Updating Subjective Probability

PERSI DIACONIS and SANDY L. ZABELL*

Jeffrey's rule for revising a probability P to a new probability P^* based on new probabilities $P^*(E_i)$ on a partition $\{E_i\}_{i=1}^n$ is

$$P^*(A) = \sum P(A \mid E_i)\, P^*(E_i).$$

Jeffrey's rule is applicable if it is judged that $P^*(A \mid E_i) = P(A \mid E_i)$ for all A and i. This article discusses some of the mathematical properties of this rule, connecting it with sufficient partitions, and maximum entropy updating of contingency tables. The main results concern simultaneous revision on two partitions.

KEY WORDS: Probability kinematics; Jeffrey conditionalization; Sufficiency; Maximum entropy; I projection; f divergence; Contingency tables; Iterated proportional fitting procedure; Coefficient of dependence.

1. INTRODUCTION

1.1 Belief Revision

The most frequently discussed method of revising a subjective probability distribution P to obtain a new distribution P^*, based on the occurrence of an event E, is Bayes' rule: $P^*(A) = P(AE)/P(E)$. Richard Jeffrey (1965, 1968) has argued persuasively that Bayes' rule is not the only reasonable way to update: use of Bayes' rule presupposes that both $P(E)$ and $P(AE)$ have been previously quantified. In many instances this will clearly not be the case (for example, the event E may not have been anticipated), and it is of interest to consider how one might proceed.

Example. Suppose we are thinking about three trials of a new surgical procedure. Under the usual circumstances a probability assignment is made on the eight possible outcomes $\Omega = \{000, 001, 010, 011, 100, 101, 110, 111\}$, where 1 denotes a successful outcome, 0 not. Suppose a colleague informs us that another hospital had performed this type of operation 100 times, with 80 successful outcomes. This is clearly relevant information and we obviously want to revise our opinion. The information *cannot* be put in terms of the occurrence of an event in the original eight-point space Ω, and the Bayes rule is not *directly* available. Among many possible approaches, four methods of incorporating the information will be discussed: (1) complete reassessment; (2) retrospective conditioning; (3) exchangeability; (4) Jeffrey's Rule.

1. *Complete Reassessment.* In the absence of further structure it is always possible to react to the new information by completely reassessing P^*, presumably using the same techniques used to quantify the original distribution P.

2. *Retrospective Conditioning.* Some subjectivists have suggested trying to analyze this kind of problem by momentarily disregarding the new information, quantifying a distribution on a space Ω^* rich enough to allow ordinary conditioning to be used, and then using Bayes' rule. For some discussion of this, see de Finetti (1972, Ch. 8) and Section 2.1. It is worth emphasizing that this type of retrospective conditioning can be an extremely difficult psychological task; see Fischoff (1975), Fischoff and Beyth (1975), Slovic and Fischoff (1977). Nor, in principle, is retrospective conditioning simpler than complete reassessment: since $P^*(A) = P(AE)/P(E)$ in this case, for each A assessment of $P(AE)$ is equivalent to reassessment of $P^*(A)$.

3. *Exchangeability.* The three future trials may be regarded as exchangeable with the 100 trials reported by our colleague. Standard Bayesian computations can then be used. However, given that the operations will have been performed at two, possibly very different, hospitals with possibly very different patient populations, this assumption might very well be judged unsatisfactory.

4. *Jeffrey's Rule.* Suppose that the original probability assignment P was *exchangeable*. That is, $P(001) = P(010) = P(100)$ and $P(110) = P(101) = P(011)$. In the situation described, the information provided contains no information about the order of the next three trials and thus we may well require that the new probability distribution remain exchangeable. This is equivalent to considering a partition $\{E_i\}_{i=0}^3$ of Ω, where $E_0 = \{000\}$, $E_1 = \{001, 010, 100\}$, $E_2 = \{110, 101, 011\}$, $E_3 = \{111\}$. Here E_i is the set of outcomes with i ones, and exchangeability implies that for any event A, and any i, $P(A \mid E_i) = P^*(A \mid E_i)$. To complete the probability assignment P^*, we need a subjective assessment of $P^*(E_i)$. Then P^* is determined by

$$P^*(A) = \sum P^*(A \mid E_i)\, P^*(E_i) = \sum P(A \mid E_i)\, P^*(E_i).$$

* Persi Diaconis is Professor, Department of Statistics, Stanford University, Stanford, CA 94305. Sandy L. Zabell is Associate Professor, Department of Mathematics, Northwestern University, Evanston, IL 60201. The authors are grateful to I.J. Good, Richard C. Jeffrey, David Freedman, Paul Meier, Jim Pitman, Sherry May, and Amos Tversky for helpful comments and suggestions.

© Journal of the American Statistical Association
December 1982, Volume 77, Number 380
Theory and Methods Section

The rule

$$P^*(A) = \sum P(A \mid E_i)\, P^*(E_i) \qquad (1.1)$$

is known in the philosophical literature as *Jeffrey's rule of conditioning*. It is valid whenever there is a partition $\{E_i\}$ of the sample space such that

$$P^*(A \mid E_i) = P(A \mid E_i) \quad \text{for all } A \text{ and } i. \qquad \text{(J)}$$

It has the practical advantage of reducing the assessment of P^* to the simpler task of assessing $P^*(E_i)$.

Approaches 2, 3, and 4 are all special routes to the requantification of approach 1; each is valid or useful under different assumptions. For example, retrospective conditioning assumes that one can do a reasonable job of assessing probabilities as if the data had not been observed; exchangeability assumes that future trials are based on the same mechanism as past ones; Jeffrey's rule assumes the availability of a partition and the validity of assumption (J).

In this article we study the assumptions and conclusions that attend Jeffrey's rule. Our main contributions are technical: In Section 2 we connect Jeffrey's rule with sufficiency; Sections 3, 4, and 5 analyze what happens when two or more partitions are considered. In Section 3 we discuss commutativity of successive updating. In Section 4 we discuss methods for dealing with two partitions simultaneously, giving a necessary and sufficient condition for two probability measures on two algebras to have a common extension. In Section 5 we discuss some other motivations for Jeffrey's rule when condition (J) has not been subjectively checked. Jeffrey's rule gives the "closest" measure to P that fixes $P^*(E_i)$, and it is related to the iterated proportional fitting procedure used in the statistical analysis of contingency tables. For ease of exposition, most of this article assumes a countable state space or a countable partition $\{E_i\}_{i=1}^{\infty}$. In Section 6 we describe the mathematical machinery needed to extend the previous results to abstract probability spaces.

1.2 Bibliographical Note on Probability Revision

From the subjectivistic perspective, the conditional probability $P(A \mid E)$ is the probability we *currently* would attribute to an event A if in addition to our present information we were also to learn E. In the language of betting, it is "the probability that we would regard as fair for a bet on A to be made immediately, but to become operative only if E occurs" (de Finetti 1972, p. 193; compare Ramsey 1931, p. 180). In this formulation, the equality $P(A \mid E) = P(AE)/P(E)$ is not a definition but follows as a theorem derived from the assumption of coherence (de Finetti 1975, Ch. 4).

If we actually *learn E* to be true, it is conventional to adopt as one's new probability

$$P^*(A) = P(A \mid E). \qquad (1.2)$$

Several authors have discussed the limitations on or justifications for this use of the Bayes rule (1.2). Ramsey put the difficulty clearly:

> [The degree of belief in p given q] is not the same as the degree to which [a subject] would believe p, if he believed q for certain; for knowledge of q might for psychological reasons profoundly alter his whole system of beliefs [Ramsey 1931, p. 180; cf. however, p. 192].

For modern discussion of this and related issues, see Hacking (1967), de Finetti (1972, p. 150; 1975, p. 203), Teller (1976), Freedman and Purves (1969). A closely related point is that our "[subjective] probabilities can change in the light of calculations or of pure thought without any change in the *empirical* data . . ." (Good 1977, p. 140). I. J. Good terms such probabilities "evolving" or "dynamic" and has discussed them in a number of papers (Good 1950, p. 49; 1968; 1977).

Other reservations about the adequacy of conditionalization as an exclusive model for belief revision center around its assumption about the form in which new information is received. Indeed, Jeffrey's original philosophical motivation for introducing "probability kinematics" was his belief that "It is rarely or never that there is a proposition for which the direct effect of an observation is to change the observer's degree of belief in that proposition to 1" (Jeffrey 1968, p. 171). Similar criticisms have been raised by Shafer (1979, 1981), whose theory of belief functions is a more radical attempt to deal with the problem. Both hold that conditioning on an event requires the assignment of an initial probability for that event, prior (in principle at least) to its observation, and for many classes of sensory experiences this seems forced, unrealistic, or impossible.

For example, suppose we are about to hear one of two recordings of Shakespeare on the radio, to be read by either Olivier or Gielgud, but are unsure of which, and have a prior with mass $\frac{1}{2}$ on Olivier, $\frac{1}{2}$ on Gielgud. After hearing the recording, one might judge it fairly likely, but by no means certain, to be by Olivier. The change in belief takes place by direct recognition of the voice; all the integration of sensory stimuli has already taken place at a subconscious level. To demand a list of objective vocal features that we condition on in order to affect the change would be a logician's parody of a complex psychological process.

Jeffrey's rule was introduced in Jeffrey (1957) and is further discussed in Jeffrey (1965, Ch. 11) and Jeffrey (1968). Isaac Levi (1967; 1970, pp. 147–152) is a vigorous critic of Jeffrey's version of probability kinematics, but has been thoroughly rebutted by Jeffrey (1970, especially pp. 173–179). Jeffrey's idea was partially anticipated by the Oxford astronomer Donkin (1851, p. 356); compare Boole (1854, pp. 251–252), Whitworth (1901, pp. 162–169, 181–182), Keynes (1921, pp. 176–177). An independent proposal of Jeffrey's rule appears in Griffeath and Snell (1974). The last few years have seen a sudden upsurge of interest in Jeffrey conditionalization; papers have appeared by Teller (1976), Field (1978), Garber (1980), Williams (1980), van Fraassen (1980), Armendt (1980), and Shafer (1981).

2. JEFFREY'S RULE OF CONDITIONING

In this section we develop some of the mathematics connected with Jeffrey's rule of conditioning. Formally: Ω is a countable set, P and P^* are probability measures on the subsets of Ω, and $\{E_i\}$ is a partition of Ω.

2.1 Bayesian Conditioning

Jeffrey's rule of conditioning is a generalization of ordinary conditioning: given the partition $\{E, E^c\}$, if $P^*(E) = 1$ and $P^*(A) = \sum P(A \mid E_i)\, P^*(E_i)$, then $P^*(A) = P(A \mid E)$. We therefore begin by investigating when one measure P^* can arise from another measure P by conditioning. To be precise, suppose P and P^* are measures on a countable space Ω. We will say that P^* *can be obtained from P by conditioning* if there exists a probability space $(\bar{\Omega}, \bar{\mathscr{A}}, Q)$, and events $\{E_\omega\}_{\omega\in\Omega}$, $E_\omega \in \bar{\mathscr{A}}$ (to be thought of as "$E_\omega = \omega$ occurred"), such that $Q(E_\omega) = P(\omega)$, and an event $E \in \bar{\mathscr{A}}$ such that $Q(E) > 0$ and $Q(E_\omega \mid E) = P^*(\omega)$.

Theorem 2.1. P^* can be obtained from P by conditioning if and only if there exists a constant $B \geq 1$ such that

$$P^*(\omega) \leq B\, P(\omega) \quad \text{for all} \quad \omega \in \Omega. \tag{2.1}$$

Proof. If P^* can be obtained from P by conditioning, let $(\bar{\Omega}, \bar{\mathscr{A}}, Q)$, $\{E_\omega\}$, E be given. Then for any $\omega \in \Omega$,

$$P^*(\omega) = Q(E_\omega \mid E) \leq \frac{Q(E_\omega)}{Q(E)} = \frac{P(\omega)}{Q(E)}.$$

This gives (2.1) with $B = 1/Q(E)$.

Conversely, suppose (2.1) is satisfied. If $B = 1$, then $P^* = P$ and the theorem is obvious. If $B > 1$, define

$$P^{**}(\omega) = \frac{B}{B-1} P(\omega) - \frac{1}{(B-1)} P^*(\omega).$$

Because of the condition, P^{**} is a probability and $P = (1/B)\, P^* + (1 - 1/B)\, P^{**}$. This suggests taking $\bar{\Omega} = \Omega \times \{a, b\}$, $E_\omega = (\omega, a) \cup (\omega, b)$, and $E = \cup_\omega(\omega, a)$. Let Q be defined by $Q(\omega, a) = (1/B)\, P^*(\omega)$ and $Q(\omega, b) = (1 - 1/B)\, P^{**}(\omega)$.

Condition (2.1) places a restriction on P, P^* when both have countable support (but not when both have finite support and $\operatorname{supp}(P^*) \subseteq \operatorname{supp}(P)$). For example, no geometric distribution can be obtained from a Poisson distribution by conditioning, but any Poisson distribution can be obtained from any geometric. If Ω is uncountable, (2.1) can be replaced by the conditions $P^* \ll P$ and $dP^*/dP \in L_\infty$; compare Section 6.

2.2 Jeffrey Conditionalization and Sufficiency

In the example discussed in Section 1, the partition $\{E_i\}$ naturally arose in the course of constructing P^* from P. But one might instead envisage being given another person's $\{P, P^*\}$ and then trying to reconstruct a possible partition $\{E_i\}$ from which the pair $\{P, P^*\}$ could have arisen via Jeffrey conditionalization. Unlike Bayesian conditionalization, this turns out to be always possible.

To apply Jeffrey's rule, it is required to find a partition $\{E_i\}$ such that

$$P(A \mid E_i) = P^*(A \mid E_i) \quad \text{for all} \quad A \quad \text{and} \quad i.$$

This is simply the problem of finding a *sufficient partition* for the two-element family $\mathscr{F} = \{P, P^*\}$; see Blackwell and Girshick (1954, Ch. 8). This simple obervation makes possible the translation of the ideas of minimal sufficiency and likelihood ratio into the language of Jeffrey's rule.

A partition $\{E_i\}$ is said to be *coarser* than a second $\{\tilde{E}_j\}$ if every E_i is a union of sets in $\{\tilde{E}_j\}$. For purposes of updating probability, a coarser partition has the advantage that P^* need be specified on fewer sets. A coarsest sufficient partition is said to be minimal sufficient. The following (well-known) theorem gives an alternative version of Jeffrey's rule and states that there is always a coarsest partition for which Jeffrey's rule is valid. Some philosophical implications of this fact are discussed by van Fraassen (1980).

Theorem 2.2. Let P, P^* be probability measures with common support on the countable set Ω. If $\{E_i\}$ is a partition of Ω such that $P(E_i) > 0$ and $P(A \mid E_i) = P^*(A \mid E_i)$ for all subsets A and elements of the partition E_i, then for each $\omega \in \Omega$,

$$P^*(\omega) = \frac{P^*(E_i)}{P(E_i)} P(\omega), \; \omega \in E_i. \tag{2.2}$$

If $R = \{x : P^*(\omega)/P(\omega) = x,\ \omega \in \Omega\}$, and $E_x = \{\omega : P^*(\omega)/P(\omega) = x,\ \omega \in \Omega\}$, then $\{E_x : x \in R\}$ is a minimal sufficient partition for $\{P, P^*\}$.

Proof. The first statement is a version of the Fisher-Neyman factorization theorem; for the second, see Blackwell and Girshick (1954, p. 221).

The following example illustrates the use of the likelihood ratio form of Jeffrey's rule.

Example 2.1. (Whitworth 1901, pp. 167–168):

> *Question 138. A, B, C* were entered for a race, and their respective chances of winning were estimated at $\frac{2}{11}$, $\frac{4}{11}$, $\frac{5}{11}$. But circumstances come to our knowledge in favour of *A*, which raise his chance to $\frac{1}{2}$; what are now the chances in favour of *B* and *C* respectively?
>
> *Answer. A* could lose in two ways, viz. either by *B* winning or by *C* winning, and the respective chances of his losing in these ways were a priori $\frac{4}{11}$ and $\frac{5}{11}$, and the chance of his losing at all was $\frac{9}{11}$. But after our accession of knowledge the chance of his losing at all becomes $\frac{1}{2}$, that is, it becomes diminished in the ratio of 18 : 11. Hence the chance of either way in which he might lose is diminished in the same ratio. Therefore the chance of *B* winning is now
>
> $$\tfrac{4}{11} \times \tfrac{11}{18}, \quad \text{or} \quad \tfrac{4}{18};$$
>
> and of *C* winning
>
> $$\tfrac{5}{11} \times \tfrac{11}{18}, \quad \text{or} \quad \tfrac{5}{18}.$$
>
> These are therefore the required chances.

3. SUCCESSIVE UPDATING

In the usual applications of subjective probability, information builds up by successive conditioning. In Bayesian conditionalization the order in which new information is incorporated is irrelevant; in Jeffrey conditionalization the situation is more complex.

3.1 The Problem

Consider an initial probability P that is Jeffrey-updated to the new probability $P_{\mathscr{E}}$ based on a partition $\mathscr{E} = \{E_i\}_{i=1}{}^e$ and new probabilities $P_{\mathscr{E}}(E_i) = p_i$, $i = 1, 2, \ldots, e$; clearly $P^*(A \mid E_i) = P_{\mathscr{E}}(A \mid E_i) = P(A \mid E_i)$ holds for our new opinion. (P^* denotes our new opinion, how- it is obtained: by Bayes' theorem, Jeffrey's rule, complete requantification, or whatever. $P_{\mathscr{E}}$ denotes the specific updated probability measure that results from Jeffrey conditionalization. Here, by assumption, $P^* = P_{\mathscr{E}}$.) Suppose we then Jeffrey-update on $\mathscr{F} = \{F_j\}_{j=1}{}^f$ with new probabilities $\{q_j\}$, and indicate this order of updating by $P_{\mathscr{E}\mathscr{F}}$. To use Jeffrey's rule at the second stage we must, of course, accept the J-condition, so $P_{\mathscr{E}}{}^*(A \mid F_j) = P_{\mathscr{E}}(A \mid F_j) = P_{\mathscr{E}\mathscr{F}}(A \mid F_j)$. Clearly, the order of updating matters, since the second opinion dominates.

Example 3.1. Suppose $\mathscr{E} = \mathscr{F}$, that is, our belief for each event E_i changes first to p_i and then to q_i. The first revision and second revision differ and we currently believe $P^*(E_i) = q_i$. If the opposite order of revision were employed, we would believe $P^*(E_i) = p_i$ after the second revision.

Example 3.2. Suppose that in a criminal case we are trying to decide which of four defendants, called a, b, c, d, is a thief. We initially think $P(a) = P(b) = P(c) = P(d) = \frac{1}{4}$. Evidence is then introduced to show that the thief was probably left-handed. The evidence does not demonstrate that the thief was definitely left-handed, but it leads us to conclude the $P(\text{thief left-handed}) \approx .8$. If a and b are the defendants who are left-handed, then $E_1 = \{a, b\}$, $E_2 = \{c, d\}$ and $P_{\mathscr{E}}(E_1) = .8$, $P_{\mathscr{E}}(E_2) = .2$. If the *only effect* of the evidence was to alter the probability of left-handedness—in the sense that $P(A \mid E_i) = P_{\mathscr{E}}(A \mid E_i)$—then $P_{\mathscr{E}}$ is obtained from Jeffrey's rule as $P_{\mathscr{E}}(a) = .4$, $P_{\mathscr{E}}(b) = .4$, $P_{\mathscr{E}}(c) = .1$, $P_{\mathscr{E}}(d) = .1$. Evidence is next presented that it is somewhat likely that the thief was a woman. If the female defendants are a and c, then $F_1 = \{a, c\}$, $F_2 = \{b, d\}$. If $P_{\mathscr{E}\mathscr{F}}(F_1) = .7$ and Jeffrey-updating is again judged acceptable, then

$$P_{\mathscr{E}\mathscr{F}}(a) = .56, \quad P_{\mathscr{E}\mathscr{F}}(b) = .24,$$

$$P_{\mathscr{E}\mathscr{F}}(c) = .14, \quad P_{\mathscr{E}\mathscr{F}}(d) = .06.$$

If instead the evidence $(F_1, .7)$, $(F_2, .3)$ is presented first and $(E_1, .8)$, $(E_2, .2)$ is presented second, is $P_{\mathscr{F}\mathscr{E}}$ equal to $P_{\mathscr{E}\mathscr{F}}$? Example 3.1 shows that in general the order matters since the currently held opinion governs; in this example the reader may check that the order does not matter. We now investigate why.

3.2 Commutativity

There are two aspects to successive updating: The updating information at each stage,

$$\{E_i, p_i\}_{i=1}{}^e, \quad \{F_j, q_j\}_{j=1}{}^f; \tag{3.1}$$

the J condition at each stage,

$$P^*(A \mid E_i) = P(A \mid E_i) \quad \text{and} \quad P_{\mathscr{E}}{}^*(A \mid F_j) = P_{\mathscr{E}}(A \mid F_j)$$

or, if updating is being considered in the other order,

$$P^*(A \mid F_j) = P(A \mid F_j) \quad \text{and} \quad P_{\mathscr{F}}{}^*(A \mid E_i) = P_{\mathscr{F}}(A \mid E_i). \tag{3.2}$$

The J condition is an internal or psychological condition that must be checked or accepted at each stage. Mathematics has nothing to offer here.

Mathematics *can* be used to check whether (3.1) is compatible with commutativity. Since Jeffrey updating fixes the probabilities on the partition (i.e., $P_{\mathscr{E}\mathscr{F}}(F_j) = q_j$ and $P_{\mathscr{F}\mathscr{E}}(E_i) = p_i$), commutativity will be possible only if

$$P_{\mathscr{E}\mathscr{F}}(E_i) = p_i \quad \text{and} \quad P_{\mathscr{F}\mathscr{E}}(F_j) = q_j, \tag{3.3}$$

for all i and j. It turns out that this condition is sufficient.

Theorem 3.1. If (3.3) holds, then $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$.

In other words, when $P_{\mathscr{F}\mathscr{E}}$ and $P_{\mathscr{E}\mathscr{F}}$ both incorporate (3.1), they actually coincide. Theorem 3.1 is an immediate consequence of Csiszár (1975, Theorem 3.2) and its proof is omitted. Csiszár's theorem implies that the common measure $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$ is the *I projection* of the original measure P onto the set of measures that incorporate (3.1). We discuss I projection further in Section 5.

3.3 Jeffrey Independence

A second approach to the mathematical aspects of commutativity of successive Jeffrey updating uses independence. Two partitions $\mathscr{E} = \{E_i\}$, $\mathscr{F} = \{F_j\}$ such that $P(E_i) > 0$, $P(F_j) > 0$ for all i and j, are P *independent* if

$$P(E_i \mid F_j) = P(E_i) \quad \text{and} \quad P(F_j \mid E_i) = P(F_j) \tag{3.4}$$

all i, j. Independence says that conditioning on $\mathscr{F}$ does not change the probabilities on $\mathscr{E}$ and vice versa. Analogously,

Definition: $\mathscr{E}$ and $\mathscr{F}$ are Jeffrey independent with respect to P, $\{p_i\}$ and $\{q_j\}$ if $P_{\mathscr{E}}(F_j) = P(F_j)$ and $P_{\mathscr{F}}(E_i) = P(E_i)$ holds for all i and j. (Briefly, "J independent with respect to $\{p_i\}$, $\{q_j\}$.") Thus Jeffrey independence says that Jeffrey updating on $\mathscr{E}$ with probabilities p_i does not change the probability on $\mathscr{F}$ and similarly with $\mathscr{E}$ and $\mathscr{F}$ interchanged. The next theorem shows the connection with commutativity.

Theorem 3.2. Let P, $\{E_i, p_i\}$ and $\{F_j, q_j\}$ be given. Then $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$ if and only if $\mathscr{E}$ and $\mathscr{F}$ are Jeffrey independent with respect to P, $\{p_i\}$, $\{q_j\}$.

Proof. Note that $P_{\mathscr{E}\mathscr{F}}(A) = P_{\mathscr{F}\mathscr{E}}(A)$ for all events A

if and only if

$$\sum_{ij} \frac{p_i q_j}{P_{\mathscr{E}}(F_j)\, P(E_i)} P(AE_iF_j) = \sum_{ij} \frac{p_i q_j}{P_{\mathscr{F}}(E_i)\, P(F_j)} P(AE_iF_j). \quad (3.5)$$

Choose $A = E_{i_0}F_{j_0}$ to get

$$P_{\mathscr{E}}(F_{j_0})\, P(E_{i_0}) = P_{\mathscr{F}}(E_{i_0})\, P(F_{j_0})$$

for all pairs i_0, j_0. Keeping i_0 fixed and summing over j_0 yields

$$P(E_{i_0}) = P_{\mathscr{F}}(E_{i_0}); \quad (3.6a)$$

similarly, fixing j_0 and summing over i_0 yields

$$P_{\mathscr{E}}(F_{j_0}) = P(F_{j_0}). \quad (3.6b)$$

Thus, $\mathscr{E}$ and $\mathscr{F}$ are Jeffrey independent with respect to P, $\{p_i\}$, $\{q_j\}$. Conversely, if (3.6) holds, then

$$P_{\mathscr{E}}(F_j)\, P(E_i) = P(F_j)\, P(E_i) = P_{\mathscr{F}}(E_i)\, P(F_j).$$

Using this equality shows that (3.5) holds and so $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$.

Theorem 3.3. Two partitions $\mathscr{E}$ and $\mathscr{F}$ are P independent if and only if $\mathscr{E}$ and $\mathscr{F}$ are Jeffrey independent with respect to *any* update probabilities $\{p_i\}$ and $\{q_j\}$.

Proof. First suppose $\mathscr{E}$ and $\mathscr{F}$ are P independent. Then

$$P_{\mathscr{E}}(F_j) = \sum_i P(F_j \mid E_i)p_i = \sum_i P(F_j)p_i = P(F_j). \quad (3.7)$$

To see the converse, suppose $\mathscr{E}$ and $\mathscr{F}$ are not P independent. Then there exist E_{i_0} and F_{j_0} such that $P(F_{j_0} \mid E_{i_0}) \neq P(F_{j_0})$. Pick p_{i_0} sufficiently close to 1. Then

$$\sum_i P(F_{j_0} \mid E_i)p_i \neq P(F_{j_0}),$$

hence it follows from (3.7) that $P_{\mathscr{E}}(F_{j_0}) \neq P(F_{j_0})$.

Example 3.3. (J independence $\neq>$ P independence). Suppose $P(E_iF_j)$ is given by the following table

	F_1	F_2	F_3	
E_1	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$
E_2	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{4}$
E_3	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	

Then $\mathscr{E}$ and $\mathscr{F}$ are not P independent, but update probabilities $\mathbf{p}$, $\mathbf{q}$ exist such that $\mathscr{E}$ and $\mathscr{F}$ are J independent with respect to them (see what follows).

An efficient algorithm for checking J independence, in this and other examples, is the following. Let r_{ij} denote W. E. Johnson's *coefficient of dependence* between E_i and F_j (compare Keynes 1921, pp. 150–155), that is,

$$r_{ij} = P(E_iF_j)/P(E_i)\, P(F_j);$$

since $\sum_i r_{ij}p_i = P_{\mathscr{E}}(F_j)/P(F_j)$ and $\sum_j r_{ij}q_j = P_{\mathscr{F}}(E_i)/P(E_i)$, it follows that $\mathscr{E}$ and $\mathscr{F}$ are J independent with respect to $\{p_i\}$, $\{q_j\}$ if and only if

$$\sum_i r_{ij}p_i = 1, \quad \text{all } j; \sum_j r_{ij}q_j = 1, \quad \text{all } i. \quad (3.8)$$

Let $\mathbf{R} = (r_{ij})$. In Example 3.3

$$\mathbf{R} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

and hence, if

$$\mathbf{p} = \left(p, \frac{1-p}{2}, \frac{1-p}{2}\right), 0 < p < 1,$$

and

$$\mathbf{q} = \left(q, \frac{1-q}{2}, \frac{1-q}{2}\right), 0 < q < 1,$$

then $\mathbf{pR} = \mathbf{1}$, $\mathbf{Rq}' = \mathbf{1}$; thus $\mathscr{E}$, $\mathscr{F}$ are J independent with respect to $\mathbf{p}$, $\mathbf{q}$.

Remark. It is not hard to show that if at least one of the two partitions $\mathscr{E}$ and $\mathscr{F}$ has only two elements, then J independence for *some* $\mathbf{p}$, $\mathbf{q}$ pair is equivalent to P independence, and hence to J independence for *all* $\mathbf{p}$, $\mathbf{q}$.

Lest the reader think that commutativity always occurs when (3.1) can be incorporated, we conclude this section with an example that has $P_{\mathscr{E}\mathscr{F}}(E_i) = p_i$ (and of course $P_{\mathscr{E}\mathscr{F}}(F_j) = q_j$), but such that $P_{\mathscr{F}\mathscr{E}}(F_j) \neq q_j$.

Example 3.4. Let $\mathscr{E} = \{E, \bar{E}\}$, $\mathscr{F} = \{F, \bar{F}\}$, and define P by

	F	$\bar{F}$	
E	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
$\bar{E}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{5}{8}$
	$\frac{1}{2}$	$\frac{1}{2}$	

Suppose $p_1 = p_2 = \frac{1}{2}$ and $q_1 = \frac{7}{15}$, $q_2 = \frac{8}{15}$. Then a simple computation shows that $P_{\mathscr{E}\mathscr{F}}(E) = \frac{1}{2} = P_{\mathscr{E}\mathscr{F}}(\bar{E})$, but $P_{\mathscr{F}\mathscr{E}}(F) \neq q_1$.

4. COMBINING SEVERAL BODIES OF EVIDENCE

Suppose we undergo a complex of experiences that result in our simultaneously adopting new degrees of belief P^* on two partitions $\mathscr{E} = \{E_i\}$ and $\mathscr{F} = \{F_j\}$, say

$$P^*(E_i) = p_i \quad \text{and} \quad P^*(F_j) = q_j. \quad (4.1)$$

How should we revise our subjective probabilities so as to incorporate these new beliefs? In general, the theory put forth by de Finetti has no neat mathematical answer to this question—you just have to think about things and quantify your opinion as best you can. In this section we discuss two reasonable routes through this quantification procedure. The routes are reasonable in the same sense that exchangeability is a reasonable thing to consider when attempting to quantify probabilities on repeated events—the circumstances that make them subjectively acceptable occur frequently. We first discuss whether measures satisfying (4.1) exist and then if so, how to uniquely select one.

4.1 Coherence of P^*

If we are to adopt the degrees of belief P^* in (4.1), they must at least be coherent; that is, P^* must be extendable to a probability measure (which we also denote by P^*). Theorem 4.1 provides a simple necessary and sufficient condition for the existence of such extensions.

Theorem 4.1. Let Ω be a countable set, $\mathscr{E} = \{E_i\}$ and $\mathscr{F} = \{F_j\}$ two partitions of Ω, and P, Q two probability measures on $\mathscr{E}$ and $\mathscr{F}$ respectively. There exists a probability measure P^* on Ω such that (4.1) holds if and only if whenever disjoint sets A and B are given, with A a union of elements of $\mathscr{E}$, B a union of elements of $\mathscr{F}$,

$$P(A) + Q(B) \leq 1. \qquad (4.2)$$

Proof. Consider the set $F = \cup\{E_i \times F_j : E_iF_j \neq \phi\}$. This is a closed set in the discrete space $\mathscr{E} \times \mathscr{F}$. Theorem 11 of Strassen (1965) gives a necessary and sufficient condition for the existence of a probability measure P^* on F with margins P and Q. Strassen's condition is easily seen to be equivalent to (4.2), and P^* may be regarded as the required measure on the partition $\{E_iF_j\}$; within a set of this partition P^* may be defined arbitrarily.

Remark. Condition (4.2) is necessary but not sufficient for Theorem 4.1 to hold if Ω is uncountable. See Diaconis and Zabell (1978) and Shortt (1982) for counterexamples and discussion.

4.2 Extending P^*

If (4.1) is coherent, it remains to

1. choose a probability P^* on the partition $\{E_iF_j\}$ that agrees with (4.1);
2. extend P^* to all of Ω.

If judged valid, the easiest way of accomplishing step 1 is to use *independence*: $P^*(E_iF_j) = P^*(E_i)P^*(F_j) = p_iq_j$; step 2 might then be achieved by Jeffrey updating on $\{E_iF_j\}$.

Richard Jeffrey (1957, Ch. 4) has advocated another route from (4.1) to a final probability assignment: successive Jeffrey updating on $\mathscr{E}$ and $\mathscr{F}$. This raises two issues:

1. When does successive updating satisfy (4.1)?
2. When is successive updating reasonable?

Question 1 arises because $P_{\mathscr{E}\mathscr{F}}$ need not equal $P_{\mathscr{F}\mathscr{E}}$. Indeed, Example 3.4 provides a situation where (4.1) is coherent (because $P_{\mathscr{E}\mathscr{F}}$ satisfies (4.1)), but $P_{\mathscr{E}\mathscr{F}} \neq P_{\mathscr{F}\mathscr{E}}$ and $P_{\mathscr{F}\mathscr{E}}$ does not satisfy (4.1). Since matters are simplified when $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$, we note that the results of Section 3 imply that the following three conditions are equivalent:

$$P_{\mathscr{E}\mathscr{F}}(A) = P_{\mathscr{F}\mathscr{E}}(A) \quad \text{for all sets } A. \qquad (4.3a)$$

$$P_{\mathscr{E}\mathscr{F}}(E_i) = P_{\mathscr{F}\mathscr{E}}(E_i) \quad \text{and} \quad P_{\mathscr{F}\mathscr{E}}(F_j) = P_{\mathscr{E}\mathscr{F}}(F_j) \quad \text{for all } i \text{ and } j. \qquad (4.3b)$$

$$P_{\mathscr{F}}(E_i) = P(E_i) \quad \text{and} \quad P_{\mathscr{E}}(F_j) = P(F_j) \quad \text{for all } i \text{ and } j. \qquad (4.3c)$$

Even when the order does not matter, we still have the responsibility of justifying the resort to successive updating, that is, question 2. One approach to this is via checking the Jeffrey condition at each stage of updating. This is a somewhat unorthodox mental exercise, given that we currently believe (4.1), a condition involving both partitions. If we update first on $\mathscr{E}$, then we must check $P(A \mid E_i) = P^*(A \mid E_i)$, which amounts to thinking as if we don't know about $\mathscr{F}$ and are only thinking about $\mathscr{E}$. At the second stage, one then checks $P_{\mathscr{E}}(A \mid F_j) = P_{\mathscr{E}}^*(A \mid F_j)$, comparing one's opinion not knowing $\mathscr{F}$ to one's opinion knowing $\mathscr{F}$. Examples such as Example 3.4 show that this can be tricky. It is a possible route, however, one more general than the route using independence suggested before.

Remark 1. There is no reason to require $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$ for successive updating to be useful and valid. If each of the (J) conditions is judged valid in forming $P_{\mathscr{E}\mathscr{F}}$ and if $P_{\mathscr{E}\mathscr{F}}$ satisfies (4.1), then $P_{\mathscr{E}\mathscr{F}}$ is a consistent quantification of current belief.

Remark 2. Condition (4.3) implies that $P_{\mathscr{E}\mathscr{F}}$ and $P_{\mathscr{F}\mathscr{E}}$ cannot both incorporate (4.1) and both be judged acceptable updates (in the sense that the (J) conditions have been checked) without $P_{\mathscr{E}\mathscr{F}} = P_{\mathscr{F}\mathscr{E}}$. Thus noncommutativity is not a real problem for successive Jeffrey updating.

Remark 3. The approach outlined in this section is an approach to the combination of evidence within the Bayesian framework. See Shafer (1976) for a related, nonadditive approach.

5. MECHANICAL UPDATING

The approach we have taken thus far to justifying Jeffrey's rule is subjective—through checking condition (J). Several authors—Griffeath and Snell (1974), May and Harper (1976), Williams (1980), and van Fraassen (1980)—have pursued a different justification. Given a prior P, partition $\{E_i\}$, and a new measure P^* on $\{E_i\}$, find the "closest" measure to P that agrees with P^* on the partition and take this as defining P^* on the whole space. Since this way of proceeding does not attempt to quantify one's new degrees of belief via introspection, we call this approach *mechanical updating*.

5.1 Minimum Distance Properties

If "close" is defined in any of several common ways, the closest measure is that given by Jeffrey's rule. We illustrate this with three common notions of closeness between measures P and Q on the countable set Ω: (1)

The variation distance

$$\| P - Q \| = \sup\{ | P(B) - Q(B) | : B \subset \Omega\}$$
$$= \frac{1}{2} \sum_{\omega} | P(\omega) - Q(\omega) |. \quad (5.1)$$

Two measures are close in variation distance if they are uniformly close on all subsets. (2) The Hellinger distance

$$H(P, Q) = \sum_{\omega} (\sqrt{P(\omega)} - \sqrt{Q(\omega)})^2. \quad (5.2)$$

(3) The Kullback-Leibler number of Q with respect to P

$$I(Q, P) = \sum_{\omega} Q(\omega) \log (Q(\omega)/P(\omega)). \quad (5.3)$$

The variation and Hellinger distances are actual metrics on the space of probability distributions, the Kullback-Leibler number is not, being asymmetric in its arguments. Kailath (1967) and Csiszár (1977) are good surveys, with bibliographies, of the properties of (5.1), (5.2), and (5.3).

Theorem 5.1. Let Ω be a countable set, P a probability on Ω, and $\{E_i\}$ a partition of Ω. Suppose $P^*(E_i) \geq 0$ are given numbers such that $\sum P^*(E_i) = 1$. Let Q be a probability on Ω such that $Q(E_i) = = P^*(E_i)$. Then

$$\| Q - P \| \geq \tfrac{1}{2} \sum | P(E_i) - P^*(E_i) |, \quad (5.4)$$

$$H(Q, P) \geq \sum (\sqrt{P(E_i)} - \sqrt{P^*(E_i)})^2, \quad (5.5)$$

$$I(Q, P) \geq \sum P^*(E_i) \log (P^*(E_i)/P(E_i)). \quad (5.6)$$

In (5.5) and (5.6) equality holds if and only if $Q(A) = \sum P(A \mid E_i) P^*(E_i)$.

Remarks. (a) Although the probability measure given by Jeffrey's rule minimizes the variation distance, it does not do so uniquely; see May (1976). (b) In Theorem 5.1, the minimum distance between P and Q is the distance between P and Q viewed as measures on the partition $\{E_i\}$. (c) A result like Theorem 5.1 holds for several other notions of distance; see Section 6, where a generalization of Theorem 5.1 is given (Theorem 6.1).

5.2 I Projections and the IPFP

Mechanical updating allows the possibility of updating on collections of sets more general than partitions. Suppose we want to adopt new degrees of belief $P^*(E_i) = p_i$, $1 \leq i \leq n$, where $\mathscr{E} = \{E_1, E_2, \ldots, E_n\}$ is not necessarily a partition of Ω. This situation is closely related to Jeffrey's proposal of updating simultaneously on several partitions, mentioned in Section 4, inasmuch as updating simultaneously on partitions $\mathscr{E}_1, \mathscr{E}_2, \ldots, \mathscr{E}_k$ is the same as updating on $\mathscr{E} = \cup_{i=1}^{k} \mathscr{E}_i$. Conversely, updating on $\mathscr{E} = \{E_1, \ldots, E_n\}$ can be viewed as updating simultaneously on the partitions $\mathscr{E}_1 = \{E_1, E_1^c\}$, $\mathscr{E}_2 = \{E_2, E_2^c\}, \ldots, \mathscr{E}_n = \{E_n, E_n^c\}$. In general, the set $C = \{Q: Q(E_i) = p_i \text{ for all } i\}$ is a convex set of probability measures on Ω that can be empty, contain a single element, or contain many elements. In the first case P^* is incoherent, in the second P^* is uniquely determined. When the third case holds, we can use the Kullback-Leibler number as a notion of "distance" to pick a unique member of C closest to P.

Theorem 5.2. Let $S(P, \infty) = \{Q: I(Q, P) < \infty\}$. If $S(P, \infty) \cap C \neq \phi$, then there exists a unique element $Q_J \in C$ such that $I(Q_J, P) = \inf\{I(Q, P): Q \in C\}$.

Proof. This is an immediate consequence of Csiszár (1975, Theorem 2.1), C being convex and closed with respect to the variation distance.

In Csiszár's terminology, Q_J is the *I projection of P onto C*. (The term is meant to suggest the projection of a vector in $\mathbf{R}^n$ onto a subspace.) The I projection is closely related to a technique widely used in the statistical analysis of contingency tables.

A standard method of adjusting an $r \times c$ contingency table so that it has the desired marginal totals is the iterated proportional fitting procedure (IPFP). In this, one first adjusts the table to have specified row sums, say (by dividing the numbers of a given row by the appropriate factor), next adjusts the new table to have the correct column sums, and then continues iteratively. It follows from Csiszár (1975, Theorem 3.2) that this procedure converges to the I projection of the initial table onto the set of tables with the specified row and column sums (provided, of course, this set is nonempty). That is, the IPFP finds the "closest" table to the original table with the prescribed margins. This is essentially the same as finding the closest measure to an initial probability with prescribed values on two partitions.

The IPFP can be used to compute Q_J of Theorem 5.2 by treating the problem as an n-dimensional contingency table with given margins $P^*(E_i)$, $1 - P^*(E_1), \ldots$.

5.3 Comparing Different Metrics

Theorem 5.1 suggests that Jeffrey's rule is an uncontroversial form of mechanical updating in the sense that it agrees with virtually every minimum-distance rule. As noted earlier, in the case of two or more partitions, the I projection or maximum-entropy solution can be viewed as a limiting form of successive Jeffrey updating. This is perhaps of some interest inasmuch as mechanical updating via the other minimum-distance methods need not, in general, yield the same answer as the I projection.

Example 5.1. (I projection $\neq$ minimum variation distance.) Consider passing from an initial table

$$\mathbf{P}^0 = \begin{array}{|c|c|} \hline \frac{1}{4} & \frac{1}{4} \\ \hline \frac{1}{4} & \frac{1}{4} \\ \hline \end{array} \quad \text{to} \quad \mathbf{P} = \begin{array}{c|c|c|} & \frac{1}{3} & \frac{2}{3} \\ \hline \frac{1}{3} & p_1 & p_2 \\ \hline \frac{2}{3} & p_3 & p_4 \\ \hline \end{array},$$

a new table with the specified margins, which is otherwise as "close" to the original table as possible, according to some notion of closeness.

1. The independent table $\mathbf{P}^I$ given by $p_1 = \frac{1}{9}$, $p_3 = p_3 = \frac{2}{9}$, $p_4 = \frac{4}{9}$ minimizes $I(\mathbf{P}, \mathbf{P}^0)$, since $\mathbf{P}^0$ is independent and I projections preserve the association factor of a 2×2 table (see, e.g., Mosteller 1968, p. 3). The variation distance for this table is

$$\| \mathbf{P}^I - \mathbf{P}^0 \| = \frac{1}{2} \sum_{i=1}^{4} | p_i - \tfrac{1}{4} |$$
$$= \tfrac{1}{2} | \tfrac{1}{9} - \tfrac{1}{4} | + | \tfrac{2}{9} - \tfrac{1}{4} | + \tfrac{1}{2} | \tfrac{4}{9} - \tfrac{1}{4} |$$
$$= \tfrac{7}{36}.$$

2. To find the table $\mathbf{P}^V$ with minimum variation distance from $\mathbf{P}^0$, subject to the margin constraints, note that given p_1, one has $p_2 = p_3 = \frac{1}{3} - p_1$ and $p_4 = p_1 + \frac{1}{3}$. Hence

$$\| \mathbf{P} - \mathbf{P}^0 \| = \tfrac{1}{2} \sum | p_i - \tfrac{1}{4} |$$
$$= \tfrac{1}{2}\{ | \tfrac{1}{4} - p_1 | + 2 | \tfrac{1}{12} - p_1 | + | -\tfrac{1}{12} - p_1 | \}$$

which is minimized by $p_1 = \frac{1}{12}$, the median of $\{-\frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{4}\}$. Hence $\mathbf{P}^V = (\frac{1}{12}, \frac{1}{4}, \frac{1}{4}, \frac{5}{12})$ and $\| \mathbf{P}^V - \mathbf{P}^0 \| = \frac{1}{6}$.

There has been considerable interest recently in maximum entropy methods, especially in the philosophical literature (Rosenkrantz 1977; Williams 1980; van Fraassen 1980). Example 5.1 suggests that any claims to the effect that maximum-entropy revision is the only correct route to probability revision should be viewed with considerable caution because of its strong dependence on the measure of closeness being used.

6. ABSTRACT PROBABILITY KINEMATICS

In this section we briefly discuss the generalization of Jeffrey's rule of conditioning from the countable setting to general spaces.

Consider a probability space $(\Omega, \mathcal{A}, P)$, thought of as describing our current subjective beliefs about the σ algebra of events $\mathcal{A}$. Let P^* be a new probability measure on $\mathcal{A}$ and $\mathcal{A}_0 \subseteq \mathcal{A}$ a sub-σ-algebra of $\mathcal{A}$. Let C be an $\mathcal{A}_0$-measurable set such that $P(C) = 0$ and $\bar{P} \ll \bar{P}^*$ on $\Omega - C$, where $\bar{P}$, $\bar{P}^*$ are the restrictions of P, P^* to $\mathcal{A}_0$. The appropriate version of Jeffrey's condition (J) is

$$\mathcal{A}_0 \quad \text{is sufficient for} \quad \{P, P^*\}. \qquad (\mathrm{J}')$$

When condition (J′) holds, Jeffrey's rule of conditioning becomes:

$$P^*(A) = \int_{\Omega - C} P(A \mid \mathcal{A}_0)\, P^*(d\omega) + P^*(A \cap C), \qquad (6.1)$$

where $P(A \mid \mathcal{A}_0)$ is the conditional probability of A given $\mathcal{A}_0$. If $P^* \ll P$, we can take $C = \phi$.

Much of the mathematical machinery for dealing with Jeffrey conditionalization in this generality has been developed (for a different purpose) by Csiszár (1967). His Lemma 2.2 translates into a likelihood-ratio version of Jeffrey's rule (compare (2.2)): Let λ be a σ-finite measure that dominates P, P^*. Let $\bar{\lambda}$, $\bar{P}$, $\bar{P}^*$ be the restrictions to $\mathcal{A}_0$. Assume $\bar{\lambda}$ is σ finite. Let $\bar{p}(x)$, $\bar{p}^*(x)$ be the densities of P, P^* with respect to $\bar{\lambda}$, and p^* the density of P^* with respect to λ. If condition (J′) holds, then

$$\begin{aligned} p^*(x) &= \bar{p}^*(x)/\bar{p}(x) \quad \text{if} \quad \bar{p}(x > 0 \\ &= p^*(x) \qquad\quad\ \ \text{if} \quad \bar{p}(x) = 0. \end{aligned} \qquad (6.2)$$

Identity (6.2) is a version of the Fisher-Neyman factorization theorem (see Halmos and Savage 1949).

Csiszár's results allow us to give a single theorem that includes Theorem 5.1, showing that the closest measure to P that agrees with P^* on $\mathcal{A}_0$ is the measure given by (6.1). Csiszár introduced the notion of f divergence, where f is a convex function defined on the interval $(0, \infty)$. If μ_1 and μ_2 are two measures on $(\Omega, \mathcal{A})$, the f divergence of μ_1 and μ_2 is

$$I_f(\mu_1, \mu_2) = \int p_2(x)\, f\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx),$$

where $\mu_i \ll \lambda$ and $p_i = d\mu_i/d\lambda$, $i = 1, 2$. Taking $f(u) = u \log u$ gives the Kullback-Leibler number, $f(u) = (u^{1/2} - 1)^2$ the Hellinger distance, $f(u) = | u - 1 |/2$ the variation distance. Csiszár shows that several other notions of distance are also f divergences for an appropriate f.

Theorem 6.1. Let C be the set of probability measures on $(\Omega, \mathcal{A})$ that agree with P^* on $\mathcal{A}_0$, and let f be a convex function on $(0, \infty)$. Then under condition (J′),

$$I_f(P^*, P) = I_f(\bar{P}^*, \bar{P}) = \inf\{I_f(Q, P) : Q \in C\}. \qquad (6.3)$$

If f is strictly convex, than P^* is the unique probability measure on $\mathcal{A}$ that minimizes the right side of (6.3).

Proof. The first equality follows from the sufficiency of $\mathcal{A}_0$ for $\{P, P^*\}$, the second from Csiszár's (1967, Sec. 3) version of the minimum information discrimination theorem of Kullback and Leibler: $I_f(Q, P) \geq I_f(\bar{Q}, \bar{P})$. Since $I_f(\bar{Q}, \bar{P}) = I_f(\bar{P}^*, \bar{P})$, (6.3) follows. If f is strictly convex, then $I_f(\cdot, P)$ is also, and the theorem follows.

[*Received October 1979. Revised April 1982.*]

REFERENCES

ARMENDT, B. (1980), "Is There a Dutch Book Argument for Probability Kinematics?" *Philosophy of Science* 47, 583–588.

BLACKWELL, D., and GIRSHICK, M.A. (1954), *Theory of Games and Statistical Decisions*, New York: John Wiley.

BOOLE, G. (1854), *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*, London: Macmillan.

CSISZÁR, I. (1967), "Information Type Measures of Difference of Probability Distributions and Indirect Observations," *Studia Scientiarum Mathematicarum Hungarica*, 2, 299–318.

——— (1975), "I-Divergence Geometry of Probability Distributions of Minimization Problems," *Annals of Probability*, 3, 146–158.

——— (1977), "Information Measures: A Critical Survey," *Transactions of the Seventh Prague Conference*, (Prague 1974), Prague: Academia, 73–86.

DE FINETTI, B. (1972), *Probability, Induction and Statistics*, New York: John Wiley.

——— (1974), *Theory of Probability* (vol. 1), New York: John Wiley.

——— (1975), *Theory of Probability* (vol. 2), New York: John Wiley.

DIACONIS, P., and ZABELL, S. (1978), "Updating Subjective Probability," Technical Report No. 136, Stanford University, Dept. of Statistics.

DONKIN, W.F. (1851), "On Certain Questions Relating to the Theory of Probabilities," *The Philosophical Magazine*, Ser. 4, 1, 353–368.

FIELD, H. (1978), "A Note on Jeffrey Conditionalization," *Philosophy of Science*, 45, 361–367.

FISCHOFF, B. (1975), "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgement Under Uncertainty," *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.

FISCHOFF, B., and BEYTH, R. (1975), "I Knew It Would Happen—Remembered Probabilities of Once Future Things," *Organizational Behavior and Human Performance*, 13, 1–16.

FREEDMAN, D., and PURVES, R. (1969), "Bayes Method for Bookies," *Annals of Mathematical Statistics*, 40, 1177–1186.

GARBER, D. (1980), "Discussion: Field and Jeffrey Conditionalization," *Philosophy of Science*, 47, 142–145.

GOOD, I.J. (1950), *Probability and the Weighing of Evidence*, New York: Hafner.

——— (1968), "Corroboration, Explanation, Evolving Probability, Simplicity, and a Sharpened Razor," *British Journal for the Philosophy of Science* 19, 123–143.

——— (1977), "Dynamic Probability, Computer Chess, and the Measurement of Knowledge," in *Machine Intelligence 8*, ed. E. W. Elcock and D. Mitchie, Ellis Horwood Ltd. and John Wylie, 139–150.

GRIFFEATH, D., and SNELL, L.J. (1974), "Optimal Stopping in the Stock Market," *Annals of Probability*, 2, 1–13.

HACKING, I. (1967), "Slightly More Realistic Personal Probability," *Philosophy of Science*, 34, 311–325.

HALMOS, P., and SAVAGE, L.J. (1949), "Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics," *Annals of Mathematical Statistics*, 20, 225–241.

JEFFREY, R. (1957), "Contributions to the Theory of Inductive Probability," Ph.D. thesis, Princeton University, Dept. of Philosophy.

——— (1965), *The Logic of Decision*, New York: McGraw-Hill.

——— (1968), "Probable Knowledge," in *The Problem of Inductive Logic*, ed. I. Lakatos, 166–180, Amsterdam: North-Holland.

——— (1970), "Dracula Meets Wolfman: Acceptance vs. Partial Belief," *Induction, Acceptance, and Rational Belief*, ed. M. Swain, 157–185, Dordrecht: D. Reidel.

KAILATH, T. (1967), "The Divergence and Bhattacharya Distance Measures in Signal Detection," *IEEE Transactions on Communications Technology*, 15, 52–60.

KEYNES, J.M. (1921), *A Treatise on Probability*, London: Macmillan.

LEVI, I. (1967), "Probability Kinematics," *Philosophy of Science*, 18, 197–209.

——— (1970), "Probability and Evidence," in *Induction, Acceptance, and Rational Belief*, ed. M. Swain, 134–156, Dordrecht: D. Reidel.

MAY, S. (1976), "Probability Kinematics: A Constrained Optimization Problem," *Journal of Philosophical Logic*, 5, 395–398.

MAY, S., and HARPER, W. (1976), "Toward an Optimization Procedure for Applying Minimum Change Principles in Probability Kinematics," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science* (vol. 1), eds. W.L. Harper and C.A. Hooker, 137–166, Dordrecht: D. Reidel.

MOSTELLER, F. (1968), "Association and Estimation in Contingency Tables," *Journal of the American Statistical Association* 63, 1–28.

RAMSEY, F.P. (1931), "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays*, ed. R.G. Braithwaite, 156–198, London: Routledge and Kegan Paul: reprinted (1964) in *Studies in Subjective Probability*, eds. H.E. Kyburg, Jr., and H.E. Smokler, 61–92, New York: John Wiley.

ROSENKRANTZ, R.D. (1977), *Inference, Method, and Decision*, Dordrecht: D. Reidel.

SAVAGE, L.J. (1967), "Difficulties in the Theory of Personal Probability," *Philosophy of Science*, 34, 305–310.

SHAFER, G. (1976), *A Mathematical Theory of Evidence*, Princeton: Princeton University Press.

——— (1979), "Two Theories of Probability," in *PSA 1978*, vol. 2, East Lansing, Michigan: Philosophy of Science Association.

——— (1981), "Jeffrey's Rule of Conditioning," *Philosophy of Science*, 48, 337–362.

SHORTT, R. (1982), "Distributions With Given Margins," Ph.D. dissertation, Massachusetts Institute of Technology, Dept. of Mathematics.

SLOVIC, P., and FISCHOFF, B. (1977), "On the Psychology of Experimental Surprises," *Journal of Experimental Psychology: Human Perception and Performance*, 3, 544–551.

STRASSEN, V. (1965), "The Existence of Probability Measures With Given Marginals," *Annals of Mathematical Statistics*, 36, 423–439.

TELLER, P. (1976), "Conditionalization, Observation, and Change of Preference," in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science* (vol. 1), eds. W.L. Harper and C.A. Hooker, 205–253, Dordrecht: D. Reidel.

VAN FRAASSEN, BAS. C. (1980), "Rational Belief and Probability Kinematics," *Philosophy of Science* 47, 165–187.

WHITWORTH, W.A. (1901), *Choice and Chance* (5th ed.), Cambridge: Deighton Bell.

WILLIAMS, P.M. (1980), "Bayesian Conditionalization and the Principle of Minimum Information," *British Journal for the Philosophy of Science*, 31, 131–144.