001 \sin t)$ and all others for which $\tilde{\theta}_t \doteq X_t$.

   But many more models are objectively ruled out by the data. The most notable case is $\tilde{\theta}_t$ = 100 percent. We can also rule out $\tilde{\theta}_t$ = 100 percent $\times$ $(1 + 0.02t)$ and $\tilde{\theta}_t$ = 100 percent $\times$ $(5 + \log t)$, ad infinitum. Here QIP and common sense are in accord—but not FI or strict SI, for which the only logically permissible models are of the type $\tilde{\theta}_t$ = constant, and which are then concerned only with estimating the constant.
3. All models consistent with the objective data will indicate a downward trend, but there is no certainty that such a trend will continue into the future. Some subjectivity must enter into estimates of the future probabilities (see Sec. III, 6). But QIP urges the actuary to investigate what *objective* features have *caused* the downward trend and whether these factors are likely to persist. Neither FI nor SI/BS provides for this sort of cause-effect relationship.
4. We must be sure to distinguish clearly among X_t, θ_t, and $\tilde{\theta}_t$. X_t is a precise objective datum. It is a measurement of the quantity θ_t. In this case θ_t is a statistical parameter, but in more simply constructed experiments it is simply a probability. The value of θ_t is not precise, for reasons detailed in

the paper, but it is measurable and not the subject of personal opinion. Finally, $\tilde{\theta}_t$ is the predicted or expected value of θ_t according to some model. If $\tilde{\theta}_t \doteq \theta_t$ for all known values of θ_t, the model is corroborated (thus far); otherwise, it is objectively incorrect.

There are plainly many subjective elements in this whole procedure, but it does have what I take to be the hallmark of objectivity: the rejection or acceptance of one's provisional assumptions based on empirical data—feedback from the external environment. It goes beyond the Bayesian canon of orderly revision of opinion in two respects. Data are used to test *models;* and, when a model is invalidated, any opinions or estimates associated with it are proved false, not merely updated. This is indeed common sense and sound practice, but set in a comprehensive conceptual framework.

It may be that some "syntheses" or "combinations" of subjectivity and objectivity achieve a similar congruence with common sense, but it is not clear that Professor Good achieves this sort of congruence. Taking account of all the relevant data is not necessarily the best approach. One could drown in all the potentially relevant data, and in the previous illustration our hypothetical actuary has deliberately ignored some. The trick is to organize the data—to structure the experiment—in a manner appropriate to the problem. This is, I readily concede, a subjective choice, but it is based on one's knowledge of the objective structure or character of the data that are known and on informed opinion as to whether it is worth continuing to search for more structure, that is, relationships between θ and experimental parameters not yet considered.

Here I am totally in sympathy with Mr. Hachemeister's comment that what is remarkable is what our theories or models assume to be irrelevant. There is, in principle, no limit to the number of parameters one could test. One ignores nearly all of them—often because one cannot conceive of any way in which the objective attributes they characterize could physically *cause* any substantial effect on the results.

Professor Dawid clearly recognizes that unbridled SI/BS can lead to probability judgments with inadequate correspondence to reality. His remarks suggest that he is working toward a modification of Bayesianism that would achieve the necessary correspondence. We can but welcome this effort to square Bayesianism with common sense.

Let me turn to some of the many other points raised in the discussions.

Professor Hickman cites the Box-Tiao paper on intervention analysis in time series. The government's decision to impose wage-price controls caused a sudden shift (intervention) in the economic data. But every economic action, be it a sudden massive act by the government or a lesser act by the government, a corporation, or an individual, is a distinct, palpable factor impacting on θ. Surely the distinction between changes in θ (interventions) and random "noise" is arbitrary.

Professor Hickman also cites the Harrison-Stevenson paper "Bayesian Forecasting," which applies the Kalman filter to time-series problems. This paper and its printed discussion are indeed noteworthy, both for the mathematics and for the "philosophy." The authors are keenly aware of the problem of time-heterogeneity in all its manifestations. "[This] leads to a dilemma: if short series are used for parameter estimation, sampling errors will be large; whereas, if long data series are used, the resultant estimates will be, at best, several years out of date, or, at worst, totally misleading." This is not far from a statement of UP. The authors agree that there is "uncertainty about the model and its parameters." When it comes to model selection, "we are not religious Bayesians; we are fully prepared to scrap or revise our original formulations if subsequent results show them to be invalid or inadequate."

It seems clear that, like Professor Dawid, Harrison and Stevenson are breaking out of the tight confines of dogmatic BS. A new paradigm is needed. Enter QIP.

Mr. Hickman is concerned about the implications of QIP for the mathematical axioms of probability. My statement of the relation between QIP and these axioms was incomplete. If we insist that P has a precise value, then the statement stands. If we allow P an appropriate measure of indeterminacy, however, the mathematical theory of probability is no longer empirically contradicted. It is not my intention that we shelve mathematical probability.

Mr. Chan points out that UP in QM embodies the notion that an experiment disturbs the physical state. The act of measurement imparts an uncertain change to the system being measured and hence precludes exact verification of the laws of classical physics. This paradox is often cited as a basis for QM.

An analogous paradox does arise in the case of probability. An insured population is not physically identical before and after a time period of exposure to risk, and tossing a coin disturbs the physical state of the coin (in its relation to the tossing apparatus). I refer to this paradox in the section discussing FI. In a series of repetitions, the change in physical conditions that admits "random change" in r—or simply permits r to occur—may also alter the value of θ, and there is no logically defensible distinction between the two types of change. To restate the paradox: If ϵ_2 is precisely identical with ϵ_1, then r_2 must equal r_1; if they are not precisely identical, we cannot be certain that $\theta_2 = \theta_1$. UP is as inescapable in probability as in physics.

But the paradox is an analogy, not an identity. Contrary to Mr. Chan's inference, I do not assert that QM "in its real physical sense" is necessary to interpreting probability.

Mr. Tilley may be under a similar misapprehension. I am not trying to *derive* QIP from QM. At most I am asserting that there are some analogies. These may render QIP more plausible, since the notion of an objective yet inherently imprecise quantity has been firmly established in QM.

Moreover, I do not intend *argument by analogy*. I am not asserting that, *because* probability is like physics in respect to A, B, and C, it is also like physics in respect to D, E, and F. The arguments for QIP could have been stated with no reference at all to QM. They stand on their own. If analogy with QM helps you to understand QIP, good; if not, please set QM aside.

For this reason, Mr. Tilley's objections to my description of QM do not apply directly to QIP. I will respond, however, to those that are most germane.

1. Whether QM "replaces" classical mechanics is a question of definition. I was careful to state that QM is necessary to make atomic phenomena intelligible. Relative to large-scale phenomena, however, the predictions of classical mechanics are utterly indistinguishable from those of QM. By the canons of empirical validation, the classical theory is not inferior to QM in this domain. It is also very much easier to use.

   In the case of probability, N is almost never large enough for the effects of UP to disappear. For a million tosses of a coin, $\Delta\theta$ is still about 0.001, which is hardly imperceptible.
2. My description of the role of uncertainty in QM states clearly that "the limits apply to the simultaneous measurement of . . . conjugate pairs" and that "one can obtain arbitrarily small uncertainty in one variable, but only at the expense of increased uncertainty in its conjugate." Mr. Tilley needlessly restates essentially the same points as a criticism. In any case, I certainly concur that any reference to uncertainty or indeterminacy, as to either QM or QIP, must be understood in the context of a complementary uncertainty.
3. My terse reference to the time-energy uncertainty relationship was indeed misleading. A full discussion would take too much space, but let me excerpt one bit: "if *any* observable plainly exhibits a rapid variation with time . . . , then the system *cannot* have a well-defined energy" [1] [italics in original). This instance of UP in QM may indeed bear more analogy to QIP than does the position-momentum relationship, but this point cannot be pursued further here.
4. That Planck's constant has a precise value is not in question. But the

quantity $h/4\pi$ is the product of the respective "uncertainties" Δp and Δx, where each uncertainty is defined as *one* standard deviation. One could as readily have selected a basis of *two* standard deviations.

Mr. Tilley goes on to discuss QIP.

1. There is nothing profound about quantization alone. Planck and Bohr had used quantization prior to the appearance of QM and UP, but it was finally QM that successfully explained those properties of atomic and subatomic systems that cannot be understood in terms of classical mechanics. Analogously, the "commonsense" attributes of ratemaking cannot be understood in terms of FI and SI/BS.
2. My explanation of uncertainty does need elaboration. I think we are accustomed to the fact that, even in a strictly time-homogeneous situation, we cannot precisely measure a probability when n is finite. The measurement is subject to an imprecision, which is quantified in terms of standard deviations. We recognize that one standard deviation is inversely proportional to $\sqrt{n}$.

   We are less accustomed to conceptualizing uncertainty in space-time. Suppose we measure the probability of death among United States males age 45 in the year 1980. This population spans a vast geographical area, a host of occupations, a wide range of health conditions, and so on. Thus the measurement is not well localized in terms of these important spacelike parameters. Furthermore, the measurement is for a full calendar year and is not specific to any time period within that year. The foregoing is one way to describe the measurement's uncertainty in space-time. Improved precision as to space-time is obtainable, obviously, only at the expense of more uncertainty in θ.

   Another perspective on the extent of uncertainty is to consider the number of persons exposed to death. The larger the population, the less specific the measurement to any one individual. This perspective does appear somewhat less meaningful in this case. On the other hand, where the experiment consists of a series of tosses of a specific coin, there is no uncertainty as to which coin. Here the uncertainty (in space-time) can refer only to the number of tosses. The larger the value of n, the less specific the measurement to any one toss or to any fixed subset of tosses.

   I would agree that UP in QIP is not perfectly analogous to UP in QM, but, as stated above, exact analogy is not essential to the argument.
3. My references to identical conditions and deterministic predictions do not refer at all to quantum mechanical systems, that is, to experiments in which *physical* uncertainties are appreciable. One of the examples dealt with a mechanical coin-tossing device. I have no difficulty imagining that a physicist or engineer could predict the outcomes of the first few tosses, given appropriate initial measurements. Theoretically, it would take many tosses before the cumulative effects of physical uncertainty rendered prediction difficult. Statistical uncertainty a la QIP is altogether different.

Two of the discussions allude to the law of large numbers. This is a mathematical theorem. It should not be confused with a scientific law, which, though expressed in mathematical terms, must be consistent with empirical data. In my earlier example of mortality rates, this "law" figures nowhere in the actuary's solution to the problem. You may pretend that if one of the calendar years of exposure were somehow repeated over and over, the frequency would converge toward a limit. My objections to this mental exercise include, but go well beyond, its impracticality. Either repeating a year is like rerunning a movie, where the ending never changes, or it is a new experiment, with altered conditions. To pick up on Mr. Hickman's point, it is a consequence *only* of the mathematical theory that, with more data, the frequency of large deviations is reduced. But this theory fits only time-homogeneous situations, which are atypical. What is more, QIP does fit these situations (see the section on radioactive decay).

Despite Professor Good's suggestions, I want to retain "quantum" in the name of this interpretation of probability. I believe that probability and physics, or chance and determinism, are even more subtly intertwined than has generally been thought. The discovery in the 1920s that randomness lies at the core of mechanics was, as Professor Hickman indicates, a rude shock, but it obviously also signaled a profound relationship between probability and physics. QIP goes further, in two respects. We can attach precise significance to the fundamental concepts of probability only in certain domains of observation. Here, large masses of data are involved; quantum effects are insubstantial. In other domains, of small mass and palpable discreteness, the concepts become blurred with uncertainty. The parallel with physics is quite direct. In addition, QIP makes the point that nearly all the "random" events that are the subject of statistical models can also be interpreted as having been caused. Each such event can be explained by a deterministic model, as, for example, a model drawn from classical physics. It appears that only two types of events are not so interpretable: those that correspond to freedom of choice—the mental activities of humans and higher animals—and those that fall in the domain of QM. Each event of these two types is mysteriously indeterminate, by whatever model.

In closing, let me thank the seven discussants for their stimulating remarks.

REFERENCE

1. GILLESPIE, D. T. *A Quantum Mechanics Primer*. New York: John Wiley & Sons, 1974.