TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 32, Núm. 3, 1981, pp. 135 a 141

NOTAS

COMMENT ON "ON SOME STATISTICAL PARADOXES AND NON-CONGLOMERABILITY" BY BRUCE HILL

Isaac Levi
Columbia University

Those who follow Harold Jeffreys in using improper priors together with likelihoods to determine posteriors have thought of the improper measures as probability measures of a deviant sort. This is a mistake. Probability measures are finite measures. Improper distributions generate σ-finite measures.

Nonetheless, we may continue to use σ-finite measures to represent probability distributions just as we use density and characteristic functions whenever it is convenient to do so. Using σ-finite measures in this way requires a manual of translation so that probability measures may be derived from σ-finite measures. A manual is supplied in Levi (1979). The probability measures determined in accordance with the manual are finitely but not necessarily countably additive.

In his concluding summary, Hill contends that the "conventional" use of improper priors to represent uniform distributions "can lead to difficulties" with regard to Heath-Sudderth coherence and admissibility. He subsequently acknowledges, however, that the same is true when finitely additive representations of uniformity are employed. In a footnote, he also concedes that "much of Jeffreys' use of improper prior distributions can in fact be justified by the finitely additive theory using de Finetti's Axiom 3." But in the penultimate sentence of his paper, he charges that Jeffreys method "although largely justifiable by the finitely additive theory, appears at present to be a merely formal approach."

The truth of the matter is that finitely additive measures and σ-finite measures are alternative formalisms which may be used to represent the same states of probability judgement. One can justify the use of finitely additive measures by starting with a commitment to the use of σ-finite measures and employing the manual of translation. And one can justify the use of σ-finite measures by starting with a commitment to finitely additive measures. Neither mode of representation is more fundamental than the other. Nor is one mode of representation more or less "formal" than the other. Above all, using De Finetti's finitely additive measures is no more and no less immune from trouble than using Jeffreys' improper distributions (i.e., using σ-finite representations).

Abandoning countable additivity together with allowing conditional probability to be defined even when the condition bears probability O allows situations to arise where probability measures lack the conglomerative property. Hill correctly points out that conglomerability fails to obtain sometimes even in contexts (such as the example concerning the sphere) where countable additivity is satisfied. Teddy Seidenfeld has pointed out hat such breakdowns of conglomerability are related to the ways in which continuous variables are transformed. In this discussion, I wish to avoid consideration of such cases. Consequently, I shall restrict attention to σ-algebras of propositions generated by finite or countably infinite sets of exclusive and exhaustive propositions. However, the account offered here could be extended to apply to propositions which fail to constitute a σ-algebra. The restriction is intended to simplify the discussion and focus attention on those cases of failure of conglomerability where no question of transforming variables is at issue.

Let U be a finite or countably infinite set of exclusive and exhaustive propositions and m a σ-finite measure defined over the algebra generated by U. The translation manual partially defines a probability measure on the σ-algebra according to the following rules.

(*a*) Let $w_1, w_2, \ldots$ be the elements of U. If $\sum_{l=1}^{\infty} m\,(w_i) = a$ where a is positive and finite, for every g in the algebra, $Pr\,(g) = m\,(g)/a$.

(*b*) If $\sum_{l=1}^{\infty} m\,(w_i) = \infty$, there are two cases to consider:

- (i) if $m\,(g)$ is finite (so that $m\,(-g) =$), $Pr\,(g) = 0$ and $Pr\,(-g) = 1$.
- (ii) if $m\,(g) = m\,(-g) = \infty$, $Pr\,(g)$ and $Pr\,(-g) = 1 - Pr\,(g)$ are not determined by the m-function. Additional stipulations are needed which may involve using another σ-finite measure.

An unconditional σ-finite measure over the algebra generate by U may be used to determine conditional σ-finite measures as follows:

(*c*) if $m\,(e)$ is finite and positive, $m\,(h; e) = m\,(h \,\&\, e)/m\,(e)$.

(*d*) if $m\,(e) = 0$, $m\,(h; e)$ must be defined by additional stipulation.

(*e*) if $m\,(e) = \infty$, $m\,(h; e) = m\,(h \,\&\, e)$.

Conditional probabilities are derivable from conditional σ-finite functions in a manner analogous to the method for deriving unconditional probabilities from unconditional σ-finite functions.

Let U consist of all propositions of the form h_i & e_j where e_j is a report of the outcome of a random experiment and h_i "state of nature". Let i and j both range over the positive integers.

Supposewe are given the conditional probability distribution $Pr\ (e_j; h_i)$ over the e_j '*s* for each h_i and that all such distributions are countably additive. Assume further that the marginal or prior distribution over the h_i '*s* is uniform.

If we rely on the use of finitely additive representations of probability distributions, we know that uniformity entails the following requirements by definition:

(1) For every $h_i, Pr\ (h_i) = 0$.

(2) For every g which is a disjuntion of finitely many of the h_i '*s*, $Pr\ (g) = 0$.

(3) If h_i is a disjunct in the g of (2) and g has n disjuncts among the h_i '*s*, $Pr\ (h_i; g) = 1/n$.

Observe that although (2) is derivable from (1), (3) is not. Hence, if we were to define uniformity as the requirement that (1) be satisfied, we could not conclude that uniform distributions satisfy (3).

From (1) and (3) we can establish that for every i and j, $Pr\ (e_j \ \&\ h_i) = 0$ and that any finite disjunction of elements of U bear 0 probability. As a general rule, this means that we may not be able to determine $Pr\ (e_j)$ or $Pr\ (h_i; e_j)$ from the information given and the properties of finitely additive priors. This does not imply that, according to De Finetti's theory, there will be no values for $Pr\ (e_j)$ and $Pr\ (h_i; e_j)$. De Finetti assumes the contrary. The point I am now emphasizing is that these values are not determined by information about the likelihood function and the prior distribution. Additional stipulations are needed.

Suppose, however, that uniformity is represented by the σ-finite function $m\ (h_i) = k$ for all i where k is a positive, finite constant. According to the translation manual, $Pr\ (e_j; h_i) = m\ (e_j; h_i)$. $m\ (e_j) = k\ \Sigma\ Pr\ (e_j; h_i)$. If $m\ (e_j) = \infty$, $Pr\ (e_j)$ is undefined. However, $Pr\ (h_i; e_j) = 0$. If $m\ (e_j)$ is positive and finite, $Pr\ (e_j) = 0$. If we were to rely on the properties of finitely additive measures alone, we would often be incapable of determining $Pr\ (h_i; e_j)$ without additional stipulation. However, because $m\ (e_j)$ is positive and finite, $m\ (h_i; e_j) = Pr\ (h_i; e_j)$ is well defined. Thus, using σ-finite representations of uniformity is sometimes a stronger characterization than that afforded by finitely additive representations.

However, in some special cases, finitely additive representations of uniform priors suffice, along with likelihoods, to determine posteriors without additional specification or the intervention of σ-finite measures. This is true, in particular, when each e_j is consistent with only a finite number of h_i '*s*. Stone's Flatworld example illustrates this point.

The h_i 's specify possible tight paths to the treasure and the e_j 's possible tight paths to the sailor's grave. Two paths are adjacent if and only if one is an extension of the other. The conjunction e_j & h_i is consistent if and only if the conjunctions specify adjacent paths. U consists only of consistent conjunctions. We are given, for every consistent e_j & h_i, that $Pr\ (e_j; h_i)$ and that the prior over the h_i 's is uniform in the sense that it satisfies (1), (2) and (3).

Let g_j be the disjunction of the four h_i 's consistent with e_j and let h_k be one of these four h_i 's. $Pr\ (h_k; g_j) = 1/4$ by condition (3) on finitely additive representations of uniform distributions. Hence, $Pr\ (e_j\ \&\ h_k; g_j) = 1/16$ and $Pr\ (e_j; g_j) = 1/4$. From this it follows that $Pr\ (h_k; e_j\ \&\ g_j) = Pr\ (h_k; e_j) = 1/4$.

Needless to say, the same result is obtainable straightforewardly by using σ-finite measures via the σ-finite version of Bayes' theorem following Jeffreys.

Let E assert that the path to the sailor's grave is an extension of the path to the treasure. Without appealing to improper priors (or to any special properties of the prior distribution), we obtain $Pr\ (E; h_i) = 3/4$ for every h_i and $Pr\ (E; e_j) = 1/4$ for every e_j.

The "stoned Bayesian" who assigns a uniform prior to the h_i 's will assign E a posterior probability of 1/4 on the information e_j. Given that information, he will accept a bet where he receives one utile if E is false and loses one utile if E is true. Yet, before discovering the path to the sailor's grave (i.e., finding out that e_j is true) he would assign a negative conditional expected utility to accepting such as bet given h_i for every value of i. The stoned Bayesian may be coherent in De Finetti's sense but he will be incoherent in the sense of Heath and Sudderth. (See Hill's reference (4).)

This result obtains regardless of whether the posterior probability for E is derived with the aid of a σ-finite (i.e., improper) representation of the uniform prior over the lengths of paths to the treasure or with the aid of a finitely additive representation.

For those who insist that decision making must be based on reasoning which is acceptable according to a prospective analysis, violation of Heath-Sudderth coherence will seem intolerable. But Bayesians favor retrospective analysis over prospective analysis whenever there is a conflict between the two–e.g., in connection with optional stopping. Hence, they should bite the bullet and dismiss incoherence in the sense of Heath and Sudderth. This applies not only to Flatworld but to the Dubins example and example 5.2 of Heath and Sudderth both mentioned by Hill.

In any case, Hill is prepared to tolerate incoherence in the sense of Heath and Sudderth in the case of the stoned Bayesian-presumably regardless of whether such incoherence is derived via appeal to finitely additive representations of the prior or σ-finite representations.

Hill worries, however, about another feature of the Flatworld example as analysed by Stone. According to Stone, the stoned Bayesian will assign E the unconditional probability of 3/4 because $Pr\ (E; h_i) = 3/4$ for every h_i. Hence, the stoned Bayesian has a negative *unconditional* expectation prior to discovering the burial place of the sailor for the policy he is obliged as a Bayesian to tollow.

For someone who is prepared to give precedence to retrospective over prospective analysis when the two conflict, it is not clear why it should be any more disturbing that the stoned Bayesian's policy prior to discovery of the burial place has a negative undonditional expectation than it is that this policy bears negative conditional expectation relative to each h_i.

These is, however, a more serious difficulty than the mere fact that $Pr\ (E)$ is assigned the value 3/4 because $Pr\ (E; h_i) = 3/4$ for every h_i. It is also the case that for every e_j, $Pr\ (E; e_j) = 1/4$. Hence, by parity of reasoning, $Pr\ (E)$ should also equal 1/4. The result is an out and out contradiction.

Hill contends that the source of the difficulty here is the appeal to conglomerability; but he seems anxious to invoke nonconglomerability to question the propriety of assigning $Pr\ (E)$ the value 3/4. But, although Hill is right it does not follow from the fact that $Pr\ (E; h_i) = 3/4$ that $Pr\ (E) = 3/4$ if we abandon conglomerability, it does not follow that $Pr\ (E)$ differs from 3/4 either. The matter is left entirely open.

What is clear, however, is that $Pr\ (E)$ cannot both equal 3/4 and 1/4 so that conglomerability must be violated one way or another.

It is important to appreciate that the breakdown of conflomerability is recognizable by using the σ-finite representation of the stoned Bayesian prior.

$$m\ (E) = \sum_{L=1}^{\infty} m\ (E; h_i)\, m\ (h_i) = \sum_{L=1}^{\infty} 3\,k/4 = \infty .$$

$$m\ (E) \text{ also equals } \sum_{j=1}^{\infty} m\ (E; e_j)\, m\ (e_j) = \sum_{j=1}^{\infty} k/4 = \infty .$$

Thus, we have two methods for calculating $m\ (E)$ both of which lead to the same value. Moreover, the value of $m\ (E)$ differs from both $m\ (E; h_i) = 3/4$ and $m\ (E; e_j) = 1/4$ so that the analogue of the conglomerative property for σ-finite representations is violated; and, since $m\ (-E)$ also equals ∞, we require extra stipulations to determine the values of $Pr\ (E)$ and $Pr\ (-E)$. According to the σ-finite analysis, on the information furnished concerning the predicament of the stoned Bayesian, there is no basis for ruling out any assignment between 0 and 1 to the probability of E. And even if we guarantee that conglomerability with respect to the finitely additive measure is satisfied one way by assigning $Pr\ (E)$ either the value 3/4 or 1/4, the σ-finite method of representation furnished a sense in which it can be said that conformity to conglomerability is pseudo conformity.

Finally, it should be noted that a very good argument is available for recommending that $Pr(E) = 3/4$. This argument does not depend on an appeal to conglomerability although it does presuppose the meaningfulness of statements of objective or statistical probability (as Stone apparently does).

According to Stone, all of the tight paths share a certain stochastic property-namely the tendency to be extended on an application of the direction selector to the endpoint with an objective or statistical probability of 3/4. Hence, even though we do not know which of these infinitely many tight paths leads to the treasure, we do know it has the stochastic property just cited . Moreover, we know this without any appeal to conglomerability.

The situation is analogous in this one respect (but not others) to one where we are dealing with a pivotal random variable whose objective or statistical probability distribution is known to be independent of the population parameter.

Given this knowledge and armed with the information that the selector will be applied to the path leading to the treasure and the sailor buried accordingly, we are entirely justified in assigning a personal or credal unconditional probability of 3/4 to E.

Thus, there is a very good argument for assigning $Pr(E)$ the value 3/4. This argument does not presuppose the conglomerative property to be satisfied in this case as a premise. It does establish that the conglomerative property is satisfied in this special case as a conclusion. The controversial assumption here is that regardless of which tight path is the path to the treasure, it has an objective or statistical probability of 3/4 of being extended by an application of the direction selector to it.

De Finetti, of course, thinks that objective probability is metaphysical moonshine and, hence, would not adopt the crucial presupposition of this argument. For him, the only route to $Pr(E) = 3/4$ appears to be via conglomerability. According to De Finetti's approach, there is no good reason for assigning 3/4 as the value of $Pr(E)$. On the other hand, *there is no good reason for prohibiting that assignment either.*

Thus, Hill can rebut Stone's contention that $Pr(E) = 3/4$ only by questioning Stone's apparent readiness to make assumptions about objective or statistical probabilities. Given Hill's apparent desire to avoid controversy concerning the intelligibility of objective or statistical probability, this maneuver does not look atractive. But even if Hill does follow De Finetti and regards objective probability as nonsense, he still lacks a good reason for prohibiting $Pr(E) = 3/4$.

In any case, as I have already explained, the respect in which deviation from conglomerability emerges in the Flatworld example is brought out very clearly through the use of σ-finite representations. Whether it is brought out more clearly this way than by relying on the use of finitely additive representations exclusively is, I suspect, largely a matter of taste. But the use of σ-finite representations is no more formal than the use of finitely additive representations, it has just as adequate a theoretical underpinning and is just as directly connected with applications.

Isaac Levi
Columbia University

REFERENCES

LEVI, Isaac: *The Enterpise of Knowledge*, Cambridge, Mass.: MIT Press (1979). The discussion of σ-finite representations of finitely additive probability distributions appears in 5.7, 5.9, 5.11, 12.4, 12.16, 14.4 and 15.4. This book was written without knowledge of the work of Dubins, Heath and Sudderth or Hill. I am grateful to Teddy Seidenfeld for having drawn my attention to these papers.

SEIDENFELD, Teddy: "Probability, Continuity and Transformations of Continuous Random Variables", *Essays in Semantics and Epistemology* ed. by H. Leblanc, R. Gumb and R. Stern, New York: Haven Press (1979).