2.5

Statistical Implications of Finitely Additive Probability

JOSEPH B. KADANE, MARK J. SCHERVISH, AND TEDDY SEIDENFELD

I. INTRODUCTION

In his classic monograph on the foundations of the theory of probability, Kolmogorov (1956) introduces a postulate [P6], equivalent to the principle of countable additivity, which he justifies as a mathematical expedient for infinite probability structures. Countable additivity requires that the denumerable union of pairwise disjoint events has probability equal to the sum of the individual probabilities, i.e. if $A = \cup_{i=1}^{\infty} A_i$ ($A_i \cap A_j = \varnothing$, $i \neq j$), then $P(A) = \Sigma_{i=1}^{\infty} P(A_i)$. His theory less P6 is hereafter described as the theory of finitely additive probability.

Our dispute with Kolmogorov's characterization of countable additivity as an "expedient" does not stem from the fear of non-measurability arising in routine problems of statistical inference. For our purposes it is enough that probability be defined over a σ-field. The questions we ask are:

1. Does countable additivity build in unexpected statistical consequences?
2. Does Bayesian statistics mandate the added restrictions countable additivity imposes?

We thank Morris DeGroot, Isaac Levi and Herman Rubin for their suggestions and criticisms.

Reprinted from *Bayesian Inference and Decision Techniques* (Amsterdam: Elsevier Science Publishers, 1986), pp. 59–76, with permission of the Editors, Prem K. Goel and Arnold Zellner.

We argue that the answers are yes and no (respectively).

Without countable additivity, finitely additive probability undergoes a failure of "conglomerability" (discussed in section II). Without conglomerability, familiar decision theoretic principles, e.g. admissibility (discussed in section III) and other forms of dominance rules (discussed in section IV) are invalid. What reason is there for taking seriously the finitely additive theory if dropping countable additivity undermines what has become accustomed statistical decision theory?

First, if one adopts the standards of coherence defended by deFinetti (1974), Savage (1974), and Lindley (1981), one's beliefs are modeled by a probability that is finitely but not necessarily countably additive. (See Seidenfeld and Schervish, 1983, for discussion of this view.) Second, as we investigate in section V, there are important connections between finite additivity and Bayesian reconstructions of standard (orthodox) statistical inference – reconstructions using "improper" priors in the fashion of Jeffreys (1961) and Lindley (1970). By interpreting improper distributions as finitely and not countably additive probabilities, the way is paved to resolve a host of anomalies thought by some to be evidence of inconsistency in Bayesian theory. In this we agree with Heath and Sudderth (1978), Hill (1980), and Levi (1980).

II. NON-CONGLOMERABILITY OF MERELY FINITELY ADDITIVE PROBABILITY

Define P to be *conglomerable in* π (see deFinetti, 1972, p. 99) when for every event E such that $P(E|h_i)$ is defined for all i, and for all constants k_1, k_2, if $k_1 \leq P(E|h_i) \leq k_2$ for all $h_i \in \pi$, then $k_1 \leq P(E) \leq k_2$. That is to say, conglomerability asserts that, for each event E, if all the conditional probabilities over a partition π are bounded by two quantities, k_1 and k_2, then the unconditional probability for that event is likewise bounded by these two quantities. DeFinetti draws attention to non-conglomerability of finitely additive probability in denumerable partitions.

Example 2.1 (due to P. Lévy, 1930; see deFinetti, 1972, p. 102). Let $P(\cdot)$ be a finitely additive probability defined over the field of all subsets of the denumerable set of points $\{\langle i, j\rangle : i, j$ are positive integers$\}$. Thus, one may think of $P(\cdot)$ defined over subsets of integer coordinate points $\langle i, j\rangle$ in the first quadrant. Constrain $P(\cdot)$ so that for any point $\langle i, j\rangle$,

$P(\langle i, j\rangle) = 0$. Hence, $P(\cdot)$ is not countably additive. Also, for any pair $\langle i, j\rangle$, let $P(\langle i, j\rangle | B) = 0$ if B is an infinite set. Hence, conditional on an infinite set B, $P(\cdot | B)$ is again not countably additive.

Now, using the finitely additive probability $P(\cdot)$ (given above), we see that conglomerability must fail with respect to the event $A = \{\langle i, j\rangle: j \geq i\}$. That is, let A be the region whose lower boundary is the diagonal "$i = j$". Consider the partition $\pi_1 = \{h_i: h_i = \{\langle i, j\rangle: j < \omega\}$ and $i < \omega\}$. That is, π_1 is the partition into "vertical" sections. Since, for each $i < \omega$, h_i is an infinite set, $P(A | h_i) = 1$ (since there are only finitely many points below the point $\langle i, i\rangle$, for each $i < \omega$). Let $\pi_2 = \{h'_j: h'_j = \{\langle i, j\rangle: i < \omega\}$ and $j < \omega\}$. That is, π_2 is the partition into "horizontal" sections. Since, for each $j < \omega$, h'_j is an infinite set, $P(A | h_j) = 0$ (since there are only finitely many points to the left of the point $\langle j, j\rangle$, for each $j < \omega$). Thus, conglomerability must fail for at least one of π_1, π_2 as $P(A)$ must differ from at least one of the two values 0, 1.

In an earlier paper (Schervish, Seidenfeld and Kadane, 1984), we investigate several general questions about the existence and magnitude of failures of deFinetti's conglomerability principle. Using a result due to Dubins (1975) (that conglomerability in a partition is equivalent to "disintegrability" in that partition), we provide least upper bounds on the failures of conglomerability with respect to *denumerable* partitions. We also show that for finitely additive probabilities, non-conglomerability in *denumerable* partitions characterizes those which fail to be countably additive, confirming a statement of deFinetti (1972, p. 99).

III. ADMISSIBILITY

Non-conglomerability of finitely additive probabilities, i.e. the phenomenon that an unconditional probability may lie outside the range of values of conditional probabilities over an exhaustive partition, quickly leads to a violation of a familiar decision-theoretic principle, admissibility.

Let O_1 and O_2 be two options in a decision and let π be a partition into states (independent of the options) such that for each state the same option, say O_1, is strictly preferred to the other option, O_2. Then O_2 is *inadmissible*, i.e. O_1 is strictly preferred to O_2 unconditionally.

However, if we consider a choice between O_1: bet on the event A

(as defined in Example 2.1) at even odds, and O_2: bet on the complementary event $\bar{A}$ at even odds, then for each $h_i \in \pi_1$, O_1 is strictly preferred (in expectation) to O_2; while for each $h'_j \in \pi_2$, O_2 is strictly preferred (in expectation) to O_1. It cannot be both that O_1 is strictly preferred to O_2 *and* that O_2 is strictly preferred to O_1. Therefore, failures of conglomerability entail failures of admissibility. This observation permits a simplification of Arrow's (1972) axiom system for preferences which include both a principle of Monotone Continuity (Villegas, 1964) and an admissibility rule, called Dominance. Since the Dominance principle implies countable additivity, Monotone Continuity is redundant.

The fundamental property of admissible procedures, namely that they alone are Bayesian procedures or limits of them, fails for finitely additive probabilities. There are reasonable acts, Bayesian with respect to a finitely additive opinion, which are inadmissible.

Historically, admissibility took on its greatest importance for statistics when Stein (1955) showed that $\bar{x}$ is inadmissible as an estimate of the mean μ of a normal vector with dimension greater than two. Stein showed that drawing in the components of $\bar{x}$ toward an arbitrary origin was a strict improvement over $\bar{x}$ in the admissibility sense. Lindley (1962) later showed in a simple Bayesian model in which μ itself is considered to be normally distributed with mean μ_0, that $\bar{x}$ should be drawn toward μ_0 by an amount determined by the prior variance of μ around μ_0, and the sampling variance of $\bar{x}$ around μ. These observations in turn led to an increased interest in empirical Bayes methods, using the data to estimate hyperparameters like μ_0. In general, the move away from the automatic use of $\bar{x}$, occasioned by Stein's and Lindley's work, has been a healthy development for statistics, we think. Statisticians have been encouraged by these results to consult their prior beliefs more systematically than they had been before.

A consequence of accepting the coherence of merely finitely additive options is to reduce the importance of the concept of admissibility. In particular, there are finitely but not countably additive, prior distributions for which $\bar{x}$ is an optimal Bayesian act regardless of the dimension of $\bar{x}$. We do not interpret this to mean that now people should go back to using $\bar{x}$ (if they ever stopped) without concern for their prior opinion. Rather, we think that our results re-emphasize the importance of using your opinion within a Bayesian paradigm that does not insist on countably additive "priors".

IV. COHERENCE AND "STRONG INCONSISTENCY" WITH FINITELY ADDITIVE PROBABILITY

Associated with a failure of conglomerability are instances of "strong inconsistencies", as Stone (1976) calls them. These are colorfully illustrated by the following game (Stone, 1981), which is a rewording of Stone's (1976) "Flatland" example.

A regular tetrahedral die is rolled a very large number of times. The faces of the die are labelled: e^+ (positron); e^- (electron); μ^+ (muon); and μ^- (antimuon). A record is kept of the outcomes subject to the constraint that if complementary events occur successively in the record, they "annihilate" each other and the record contracts, without trace of the "annihilation". Thus, the sequences $\ldots e^+e^- \ldots, \ldots e^-e^+ \ldots$, $\ldots \mu^+\mu^- \ldots$, and $\ldots \mu^-\mu^+ \ldots$, cannot occur in the record. At some (arbitrary) point in the sequence of rolls the player, who is ignorant of the outcome to date, calls for one last roll after which he is shown the final record. He is then asked to gamble on the outcome of the last toss of the die.

Consider a finitely additive prior over the countable set S of all possible states of the record prior to the final toss defined as follows. First arrange all elements of S into a single sequence, and let B_n be the set consisting of the first n of them. For $B \subseteq S$, let $\lambda_n(B) = \#(B \cap B_n)/n$. Take any limit point of the sequence λ_n as a prior λ. Since every finite subset of S has zero probability under λ, define for any finite set C, $\lambda(A|C) = \#(A \cap C)/\#(C)$. This set of conditional probabilities is consistent in the sense that if A and B are any subsets of S, and C is any finite subset such that $A \cap C \neq \varnothing$, $\lambda(B|A \cap C)\, \lambda(A|C) = \lambda(B \cap A|C)$. If the player is a Bayesian who adopts this prior distribution, then upon seeing the record he will assign equal probability to the four possible outcomes compatible with the record. For example, if the record ends $\ldots e^+\mu^+$, he will assign probability 1/4 to the four possible outcomes of the final toss. But then the player assigns probability 1/4 to each of the following four states of the record as it existed immediately before the final roll of the die:

$\ldots e^+$ (corresponding to a final toss landing μ^+),
$\ldots e^+\mu^+e^+$ (corresponding to a final toss landing e^-),
$\ldots e^+\mu^+e^-$ (corresponding to a final toss landing e^+) and
$\ldots e^+\mu^+\mu^+$ (corresponding to a final toss landing μ^-).

Thus, he assigns probability 3/4 to the event that the final toss resulted in an "annihilation" in the record for the final entry. Moreover, with the prior described above, the player assigns probability 3/4 to the event "annihilation" in the record for the final entry (call this event A) for each observation of a non-vacuous record. If he were to observe a blank record, then for certain the last roll resulted in A. Hence,

$$3/4 \leq P(A|x) \leq 1, \quad \text{for each observation } x. \tag{1}$$

Let us call the state of the record just prior to the final toss, θ, the parameter. Then, on the assumption that the die is fair,

$$0 \leq P(A|\theta) \leq 1/4, \quad \text{for each } \theta. \tag{2}$$

Note that $P(A \mid \theta) = 0$ only when $\theta = \theta_0$, corresponding to a blank record just prior to the final toss. $P(A \mid \theta) = 1/4$ for all other parameter states.

If conglomerability applies, then incoherence results as $3/4 \leq P(A)$ by (1) and $P(A) \leq 1/4$ by (2). It is trivial to "make book" in such circumstances, by betting at, say, 1:3 odds both against and for the event A. The player who (incoherently) adopts conglomerability is led to accept these gambles, as shown by the two (inconsistent) inequalities for $P(A)$.

What if the player adopts conglomerability in only one of the two partitions? For instance, Levi (1980, pp. 284–287) requires conglomerability in the $\pi_\theta = \{\theta_i\}$ partition, as a consequence of his theory of "direct inference". It appears that most other writers assume conglomerability in this, the margin of the "parameter". The result is a finitely additive distribution which is coherent, in the sense that no finite collection of bets suffice to "make book". However, on pain of a sure loss, the player with such a coherent, finitely additive distribution will decline the offer of an infinite class of "fair" bets, where he is agreeable to every finite subclass of wagers. This is, according to Stone, a "strong inconsistency".

For example, let conglomerability apply in $\pi_\theta = \{\theta_i, i = 0, \ldots\}$. Then prior to the game the player has odds of 1:3 on A (an "annihilation" on the final roll), as $P(A) = 1/4$ (since $P(\theta_0) = 0$). Prior to the game, the player also holds the infinite collection of conditional odds, given $x = x$ $(i = 0, \ldots)$, of (at least) 3:1 on A, for each possible observation x_i. But the player will not accept all the denumerably many called-off bets (called off in case $x = x_i$ fails to occur) at the conditional odds of 3:1 on A, while also agreeing to wager on A at the unconditional odds of 1:3. To do so would expose him to a sure loss. Of course, the player is willing to accept any finite subset of this infinite set of wagers. No finite subset is sufficient to fix a sure loss. Hence, the player is coherent in deFinetti's sense (1974, ch. 3).

A similar argument applies in case the player holds conglomerability only in $\pi_x = \{x_i\}$, the margin of the "observable". Then $P(A) = 3/4$. The requisite infinite class of "fair" gambles is constructed by considering the conditional odds on the event A, given the state of the record immediately prior to the final toss (θ_i). The player is agreeable to each of the denumerably many called-off bets on A, called-off in case $\theta = \theta_i$ fails, at odds of 1:3, as $P(A \mid \theta_i) \leq 1/4$ ($i = 0, \ldots$). Once again, the player will not accept all such bets at once. Instances of "strong inconsistencies" are thus not violations of coherence.

Additional cases of "strong inconsistency" (without failure of coherence in deFinetti's sense) can arise when a posterior probability is calculated by Bayes' Theorem for densities with an "improper" prior, subject to the assumption of conglomerability in the margin of the parameter (the unobserved quantity). Heath and Sudderth (1978) illustrate this with their example 5.2. We discuss their theory in the next section.

V. "IMPROPER" PRIORS, COHERENT POSTERIORS, AND ATTEMPTS AT CURTAILING NON-CONGLOMERABILITY

The use of improper distributions to represent ignorance has a long history (see, for example, Jeffreys, 1961 [first edition 1939], Lindley, 1970 [first edition 1965], and Hartigan, 1964). Improper distributions are not probabilities because they assign infinite mass to the universal set, e.g. Lebesgue measure on the real line. An improper distribution may be translated directly into a set of finitely additive probabilities, as does Levi (1980, pp. 125–131), who treats improper distributions as σ-finite representations of finitely additive measures. Alternatively, an improper distribution may arise as a limit of finite measures which can be normalized to be probabilities (see Renyi, 1955).

The limits of the normalized measures often do not exist, however. For example, if the universal set is the real numbers, let λ_n be Lebesgue measure restricted to the interval $[-n, n]$. Each λ_n can be normalized to $\mu_n = (2n)^{-1}\lambda_n$ which is a probability. Whereas λ_n converges to Lebesgue measure on the entire line, the sequence $\{\mu_n\}_{n=1}^{\infty}$ does not converge to a countably additive measure. There exist subsequences of

$\{\mu_n\}$, however, which converge to finitely additive probabilities. This can be seen as follows. Each probability μ_n is a function from a field of events $\mathcal{F}$ into the interval [0, 1]. The collection of all such functions is a compact space (in the product topology). Hence, every sequence $\{\mu_n\}$ in this space has a limit point. Since limits preserve finite additivity, each such limit point μ will be a finitely additive probability.

The same reasoning can be applied to any improper prior to produce a collection of finitely additive probabilities associated with it, as follows. For each improper prior λ there exist finitely additive probabilities μ, each of which is a limit point of a sequence of probabilities $\{\mu_n\}$, and each μ_n is λ restricted to a set of finite measures and normalized to be a probability. This establishes a connection between improper priors and finitely additive probabilities.

Having established the connection, we can ask if the inferences made using improper distributions remain valid in the finitely additive theory. For example, suppose the distribution of $X = (\bar{X}, S)$ given $\theta = (\mu, \sigma)$ is that $\bar{X}$ and S are independent, $\bar{X}$ has a normal $N(\mu, \sigma^2/n)$ distribution and nS^2/σ^2 has a chi-squared distribution with $n - 1$ degrees of freedom. If we pretend that the improper prior $1/\sigma$ is a density for (μ, σ), then use of Bayes' Theorem for densities leads to the conclusion that the posterior distribution of μ given X is such that $(n-1)^{1/2}(\mu - \bar{X})/S$ has a t distribution with $n - 1$ degrees of freedom, a proper distribution.

In their important paper, Heath and Sudderth (1978) take the following approach to inferences involving improper distributions. They define the posterior distribution of a parameter θ given the data X as the conditional distribution necessary to make the joint distribution of (X, θ) conglomerable in *both* the X and θ margins if such a posterior distribution exists. That is to say, if the conditional distribution of X given θ (i.e. the likelihood) is $p(\mathrm{d}x|\theta)$ and the prior for θ is $\pi(\mathrm{d}\theta)$, then the marginal for X is $m(\mathrm{d}x) = \int p(\mathrm{d}x|\theta)\pi(\mathrm{d}\theta)$ and the posterior $q(\mathrm{d}\theta|x)$, if it exists, will satisfy

$$\int\int \phi(x,\theta)q(\mathrm{d}\theta|x)m(\mathrm{d}x) = \int\int \phi(x,\theta)p(\mathrm{d}x|\theta)\pi(\mathrm{d}\theta), \qquad (3)$$

for all bounded measurable ϕ. In the above example concerning (μ, σ), Heath and Sudderth find a class of measures, to which $\mathrm{d}\lambda = \mathrm{d}\mu\mathrm{d}\sigma/\sigma$ belongs, which have the property, among others, that there exists a sequence $\{B_n\}_{n=1}^{\infty}$ of sets with $0 < \lambda(B_n) < \infty$, and $\cup_{n=1}^{\infty} B_n$ equal to the space of all (μ, σ) pairs. They then form the sequence of probabilities

$\lambda_n(\cdot) = \lambda(\cdot \cap B_n)/\lambda(B_n)$. Each limit point π of this sequence is a distinct finitely additive prior for (μ, σ) which has a "posterior" distribution for μ given X agreeing with the one obtained above by use of Bayes' Theorem for densities applied to the improper prior $d\lambda$.

However, Heath and Sudderth (H–S) use the term "posterior" in an overly restricted sense, in our opinion. The implication of their definition is that if no such q exists, there is no "posterior". The work of Schervish, Seidenfeld and Kadane (1984) shows that under mild conditions, each merely finitely additive probability, even one that satisfies (3), will have countable partitions (margins) in which it is not conglomerable. The requirement that a probability be conglomerable in a pair of given (albeit uncountable) partitions before admitting the existence of a posterior is, perhaps, too harsh. After all, the prior $\pi(d\theta)$, if it is merely finitely additive, will not be conglomerable in all countable partitions which result from coarsening the parameter space. Consider the following example:

Example 5.1. Assume $\theta \in \{0, 1\}$, $X \in \{1, 2, 3, \ldots\}$ with a distribution satisfying $p(x|\theta) = 2^{-(x+\theta)}$ and $\pi(\theta) = \frac{1}{2}$. There are many such finitely additive distributions. Each has some probability adherent along the sequence of (x, θ) values $\{(n, 1)\}_{n=1}^{\infty}$ in the sense of deFinetti (1974, p. 240). It follows that the marginal for X is $m(x) = (3)2^{-(x+2)}$ and Bayes' Theorem gives the posterior of θ given X to be $q(\theta|x) = (2 - \theta)/3$ for all x. Let $\phi(x, \theta)$ equal 1 if $\theta = 1$ and zero otherwise. Then

$$\int\int \phi(x,\theta)p(dx\,|\,\theta)\pi(d\theta) = \frac{1}{2}$$

and

$$\int\int \phi(x,\theta)q(d\theta|x)m(dx) = \frac{1}{3},$$

where the integrals are defined as in Dunford and Schwartz (1958). Yet it is reasonable to claim that q is the posterior for θ given X once finite additivity is accepted. This example makes clear the need for a less restrictive definition of posterior distribution that will allow inference even when a probability cannot be made conglomerable in a specific partition.

Another problem with requiring probabilities to be conglomerable in a specific partition is illustrated by this example. Schervish, Seidenfeld and Kadane (1984) show that there are finitely additive probabilities on the space of all (x, θ) pairs which assign probability one to the points $\{(x, 1)\}_{x=1}^{\infty}$, are conglomerable in the x margin, and assign zero probability to each individual point. Let P_1 be such a probability. There is a countably additive probability P_0 which assigns probability 2^{-x} to the point $(x, 0)$ for each x. Of course, this probability is conglomerable in the x margin. But $P = \frac{1}{2}P_0 + \frac{1}{2}P_1$ has exactly the form of the probability described in Example 2.1, where we showed that P is not conglomerable in the x margin. Hence, the collection of probabilities conglomerable in a specified partition is not closed under convex combination, and H–S coherence is not preserved under mixtures of H–S coherent distributions.

Finitely additive distributions that are coherent in the usual sense but not H–S coherent reflect "strong inconsistencies" similar to the anomaly of Stone's example (section IV). Likewise, coherent but H–S incoherent distributions fail Robinson's (1979) criterion that there be no "super-relevant" betting procedures.

In contrast to H–S coherence, where "posteriors" are free of strong inconsistency in the two canonical margins (π_x, and π_θ), Levi's (1980) theory requires little more than coherence in deFinetti's sense, and suggests one view of how to calculate posterior probability with an "improper" prior. Levi's (1980, p. 129) analysis supports the familiar manuever with the Bayes formula if the likelihood function is countably additive.

Whether or not one is prepared to insist on H–S coherence (we are not), it seems reasonable to investigate "conditional properties" of coherent distributions. That is, from a Bayesian point of view the distinction between "absolute" and "conditional" probability is tenuous, at best. What is the effect of strengthening conglomerability to fix conditional probability values?

Conditional Conglomerability

Let F be an event for which $P(\cdot|F)$ is defined, and let $\pi = \{h_i\}$ be a partition. Let $h_{iF} = h_i \cap F$, and assume $P(\cdot|h_{iF})$ is defined for all i. Then P is *conditionally conglomerable in π with respect to (given) F* if for each $E \subseteq F$, and for all constants k_1 and k_2 such that

$$k_1 \leq P(E|h_{iF}) \leq k_2, \quad \text{for all } i; \quad k_1 \leq P(E|F) \leq k_2.$$

Trivially, if conditional conglomerability with respect to all events is satisfied in a specified partition, then conglomerability also holds in that margin. The converse is false, however, as we show using the example constructed by Buehler and Feddersen (1963), which was a rebuttal to arguments of Fisher (1956). Thus, H–S coherence admits distributions which are conditionally H–S incoherent.

Let (x_1, x_2) be i.i.d. $N(\theta, \sigma^2)$, with both parameters "unknown". It is obvious that

$$P(x_{\min} \leq \mu \leq x_{\max} | (\mu, \sigma^2)) = 0.5, \quad \text{for each pair}\, (\mu, \sigma^2). \tag{4}$$

Let $t = (x_1 + x_2)/(x_1 - x_2)$. Buehler and Feddersen (1963) show that

$$P(x_{\min} \leq \mu \leq x_{\max} | (\mu, \sigma^2), \quad |t| \leq 1.5) > 0.518, \tag{5}$$

for each pair (θ, σ^2), despite (4).

Moreover, Heath and Sudderth (1978) show there is a class of finitely additive "prior" probabilities over pairs (μ, σ^2) such that for each "prior" conglomerability is satisfied in *both* partitions $\pi_{(x_1,x_2)}$ and $\pi_{(\mu,\sigma^2)}$. (Note that each partition has cardinality of the continuum.) Also, they show that each finitely additive "prior" probability induces a familiar countably additive "posterior" distribution, where

$$P(x_{\min} \leq \mu \leq x_{\max}; (x_1, x_2)) = 0.5, \quad \text{for each pair}\, (x_1, x_2). \tag{6}$$

Now, since t is a function of the (x_1, x_2) pairs, we can partition the event $|t| \leq 1.5$ by the set of pairs (x_1, x_2) for which this inequality holds. Thus, by conditional conglomerability applied to this partition of $|t| \leq 1.5$, (7) is a consequence of (6):

$$P(x_{\min} \leq \mu \leq x_{\max} | \, |t| \leq 1.5) = 0.5. \tag{7}$$

But, also we may partition the event $|t| \leq 1.5$ into the continuum of states $(|t| \leq 1.5, \mu, \sigma^2)$, so that with conditional conglomerability applied to this partition of $|t| \leq 1.5$, (8) is a consequence of (5):

$$P(x_{\min} \leq \mu \leq x_{\max} | \, |t| \leq 1.5) > 0.518. \tag{8}$$

At least one of (7) and (8) must fail, hence conditional conglomerability fails in at least one partition where conglomerability holds (see Seidenfeld, 1979).

The failure of conditional conglomerability in a partition that admits conglomerability is restricted to events that have zero probability. We show this as follows. Let $P(F) \neq 0$, and argue that

$$\begin{aligned}\int_\theta P(E|F,\theta)\mathrm{d}P(\theta|F) &= \int_\theta P(E|F,\theta)\frac{P(F|\theta)\mathrm{d}P(\theta)}{P(F)}\\ &= \int_\theta \frac{P(E\cap F|\theta)\mathrm{d}P(\theta)}{P(F)}\\ &= P(E\cap F)/P(F)\\ &= P(E|F),\end{aligned}$$

where the first equality is by conglomerability in π_θ and Bayes' Theorem, the second and fourth equalities are by the multiplication theorem for conditional probability, and the third equality is again by conglomerability in π_θ. Hence, in the Fisher–Buehler–Feddersen problem, it must be that $P(|t| \leq 1.5) = 0$.

Examples such as the Fisher–Buehler–Feddersen paradox, where conditional conglomerability fails in a margin for which conglomerability holds, illustrate the violation of Robinson's (1979) criterion that there be no "relevant" betting procedures. Thus, H–S coherence is insufficient to preclude "relevant" betting schemes, in Robinson's sense. Also, subject to conditional conglomerability, advantage may switch repeatedly between two gamblers by an iteration of who has last say in determining when a bet is "on". This is illustrated by Fraser's (1977) "balanced procedure".

In contrast to the anomaly of "strong inconsistency" that is a consequence of non-conglomerability, we can identify a "weak inconsistency" of finitely additive distributions that arises from failure of a simple dominance (admissibility) rule even where conglomerability applies. Consider the choice between "no-bet" and a gamble that pays off $1/i$ if θ_i occurs ($i = 1, \ldots$). Clearly, for each value of θ the gamble is strictly preferred to "no-bet". However, if one imposes a merely finitely additive probability distribution over θ, such that $P(\theta_i) = 0$ (all i), E(no-bet) $= E$(gamble) $= 0$. That is, one becomes indifferent between the two options even though simple dominance obtains.

We can express this failure of dominance as a violation of a strengthened version of conglomerability. That is, "weak inconsistency" results from violating the following:

If there exists a constant k, such that $P(E|\theta_i) \gtrless k$ for all i, then $P(E) \gtrless k$. (9)

In terms of the above example, if $P(E|\theta_i) = 1/i$ and we receive \$1 if E occurs when we choose to gamble, then for each θ_i we gain (in expectation) $1/i$ if θ_i occurs. However, though $P(E|\theta_i) > 0$ (all i), $P(E) = 0$ under the merely finitely additive prior described above, in violation of (9).

A somewhat different example is as follows. Let $x \sim N(\mu, 1)$. As Heath and Sudderth (1978) show, suitably invariant "priors" over μ exist for which conglomerability is satisfied simultaneously in π_μ and π_x, and where (for each x) $P(\mu|x)$ is $N(x, 1)$. Hence, for each x, $P(x \le \mu|x) = 0.5$. Assume that the prior chosen assigns positive probability to $\{x: x < 0\} = X^-$, i.e. $P(X^-) \ne 0$. (We make this assumption to avoid unnecessary questions about probability conditional upon an event of zero probability.) Since we know conglomerability holds in π_x (and since $P(X^-) \ne 0$) we have that $P(x \le \mu|X^-) = 0.5$. However, we note that, for each μ, $P(x \le \mu|X^-, \mu) > 0.5$. (In fact, $P(x \le \mu|X^-, \mu) = 1$ for all $\mu \ge 0$.) Thus, $P(\cdot|X^-)$ fails (9) in π_μ, though conditional conglomerability, given X^-, holds.

Finally, we come to those "marginalization paradoxes" involving the transformation of continuous random variables. These include all but the first of those of Dawid, Stone and Zidek (1973) and Stone and Dawid (1972). Here, by transformation, e.g. changing from $N(\mu, \sigma)$ to (τ, σ), where $\tau = \mu/\sigma$, we create the situation that:

i. $P(\tau|x) = P(\tau|t)$; for t a function of $x = \{x_1, \ldots, x_n\}$, $x_i \sim N(\mu, \sigma)$;
ii. $P(t|(\mu, \sigma)) = P(t|\tau)$, i.e. t depends solely on τ;
iii. conglomerability holds simultaneously in $\tau_{(\mu,\sigma)}$ and π_x;
iv. yet the probability density $p(t|\tau)$ [from (ii)] does not factor the density $P(\tau|t)$ [from (i)] as a function of τ.

As is argued by Seidenfeld (1981) and Sudderth (1980), no violation of conglomerability is present here. Only (9) is violated. But we saw above that (9) can fail in partitions where conglomerability holds. Where conditional conglomerability obtains, but (9) does not, we are faced with betting schemes that violate Robinson's (1979) criterion of "semi-relevance", though satisfying his requirement that there be no "super-relevant" or "relevant" betting policies.

In conclusion, we have noted several strengthened versions of conglomerability and have described statistical anomalies that reflect violation of each. Since each coherent finitely additive distribution must fail conglomerability in some denumerable partition (unless the distribution is countably additive), our reaction to these anomalies is to see them as further evidence that the attempt to modulate non-conglomerability (as in Robinson's criteria prohibiting betting schemes with "super-relevant" or "relevant" selections) is misguided. If conglomerability is necessary for "consistency", then nothing less than countably additive distributions suffice with denumerable partitions, and even "proper" priors may suffer non-conglomerability in non-denumerable partition, as we discuss in the next section.

VI. COUNTABLE ADDITIVITY AND NON-CONGLOMERABILITY IN NON-DENUMERABLE PARTITIONS

Since non-conglomerability in denumerable partitions characterizes merely finitely additive probability, it is a mistake to advocate *both* a finitely additive theory of probability *and* standards of coherence entailing dominance or disintegrability in particular partitions (margins). One may cite non-conglomerability as reason enough to reinstate countable additivity. Non-conglomerability may be too high a price to pay for the convenience of "uniform" distributions.

However, Example 6.1 shows that non-conglomerability, which is characteristic of merely finitely additive probability in denumerable partitions, can occur with countably additive measures in non-denumerable partitions.

Example 6.1. Suppose X_1 and X_2 are independent standard normal random variables except that $X_2 = 0$ is impossible (this does not change the joint distribution function). Since X_1 and X_2 are independent, the distribution of X_2 given $A = \{X_1 = 0\}$ is standard normal. Using the usual transformation of variables technique, the conditional distribution of X_2 given $B = \{X_1/X_2 = 0\}$ is that of $(-1)^Y Z$, where $P[Y = 1] = P[Y = 2] = \frac{1}{2}$, Z has chi-squared distribution with two degrees of freedom, and Y and Z are independent. The two events A and B are identical, however.

The fact that the conditional distribution of X_2 given A differs from that given B is an example of the well-known Borel paradox (Kolmogorov, 1956). This paradox was also discussed by Hill (1980, p. 44). The seeming contradiction is often resolved by claiming that the transformation of variables only yields conditional probability given the sigma field of events determined by the random variable X_1/X_2, not given individual events in the sigma field. This approach is unacceptable from the point of view of the statistician who, when given the information that $A = B$ has occurred, must determine the conditional distribution of X_2. A more reasonable approach is to consider the theory of conditional probability spaces as defined by Dubins (1975), among others. In such a theory $P[X_2 \in E|A]$ means the conditional probability that X_2 is in the set E given the event A and is not relative to a sigma field. This, then, is the meaning of conditional probability one assumes when one conditions on the occurrence of a particular event.

To conclude Example 6.1, suppose one must determine conditional probabilities given all events of the form $A_{a,c} = \{X_1 - cX_2 = a\}$. The distribution of X_2 given $A_{a,c}$ can be calculated from the transformation of variables technique, if one wishes, as

$$\text{Normal } N\left(\frac{-ac}{1+c^2}, \frac{1}{1+c^2}\right).$$

In particular, if $a = 0$ the distribution of X_2 is normal with mean zero and variance $(1 + c^2)^{-1}$. Consider the uncountable partition of the sample space via the events $\{A_{0,c} | c \text{ a real number}\} = \theta$. Let Φ denote the standard normal distribution function and $E = \{X_2 > 1, X_1 \neq 0\}$. Then $P[E | A_{0,c}] = 1 - \Phi((1 + c^2)^{1/2}) < 1 - \Phi(1) = P(E)$, for all c. Hence, P is not conglomerable in the partition θ. Theorem A1 in the Appendix shows that non-conglomerability in uncountable margins is quite common for countably additive probabilities with continuous distributions, just as non-conglomerability in countable margins is quite common for merely finitely additive probabilities. In fact, the theorem implies that no matter how one defines conditional probability given $A_{a,c}$ in the above example, conglomerability will fail in some partition.

Hence, it appears that finite additivity cannot be rejected merely on the grounds that it allows failures of conglomerability, unless countable additivity is also to be rejected. Since few, if any, statisticians are willing to reject countable additivity, we suggest that finite additivity be judged on issues other than conglomerability.

VII. CONCLUSION

What attitude does our research lead us to with respect to finite additivity? We are comforted to know that various seeming paradoxes in statistical theory are understood once the lack of conglomerability of merely finitely additive probabilities is accounted for. Thus, marginalization, the Fisher–Buehler–Feddersen paradox, etc. no longer are troublesome as problems illustrating a failure in the foundations of probability and (subjective Bayesian) statistical theory.

Should we then advocate the use of merely finitely additive probabilities as likelihoods and/or prior distributions? So much is unknown, or known only to a very few, about the probability theory of merely finitely additive probabilities that, as of 1985, it is very difficult to make such a judgment. On the one hand, merely finitely additive distributions do allow certain kinds of invariant distributions not allowed under the requirement of countable additivity, such as uniform distributions on countable sets, and translation invariant measures on the real line. For certain purposes these may be useful objects (Jeffreys, 1961; Zellner, 1971; Box and Tiao, 1973). Yet countably additive distributions do offer a wide range of expression of opinion as it is. For the moment, then, we propose the dual strategy of encouraging the development of the probability theory of finite additivity, while leaving open the matter of whether the extra generality permitted by going beyond countably additive opinions is worth the cost.

We conjecture that, ultimately, the combined force of deFinetti's argument that only finitely additive probabilities are required for coherence, and the kind of considerations that led Dubins and Savage (1965) to write their book in a finitely additive framework, will lead to a recognition that finitely additive probability is the proper setting for subjective Bayesian inference. The pace at which this innovation occurs will be governed largely, we think, by the development of the necessary probability theory.

APPENDIX: NON-CONGLOMERABILITY OF COUNTABLY ADDITIVE PROBABILITIES

To continue Example 6.1, consider two independent standard normal random variables X_1 and X_2 with $X_2 = 0$ impossible. Define $Y_c = X_1 - cX_2$ and $Z_a = (X_1 - a)/X_2$. Then $A_{a,c} = \{Y_c = a\} = \{Z_a = c\}$. The conditional density of X_2 given $Y_c = a, f_c(X_2|a)$, is proportional to

$$\exp\left\{-\tfrac{1}{2}(1+c^2)\left[x_2 + ca(1+c^2)^{-1}\right]^2\right\},$$

and the conditional density of X_2 given $Z_a = c$, $f^a(x_2|c)$, is proportional to

$$|x_2|\exp\left\{-\tfrac{1}{2}(1+c^2)\left[x_2+ca(1+c^2)^{-1}\right]^2\right\}.$$

The set $\{(a, c, x): f^a(x|c) = f_c(x|a)\}$ has zero three-dimensional Lebesgue measure. If we assume that the conditional density of X_2, given $A_{a,c}$, $f(x_2|a, c)$, is defined for all a and c and is a measurable function of (x_2, a, c), then the conditions of Theorem A1 below are satisfied. The conclusion is that conglomerability fails in some uncountable partition. In fact, the partition will be the one determined by one of the random variables Y_c or one of the Z_a.

Theorem A1. Suppose two random variables (X_1, X_2) have a joint density $f(x_1, x_2)$ which is strictly positive over some measurable set G of positive Lebesgue measure. Let $\mathbb{R}$ denote the real numbers. Suppose there exist two sets of random variables $\{Y_c: c \in \mathbb{R}\}$ and $\{Z_a: a \in \mathbb{R}\}$, all measurable functions of (X_1, X_2), such that $Y_c = a$ if and only if $Z_a = c$, and (Y_c, X_2), and (Z_a, X_2) are each one-to-one functions of (X_1, X_2) for all a and c. Suppose the conditional density of X_2, given $A_{a,c} = \{Y_c = a\} = \{Z_a = c\}$, is $f(x_2|a, c)$ and is a measurable function of (x_2, a, c). Suppose that the transformation of variables technique gives the conditional density of X_2, given $Y_c = a$, as $f_c(x_2|a)$, and given $Z_a = c$, as $f^a(x_2|c)$, with the three-dimensional Lebesgue measure of $\{(a, c, x): f^a(x|c) = f_c(x|a)\}$ equal to zero. Then there exists $\delta > 0$, an event E, and a partition $\pi = \{h_b: b \in \mathbb{R}\}$ such that

$$P(E) - \int_\pi P(E|h_b)\,\mathrm{d}P(h_b) > \delta\,(or < -\delta) \qquad \text{(A1)}$$

and conglomerability fails in the partition π.

Before proving Theorem A1, we state and prove the following lemma.

Lemma A2. Under the conditions of Theorem A1, there exists $\varepsilon > 0$, such that for some c the intersection of one of the following two sets with G has positive Lebesgue measure:

$$G_1(c, \varepsilon) = g_c^{-1}\{(a, x_2): f(x_2|a, c) - f_c(x_2|a) > \varepsilon\},$$

$$G_2(c, \varepsilon) = g_c^{-1}\{(a, x_2): f(x_2|a, c) - f_c(x_2|a) > -\varepsilon\}.$$

or for some a, the intersection of one of the following two sets with G has positive Lebesgue measure:

$$G_3(a, \varepsilon) = (g^a)^{-1}\{(c, x_2) : f(x_2|a, c) - f^a(x_2|c) > \varepsilon\},$$

$$G_4(c, \varepsilon) = (g^a)^{-1}\{(c, x_2) : f(x_2|a, c) - f^a(x_2|c) < -\varepsilon\}.$$

Proof of Lemma A2. Let $(Y_c, X_2) = g_c(X_1, X_2)$ and $(Z_a, X_2) = g^a(X_1, X_2)$ be the one-to-one functions guaranteed by the hypothesis of Theorem A1. Since g^a and g_c are both one-to-one measurable functions, if D is any subset of G, then D has positive Lebesgue measure if and only if $g_c(D)$ has positive measure for all c and $g^a(D)$ has positive measure for all a. Assume that for all $\varepsilon > 0$, and a the intersections of $G_3(a, \varepsilon)$ and $G_4(a, \varepsilon)$ with G have zero Lebesgue measure. It follows that the intersection of G with

$$(g^a)^{-1}\{(c, x_2) : f(x_2|a, c) \neq f^a(x_2|c)\}$$

has zero Lebesgue measure for all a. It follows that the Lebesgue measure of

$$g^a(G) \cap \{(c, x_2) : f(x_2|a, c) \neq f^a(x_2|c)\}$$

is zero for all a. Integrating the measure of this set with respect to one-dimensional Lebesgue measure da gives that the three-dimensional measure of

$$\left[\bigcup_a \{(a, c, x_2) : (c, x_2) \in g^a(G)\}\right] \cap \{(a, c, x_2) : f(x_2|a, c) \neq f^a(x_2|c)\} \tag{A2}$$

is zero. Call the set on the left of the intersect symbol in equation (A2) A and call the set on the right B. Next, consider the set

$$A \cap \{(a, c, x_2) : f(x_2|a, c) = f_c(x_2|a)\}. \tag{A3}$$

Call the set on the right of the intersect symbol in equation (A3) C. It follows from the hypotheses of Theorem A1 that $B^c \cap C$ has zero measure, where B^c is the complement of B. Write

$$A \cap C = (A \cap C \cap B) \cup (A \cap C \cap B^c).$$

Since $A \cap B$ and $C \cap B^c$ both have zero measure, so does $A \cap C$. Since $g_c(x_1, x_2) = (a, x_2)$ if and only if $g^a(x_1, x_2) = (a, x_2)$, it follows that we can also write A as

$$\bigcup_c \{(a, c, x_2) : (a, x_2) \in g_c(G)\}.$$

The three-dimensional measure of $A \cap C$ equals the integral over $\mathbb{R}$ with respect to dc of the two-dimensional measure of

$$g_c(G) \cap \{(a, x_2) : f(x_2 | a, c) = f_c(x_2 | a)\},$$

which must be zero for almost all c by Tonelli's Theorem. Since $g_c(G)$ has positive measure, the intersection of $g_c(G)$ with either

$$\{(a, x_2) : f(x_2 | a, c) > f_c(x_2 | a)\}$$

or

$$\{(a, x_2) : f(x_2 | a, c) < f_c(x_2 | a)\}$$

has positive measure for some c. Hence, there exists $\varepsilon > 0$ and c such that the intersection of G with one of the sets $G_1(c, \varepsilon)$ or $G_2(c, \varepsilon)$ has positive measure. □

Proof of Theorem A1. In the notation of Lemma A2, suppose that the intersection of $G_1 = G_1(c_0, \varepsilon)$ with G has positive Lebesgue measure. The proof is similar for the others. Let $f_{c_0}(a)$ denote the marginal density of Y_{c_0}. Consider the partition $\pi = \{h_a : a \in \mathbb{R}\}$, where $h_a = A_{a,c_0}$. This is the partition determined by Y_{c_0}, hence d$P(h_a) = f_{c_0}(a)$da. If E is any event of the form $\{(X_1, X_2) \in D\}$, then

$$P(E) = \int\int_D f(x_1, x_2) \mathrm{d}x_1 \, \mathrm{d}x_2 = \int\int_{g_{c_0}(D)} f_{c_0}(x_2 | a) f_{c_0}(a) \mathrm{d}x_2 \, \mathrm{d}a. \qquad \text{(A4)}$$

Let $D = G_1$ and $E = \{(X_1, X_2) \in D\}$. Note that $P(E)$ is positive since the intersection of G and D has positive Lebesgue measure. Let $G_0 = g_{c_0}(G_1)$. Since G_{c_0} is one-to-one, equation (A4) implies:

$$P(E) = \int\int_{G_0} f_{c_0}(x_2 | a) f_{c_0}(a) \mathrm{d}x_2 \mathrm{d}a. \qquad \text{(A5)}$$

From the definitions of G_0 and G_1 and equation (A5), it follows that

$$P(E) < \int\int_{G_0} f(x_2 | a, c_0) f_{c_0}(a) \mathrm{d}x_2 \, \mathrm{d}a - \varepsilon \int\int_{G_0} f_{c_0}(a) \mathrm{d}x_2 \, \mathrm{d}a. \qquad \text{(A6)}$$

Since $P(E)$ is positive, it follows from equation (A6) that f_{c_0} cannot be zero almost everywhere in G_0, hence the second integral on the right-hand side of equation (A6) is positive. Let $\delta = \varepsilon \int\int_{G_0} f_{c_0}(a) \, \mathrm{d}x_2 \mathrm{d}a > 0$. If we let $G(a) = \{x_2 : (a, x_2) \in G_0\}$, then $E \cap A_{a,c_0}$ equals $\{a\} \times G(a)$ and

$$P(E | A_{a,c_0}) = \int_{G(a)} f(x_2 | a, c_0) \mathrm{d}x_2, \qquad \text{(A7)}$$

by the definition of $f(x_2|a, c_0)$. The first integral on the right-hand side of (A6) can be rewritten, using equation (A7), as

$$\int_{\mathbb{R}}\int_{G(a)} f(x_2|a, c_0)f_{c_0}(a)\mathrm{d}x_2\mathrm{d}a = \int_{\mathbb{R}} P(E|A_{a,c_0})f_{c_0}(a)\mathrm{d}a$$
$$= \int_{\pi} P(E|h_a)\mathrm{d}P(h_a). \qquad \text{(A8)}$$

Together equation (A6), equation (A8), and the definition of δ imply equation (A1) (with $< -\delta$). In the terminology of Dubins (1975), we have proven that the joint distribution of X_1 and X_2 is not disintegrable in π. Dubins proves that this implies that conglomerability fails in π also. □

To conclude Example 6.1, assume that the conditional density of X_2 given $A_{a,c}$ is given by $f(x_2|a, c)$. As long as this is a measurable function of (x_2, a, c), Theorem A1 shows that conglomerability fails in some partition determined by one of the Y_c or one of the Z_a.

REFERENCES

Arrow, K.J. (1972) "Exposition of the theory of choice under uncertainty", in: C.B. McGuire and R. Radner, eds., *Decision and Organization* (North Holland, Amsterdam).

Box, G.E.P. and G. Tiao (1973) *Bayesian Inference in Statistical Analysis* (Addison-Wesley Publishing Co., Reading).

Buehler, R.J. and A.P. Feddersen (1963) "Note on a conditional property of Student's *t*", *Ann. Math. Stat.* 34, 1098–1100.

Dawid, A.P., M. Stone and J.V. Zidek (1973) "Marginalization paradoxes in Bayesian and structural inference", *J.R.S.S.*, Ser. B 35, 189–233 (with discussion).

deFinetti, B. (1972) *Probability, Induction and Statistics* (Wiley, New York).

deFinetti, B. (1974) *Theory of Probability*, Vol. 1 (Wiley, New York).

Dubins, L. (1975) "Finitely additive conditional probabilities, conglomerability and disintegrations", *Ann. Probability*, 3, 89–99.

Dubins, L. and L.J. Savage (1965) *How to Gamble if You Must: Inequalities for Stochastic Processes* (McGraw-Hill, New York).

Dunford, N. and J.T. Schwartz (1958) *Linear Operations, Part I: General Theory* (Interscience Publishers).

Fisher, R.A. (1956) "On a test of significance in Pearson's Biometrika Tables (No. 11)", *J. Roy. Statist. Soc.*, Ser. B 18, 56–60.

Fraser, D.A.S. (1977) "Confidence, posterior probability, and the Buehler example", *Ann. Stat.*, 5, 892–898.

Hartigan, J. (1964) "Invariant prior distributions", *Ann. Math. Statist.*, 35, 836–845.

Heath, J.D. and W. Sudderth (1978) "On finitely additive priors, coherence, and extended admissibility", *Ann. Stat.*, 6, 333–345.

Hill, B. (1980) "On some statistical paradoxes and non-conglomerability", in: Bernardo, DeGroot, Lindley and Smith, eds., *Bayesian Statistics* (University Press, Valencia, Spain), pp. 39–49 (with discussion).

Jeffreys, H. (1961) *Theory of Probability* (Oxford University Press).

Kolmogorov, A. (1956) *Foundations of the Theory of Probability* (Chelsea, New York).

Levi, I. (1980) *The Enterprise of Knowledge* (The MIT Press, Boston).

Lindley, D.V. (1962) "Discussion of C. Stein's paper, 'Confidence sets for the mean of a multivariate normal distribution'", *J. Roy. Statist. Soc.*, Ser. B 24. 285–287.

Lindley, D.V. (1970) *Introduction to Probability and Statistics* (Cambridge University Press).

Lindley, D.V. (1981) "Scoring rules and the inevitability of probability", unpublished report ORC 81-1. Operations Research Center, University of California at Berkeley.

Renyi, A. (1955) "On a new axiomatic theory of probability", *Acta Math. Acad. Sci. Hung.*, 6, 285–335.

Robinson, G.K. (1979) "Conditional properties of statistical procedures", *Ann. Stat.*, 7, 742–755.

Savage, L.J. (1974) *Foundations of Statistics* (Wiley, New York).

Schervish, M., T. Seidenfeld and J. Kadane (1984) "The extent of non-conglomerability of finitely additive probabilities", *Z. f. Wahrscheinlichkeitstheorie*, 66, 205–226.

Seidenfeld, T. (1979) *Philosophical Problems of Statistical Inference* (D. Reidel, Dordrecht, Holland).

Seidenfeld, T. (1981) "Paradoxes of conglomerability and fiducial inference", in: J. Los and H. Pfeiffer, eds., *Proceedings of the 6th International Congress on Logic, Methodology and Philosophy of Science* (North Holland, Amsterdam).

Seidenfeld, T. and M.J. Schervish (1983) "A conflict between finite additivity and avoiding Dutch Book", *Phil. Sci.*, 50, 398–412. [Chapter 2.4, this volume]

Stein, C. (1955) "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", in: *Third Berkeley Symposium* (University of California Press, Berkeley), pp. 197–206.

Stone, M. (1976) "Strong inconsistency from uniform priors", *J.A.S.A.*, 71, 114–116 (with discussion, 117–125).

Stone, M. (1981) "Review and analysis of some inconsistencies related to improper priors and finite additivity", in: J. Los and H. Pfeiffer, eds., *Proceedings of the 6th International Congress on Logic, Methodology and Philosophy of Science* (North Holland, Amsterdam).

Stone, M. and A.P. Dawid (1972) "Un-Bayesian implications of improper Bayes inference in routine statistical problems", *Biometrika*, 59, 369–375.

Sudderth, W. (1980) "Finitely additive priors, coherence and the marginalization paradox", *J. Roy. Statist. Soc.*, Ser B 42, 339–341.