COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.
DIACONIS, P. and FREEDMAN, D. (1986a). On the consistency of Bayes estimates. *Ann. Statist.* **14** 1–26.
DIACONIS, P. and FREEDMAN, D. (1986b). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68–87.
HODGES, J. L., JR. and LEHMANN, E. L. (1963). Estimates of location based on ranks. *Ann. Math. Statist.* **34** 598–611.
HOEFFDING, W. (1950). "Optimum" nonparametric tests. *Proc. Second Berkeley Symp.* 83–92. Univ. California Press.
KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihood based on Cox's regression and life model. *Biometrika* **60** 267–278.
LO, A. Y. (1984). Consistency in the location model: The undominated case. *Ann. Statist.* **12** 1584–1587.
PETTITT, A. N. (1983). Likelihood based inference using signed ranks for matched pairs. *J. Roy. Statist. Soc. B* **45** 287–296.
SAVAGE, I. R. (1969). Nonparametric statistics: A personal review. *Sankhyā* **31** 107–144.
STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp.* **1** 187–196.
STONE, C. J. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.* **3** 267–284.
TUKEY, J. W. (1954). Unsolved problems of experimental statistics. *J. Amer. Statist. Assoc.* **49** 706–731.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

DEPARTMENT OF STATISTICS
SUNY AT BUFFALO
4230 RIDGE LEA ROAD
BUFFALO, NEW YORK 14226

HANI DOSS

Florida State University

It is generally acknowledged that it is hard to think about priors in high (but finite) dimensional spaces. Subjective Bayesians know that it is hard to elicit a prior from an individual when the dimension is 3 or 4. Diaconis and Freedman's interesting results on an inconsistent Bayes rule involving a reasonably natural prior show how far off our intuition can be when we pass to an infinite dimensional setting. In this discussion, we present other peculiarities, in addition to the inconsistent behavior, that arise when one uses the symmetrized Dirichlet prior. The discussion concludes with a few remarks on an alternative way of constructing priors on c.d.f.'s.

1. The symmetrized Dirichlet priors. The setup considered by Diaconis and Freedman is the following:

$$X_i = \theta + \varepsilon_i, \qquad i = 1, 2, \dots, n, \qquad \varepsilon_i \text{ are i.i.d.} \sim F.$$

The parameters θ and F are independent, θ having a density f, and F being distributed according to $\bar{\mathscr{D}}_\alpha$, with α absolutely continuous.

Let $\theta_{ij} = \frac{1}{2}(X_i + X_j)$, and let $\#(\theta_{ij})$ denote the number of distinct pairs (X_k, X_l) such that $\frac{1}{2}(X_k + X_l) = \theta_{ij}$. (The pairs (X_k, X_l) and (X_s, X_r) are called distinct if the sets $\{X_k, X_l\}$ and $\{X_s, X_r\}$ are distinct.) The number $\#(\theta_{ij})$ will be called the multiplicity of θ_{ij}. The posterior distribution of θ given $X_1, \dots, X_n$ is denoted $\bar{\pi}_n$.

If $\max_{i \neq j} \#(\theta_{ij}) = 1$, then $\bar{\pi}_n$ has been described by Diaconis and Freedman (1986, Lemma 3.1).

Doss (1984) shows that if $\max_{i \neq j} \#(\theta_{ij}) > 1$, then $\bar{\pi}_n$ is a discrete measure, concentrating all its mass on the points θ_{ij} of highest multiplicity. (An intuitive explanation of this is in Section 3 of Doss (1984); the corollary in that section gives an explicit formula for $\bar{\pi}_n$.) In particular, if there is a unique point of highest multiplicity, then $\bar{\pi}_n$ is a point mass at that point. This qualitative result is true independent of the parameter α.

Several observations can be made immediately from this result; it is easiest to proceed by way of example. Consider a data set consisting of 104 points, as follows: $X_1, \dots, X_{100}$ are between -1 and 1, and $\#(\theta_{ij}) = 1$ for $1 \leq i, j \leq 100$; the last 4 points are 3, 6, 10, and 13. Under reasonable conditions on α and f (e.g., α and f are both standard normal), $\bar{\pi}_{100}$ will have most of its mass between -1 and 1. Note, however, that $\bar{\pi}_{104}$ is a point mass at 8 because 8 has multiplicity 2. Not only does the posterior undergo a drastic change because of a few additional observations, but we have a situation where the posterior ($\bar{\pi}_{104}$) is singular with respect to the prior. To a subjective Bayesian, this is very counterintuitive: A finite data set completely changes the opinion of the statistician.

Next, suppose that one of the last four points is perturbed very slightly. The result is that the posterior is no longer a point mass at 4. Thus, we see that the map $\mathbf{X} \mapsto \bar{\pi}_n$ is highly discontinuous. (Compare with Diaconis and Freedman's discussion of the derivative of the map taking priors to posteriors.) This is unusual behavior and is disturbing when one considers the possibility of rounding and/or grouping of the data. In light of the "what if" method discussed by Diaconis and Freedman, this raises questions about the use of the prior $\bar{\mathscr{D}}_\alpha$.

Diaconis and Freedman have examined the asymptotics of $\bar{\pi}_n$ when the data $X_1, \dots, X_n$ are i.i.d. from a continuous distribution. The form of the posterior when $\max_{i \neq j} \#(\theta_{ij}) > 1$ and the observations made above raise the question of how $\bar{\pi}_n$ behaves when $X_1, \dots, X_n$ are i.i.d. from a symmetric *discrete* distribution. The answer is that $\bar{\pi}_n$ can perform extremely well.

Consider a discrete distribution H which is symmetric about θ, and suppose that H has just a finite number of atoms, say at $\theta \pm a_i$, for $i = 1, \dots, k$. Let $\{X_i\}$ be i.i.d. $\sim H$. As soon as the values $\theta \pm a_i$, $i = 1, \dots, k$ have all been observed, θ is a midpoint of multiplicity k, while all other midpoints have multiplicity less than k. The posterior $\bar{\pi}_n$ is then a point mass at θ, and no additional observations can change it. Furthermore, the probability that at least one of the values $\theta \pm a_i$ has not been observed by time n goes down exponentially with n. This can be generalized to other types of discrete symmetric distributions. For example, the result is still true (with probability 1) if H is picked according to $\bar{\mathscr{D}}_\alpha$; see Doss (1984). Thus, the posterior (and its mean) can behave extremely well if the data come from a discrete distribution.

Suppose that $X_1, X_2, \dots, X_n$ have a unique midpoint θ^* of highest multiplicity. Since the posterior distribution of θ is a point mass at θ^* *regardless of the parameter* α, it follows that if the prior on F is a mixture of symmetrized Dirichlet priors, then the posterior distribution of θ is still a point mass at θ^*. Thus, the entire discussion above applies to mixtures of symmetrized Dirichlet priors. Dalal (see, e.g., 1979) has shown that such mixtures are dense in the set of all priors on symmetric F's, with the weak topology. These considerations raise questions as to the appropriateness of approximating priors on symmetric F's with mixtures of symmetrized Dirichlet priors, since it is obvious that it is the "details" of these mixtures (viz. the fact that they give probability 1 to *discrete* symmetric c.d.f.'s) that is causing the peculiar behavior.

2. An alternative construction of random c.d.f.'s. The atoms in the posterior distribution of θ arise because the construction of random c.d.f.'s used by Diaconis and Freedman produces symmetric discrete c.d.f.'s with probability 1. The atoms vanish if instead we proceed as follows. Let α be a finite symmetric (absolutely continuous) measure on $\mathscr{R}$, and let α_+ and α_- denote the restrictions of α to $(0, \infty)$ and $(-\infty, 0)$, respectively. Choose F_1 and F_2 *independently* from $\mathscr{D}_{\alpha_+}$ and $\mathscr{D}_{\alpha_-}$, respectively, and form

$$F(t) = \tfrac{1}{2}F_1(t) + \tfrac{1}{2}F_2(t).$$

This F has median equal to 0 and with probability 1 is not symmetric, although it is symmetric "on the average," e.g.,

$$EF(t) = \frac{\alpha(t)}{\|\alpha\|}.$$

If we use this construction, the problem becomes one of estimating the median. This is done in Doss (1985a, b), which contain the details of the rest of the discussion. The Bayes estimate $\hat{\theta}^B$ of θ can be described as follows. Let $\hat{\theta}^m$ denote the maximum likelihood estimate of θ when the distribution of the ε_i's is assumed to be $\alpha/\|\alpha\|$. Then, $\hat{\theta}^B$ is essentially a convex combination of $\hat{\theta}^m$ and of the sample median, with the weights depending on the sample.

Suppose that $\{X_i\}$ are i.i.d. from a distribution function H with unique median equal to θ. If H is discrete, then the Bayes estimate $\hat{\theta}^B$ is consistent. However, if H is continuous, then the Bayes estimate can be inconsistent: It can converge to a wrong value, it can oscillate indefinitely between two wrong values, or the set $\{\hat{\theta}_n\}$ can be dense in $\mathscr{R}$. As before, this behavior can be traced to the fact that the Dirichlet priors give probability 1 to discrete distributions.

REFERENCES

DALAL, S. R. (1979). Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions. *Stochastic Process. Appl.* **9** 99–107.

DIACONIS, P. and FREEDMAN, D. (1986). On inconsistent Bayes estimates of location. *Ann. Statist.* **14** 68–87.

DOSS, H. (1984). Bayesian estimation in the symmetric location problem. *Z. Wahrsch. verw. Gebiete* **68** 127–147.

DOSS, H. (1985a). Bayesian nonparametric estimation of the median; Part I: Computation of the estimates. *Ann. Statist.* **13** 1432–1444.

DOSS, H. (1985b). Bayesian nonparametric estimation of the median; Part II: Asymptotic properties of the estimates. *Ann. Statist.* **13** 1445–1464.

DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA 32306