L. LE CAM

University of California, Berkeley

Our distinguished colleagues deserve congratulations for contributing yet another important study on the behavior of Bayes estimates. Looking over the main thrust of their essay, I find it ironic that two self-confessed former Bayesians have spent so much ingenuity showing that Bayes estimates can behave very badly, while the present writer, a staunch former and present anti-Bayesian, made efforts to emphasize the good properties of Bayes estimates. One could perhaps summarize the situation as follows: Take a family $\{P_\theta: \theta \in \Theta\}$ of probability measures P_θ on a Polish space $\mathscr{X}$ and suppose that Θ itself is either Polish or at least Borelian in a Polish space. Then, according to Wald and others, for any decision problem, Bayes and approximate Bayes procedures form complete classes. If μ is a positive finite measure on Θ and if $\theta \rightsquigarrow P_\theta(A)$ is measurable for each Borel $A \subset \mathscr{X}$, one can form a marginal measure $\mu \cdot P$ and a joint "semidirect product" measure $\mu \otimes P$ by $(\mu \otimes P)(B \times A) = \int_B P_\theta(A)\mu(d\theta)$. If one takes seriously the principle, call it Principle Π, that sets of very small $\mu \otimes P$ probability are practically negligible, then Bayes procedures for $\mu \otimes P$ are good. If, however, one induces the distributions on $\mathscr{X}$ through some other measure, say Q, Bayes procedures can behave in a most unpredictable fashion. This is so, as shown by our colleagues, even if μ is itself a direct product of two terms (Dirichlet $\times$ Gaussian) that, separately, lead to excellent behavior.

Under Principle Π one obtains theorems such as Doob's theorem of 1949 and a variety of other results. For instance, in the i.i.d. case, and many other ones, anything that is asymptotically Bayes for a prior measure μ is also asymptotically Bayes for any $\nu \geq 0$ dominated by μ. In the most general case, with all the items in sight depending on some n that tends to infinity, suppose Θ metrized by a distance d and look at balls $B(t, r)$ of center t and radius r depending on x. Select, among balls whose posterior probability is $> \frac{1}{2}$, one that has almost the smallest possible radius. Let $\hat{\theta}_n$ be its center. Then if for the joint measures $\mu \otimes P$ there are estimates T_n that converge at a rate δ_n (in the sense that for $\varepsilon > 0$ there is a $b < \infty$ such that $[\mu \otimes P][d(T_n \theta) \geq bd_n] < \varepsilon$ for n large), then $\hat{\theta}_n$ enjoys the same properties. The "tails" $(\mu \otimes P)[d(\hat{\theta}, \theta) \geq b\delta_n]$ also tend to zero at the best possible rate.

There are many more properties of this general nature. Unfortunately, they give little information about what happens for observations X generated from a probability measure Q, unless it happens that Q is close to an average $P_\nu = \int_V P_\theta \mu(d\theta)/\mu(V)$ for sets V whose μ measure is not too small, or, if there is an n involved, for sets such that $\mu(V)$ does not tend to zero too rapidly.

In a paper (Le Cam, 1982) cited by Diaconis and Freedman, the present writer attempted to obtain bounds on the maximum risk of Bayes estimates in a situation describable as follows: One has independent observations X_j, $j = 1, 2, \ldots,$ where X_j has distribution $p_{\theta, j}$, $\theta \in \Theta$. One introduces a distance H by $H^2(s, t) = \frac{1}{2}\Sigma_j \int(\sqrt{dp_{s,j}} - \sqrt{dp_{t,j}})^2$. Then, letting $D(\tau)$ be the metric dimension of the space $\{\Theta, H\}$ at the level τ, one can show that there exist estimates T_n such that $E_\theta H^2(T_n, \theta) \leq CD(a)$ where a is a number such that (for $D(a)$ large) one

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

has approximately $D(a)/a^2 \sim 324$ and where C is a universal constant (≤ 10 for large $D(a)$).

One would then expect that, for a loss function $H^2(t, \theta)$, and for prior measures μ that are sufficiently well spread out, the Bayes estimates β_n would satisfy a similar inequality: $E_\theta H^2(\beta_n, \theta) \leq C'D(a)$. This is indeed the case. However, we could not find measures μ that are sufficiently well spread out except under a severe growth restriction on $D(\tau)$ as $\tau \to 0$. Roughly, the growth restriction is that $D(\tau)$ increases slower than $\tau^{-1/3}$ as $\tau \to 0$. This rules out interesting cases, such as the case where Θ is the set of bounded densities satisfying a Lipschitz condition on the unit square of the plane. The nonparametric sets used by Diaconis and Freedman have dimensions that increase very rapidly as $\tau \to 0$, even if the distances used are much weaker than our H. Most small open sets have positive but essentially negligible probabilities.

To obtain better results, it seems necessary to take into account features of the statistical problem that are not summarized by the distance H. Which features are most important is presently a matter of conjecture. Here, Diaconis and Freedman suggest a direction of study that may be very important: They investigate the derivative of the posterior measure viewed as a function of the prior measure. Now, let $\mu \cdot P$ be the marginal measure $\int P_\theta \mu(d\theta)$, let $\mu \otimes P$ be the joint distribution, and let K_x be the conditional distribution of θ given x. Then, with the present symbolism

$$(\mu \cdot P) \otimes K(\mu, P) = \mu \otimes P.$$

This relation can be differentiated not only in μ but also in P. For instance, retaining only first order terms in ε, one would have

$$(\mu \cdot P) \otimes \{K(\mu, P + \varepsilon\Delta) - K(\mu, P)\} \sim \varepsilon\{\mu \otimes \Delta - \Delta \otimes K(\mu, P)\},$$

a relation analogous to the one given by Diaconis and Freedman. It may be feasible from such relations to find out which features of μ or $\{P_\theta\colon \theta \in \Theta\}$ influence the posterior distributions and the attached risks. However, as far as we know the subject has not yet been studied in sufficient detail.

Perhaps my formerly Bayesian colleagues will tell us in the near future what pairs (μ, P) are "safe" and what pairs are bound to give trouble.

Department of Statistics
University of California
Berkeley, California 94720

Dennis V. Lindley

Somerset, England and Monash University

My own view of statistics is that it is a way of studying some aspects of the real world, namely the uncertainty present in any study, and of expressing my beliefs about the world. The subject is not primarily mathematical but mathematics plays an essential role because it enables me to pursue the logical