BAYESIAN NONPARAMETRIC INFERENCE

Thomas S. Ferguson, University of California at Los Angeles,

Eswar G. Phadia, Willian Paterson College,

and

Ram C. Tiwari, University of North Carolina

Introduction

Problems of statistical inference with an infinite dimensional parameter space, usually a space of probability distributions over a set, are of great importance both theoretically and practically. The Bayesian approach to such *nonparametric* problems requires that a probability distribution be placed over this space. Much progress has been made in the past 15 years and the results have been scattered throughout the statistical and probability literature. It is the purpose of this paper to review the progress in this area to date with special emphasis on random probability measures and on results that have appeared since the review article of Ferguson (1974).

The central class of distributions for use in these problems is the class of Dirichlet processes. Developments in the basic theory of such processes are reviewed in the next section. The settling of Doksum's conjecture by James and Mosimann is observed in the third section on tailfree and neutral processes. Progress in the application of mixtures of Dirichlet processes to the Bayesian analysis of empirical Bayes problems, bio-assay and density estimation is presented in the fourth section. The far-reaching extension of the basic techniques to problems with partially censored data is reviewed in the fifth section, with application to reliability and the Cox proportional hazard model. The use of random distributions in empirical Bayes estimation, initiated by Hollander and Korwar, has been extensively developed and is reviewed in the sixth section. In the seventh section, the problems of inconsistency of the Bayes estimates in Dalal's symmetric Dirichlet model, discovered by Diaconis and Freedman, are presented. In the final section, various other Bayesian nonparametric techniques and applications are briefly touched upon.

The Dirichlet Process

Let $\mathfrak{X}$ be a set, let $\mathcal{A}$ be a σ-field of subsets of $\mathfrak{X}$, and let α be a finite nonnull measure on $(\mathfrak{X}, \mathcal{A})$. Among the various methods for putting prior distributions on the set of all probability distributions over $(\mathfrak{X}, \mathcal{A})$, the Dirichlet process is still central. As defined in Ferguson (1973), a *Dirichlet process with parameter* α, denoted $\mathfrak{D}(\alpha)$, is a random process, P, indexed by elements of $\mathcal{A}$ with the property that for all positive integers k, and every measurable partition

$A_1,\ldots, A_k$ of $\mathfrak{X}$, the random vector $(P(A_1),\ldots,P(A_k))$ has a k-dimensional Dirichlet distribution with parameter $(\alpha(A_1),\ldots,\alpha(A_k))$. The basic result for this process is:

Theorem 1 (Ferguson, 1973)

If P is a Dirichlet process with parameter α, and if, given P, $X_1,\ldots,X_n$ is a sample from P, then the posterior distribution of P given $X_1,\ldots,X_n$ is a Dirichlet process with parameter $\alpha + \Sigma\delta(X_j)$, where $\delta(x)$ represents the distribution giving mass one to the point x.

Two proofs of the existence of such a process were given, one non-constructive using the Kolmogorov consistency conditions, and the other constructive, in which P is a sum of a countable number of point masses whatever be α. That a Dirichlet process has a representation that is discrete a.s. even if α is continuous is a striking fact that has been the subject of several papers, e.g., Blackwell (1973), Berk and Savage (1979), Basu and Tiwari (1982). A new construction simpler than that of Ferguson has been given by Sethuraman and Tiwari (1982).

Theorem 2 (Sethuraman and Tiwari, 1982)

Let Y_1, $Y_2,\ldots$ be i.i.d. with a beta distribution, $\mathbb{B}e(M,1)$ $M > 0$, let Z_1, $Z_2,\ldots$ be i.i.d. F_0, and let $\{Y_j\}$ and $\{Z_j\}$ be independent. Define $P_1 = (1 - Y_1)$, and $P_n = Y_1 \ldots Y_{n-1}(1 - Y_n)$ for $n > 1$. Then, $P = \Sigma P_j\delta(Z_j)$ is a Dirichlet process with parameter $\alpha = MF_0$.

Throughout, we shall use $M = \alpha(\mathfrak{X})$ to represent the total mass of α, and $F_0 = \alpha/M$ to be the *prior guess at P*. The latter phrase stems from the fact that from the definition, $P(A)$ has a beta distribution, $\mathbb{B}e(\alpha(A), M-\alpha(A))$, so that $EP(A) = \alpha(A)/M = F_0(A)$. In particular, the posterior guess at P given a sample from P is, according to Theorem 1, $F_n = p_nF_0 + (1 - p_n)\hat{F}_n$, where $\hat{F}_n$ is the empirical process and $p_n = M/(M + n)$. As a consequence, suppose that it is required to estimate with squared error loss the mean $\mu = \int xdP(x)$ of an unknown distribution P on the real line based on a sample $X_1,\ldots,X_n$, with prior $P \in \mathfrak{D}(MF_0)$, where F_0 has finite first moment. Then, μ is finite a.s. and

$$E(\mu \mid X_1,\ldots,X_n) = p_n\mu_0 + (1 - p_n)\bar{X}_n,$$

where μ_0 is the mean of F_0, and $\bar{X}_n$ is the sample mean. (In subsequent discussions, Bayes procedures are assumed to be taken with respect to squared error loss, unless stated otherwise.)

In regard to this simple problem, there was an error in Ferguson (1974) in stating that μ is finite a.s. if and only if F_0 has a finite first moment. That the *only if* part is false was pointed out in Doss and Sellke (1982), who obtain the following results on the tail behavior of P. Let $F(t) = P((-\infty, t])$.

Theorem 3 (Doss and Sellke, 1982)

If $F \in \mathfrak{D}(MF_0)$, then

$$exp(-h_1(t)) \leq 1 - F(t) \leq exp(-h_2(t))$$

for sufficiently large t a.s.

where $h_1(t) = 2 \ log \mid log(1 - F_0(t))|/(1 - F(t))$ and $h_2(t) = \{(1 - F_0(t)) \times [log(1 - F_0(t))]^2\}^{-1}$.

As an example of this behavior, Yamato (1984) obtains the distribution of μ when F_0 is a Cauchy distribution.

Theorem 4 (Yamato, 1984)

If $F \in \mathfrak{D}(MF_0)$ where F_0 is a Cauchy distribution, then the random variable $\mu = \int x dF(x)$ has the same Cauchy distribution.

In Cifarelli and Regazzini (1979) and in Hannum, Hollander and Langberg (1981), methods of finding the distribution of the mean of a Dirichlet process are reported.

A number of simple applications were presented in Ferguson (1973) such as estimating a distribution function or a median, mean or variance. In the two-sample problem of estimating $P(X > Y)$, the Mann-Whitney-Wilcoxon rank-sum statistic was seen to appear naturally. A number of other similar applications have appeared since that time. We mention a few.

Yamato (1975) obtains a Bayes estimate for $d(F, G) = \int (F(x) - G(x))^2 d(F(x) + G(x))/2$, based on independent samples from F and G which are given independent Dirichlet priors. Campbell and Hollander (1978) provide estimates of the rank of X_1 among $X_1,\ldots,X_n$ based on $X_1,\ldots,X_s$, $s < n$, when sampling from a Dirichlet process F. Hollander and Korwar (1980) find a Bayes estimate of $\Delta(x) = G^{-1}(F(x)) - x$, a measure of the difference between F and G at x, based on independent samples from each, with G known and F having a Dirichlet prior. Dalal and Phadia (1983) consider the problem of estimating $\tau = E\{sign((X - X')(Y - Y'))\}$, a measure of dependence for a bivariate distribution, where (X, Y) and (X', Y') are independent samples from the distribution. The Bayes estimate is computed using a Dirichlet prior in 2-dimensions, and Kendall's tau is seen to appear naturally. Zalkikar, Tiwari and Jammalamadaka (1986) obtain a Bayes estimate for $\Delta(F) = P(Z > X + Y)$, where X, Y, Z are i.i.d. chosen from F, based on a sample from F, which is given a Dirichlet prior.

These are all examples of estimation problems. The difficulty of using Dirichlet priors in hypothesis testing problems was mentioned in Ferguson (1973), but Susarla and Phadia (1976) show how to test H_0: $F \leq F_0$ for a given distribution function F_0 using a Bayes approach. The idea is to replace the usual zero/one loss function with the smoother loss $L(F, a_0) = \int (F - F_0)^+ dW$ and

$L(F, a_1) = \int (F - F_1)^{-} dW$, where a_0 (resp. a_1) is the action accept (resp. reject) H_0, and W is an arbitrary weighting measure. This idea also extends to multiple decision problems.

Relation to Tailfree and Neutral Processes

Let $\mathcal{P}_1, \mathcal{P}_2, \ldots$ be a sequence of finite measurable partitions of $\mathfrak{X}$ such that for all $n \geq 1$, $\mathcal{P}_{n+1}$ is a refinement of $\mathcal{P}_n$. We say that a random probability measure P on $(\mathfrak{X}, \mathcal{A})$ is *tail-free w.r.t. the sequence* $\{\mathcal{P}_n\}$ if the sets of random variables $\{P(B \mid A)$: $A \in \mathcal{P}_{n-1}$, $B \in \mathcal{P}_n\}$ for $n = 1, 2, \ldots$ are independent. (Here $\mathcal{P}_0 = \{\mathfrak{X}\}$.) The notion of tailfree processes goes back to Freedman (1963), Fabius (1964) and Kraft (1964). In the dyadic tailfree process, each set of the partition $\mathcal{P}_n$ is cut into two pieces in the partition $\mathcal{P}_{n+1}$.

One drawback of using a tailfree process as a prior is that the behavior of the estimates depends on the choice of the partitions used to describe the process. This is true with one notable exception. The Dirichlet process is tailfree with respect to every sequence of partitions. Moreover, if a process is tailfree with respect to every sequence of partitions then it is either a Dirichlet process or a limit of Dirichlet processes or concentrated on two nonrandom points (Fabius, 1973).

There is another class of prior distributions that shares this property to a lesser degree, the processes neutral to the right, introduced by Doksum (1974). A random distribution function $F(t)$ on the real line is said to be *neutral to the right* if for every m and $t_1 < t_2 < \ldots < t_m$, the random variables $1 - F(t_1)$, $(1 - F(t_2))/(1 - F(t_1)), \ldots, (1 - F(t_m))/(1 - F(t_{m-1}))$ are independent. This is equivalent to saying $Y(t) = -log(1 - F(t))$ has nonnegative independent increments. The basic theorem is:

Theorem 5 (Doksum, 1974)

If F is neutral to the right, and if $X_1, \ldots, X_n$ is a sample from F, then the posterior distribution of F given $X_1, \ldots, X_n$ is neutral to the right.

Basically, a process neutral to the right is tailfree with respect to every sequence of partitions $\{\mathcal{P}_n\}$ such that $\mathcal{P}_{n+1}$ is obtained from $\mathcal{P}_n$ by splitting the rightmost element, (t_n, ∞) into two pieces, $(t_n, t_{n+1}]$, (t_{n+1}, ∞). Thus, a Dirichlet process on the real line is neutral to the right, and neutral to the left, etc. Doksum (1974) conjectured that this property characterizes the Dirichlet process. This has been settled affirmatively.

Theorem 6 (James and Mosimann, 1980)

If F is neutral to the right and neutral to the left, then F is a Dirichlet process or a limit of Dirichlet processes or concentrated on two nonrandom points.

For another characterization of the Dirichlet process in terms of Johnson's sufficiency postulate or *learn-merge invariance*, see Böge and Möcks (1986).

Applications of Mixtures of Dirichlet Processes

In the paper of Antoniak (1974), a number of Bayesian statistical problems with Dirichlet process priors were discussed whose solution involved posterior mixtures of Dirichlet processes, in particular empirical Bayes, bio-assay, regression, discrimination, and classification problems. The computational difficulties involved were such that Antoniak treated only very small size problems. Since then, Monte Carlo methods due to Kuo (1986) have been developed making Bayes solutions to these problems feasible. See Dalal (1978) and Dalal and Hall (1980) for a discussion of approximation of arbitrary random probability measures by mixtures of Dirichlets.

1. Bayes empirical Bayes

Consider first the Bayes empirical Bayes problem. In the usual empirical Bayes setting, it is assumed that unobservable parameters θ_j, $j = 1,\ldots,n$ are taken independently from an unknown distribution G, and that associated with each θ_j, a random variable X_j is chosen independently from a distribution with density $f_j(x|\theta_j)$ for $j = 1,\ldots,n$. The problem is to estimate one or more of the θ_j. Most procedures use $X_1,\ldots,X_n$ to obtain an estimate G_n of G first and then estimate θ_j as the Bayes estimate with respect to the prior G_n. In the Bayes approach to the empirical Bayes problem, a prior distribution is placed on G. Berry and Christensen (1979) take G to be a Dirichlet process, $\mathfrak{D}(\alpha)$. Following Antoniak, the posterior distribution of G is a mixture of Dirichlet processes with parameter $\alpha + \Sigma\delta(\theta_j)$ and mixing distribution $H(\underline{\theta} \mid \underline{X})$, the posterior distribution of the θ_j given the X_j, in symbols,

$$G \mid \underline{X} \in \int \mathfrak{D}\left(\alpha + \sum_{j=1}^{n} \delta(\theta_j)\right) dH(\underline{\theta} \mid \underline{X}). \tag{1}$$

In view of the computation difficulties, even in the simple case where $f_j(x|\theta)$ is a binomial distribution with probability of success θ and sample size depending on j, Berry and Christensen suggest a couple of rough approximations to the Bayes rule that are easy to evaluate.

Monte Carlo approximation of the exact Bayes estimate was considered by Kuo (1986a, 1986b). Let $H(\underline{\theta})$ denote the unconditional marginal distribution of $\underline{\theta}$,

$$dH(\underline{\theta}) = \prod_{j=1}^{n} (M + j - 1)^{-1}\left(\alpha + \sum_{i=1}^{j-1} \delta(\theta_i)\right)(d\theta_j), \tag{2}$$

as given in Blackwell and MacQueen (1973). Then, from a formula of Lo (1984) for the posterior distribution of $\underline{\theta}$ given $\underline{X}$,

$$dH(\underline{\theta}|\underline{X}) = \prod_{j=1}^{n} f_j(X_j|\theta_j)\, dH(\underline{\theta}) \Big/ \left[\int \prod_{j=1}^{n} f(X_j|\theta_j)\, dH(\underline{\theta}) \right], \tag{3}$$

the exact Bayes estimate of θ_n, say, may be written,

$$\hat{\theta}_n(\underline{X}) = \int \theta_n dH(\underline{\theta}|\underline{X}) = \frac{\int \theta_n \Pi f_j(X_j|\theta_j)\, dH(\theta_j)}{\int \Pi f_j(X_j|\theta_j)\, dH(\theta_j)}. \tag{4}$$

The obvious Monte Carlo method in which vectors $\underline{\theta}^1, \ldots, \underline{\theta}^N$ are generated i.i.d. from the distribution (2) and then used to approximate the two integrals in the right side of (4) does not work well. In the method of Kuo, Monte Carlo is used only to decide which of the θ_j are equal to which others according to (2). Then the n-dimensional integrals in the right side of (4) reduce to a product of 1-dimensional integrals $dF_0(\theta)$, which can often be integrated exactly, for example, if $F_0(\theta)$ is taken as a conjugate prior of $f(x|\theta)$.

2. Bayesian bio-assay

As another application, consider the bio-assay problem. Let $F(t)$ denote the probability of a positive response for a subject treated at dose level t. It is assumed that $F(t)$ increases with t. Suppose that n_j subjects are treated at dose level t_j and that Y_j is the number of positive responses, $j = 1, \ldots, L$. It is assumed that the Y_j are independent binomial variables with probability $F(t_j)$ of success. The problem is to estimate F. The Bayes approach to this problem goes back to Kraft and Van Eeden (1964) who use a dyadic tailfree process as a prior. Ramsey (1972) uses a Dirichlet process prior and obtains the modal estimates of F by maximizing the finite dimensional joint density of the posterior distribution. (This seems to be the first description of the Dirichlet process; unfortunately, it is in a problem where the posteriors are not Dirichlet.)

Bhattacharya (1981) develops a large sample procedure for approximating the finite-dimensional distributions of the posteriors as a normal mixture of Dirichlet distributions. Disch (1981) considers the problem of estimating quantiles of a potency curve with Dirichlet process priors, and avoids the difficult computational tasks by suggesting approximations similar to those made by Berry and Christensen in the empirical Bayes problem. However, the methods of Kuo may be applied to this problem as well. For related work, see Kuo (1983, 1988) and Ammann (1984).

3. Bayesian density estimation

Another application of mixtures of Dirichlet processes is to estimate a density, $f(x)$, based on a sample of size n from f. Lo (1984) puts a prior on f, by writing $f(x) = \int K(x, u)dG(u)$ and letting G have a Dirichlet process prior, $\mathcal{D}(\alpha)$. He obtains the posterior distribution of G as a mixture of Dirichlet processes and uses this to obtain formulas for the Bayes estimate of f. One of his applications is to the two-parameter normal kernel $K = \phi(x|\mu,\sigma)$. This example was expanded in Ferguson (1983) who, using the representation of Sethuraman and Tiwari (1982), described $f(x)$ as a mixture of normal densities, $\Sigma P_j \phi(X|\mu_j,\sigma_j)$, where the P_j are as in Theorem 2 and the (μ_j,σ_j) are a sample from the four-parameter conjugate prior for the normal. Kuo's method was seen to provide a simple and effective means of performing the computations for large data sets. The estimates are seen to provide evidence for two suggestions: (1) for using a variable kernel estimate with wider windows at the tails, and (2) for using shrinkage estimates on the observations, namely bringing observations in toward the center, proportional to their distance from the center. In the paper of Kumar and Tiwari (1989), Kuo's method is applied to estimating a mixture of exponential densities.

Gaussian processes may also be used to generate densities. In the approach of Leonard (1978), a density on the interval (a, b) is written as $exp\{g(t)\}/ \int exp\{g(x)\}dx$ where g is a given Gaussian process. An alternate approach is provided by Thorburn (1986), in which the density is written as $exp\{g(t)\}$ where $g(t)$ is a Gaussian process conditional on $\int exp\{g(x)\}dx = 1$.

Application to Censored Data and Reliability

An important extension of nonparametric Bayes theory is to the treatment of censored data. The problem of estimating an unknown cdf F based on censored data is usually formulated as follows. Let $X_1,\ldots,X_n$ be a sample from F, and let the censoring points, $Y_1,\ldots,Y_n$, be random variables independent of the X's. The observations are $Z_j = min(X_j, Y_j)$, and $d_j = I(X_j \leq Y_j)$, $j = 1,\ldots,n$, where $I(A)$ represents the indicator function of the set A. The problem is to estimate F based on the observations. The usual nonparametric estimate is the product limit estimate, due to Kaplan and Meier (1958).

The first completely Bayes approach to this problem was made by Susarla and Van Ryzin (1976) who use a Dirichlet process as a prior for F. Let $u_1 < u_2 < \ldots < u_k$ be the distinct observations among $Z_1,\ldots,Z_n$; let λ_j denote the number of censored observations at u_j; let $k(t)$ denote the number of $u_j \leq t$; and let h_k be the number of $Z_j > u_k$.

Theorem 7 (Susarla and Van Ryzin, 1976)

If $F \in \mathcal{D}(\alpha)$, then the posterior expectation of the survival function, $1 - F(t)$, given the observations is

$$E(1 - F(t)|data) = \frac{\alpha(t) + h_{k(t)}}{M + n} \prod_{j=1}^{k(t)} \frac{\alpha(u_j) + h_j + \lambda_j}{\alpha(u_j) + h_j} \tag{5}$$

where $\alpha(t) = \alpha(t, \infty)$, and $M = \alpha(\mathbb{R})$.

This estimate reduces to the Kaplan-Meier estimate as the prior information, M, goes to zero. If there are no censored observations, the product term vanishes and we get the Bayes estimator of Ferguson (1973). Blum and Susarla (1977) complemented this result by showing that the posterior distribution of F given the data is a mixture of Dirichlet processes with specified transition and mixing measures.

This research was generalized to prior distributions neutral to the right by Ferguson and Phadia (1979). With Dirichlet process priors, the updating mechanism of going from prior to posterior is easy for uncensored observations and difficult for censored observations. For prior processes neutral to the right, it is the other way around. Thus, the generality provided by priors neutral to the right make them the natural priors to use for censoring problems. Also, it should be noted that the estimate in Theorem 1 does not depend on the distributions of the Y_j. Indeed, this should be the case when a Bayesian analysis is performed; in fact, as Ferguson and Phadia point out, the Y_j may be considered as constants, allowing treatment of problems in which future Y_j may depend upon past observations.

However, if X_j and Y_j are allowed to be dependent, the marginal distribution of X may not be identifiable. Nevertheless, a Bayesian treatment of the problem is possible and has been carried out by Phadia and Susarla (1983), by assuming a Dirichlet process prior for the joint distribution of (X, Y). They derive the Bayes estimate of the joint distribution, which of course need not be consistent. See also Arnold et al. (1984). Tsai (1986) adopts a different approach by taking the joint distribution of (Z, d) to be a Dirichlet process on $\mathbb{R} \times \{0,1\}$, and making an independence-like assumption that makes the marginal distribution of X identifiable, and the Bayes estimate of F consistent. Since the marginal distribution of F is not Dirichlet under this assumption, his resulting Bayes estimate is quite distinct from that of Susarla and Van Ryzin in the independent case.

For a review of the area up to 1980, see Phadia (1980b). For consistency of (5) and the product limit estimate, see Susarla and Van Ryzin (1978b) and Phadia and Van Ryzin (1980). For related results, see Gardiner and Susarla (1982, 1983), Colombo, Costantini and Jaarsma (1985), Rao and Tiwari (1985), Johnson and Christensen (1986) and Berliner and Hill (1988).

1. Application to reliability theory

A useful generalization of the gamma process for statistical problems has been introduced independently by Dykstra and Laud (1981) and Lo (1982). Given a nondecreasing left-continuous function α on $[0, \infty)$ with $\alpha(0) = 0$, $V(t)$ is said to be a gamma process with parameter α if $V(t)$ is a process with independent increments such that for all $t > 0$ the distribution of $V(t)$ is $\mathcal{G}(\alpha(t), 1)$, the gamma distribution with shape parameter $\alpha(t)$ and scale parameter 1. Given a nonnegative function β on $[0, \infty)$, the weighted gamma process with parameters α and β, $\mathcal{G}(\alpha, \beta)$, is then defined as the process $r(t) = \int_{[0,t]}\beta(s)dV(s)$. Its elementary properties include

Theorem 8 (Dykstra, Laud and Lo)

If $r \in \mathcal{G}(\alpha, \beta)$, then r is a process with independent increments, $E(r(t)) = \int_{[0,t]}\beta(s)d\alpha(s)$, and $Var(r(t)) = \int_{[0,t]}\beta^2(s)d\alpha(s)$.

Dykstra and Laud use this process (which they call an extended gamma process) as a prior distribution on the hazard rate function in nonparametric reliability problems; that is, they assume that the survival function, $S(t) = 1 - F(t)$, has the form $S(t) = exp\{-\int_{[0,t]}r(s)ds\}$, where $r \in \mathcal{G}(\alpha, \beta)$.

Theorem 9 (Dykstra and Laud)

If $r \in \mathcal{G}(\alpha, \beta)$, then $ES(t) = exp\{-\int_{[0,t]}log(1 + \beta(s)(t-s))d\alpha(s)\}$. If $X_1,\ldots,X_n$ is a sample from S, then the posterior distribution of r given the censored data $X_1 \geq x_1,\ldots,X_n \geq x_n$ is $\mathcal{G}(\alpha, \beta^*)$, where $\beta^*(t) = \beta(t)/(1 + \beta(t)\Sigma_{i=1}^n(x_j - t)^+)$.

They also show that the distribution of r given an uncensored sample is a mixture of weighted gamma processes, and examples are given showing the computational problems involved can be solved. This approach gives probability one to the absolutely continuous distributions, and Bayes estimates of the hazard rate and the cdf are derived.

Since in the above construction the gamma process has nondecreasing sample paths, the resulting survival distribution has increasing failure rate (IFR). Ammann (1984, 1985) puts this approach in a more general setting by representing the hazard rate as a function of the sample paths of nonnegative processes with independent increments which consist of an increasing component as well as a decreasing component. This results in a broad class of priors over a space of absolutely continuous distributions which contain IFR, DFR and U-shaped failure rate survival distributions. Ammann finds the posterior Laplace transforms of these processes based on data that may contain censored observations, and applies his approach to the competing risk model as well.

The Bayesian analysis discussed above may be extended to incorporate a covariate using the Cox proportional hazard model as was done by Kalbfleisch (1978). Independent observations $X_1,\ldots,X_n$ are made with respective covariate vectors $w_1,\ldots,w_n$ according to the survival distribution,

$$S(x \mid w) = S_0(x)^{\beta' w}$$

where β is the vector of regression parameters, and $S_0(x)$ is the baseline survival distribution. While the main interest in covariate analysis centers around the estimation and hypothesis testing of β, considering $S_0(x)$ as a nuisance parameter, it is still of interest to estimate $S_0(x)$ by itself. Writing $S_0(x) = exp\{-\Lambda(x)\}$, Kalbfleisch takes $\Lambda(x)$ to have a gamma process prior, and carries out the estimation of β by determining the marginal distribution of the observations as a function of β with $S_0(x)$ integrated out. Thus, the treatment is semi-parametric and semi-Bayesian. This approach was generalized to allow $1 - S_0(x)$ to be an arbitrary process neutral to the right by Wild and Kalbfleisch (1981). For related results, see Padgett and Wei (1981) and Mazzuchi and Singpurwalla (1985).

Empirical Bayes Estimation

Bayesian methods have been found to be useful in the non-Bayesian treatment of empirical Bayes problems. Suppose we are at the $n + 1^{st}$ stage of an experiment, and information is available not only from the current stage but also from the n previous stages. Let $F_1, F_2, \ldots, F_{n+1}$ be n+1 distributions on the real line, and for $j = 1, \ldots, n+1$, let $\underline{X}_j = (X_{j1}, \ldots, X_{jm_j})$ be a sample of size m_j from F_j. As a prior, we assume that $F_1, \ldots, F_{n+1}$ are a sample from the Dirichlet $\mathcal{D}(\alpha)$ where $\alpha = MG_0$. We wish to estimate $F_{n+1}(t)$ with squared error loss.

$$L(F_{n+1}, \tilde{F}) = \int (F_{n+1}(x) - \tilde{F}(x))^2 dW(x) \tag{6}$$

for some finite measure W. If we know M and G_0, this becomes a straightforward Bayes problem whose solution is

$$\tilde{F}_{n+1}(t) = q_{n+1} G_0(t) + (1 - q_{n+1}) \hat{F}_{n+1}(t) \tag{7}$$

where $q_j = M/(M + m_j)$ and $\hat{F}_j(t)$ is the sample distribution function based on $\underline{X}_j$. If α is unknown, we cannot use this estimate, but we may use $\underline{X}_1, \ldots, \underline{X}_n$ to help estimate M and G_0.

Korwar and Hollander (1976) and Hollander and Korwar (1977) consider the case where M is known and G_0 is unknown. They estimate $G_0(t)$ by the average of the sample distribution functions of $\underline{X}_1, \ldots, \underline{X}_n$, and propose the following empirical Bayes estimator of F_{n+1}:

$$H_{n+1}(t) = q_{n+1} \sum_{j=1}^{n} \hat{F}_j(t)/n + (1 - q_{n+1}) \hat{F}_{n+1}(t). \tag{8}$$

We say that this sequence of estimates is asymptotically optimal relative to a class of Dirichlet process priors if the Bayes risk of H_{n+1} given α, call it $r(\alpha, H_{n+1})$, converges to the Bayes risk of the Bayes estimate (7), call it $r(\alpha)$, whatever be α in the class. Since asymptotic optimality is a weak property, one wants rates of convergences. Korwar and Hollander prove:

Theorem 10 (Hollander and Korwar, 1977)

$$r(\alpha,\, H_{n+1}) = r(\alpha)\Big[1 + q_{n+1}\sum_{j=1}^{n}(1 - q_j)^{-1}/n^2\Big]$$

When all the m_j are equal, say to m, this reduces to $r(\alpha)(1 + M/(mn))$. Thus, $\{H_{n+1}\}$ is asymptotically optimal relative to the class of Dirichlet priors with fixed M, and the rate of convergence is $0(1/n)$. Hollander and Korwar also treat the empirical Bayes estimation of a mean, with similar results.

In their paper on testing hypotheses, Susarla and Phadia (1976) also consider the empirical Bayes extension of their problem using the method of Hollander and Korwar. In addition, they allow M as well as G_0 to be unknown, and, using an estimate of M based on the estimate of Korwar and Hollander (1973), exhibit an empirical Bayes estimate that is asymptotically optimal relative to the class of all Dirichlet priors. The extension of the Hollander and Korwar result to unknown M was made in the equal sample size case by Zehnwirth (1981), using a new estimate of M. The estimate is as follows. Let F_n denote the F-statistic in the one-way analysis of variance based on $\underline{X}_1,\ldots,\underline{X}_n$, ($F_n$ = ratio of the mean sum of squares between populations to the mean sum of squares within populations).

Theorem 11 (Zehnwirth, 1981)

$$m/(1 - F_n) \rightarrow M \text{ in probability as } n \rightarrow \infty.$$

The extension to empirical Bayes estimation of a distribution function based on censored data was made by Susarla and Van Ryzin (1978a) when all sample sizes, m_j, are 1, obtaining asymptotically optimal estimates at rate $0(1/n)$. Since the proposed estimate was not necessarily nondecreasing, Phadia (1980a) suggested using a simpler somewhat better estimate of G_0, which has the desirable property that the resulting empirical Bayes estimate is nondecreasing. This problem has also been treated by Ghorai (1981), taking a gamma process for $-log(1 - F(t))$ and obtaining asymptotically optimal estimates at rate $0(1/n)$.

In the uncensored case, Ghosh, Lahiri and Tiwari (1989) propose an empirical Bayes estimator of F_{n+1} that uses both the past as well as the current data for estimating G_0. Their proposed estimator is given by (7) with G_0 replaced by

$$\hat{G}_0(t) = \sum_{j=1}^{n+1}(1 - q_j)\hat{F}_j(t)\Big/\sum_{j=1}^{n+1}(1 - q_j). \tag{9}$$

Letting $\tilde{H}_{n+1}$ denote the resulting estimator, they derive the following result.

Theorem 12 (Ghosh et al., 1989)

$$r(\tilde{H}_{n+1}, \alpha) = r(\alpha)\left[1 + q_{n+1}\left(\sum_{j=1}^{n+1}(1 - q_j)\right)^{-1}\right]. \tag{10}$$

That this is a uniform improvement on the estimator in Theorem 10 is easily seen using Schwartz' inequality. Moreover Ghosh et al. have established the optimality of the weights used in (9), namely that the Bayes risk of $\tilde{H}_{n+1}$ is smaller than the Bayes risk of any other estimator that is a linear combination of the $\hat{F}_j$. In addition, they make a similar improvement to Zehnwirth's estimator of M by allowing it to depend upon $\underline{X}_{n+1}$ as well as by allowing the sample sizes to differ.

We comment briefly on other papers in the area. Hollander and Korwar (1976) treats the empirical Bayes estimation of $P(X > Y)$ in a two-sample problem. Phadia and Susarla (1979) treat the same problem allowing right censored data, Ghorai and Susarla (1982) consider the empirical Bayes estimation of a density using Lo's estimate. Ghosh (1985) and Tiwari and Zalkikar (1985a, b) consider empirical Bayes estimation problems for general estimable parameters of degree one and two. Tiwari, Jammalamadaka and Zalkikar (1988) treat the empirical Bayes version of the paper of Gardiner and Susarla (1983).

Random Symmetric Distributions; Problems of Consistency

An extension of the family of Dirichlet processes to the family of Dirichlet invariant processes was introduced by Dalal (1979a). Let $\mathcal{G} = \{g_1,\ldots,g_k\}$ be a fixed finite group of measurable transformations from $\mathfrak{X}$ into itself. Let α be a $\mathcal{G}$-invariant finite non-null measure on $\mathfrak{X}$. A random probability measure P on $(\mathfrak{X}, \mathcal{A})$ is said to be a *Dirichlet invariant process with parameter* α, in symbols $P \in \mathcal{DG}(\alpha)$, if P is $\mathcal{G}$-invariant (surely) and if for every partition $(A_1,\ldots,A_m)$ of $\mathfrak{X}$ made up of measurable invariant sets, $(P(A_1),\ldots,P(A_m)) \in \mathcal{D}(\alpha(A_1),\ldots,\alpha(A_m))$. Dalal and others (Tiwari, 1988; Hannum and Hollander, 1983) give constructive definitions along the following lines. Let $P \in \mathcal{D}(\alpha)$ and define P^* as $P^*(A) = (1/k)\Sigma_{g \in \mathcal{G}}P(gA)$. Then the distribution of P^* depends only upon α^*, where $\alpha^*(A) = (1/k)\Sigma_{g \in \mathcal{G}}\alpha(gA)$, and $P^* \in \mathcal{DG}(\alpha^*)$.

When $\mathcal{G}$ consists of only the identity transformation, $\mathcal{DG}(\alpha)$ corresponds to the usual Dirichlet process, $\mathcal{D}(\alpha)$. When $\mathcal{G}$ is generated by $g(x) = -x$, $\mathcal{DG}(\alpha)$ gives probability one to distributions that are symmetric about zero. Dalal (1979a) derives several properties of the Dirichlet invariant process and applies the theory to the estimation of a distribution function known to be symmetric about a known point, θ. The analysis is extended in Dalal (1979b) to the case where θ is unknown but given a prior distribution ν independent of P. See Dalal (1980) for an expository article on these problems.

An important analysis of these results, both theoretically and practically, has been given by Diaconis and Freedman (1986a, b). Such estimates may not be consistent throughout the support of the prior, as detailed in Theorem 13 below. The first example of an inconsistent Bayes estimate was given by Freedman (1963). A simple example of this phenomenon, Ferguson (1973), may be described as follows.

Let the prior distribution of F be the mixture, $F = p_0 H + (1 - p_0)\mathfrak{D}(\alpha)$, where p_0, the prior probability of H, is 1/2, where H is the uniform distribution on the interval (0, 1), and where $\alpha = MH$ with $M = 1$. The support of F is the set of all distributions on [0, 1]. The distribution of the distinct observations among a sample $X_1,\ldots,X_n$ from F is the same when $F = \mathfrak{D}(\alpha)$ as when $F = H$. Thus, as long as the observations are distinct, the posterior distribution of F given $X_1,\ldots,X_n$ is $p_n H + (1 - p_n)\mathfrak{D}(\alpha + \Sigma\delta(X_j))$, where p_n, the posterior probability of H, is easily computed to be $p_n = n!/(n! + 1)$. If ever two observations are exactly equal, then the possibility of H disappears and F has the posterior distribution $\mathfrak{D}(\alpha + \Sigma\delta(X_j))$. Now, suppose that the true distribution is continuous on (0, 1). No matter how non-uniform this distribution may be, the Bayes estimate of F converges to $\mathfrak{U}(0, 1)$.

Freedman and Diaconis (1983) have a positive result along the lines of this example: If F is a mixture of $\mathfrak{D}(\alpha_j)$ with $\alpha_j = M_j F_j$, and if the M_j are bounded, then the Bayes estimate of F is consistent. In the example above, one can think of H as a Dirichlet process with $M = \infty$, so although F is a mixture of Dirichlets, the M_j are not bounded. In Dalal's model, even if the true distribution is symmetric, the Bayes estimate may oscillate indefinitely between two wrong values.

Theorem 13 (Diaconis and Freedman, 1986a, b)

Let θ and F be independent, with θ having a standard normal distribution, and $F \in \mathfrak{D}\mathfrak{G}(\alpha)$ symmetric about zero, where $\alpha = MF_0$ with F_0 the standard Cauchy distribution. Then there exists a symmetric density, $h(x)$, with a maximum at zero and bounded support, such that if the true distribution of the X_j has density h, then the Bayes estimate of θ does not converge.

Doss (1984) provides a deep extension of the analysis of these problems from symmetric Dirichlet priors to symmetric priors neutral to the right. Doss (1985a, b) considers the problem of estimating a median in a different nonparametric Bayes framework. Let $F(x)$ be a distribution function with median zero, let θ be a real number, and let $X_1,\ldots,X_n$ be a sample from $F(x - \theta)$. To place a prior distribution on F that chooses median zero distributions with probability one, let α be a finite non-null measure, written as $\alpha = MF_0$, where F_0 is a distribution function with median zero, and suppose for simplicity that F_0 has no mass at zero. Let α_- and α_+ denote the restrictions of α to $(-\infty, 0)$ and $(0, \infty)$ respectively. Choose F_- and F_+ independently from $\mathfrak{D}(\alpha_-)$ and $\mathfrak{D}(\alpha_+)$

respectively, and let $F(t) = (F_-(t) + F_+(t))/2$. Thus, F is a random distribution function such that $F(0) = 1/2$; denote the distribution of F by $\mathcal{D}^*(\alpha)$.

Theorem 14 (Doss, 1985a)

Let θ and F be independent, with $\theta \in \nu$ and $F \in \mathcal{D}^*(\alpha)$, and assume that F_0 has continuous density f_0. Given θ and F, let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ be a sample from $F(x - \theta)$. Then the posterior distribution of θ given $\underset{\sim}{X}$ is

$$d\nu(\theta|\underset{\sim}{X}) = c(\underset{\sim}{X})[\Pi^* f_0(X_i - \theta)] \; M(\underset{\sim}{X}, \theta) d\nu(\theta),$$

where $M(\underset{\sim}{X}, \theta)^{-1} = \Gamma(M/2 + n\hat{F}_n(\theta))\Gamma(M/2 + n(1 - \hat{F}_n(\theta)))$, $\hat{F}_n$ is the empirical distribution function of $\underset{\sim}{X}$, Π^* represents the product over the distinct X_i, and $c(\underset{\sim}{X})$ is a normalizing constant.

Doss shows that if the true distribution of the X_j is discrete, the Bayes estimate of θ is consistent. However, if it is continuous, then the Bayes estimate can converge to a wrong value, it can oscillate indefinitely between two wrong values, or the set of its limit points can be dense in $\mathbb{R}$.

Hannum and Hollander (1983) have derived the Bayes risk of Dalal's (1979a) estimate of the distribution function under $\mathcal{D}\mathcal{G}(\alpha)$, and have compared it to the risk of Ferguson's (1973) estimator under $\mathcal{D}(\alpha)$. This enables them to (i) assess the savings in risk obtained by incorporating known symmetry structure in the model, and (ii) provide information about the robustness of Ferguson's estimator against a prior for which it is not Bayes. Yamato (1986, 1987) and Tiwari (1988) used the Dirichlet invariant process prior to derive the Bayes estimator of estimable parameters of an arbitrary degree.

Other Applications

Our survey is by no means complete. We mention a few other selected results and applications in this final section. Binder (1982) considers finite population models in which a population $\{Y_1, \ldots, Y_N\}$ consists of a sample from $F \in \mathcal{D}(\alpha)$. A sample, $y_1, \ldots, y_n$, is then taken from $\{Y_1, \ldots, Y_N\}$, and the Bayes estimate of ΣY_j is derived. The asymptotic distributions are found in Lo (1986). Problems of finding confidence bounds for a distribution function have been considered by Breth (1978), who finds recursive methods for computing $\mathcal{P}(u_j \leq F(t_j) \leq v_j$ for $j = 1, \ldots, m)$ for fixed numbers $\{u_j\}$, $\{v_j\}$ and $\{t_j\}$ when F is a Dirichlet process. In a continuation paper, Breth (1979) applies the method to finding confidence intervals for quantiles and the mean, and also treats Bayesian tolerance intervals. Tamura (1988) applies Dirichlet process methods to auditing problems.

1. Linear Bayes estimation

The useful idea of restricting attention to a linear space of estimates in Bayesian nonparametric problems is due to Goldstein (1975a, b). Such estimates may require less knowledge of the prior and be much easier to compute than Bayes estimates without much loss of efficiency. As an example, consider the problem of estimating a mean $\mu = \int x dP(x)$ within the class of linear functions, $\hat{\mu} = a + \Sigma b_j X_j$. The Bayes solution is

$$\hat{\mu} = \frac{M}{M+n}\mu_0 + \frac{n}{M+n}\bar{X}_n, \text{ where}$$

$$\mu_0 = E(\mu) \quad \text{and} \quad M = \frac{E(\sigma^2)}{E(\mu^2) - (\mu_0)^2}.$$

Here, $\sigma^2 = \int x^2 dP(x) - \mu^2$ is the variance of the random distribution. This estimate is formally identical to the Bayes estimate with the Dirichlet prior, Theorem 1, with however a new interpretation for the parameter M. In addition, the only information needed to be elicited from the prior are the three quantities, $E(\mu)$, $E(\mu^2)$ and $E(\sigma^2)$. These ideas were further developed by Zehnwirth (1985) in treating estimation with censored data, by Poli (1985), who finds the best linear predictor in a multivariate regression model and specializes to a Dirichlet prior and to a normal/Wishart mixture of Dirichlets, and by Kuo (1988) in estimating the potency curve in Bayesian bio-assay.

2. Sequential problems

A number of papers treat sequential nonparametric problems from a Bayesian viewpoint. Hall (1976, 1977) in treating sequential search problems with random overlook probabilities allows the distributions of the overlook probabilities to be Dirichlet or a mixture of Dirichlet. Ferguson (1982) discusses k-stage lookahead rules and modified rules in some nonparametric sequential estimation problems with Dirichlet priors. Clayton and Berry (1985) treat the finite horizon one-armed bandit with the unknown arm producing observations from a Dirichlet process. In a sequential testing problem, Clayton (1985) assumes that in sampling from $F \in \mathcal{D}(\alpha)$, the payoff if you stop at n is $max(E(X|X_1,\ldots,X_n), \nu) - nc$, where ν and $c > 0$ are constants. He shows that the optimal stopping rule is bounded if the support of α is bounded, and he conjectures that this is true even if the support of α is unbounded. Christensen (1986) obtains a similar result for the problem of sampling without recall from a distribution $F \in \mathcal{D}(\alpha)$ and constant cost of observation. Betró and Schoen (1987) consider the problem of sampling with recall and constant cost from a distribution F assumed to be a simple homogeneous process neutral to the right.

3. Point processes

Lo (1982) considers the problem of estimation of the intensity measure γ of a nonhomogeneous Poisson point process based on a random sample from this process. He shows that if the prior distribution for γ is a weighted gamma distribution $\mathcal{G}(\alpha, \beta)$, then given a sample $N_1,\ldots,N_n$ of n functions from this process, the posterior distribution of γ is again gamma, $\mathcal{G}(\alpha + \Sigma N_j, \beta/(n\beta + 1))$. Lo also shows that the posterior process converges weakly to the Brownian bridge.

Another paper of Lo (1981) describes an application to shock models and wear processes. A device is subject to shocks occurring randomly at times according to a homogeneous Poisson point process $N(t)$ with intensity γ. The i^{th} shock causes a random amount X_i of damage, assumed to be i.i.d. F on $[0, \infty)$. For the prior distribution, γ and F are chosen to be independent, with $\gamma \in$ a gamma distribution $\mathcal{G}(\lambda, \theta)$, and $F \in \mathcal{D}(\alpha)$. In the posterior distribution based on a single observation of N up to time T, γ and F are still independent, with $\gamma \in \mathcal{G}(\lambda + N(T), \theta + T)$ and $F \in \mathcal{D}(\alpha + N)$. This readily yields Bayes estimates of γ and F.

Johnson, Susarla and Van Ryzin (1979) present an application to the Bellman-Harris age-dependent branching process. Each individual x born has a random length of life λ_x and reproduces at death a random number ξ_x of offspring, where the (λ_x, ξ_x) are i.i.d. from $G \times P$. The prior distribution of G and P are taken to be independent Dirichlet processes with parameters α_1 and α_2, and Bayes estimates of G and P are developed based on an observation of the process through time T starting with one individual.

References

Ammann, L. P. (1984): Bayesian nonparametric inference for quantal response data, *Ann. Statist.* 12, 636-645.

Ammann, L. P. (1985): Conditional Laplace transforms for Bayesian nonparametric inference in reliability theory, *Stoch. Proc. and Their Appl.* 20, 197-212.

Antoniak, C. (1974): Mixtures of Dirichlet processes with application to Bayesian nonparametric problems, *Ann. Statist.* 2, 1152-1174.

Arnold, B. C., Brockett, P. L., Torrez, W. and Wright, A. L. (1984): On the inconsistency of Bayesian non-parametric estimators in competing risk/multiple decrement models, *Insurance: Mathematics and Economics* 3, 49-55.

Basu, D. and Tiwari, R. C. (1982): A note on the Dirichlet process, *Statistics and Probability: Essays in Honor of C. R. Rao*, Kallianpur, Krishnaiah and Ghosh, Eds., 89-103.

Berk, R. H. and Savage, I. R. (1979): Dirichlet processes produce discrete measures: an elementary proof, *Contributions to Statistics. Jaroslav Hajek Memorial Volume*, Academia, North Holland, Prague, 25-31.

Berliner, L. M. and Hill, B. M. (1988): Bayesian nonparametric survival analysis, *J. Amer. Statist. Assoc.* 83, 772-779.

Berry, D. A. and Christensen, R. (1979): Empirical Bayes estimation of a binomial parameter via mixtures of Dirichlet process, *Ann. Statist.* 7, 558-568.

Betró, B. and Schoen, F. (1987): Sequential stopping rules for the multistart algorithm in global optimization, *Math. Programming* 38, 271-286.

Bhattacharya, P.K. (1981): Posterior distribution of a Dirichlet process from quantal response data, *Ann. Statist.* 9, 803-811.

Binder, D. A. (1982): Non-parametric Bayesian models for samples from finite populations, *J. Roy. Statist. Soc. B* 44, 388-393.

Blackwell, D. (1973): Discreteness of Ferguson selections, *Ann. Statist.* 1, 356-358.

Blackwell, D. and MacQueen, J. B. (1973): Ferguson distributions via Polya urn schemes, *Ann. Statist.* 1, 353-355.

Blum, J. and Susarla, V. (1977): On the posterior distribution of a Dirichlet process given randomly right censored observations, *Stoch. Processes Appl.* 5, 207-211.

Böge, W. and Möcks, J. (1986): Learn-merge invariance of priors: a characterization of Dirichlet distributions and processes, *J. Mult. Anal.* 18, 83-92.

Breth, M. (1978): Bayesian confidence bands for a distribution function, *Ann. Statist.* 6, 649-657.

Breth, M. (1979): Nonparametric Bayesian interval estimation, *Biometrika* 66, 641-644.

Campbell, G. and Hollander, M. (1978): Rank order estimation with the Dirichlet prior, *Ann. Statist.* 6, 142-153.

Christensen, R. (1986): Finite stopping in sequential sampling without recall from a Dirichlet process, *Ann. Statist.* 14, 275-282.

Cifarelli, D. M. and Regazzini, E. (1979): Considerazioni generali sull'impostazione bayesiana di problemi non parametrici, *Rivista di Matematica per le Scienze Economiche e Socialli* 2 Part I 39- 52, Part II 95-111.

Clayton, M. K. (1985): A Bayesian nonparametric sequential test for the mean of a population, *Ann. Statist.* 13, 1129-1139.

Clayton, M. K. and Berry, D. A. (1985): Bayesian nonparametric bandits, *Ann. Statist.* 13, 1523-1534.

Colombo, A. G., Costantini, D. and Jaarsma, R. J. (1985): Bayes nonparametric estimation of time-dependent failure rate, *IEEE Trans. Reliability* R-34, 109-112.

Dalal, S. R. (1978): A note on the adequacy of mixtures of Dirichlet processes, *Sankhya Ser. A* 40, 185-191.

Dalal, S. R. (1979a): Dirichlet invariant processes and applications to nonparametric estimation of symmetric distribution functions, *Stoch. Proc. Appl.* 9, 99-107.

Dalal, S. R. (1979b): Nonparametric and robust Bayes estimation of location, in *Optimizing Methods in Statistics*, ed. J. Rustagi, Academic Press Inc., 141-166.

Dalal, S. R. (1980): Bayesian nonparametric theory, in *Bayesian Statistics*, eds. J. M. Bernardo et al., University Press, Valencia, Spain, 521-534.

Dalal, S. R. and Hall, G. J., Jr. (1980): On approximating parametric Bayes models by nonparametric Bayes models, *Ann. Statist.* 8, 664-672.

Dalal, S. R. and Phadia, E. G. (1983): Nonparametric Bayes inference for concordance in bivariate distributions, *Comm. Statist. Theor. Meth.* 12, 947-963.

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.

Disch, D. (1981): Bayesian nonparametric inference for effective doses in a quantal response experiment, *Biometrics* 37, 713-722.

Doksum, K. (1974): Tailfree and neutral random probabilities and their posterior distributions, *Ann. Prob.* 2, 183-201.

Doss, H. (1984): Bayesian estimation in the symmetric location problem, *Z. W.* 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.

Doss, H. and Sellke, T. (1982): The tails of probabilities chosen from a Dirichlet prior, *Ann. Statist.* 10, 1302-1305.

Dykstra, R. L. and Laud, P. (1981): A Bayesian nonparametric approach to reliability, *Ann. Statist.* 9, 356-367.

Fabius, J. (1964): Asymptotic behavior of Bayes estimates, *Ann. Math. Statist.* 35, 846-856.

Fabius, J. (1973): Two characterizations of the Dirichlet distributions, *Ann. Statist.* 1, 583-587.

Ferguson, T. S. (1973): A Bayesian analysis of some nonparametric problems, *Ann. Statist.* 1, 209-230.

Ferguson, T. S. (1974): Prior distributions on spaces of probability measures, *Ann. Statist.* 2, 615-629.

Ferguson, T. S. and Phadia, E. G. (1979): Bayesian nonparametric estimation based on censored data, *Ann. Statist.* 7, 163-186.

Ferguson, T. S. (1982): Sequential estimation with Dirichlet process priors, in *Statistical Decision Theory and Related Topics III*, Vol. 1, eds. S. Gupta and J. Berger, 385-401.

Ferguson, T. S. (1983): Bayesian density estimation by mixtures of normal distributions, in *Recent Advances in Statistics*, eds. H. Rizvi and Rustagi, Academic Press, New York, 287-302.

Freedman, D. (1973): On the asymptotic behavior of Bayes estimates in the discrete case, *Ann. Math. Statist.* 34, 1386-1403.

Freedman, D. and Diaconis, P. (1983): On inconsistent Bayes estimates in the discrete case, *Ann. Statist.* 11, 1109-1118.

Gardiner, J. C. and Susarla, V. (1982): A nonparametric estimator of the survival function under progressive censoring, *Survival Analysis, IMS Lecture Notes-Monograph Series* Vol. 2, eds. J. Crowley and R. A. Johnson, 26-40.

Gardiner, J. C. and Susarla, V. (1983): Weak convergence of a Bayesian nonparametric estimator of the survival function under progressive censoring, *Statistics and Decisions* 1, 257-263.

Ghorai, J. (1981): Empirical Bayes estimation of a distribution function with a gamma process prior, *Comm. Statist. Theory Meth.* A10(12), 1239-1248.

Ghorai, J. K. and Susarla, V. (1982): Empirical Bayes estimation of probability density functions with Dirichlet process prior, Probability and Statistical Inference, *Proc. of the 2nd Pannonian Symp. on Math. Stat.*, eds. Grossmann et al., 101-114.

Ghosh, M. (1985): Nonparametric empirical Bayes estimation of certain functionals, *Comm. Statist. Theory Meth.* 14(9), 2081-2094.

Ghosh, M., Lahari, P. and Tiwari, R. C. (1989): Nonparametric empirical Bayes estimation of the distribution function and the mean, *Comm. Statist. Theory Meth.* 18(1), 121-146.

Goldstein, M. (1975a): Approximate Bayes solutions to some nonparametric problems, *Ann. Statist.* 3, 512-517.

Goldstein, M. (1975b): A note on some Bayesian nonparametric estimates, *Ann. Statist.* 3, 736-740.

Hall, G. J., Jr. (1976): Sequential search with random overlook probabilities, *Ann. Statist.* 4, 807-816.

Hall, G. J., Jr. (1977): Strongly optimal policies in sequential search with random overlook probabilities, *Ann. Statist.* 5, 124-135.

Hannum, R. C., Hollander, M. and Langberg, N. A. (1981): Distributional results for random functionals of a Dirichlet process, *Ann. Prob.* 9, 665-670.

Hannum, R. C. and Hollander, M. (1983): Robustness of Ferguson's Bayes estimator of a distribution function, *Ann. Statist.* 11, 632-639, 1267.

Hollander, M. and Korwar, R. M. (1976): Nonparametric empirical Bayes estimation of the probability that X $\leq$ Y, *Commun. Statist. Theor. Meth.* A5(14), 1369-1383.

Hollander, M. and Korwar, R. M. (1977): Nonparametric estimation of distribution functions, in *Thy. and Appl. of Reliability 1*, Academic Press, New York, 85-107.

Hollander, M. and Korwar, R. M. (1980): Nonparametric Bayesian estimation of the horizontal distance between two populations, *Colloq. Math. Soc. Janos Bolyai 32. Nonparametric Statistical Inference*, Budapest, 409-415.

James, I. R. and Mosimann, J. E. (1980): A new characterization of the Dirichlet distribution through neutrality, *Ann. Statist.* 8, 183-189.

Johnson, R. A., Susarla, V. and Van Ryzin, J. (1979): Bayesian non-parametric estimation for age-dependent branching processes, *Stoch. Proc. & Appl.* 9, 307-318.

Johnson, W. and Christensen, R. (1986): Bayesian nonparametric survival analysis for grouped data, *Can. J. Statist.* 14, 307-314.

Kalbfleisch, J. D. (1978): Nonparametric Bayesian analysis of survival time data, *J. Royal Statist. Soc.* B40, 214-222.

Kaplan, E. L. and Meier, P. (1958): Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* 53, 457-481.

Korwar R. M. and Hollander, M. (1973): Contributions to the theory of Dirichlet processes, *Ann. Prob.* 1, 705-711.

Korwar, R. M. and Hollander, M. (1976): Empirical Bayes estimation of a distribution function, *Ann. Statist.* 4, 581-588.

Kraft, C. H. (1964): A class of distribution function processes which have derivatives, *J. Appl. Prob.* 1, 385-388.

Kraft, C. H. and van Eeden, C. (1964): Bayesian bio-assay, *Ann. Math. Statist.* 35, 886-890.

Kumar, S. and Tiwari, R. C. (1989): Bayes reliability estimation under a random environment governed by a Dirichlet process, *IEEE Transactions on Reliability*, Aug.

Kuo, L. (1983): Bayesian bio-assay design, *Ann. Statist.* 11, 886-895.

Kuo, L. (1986a): A note on Bayes empirical Bayes estimation by means of Dirichlet processes, *Statist. Prob. Letters* 4, 145-150.

Kuo, L. (1986b): Computations of mixtures of Dirichlet processes, *SIAM J. Sci. Statist. Comput.* 7, 60-71.

Kuo, L. (1988): Linear Bayes estimators of the potency curve in bioassay, *Biometrika* 75, 91-96.

Leonard, T. (1978): Density estimation, stochastic processes and prior information, *J. R. Statist. Soc.* B40, 113-146.

Lo, A. (1981): Bayesian nonparametric statistical inference for shock models and wear processes, *Scand. J. Statist.* 8, 237-242.

Lo, A. (1982): Bayesian nonparametric statistical inference for Poisson point processes, *Z. Wahrsch. verw. Gebiete* 59, 55-66.

Lo, A. (1983): Weak convergence for Dirichlet processes, *Sankhya* 45, 105-111.

Lo, A. (1984): On a class of Bayesian nonparametric estimates: I. Density estimates, *Ann. Statist.* 12, 351-357.

Lo, A. (1986): Bayesian statistical inference for sampling a finite population, *Ann. Statist.* 14, 1226-1233.

Mazzuchi, T. A. and Singpurwalla, N. D. (1985): A Bayesian approach to inference for monotone failure rates, *Statist. & Probab. Letters* 3, 135-141.

Padgett, W. J. and Wei, L. J. (1981): A Bayesian nonparametric estimator of survival probability assuming increasing failure rate, *Comm. Statist. Theor. Meth.* A10(1), 49-63.

Phadia, E. G. (1980a): A note on empirical Bayes estimation of a distribution function based on censored data, *Ann. Statist.* 8, 226-229.

Phadia, E. G. (1980b): Nonparametric Bayesian inference based on censored data–an overview, *Coll. Math. Societ. Janos Bolyai 32. Nonparametric Statistical Inference*, Budapest, 667-686.

Phadia, E. G. and Susarla, V. (1979): An empirical Bayes approach to two-sample problems with censored data, *Comm. Statist. Theor. Meth.* A8(13), 1327-1351.

Phadia, E. G. and Susarla, V. (1983): Nonparametric Bayesian estimation of a survival curve with dependent censoring mechanism, *Ann. Inst. Statist. Math.* 35A, 389-400.

Phadia, E. G. and Van Ryzin, J. (1980): A note on convergence rates for the product limit estimator, *Ann. Statist.* 8, 673-678.

Poli, I. (1985): A Bayesian non-parametric estimate for multivariate regression, *J. Econometrics* 28, 171-182.

Ramsey, F. (1972): A Bayesian approach to bio-assay, *Biometrics* 28, 841-848.

Rao, J. S. and Tiwari, R. C. (1985): Estimation of survival function and failure rate, *Statistics* 16, 535-540.

Sethuraman, J. and Tiwari, R. C. (1982): Convergence of Dirichlet measures and the interpretation of their parameter, in *Statistical Decision Theory and Related Topics III* 2, eds. Gupta and Berger, Academic Press, New York, 305-315.

Susarla, V. and Phadia, E. G. (1976): Empirical Bayes testing of a distribution function with Dirichlet process priors, *Commun. Statist. Theor. Meth.* A5(5), 455-469.

Susarla, V. and Van Ryzin, J. (1976): Nonparametric Bayesian estimation of survival curves from incomplete observations, *J. Amer. Statist. Soc.* 71, 897-902.

Susarla, V. and Van Ryzin, J. (1978a): Empirical Bayes estimation of a distribution (survival) function from right censored observations, *Ann. Statist.* 6, 740-754.

Susarla, V. and Van Ryzin, J. (1978b): Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples, *Ann. Statist.* 6, 755-768. Addendum (1980) *Ann. Statist.* 8, 693.

Tamura, H. (1988): Estimation of rare errors using expert judgement, *Biometrika* 75, 1-9.

Thorburn, D. (1986): A Bayesian approach to density estimation, *Biometrika* 73, 65-75.

Tiwari, R. C. (1988): Convergence of Dirichlet invariant measures and the limits of Bayes estimates, *Commun. Statist. Theory Meth.* A17(12), 375-393.

Tiwari, R. C. and Zalkikar, J. N. (1985a): Empirical Bayes estimation of functionals of unknown probability measures, *Comm. Statist. Theor. Meth.* 14, 2963-2996.

Tiwari, R. C. and Zalkikar, J. N. (1985b): Empirical Bayes estimate of certain estimable parameters of degree two, *Calcutta Statist. Assoc. Bull.* 34, 179-188.

Tiwari, R. C., Jammalamadaka, S. R. and Zalkikar, J. N. (1988): Bayes and empirical Bayes estimation of survival function under progressive censoring, *Comm. Statist. Theory Meth.* A17(10): 3591-3606.

Tsai, W.-Y. (1986): Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application, *Ann. Statist.* 14, 238-249.

Wild, C. J. and Kalbfleisch, J. D. (1981): A note on a paper by Ferguson and Phadia, *Ann. Statist.* 9, 1061-1065.

Yamato, H. (1975): A Bayesian estimation of a measure of the difference between two continuous distributions, *Rep. Fac. Sci. Kagoshima University* 8, 29-38.

Yamato, H. (1984): Characteristic functions of means of distributions chosen from a Dirichlet process, *Ann. Probab.* 12, 262-267.

Yamato, H. (1986): Bayes estimates of estimable parameters with a Dirichlet invariant process, *Comm. Statist. Theor. Meth.* 15(8), 2383-2390.

Yamato, H. (1987): Nonparametric Bayes estimates of estimable parameters with a Dirichlet invariant process and invariant U-statistics, *Comm. Statist. Theor. Meth.* 16, 525-543.

Zalkikar, J. N., Tiwari, R. C. and Jammalamadaka, S. R. (1986): Bayes and empirical Bayes estimation of the probability that $Z > X + Y$, *Comm. Statist. Theor. Meth.* 15(10), 3079-3101.

Zehnwirth, B. (1981): A note on the asymptotic optimality of the empirical Bayes distribution function, *Ann. Statist.* 9, 221-224.

Zehnwirth, B. (1985): Nonparametric linear Bayes estimation of survival curves from incomplete observations, *Comm. Statist. Theor. Meth.* 14(8), 1769-1778.