An Elementary Continuous-Type Nonparametric Distribution Estimate

by

James M. Dickey*

University of Minnesota, School of Statistics

Technical Report No. 525
January 1989

* Research was supported in part by National Science Foundation Grant No. DMS-8614793.

An Elementary Continuous-Type Nonparametric Distribution Estimate

JAMES M. DICKEY*

1. INTRODUCTION

In problems of nonparametric inference from a sample, parameterized models for the population distribution are avoided. This is natural in a wide range of applications where such assumptions would be unjustified, for example in one dimension, in the survival analysis of a new disease, or in many dimensions, in a physiologic profile analysis of a new species from sample organisms. Because of the apparent intractability of nonparametric problems to a Bayesian approach, in which a whole likelihood function must be used and not merely some nonsufficient statistic, such problems were considered by L. J. Savage (1964) to be an embarrassment for Bayesian statistics. Since the population distribution is not known to within only a few parameters, which could then be given a prior distribution, a whole prior random process for the unknown distribution must be constructed. The often considered Dirichlet prior random process is unsatisfactory in having discrete distributions as its outcomes and leading to an noncontinuous mixed-type distribution estimate as the mean distribution of the posterior process (Ferguson 1973, Blackwell 1973). Worse yet, perhaps, the Dirichlet prior process is prejudiced in favor of highly nonsmooth distributions, in the sense that the prior correlations between the unknown probabilities of disjoint neighboring interval events are all negative. The Bayesian literature on the subject is extensive and could be said to include Good (1950,1965), Whittle (1958), Hill (1968,1987ab), Dickey (1968ab,1969), Good and Gaskins (1971,1980), Ferguson (1973), Leonard (1973,1978), Lo (1984), Lenk (1984,1988), Titterington (1986), and Olkin and Spiegelman (1987).

We will propose an elementary new nonparametric distribution estimate that will not be motivated by a Bayesian derivation. It is a continuous type distribution and even has a continuous density function. However, the distribution will not be defined in terms of its density. In general, the density function can be expensive to compute, and so the method may not be useful in some density-estimation contexts. The distribution will be identified, rather, by synthetically defining its random variable or random vector as a function of a finite dimensional Dirichlet random vector. (It will not be a "mixture of Dirichlet distributions," since only one Dirichlet random vector will be involved.) Thus, the method can be used readily to draw Monte Carlo samples from the estimated distribution. These can be useful for generalized bootstrapping, cross-validation, or in simulation studies where a population distribution is not fully known. Versions of the proposed distribution can serve as a predictive distribution of further data, based on the observed sample.

The moments of the distribution are easy to calculate. Indeed, the distribution can be constructed to have its low-order moments identical to the sample moments. Limiting forms of this continuous-type distribution include the so called sample distribution, the discrete empirical distribution of the

*Research was supported in part by National Science Foundation Grant No. DMS-8614793

data, itself. The method produces a distribution from a sample of n data vectors of any dimension k, and a marginalization property is satisfied, as follows. Applying the method directly to k_1 coordinates of the n data vectors, $k_1 \le k$, yields a distribution estimate that is identical to the k_1-dimensional marginal distribution of the full k-dimensional distribution estimate.

2. A FIRST ATTEMPT

The simple idea involved in constructing the proposed distribution is that a linear combination,

$$\xi = \mathbf{X}\alpha$$
$$= \alpha_1 \mathbf{x}_1 + \ldots + \alpha_n \mathbf{x}_n , \qquad (2.1)$$

of sample column vectors, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $k \times n$, can be given a post-sampling distribution with controllable properties by taking the data vectors as fixed and the coefficients vector α to be random with a parameterized distribution. Such a vector ξ can also be described as a linear function of α, and we will call the distribution of ξ a **filtered-variate** modification of the distribution of α. If the distribution of α is of continuous type over a region, Support(α), in R^n, then ξ will have a continuous distribution, too, over a linear transformation of the region, Support(ξ) = $\mathbf{X}$ Support(α), in R^k. In particular, if α has a Dirichlet distribution $\alpha \sim D(\mathbf{a})$ with the density in $\alpha_1, \ldots, \alpha_{n-1}$ on the probability simplex, Support(α) = Δ^{n-1}, $\Delta^{n-1} = \{\alpha : \text{each } \alpha_j \ge 0 \text{ and } \alpha_1 + \ldots + \alpha_n = 1\}$,

$$p(\alpha) = B(\mathbf{a})^{-1} \Pi_{j=1}^{n} \alpha_j^{a_j - 1}, \qquad (2.2)$$

where each $0 < a_j < \infty$ and $B(\mathbf{a}) = [\Pi\, \Gamma(a_j)] / \Gamma(\Sigma\, a_j)$, then the distribution of ξ will be said to have a **filtered-variate Dirichlet distribution.** Its region of support is the convex hull of the sample points, Support(ξ) = Hull($\mathbf{x}_1, \ldots, \mathbf{x}_n$). Under the usual assumption of full-rank data, $n \ge k+1$ and $\text{rank}(\mathbf{X} - \mathbf{x}_n \mathbf{1}_n^T) = k$, where $\mathbf{1}_n = (1, \ldots, 1)^T$, $n \times 1$, this support is a k-dimensional solid polytope whose vertices are a subset of the sample points. In one dimension (k = 1), this gives a distribution that has a positive density function over the observed range interval of the data, [min x_j, max x_j].

Known properties of the Dirichlet distribution establish the character of the tentative distribution estimate, (2.1) with (2.2). For simplicity, assume a symmetric Dirichlet distribution with all $a_j \equiv a$, $j = 1, \ldots, n$, and thus the centered mean vector $\mathbf{a}/\Sigma a_j = n^{-1}\mathbf{1}_n$. As a grows without bound, the Dirichlet degenerates to a one-point distribution concentrated at its mean point, $P[\alpha = n^{-1}\mathbf{1}_n] = 1$, and then our random vector ξ concentrates right at the sample mean, $P[\xi = \bar{\mathbf{x}}] = 1$, where $\bar{\mathbf{x}} = n^{-1} \Sigma\, \mathbf{x}_j$. In the other extreme, as the common parameter value a shrinks to zero, the distribution of α becomes a discrete uniform distribution over the set of unit coordinate vectors $\delta_{(j)}$, $j = 1, \ldots, n$, where $\delta_{(j)} = (\delta_{(j)1}, \ldots, \delta_{(j)n})^T$, $\delta_{(j)j}=1$, and $\delta_{(j)i}=0$ for $i \ne j$. Then $P[\alpha = \delta_{(j)}] = 1/n$, $j = 1, \ldots, n$. And so in this case, ξ has the discrete uniform distribution over the set of sample points, the empirical distribution of the sample: $P[\xi = \mathbf{x}_j] = 1/n$, $j = 1, \ldots, n$. Finally, in general when each $0 \le a_j \le \infty$, as we shall see below, the low order moments for ξ are

$$E\,\xi = \bar{\mathbf{x}} \quad \text{and} \quad \text{Var}\,\xi = c\,\mathbf{S_X}, \qquad (2.3)$$

where $c = (na + 1)^{-1}$ and $\mathbf{S_X} = n^{-1} \Sigma\, (\mathbf{x}_j - \bar{\mathbf{x}})\, (\mathbf{x}_j - \bar{\mathbf{x}})^T$.

So what we have in the tentative distribution estimate, (2.1) with (2.2) and $a_j \equiv a$, is a continuous-type distribution (if $0 < a < \infty$) with interesting limiting forms, whose mean is the sample mean and whose variance matrix is proportional to the sample variance matrix. Note, however, that the constant of proportionality in (2.3) cannot be greater than unity, $0 \le c \le 1$, and its upper bound, $c = 1$ for $a = 0$, would be achieved at the expense of degeneracy of the distribution to the sample empirical distribution. This, together with the limitation that the support cannot extend beyond the observed data points, leads us to speculate that both these restrictions can be removed by a variance-dilation or proportional extension of the data outward from its mean point. We will see that this can work, using a generalization of the filtered-variate Dirichlet distribution, following a discovery that the filtered-variate Dirichlet family is not unique for our purpose.

3. VARIATE FILTERS OF DISTRIBUTIONS HAVING A MEAN-STRUCTURED VARIANCE

By the form of the Dirichlet density (2.2), the Dirichlet moments are obvious ratios of products of gamma functions. These give the Dirichlet mean and variance as

$$E\alpha = \mathbf{u},$$

$$\mathrm{Var}\ \alpha = c\,[\mathrm{Diag}(\mathbf{u}) - \mathbf{u}\,\mathbf{u}^T] \qquad (3.1)$$

where the mean vector $\mathbf{u} = a/\Sigma a_j$ and the proportionality factor $c = (\Sigma a_j + 1)^{-1}$. Let us say that any distribution on Δ^{n-1} satisfying (3.1) for some vector $\mathbf{u}$ and some proportionality value c, where $0 \le c \le 1$, has a **mean-structured variance (MSV)**. Denote this by $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$. Another familiar MSV distribution is the normalized multinomial: if $\mathbf{n} \sim \mathrm{multinomial}(N, \mathbf{u})$ and $\alpha = \mathbf{n}/N$, then $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$ with $c = 1/N$. Many other families of distributions on the probability simplex are MSV. Indeed, we find that the MSV property is preserved under: (i) mixing over the mean by a mixing distribution that is itself MSV; (ii) partitioning and grouping of coordinate indices and corresponding summation of coordinates; and (iii) weighted averaging between independent (or uncorrelated) MSV vectors. For example, (i) implies that the normalized form of the Dirichlet-multinomial, the conjugate Bayesian predictive distribution for multinomial sampling, is MSV with $c = (N + \Sigma a_j) / [N (\Sigma a_j + 1)]$, in an obvious choice of notation.

Lemma 3.1 As $c \downarrow 0$, an MSV distribution $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$ on Δ^{n-1} approaches the singular distribution with unit mass at $\alpha = \mathbf{u}$. As $c \uparrow 1$, the limiting distribution is the discrete distribution of a random vector α^* supported on the extreme points of Δ^{n-1} in which each $P[\alpha^* = \delta_{(j)}] = u_j$, $j = 1, \ldots, n$.

Now, consider any filtered-variate distribution with $\xi = \mathbf{Z}\alpha$, for some matrix $\mathbf{Z}$, in which $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$. The limiting distributions of ξ with c are immediate from the limiting forms of $\mathrm{MSV}(\mathbf{u}, c)$.

Theorem 3.1 As $c \downarrow 0$, the induced filtered-variate distribution of $\xi = \mathbf{Z}\alpha$, where $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$, approaches the singular distribution with unit mass at $\xi = \mathbf{Z}\,\mathbf{u}$. As $c \uparrow 1$, the limiting distribution is the discrete distribution of the random vector ξ^* on the set of column vectors of $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ in Δ^k, with each $P(\xi^* = \mathbf{z}_j) = u_j$, $j = 1, \ldots, n$.

The moments of $\xi = \mathbf{Z}\alpha$ must be the usual transform of the moments of α,

$$E\xi = \mathbf{Z}\,(E\alpha)$$

$$\mathrm{Var}\ \xi = \mathbf{Z}\,(\mathrm{Var}\ \alpha)\,\mathbf{Z}^T. \qquad (3.2)$$

By eqs. (3.1), (3.2) and Theorem 3.1, if $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$, the low-order moments of $\xi = \mathbf{Z}\alpha$ must be the following simple functions of the moments of ξ^*, or (weighted) empirical moments of the list of column vectors of $\mathbf{Z}$,

$$E\xi = E\xi^* = \Sigma u_j \mathbf{z}_j$$

$$\mathrm{Var}\,\xi = c\,\mathrm{Var}\,\xi^*$$

$$= c\ \Sigma u_j (\mathbf{z}_j - E\xi^*)(\mathbf{z}_j - E\xi^*)^T. \qquad (3.3)$$

This yields the following result, in the symmetric case $u_j \equiv 1/n$.

Theorem 3.2 The low order moments of a filtered-variate symmetric MSV distribution are simply expressed in terms of the empirical moments of the column vectors in the filter matrix. If $\xi = \mathbf{Z}\alpha$ and $\alpha \sim \mathrm{MSV}(\mathbf{u}, c)$ with $\mathbf{u} = n^{-1}\mathbf{1}_n$, then

$$E\xi = \bar{\mathbf{z}} \quad \text{and} \quad \mathrm{Var}\,\xi = c\,S_Z, \qquad (3.4)$$

where $\bar{\mathbf{z}} = n^{-1}\Sigma \mathbf{z}_j$ and $S_Z = n^{-1}\Sigma(\mathbf{z}_j - \bar{\mathbf{z}})(\mathbf{z}_j - \bar{\mathbf{z}})^T$.

4. THE DISTRIBUTION ESTIMATE

Given a sample $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, $k \times n$, and a proportionality constant $d \geq 1$, we define the set $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ of outward proportionally extended points,

$$\mathbf{z}_j = d(\mathbf{x}_j - \bar{\mathbf{x}}) + \bar{\mathbf{x}}, \qquad (4.1)$$

$j = 1, \ldots, n$. These will then satisfy

$$\bar{\mathbf{z}} = \bar{\mathbf{x}} \quad \text{and} \quad S_Z = d^2 S_X. \qquad (4.2)$$

Since the inverse transformation is a convex combination, $\mathbf{x}_j = d^{-1}\mathbf{z}_j + (1 - d^{-1})\,\bar{\mathbf{z}}$, $j = 1, \ldots, n$, the original data points are all contained within the convex hull of the new points, $\mathrm{Hull}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$. We let α have a symmetric mean-structured variance, $\alpha \sim \mathrm{MSV}(n^{-1}\mathbf{1}_n, c)$, and then we use α and the new points to define the filtered-variate distribution with random vector,

$$\xi = \mathbf{Z}\alpha = \alpha_1 \mathbf{z}_1 + \ldots + \alpha_n \mathbf{z}_n$$

$$= \bar{\mathbf{x}} + \alpha_1 d(\mathbf{x}_1 - \bar{\mathbf{x}}) + \ldots + \alpha_n d(\mathbf{x}_n - \bar{\mathbf{x}}), \qquad (4.3)$$

thus replacing the tentative definition (2.1). Our newly constructed distribution for ξ is restricted to $\mathrm{Hull}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$, and by (3.4) and (4.2), it has the moments

$$E\xi = \bar{\mathbf{x}} \quad \text{and} \quad \mathrm{Var}\,\xi = c\,d^2 S_X. \qquad (4.4)$$

An interesting choice of parameters is

$$c\,d^2 = 1, \qquad (4.5)$$

for which the variance of our distribution matches the sample variance, $\mathrm{Var}\,\xi = S_X$. This leaves only one parameter to tune in the distribution estimate, one extreme of which, $d \downarrow 1$, $c \uparrow 1$, gives again the sample empirical distribution. In the other extreme, $d \uparrow \infty$, $c \downarrow 0$, our distribution remains a distribution of continuous type, and by the following lemma, if $\mathrm{rank}(\mathbf{X} - \mathbf{x}_n \mathbf{1}_n^T) = k$, the distribution has the whole space R^k as its support.

Lemma 4.1 (Grunbaum 1967, p 3.) The dimension in R^k of the region $H = \mathrm{Hull}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is equal to $\mathrm{rank}(\mathbf{x}_1 - \mathbf{t}, \ldots, \mathbf{x}_n - \mathbf{t})$ where $\mathbf{t}$ is any point in H.

In the case of the symmetric Dirichlet distribution for the random weights, $\alpha \sim D(a\mathbf{1}_n)$, for which

$$c = (na + 1)^{-1}, \qquad (4.6)$$

the constructed distribution will be of continuous type on $\mathrm{Hull}(\mathbf{z}_1, \ldots, \mathbf{z}_n)$ if $0 < a < \infty$. The relation (4.6) explicitly shows the effect of sample size. If c and d are held fixed with increasing n, the parameter a must shrink to zero in the underlying Dirichlet distribution. We will take special interest, first, in fixed a, for which (4.5) yields $d = (na + 1)^{1/2}$, and an indefinitely expanding support set, $d \uparrow \infty$, $c \downarrow 0$, as $n \to \infty$.

In many problems, a sampling distribution is considered to have its support resticted to a subregion of the space R^k. But there is often a natural invertible transformation that maps the support onto the whole space. This method has the feature that a distribution estimate for such transformed data can yield, by a change of variable, a distribution on the subregion involved. If the data vectors y_j, $j = 1, \ldots, n$, have all their coordinates positive, then the method can be applied to their logarithmns $\mathbf{x}_j = \log(\mathbf{y}_j)$ (evaluated coordinate-by-coordinate), to produce a random vector ξ, which then yields a **log filtered-variate Dirichlet distribution** with random vector $\eta = \exp(\xi)$. Such a distribution for η is as easy to use in a simulation as a filtered-variate Dirichlet.

A Bayesian posterior predictive distribution is typically more diffuse than the sampling distribution. Consider that $\mathrm{Var}(x_{n+1} \mid x_1, \ldots, x_n) = E\,[\mathrm{Var}(x_{n+1} \mid \theta) \mid x_1, \ldots, x_n] + \mathrm{Var}\,[E\,(x_{n+1} \mid \theta) \mid x_1, \ldots, x_n]$. In the problem of constructing a predictive distribution, Morris Eaton has suggested privately that, unlike the problem of estimating a population distribution, one may wish to assert a more diffuse distribution than the sample empirical distribution, choosing $c\,d^2 = 1 + \varepsilon$, say, for which $\mathrm{Var}\,\xi = (1 + \varepsilon)\,\mathbf{S}_\mathbf{X}$.

5. THE COLLAPSED-HYPERVOLUME DISTRIBUTION

Clearly, a linear transformation of a filtered-variate distribution will again be filtered-variate with the same underlying random vector α. If $\xi = \mathbf{Z}\alpha$, then $B\,\xi = (B\,\mathbf{Z})\,\alpha$. In particular, a coordinate projection preserves the filter structure: if $\xi = (\xi^{(1)T}, \xi^{(2)T})^T$ and $\mathbf{Z} = (\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)})$ are partitioned conformably, and $\xi = \mathbf{Z}\alpha$, then $\xi^{(1)} = \mathbf{Z}^{(1)}\alpha$. Thus, in a very strong sense, the method is invariant to the dimensionality k of the sample vectors. The distribution estimate obtained directly from a few coordinates of the data vectors is exactly the same as the corresponding marginal distribution of the distribution estimate obtained from the full dimensional data vectors. Beyond the logical attraction of this feature, it will be important as a conceptual and computational aid in working with an interesting particular choice of distribution estimate.

Given a set of n points in R^k, $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$, such as an outward dilation of sample data by (4.1), we define the **collapsed-hypervolume distribution** as the distribution of the random k-dimensional vector $\xi = \mathbf{Z}\alpha$ where α has the uniform distribution on Δ^{n-1}. That is, α has the symmetric Dirichlet distribution with unit parameter $(a = 1)$, $\alpha \sim D(\mathbf{1}_n)$.

To motivate this definition, when $n \geq k + 1$, just imagine augmenting the matrix $\mathbf{Z}$ by appending $n\text{-}1\text{-}k$ further rows, if necessary, to obtain a matrix of $n\text{-}1$ rows, $\tilde{\mathbf{Z}} = (\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_n)$, for which $\mathrm{rank}(\tilde{\mathbf{Z}} - \tilde{\mathbf{z}}_n \mathbf{1}_n^T) = n\text{-}1$. Then, by Lemma 4.1, the region $\mathrm{Hull}(\tilde{\mathbf{z}}_1, \ldots, \tilde{\mathbf{z}}_n)$ is of full dimension in R^{n-1}, and $\tilde{\xi} = \tilde{\mathbf{Z}}\alpha$ is uniformly distributed over this polytope. So $\tilde{\xi}$ has an (n-1)-dimensional hypervolume distribution, and its subvector ξ, more generally, has a k-dimensional collapsed-hypervolume distribution. If, as is commonly assumed, the original data x_j's, and hence the z_j's, are of full rank, then it is possible, by the following lemma, to construct such augmented $\tilde{Z}$. For example, in one dimension ($k=1$), we require only that the data not have an n-way tie at a single numerical value.

Lemma 5.1 Assume $Z = (z_1, \ldots, z_n)$, $k \times n$, $n \geq k+1$, and $\text{rank}(Z - z_n 1_n^T) = k$. Without loss of generality, the n vectors can be reordered so that the first k vectors and the nth vector satisfy $\text{rank}(z_1 - z_n, \ldots, z_k - z_n) = k$. If $\tilde{Z} = (Z^T, J^T)^T$, where $J = (0_{h \times k}, I_h, 0_{h \times 1})$ with $h = n-1-k$, then $\text{rank}(\tilde{Z} - \tilde{z}_n 1_n^T) = n-1$.

An amazing feature of a collapsed-hypervolume distribution is that it depends in no way on which (n-1)-dimensional hypervolume is collapsed to form it. The distribution of $\xi = Z\alpha$ on $\text{Hull}(z_1, \ldots, z_n)$ is the marginal distribution of the uniform distribution on $\text{Hull}(\tilde{z}_1, \ldots, \tilde{z}_n)$ for any augmentation $\tilde{Z}$ of Z for which $\text{rank}(\tilde{Z} - \tilde{z}_n 1_n^T) = n-1$. Figure 1 illustrates a hypervolume collapse for one-dimensional data, $k=1$, consisting of $n=4$ data points including one tie. The solid body in $n-1 = 3$ dimensions is depicted relative to the first two axes in the plane of the page and the third axis rising perpendicularly from the page. It has $n = 4$ vertices, $\tilde{z}_1$, $\tilde{z}_2$, $\tilde{z}_3$, and $\tilde{z}_4$, which can be placed arbitrarily, except for their first coordinates , x_1, x_2, x_3, and x_4, which are fixed at 0, 2, 2, and 4, respectively. Alternative point values are shown for $\tilde{z}_4$. The collapsed-volume density plotted here is proportional to the cross-sectional area of the solid figure. It is easily seen to be the piecewise quadratic: $f(x) = (3/16) x^2$, if $0 \leq x \leq 2$, and $(3/16)(4-x)^2$, if $2 \leq x \leq 4$.

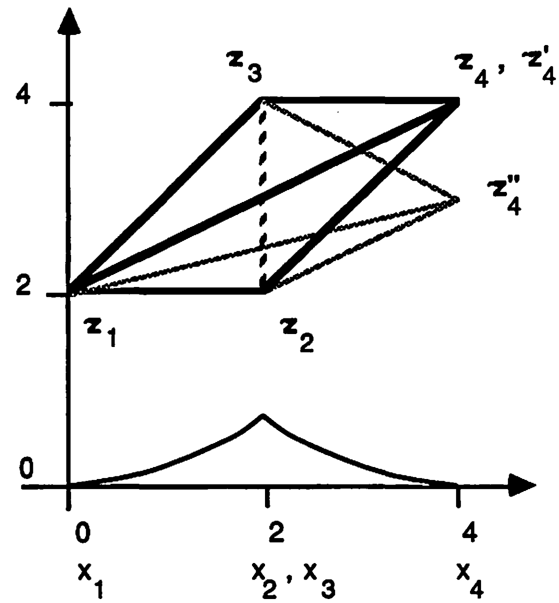

Figure 1. A collapsed-volume distribution, for $k = 1$ and $n = 4$. (The third axis rises perpendicularly from the plane of the page.)

6. METHODS FOR THE DENSITY

The distribution of a linear function of a Dirichlet vector has been studied since, at least, Bloch and Watson (1967) and interest continues in recent work of Jiang (1984) and others. For the extension to the distribution of a Stieltjes integral of a Dirichlet process, see, for example, Jiang (1988) and Cifarelli and Regazzini (1988). The density is especially easy in one particular case. This is the full-dimensional or uncollapsed case of $\xi = Z\alpha$, $\alpha \sim D(a)$, Z $k \times n$, where $k = n-1$ and $\text{rank}(Z - z_n) = n-1$. Then the filter transformation is invertible and the density of ξ can be obtained from the density of α, just by a simple substitution with constant Jacobian,

$$p_\xi(\xi) = p_{\alpha_1,\ldots,\alpha_{n-1}} [(\mathbf{z}_1-\mathbf{z}_n, \ldots, \mathbf{z}_{n-1}-\mathbf{z}_n)^{-1} \xi] \, / \, | \det(\mathbf{z}_1-\mathbf{z}_n, \ldots, \mathbf{z}_{n-1}-\mathbf{z}_n) | . \tag{6.1}$$

Although a value of n as small as k+1 seems rarely to be of interest *per se*, this density is useful in the general case as the initial function in two iterative methods: iterating by a collapse of the dimension k; or iterating by a growth of the sample size n. To collapse the dimension is just the usual marginalization operation from a joint density and seems not to have special properties in our setting. We give the iteration with respect to sample size in the following theorem.

Lemma 6.1 (Wilks 1962, pp 180-181.) Write, in terms of the scalers α_n and a_n,

$$\alpha^{(n)} = \begin{pmatrix}(1-\alpha_n)\alpha^{(n-1)} \\ \alpha_n\end{pmatrix},$$

$$\mathbf{a}^{(n)} = \begin{pmatrix}\mathbf{a}^{(n-1)} \\ a_n\end{pmatrix}. \tag{6.2}$$

Then $\alpha^{(n)} \sim D(\mathbf{a}^{(n)})$ if and only if

$$\alpha^{(n-1)} \sim D(\mathbf{a}^{(n-1)}),$$

$$\alpha_n \sim \text{beta}(a_n, a_1+\ldots+a_{n-1}), \tag{6.3}$$

and $\alpha^{(n-1)}$ and α_n are independent.

Theorem 6.2 If $\xi^{(n-1)} = (\mathbf{z}_1, \ldots, \mathbf{z}_{n-1})\, \alpha^{(n-1)}$ and $\xi^{(n)} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)\, \alpha^{(n)}$ where $\alpha^{(n-1)}$, α_n, and $\alpha^{(n)}$, satisfy the conditions of Lemma 6.1, then

$$\xi^{(n)} = (1-\alpha_n)\, \xi^{(n-1)} + \alpha_n \mathbf{z}_n, \tag{6.4}$$

where $\alpha_n \sim \text{beta}(a_n, a_1+\ldots+a_{n-1})$ independently of $\xi^{(n-1)}$. The densities of $\xi^{(n)}$ and $\xi^{(n-1)}$ are related by

$$p_{\xi^{(n)}}(\xi) = \int_0^1 p_{\xi^{(n-1)}} [(1-\alpha)^{-1}(\xi - \alpha\, \mathbf{z}_n)] \; b(\alpha;\, a_n,\, a_1+\ldots+a_{n-1})\, d\alpha. \tag{6.5}$$

The first factor in the integrand of (6.5) is nonzero only when its argument lies in Hull($\mathbf{z}_1, \ldots, \mathbf{z}_{n-1}$). This restriction can be difficult to handle for a moderate or large dimension k. But for $k = 1$, the range of ξ and α is

$$\alpha z_n + (1-\alpha) \min\{z_1, \ldots, z_{n-1}\} \le \xi \le \alpha z_n + (1-\alpha) \max\{z_1, \ldots, z_{n-1}\} \tag{6.6}$$

and $0 \le \alpha \le 1$.

REFERENCES

Blackwell, D. (1973), "Discreteness of Ferguson selection," *Annals of Statistics,* 1, 356-358.

Bloch, Daniel A., and Watson, Geoffrey S. (1967), "A Bayesian study of the multinomial distribution," *Annals of Mathematical Statistics,* 38, 1423-1434.

Cifarelli, Donato Michele, and Regazzini, Eugenio (1988), "Distribution functions of means of a Dirichlet process," Typescript, University of Milan.

Dickey, James M. (1968a), "Estimation of disease probabilities conditioned on symptom variables," *Mathematical Biosciences,* 3, 249-265.

__________ (1968b), "Smoothed estimates for multinomial cell probabilities," *Ann. Math. Statist.*, 39, 561-566.

__________ (1969), "Smoothing by cheating," *Ann. Mathl. Statist.* , Vol 40, No 4, 1477-82.

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

Good, I. J., and Gaskins, R. A. (1971), "Nonparametric roughness penalties for probability densities," *Biometrika* , 58, 255-277.

__________ (1980), "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data," *J. Amer. Statistical Assoc.* , 75, 42-56.

Grunbaum, Branko (1967), *Convex Polytopes* , New York: John Wiley, Interscience.

Hill, B. M. (1968), "Posterior distribution of percentiles: Bayes' theorem for sampling from a finite population," *J. Amer. Statistical Assoc.* 63, 677-691.

Hill, B. M. (1987a), "De Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference" (With discussion), To appear in *Bayesian Statistics 3,* J. M. Bernardo, M. H. DeGroot, D. V. Lindley, A. F. M. Smith, eds., Oxford: Oxford Univ. Press.

Hill, B. M. (1987b), "Parametric models for A_n: Splitting processes and mixtures," Manuscript, Dept. of Statistics, Univ. of Michigan, Ann Arbor, Mich.

Jiang, Jyh-Ming (1984), *Distributional properties of linear forms in a Dirichlet vector and applications,* PhD dissertation, Res. Rept. 7/84-#14, Dept. of Mathematics and Statistics, State Univ. of N. Y. at Albany.

__________ (1988), "Starlike functions and linear functions of a Dirichlet distributed vector," *SIAM J. Math. Anal.,* Vol 19, No 2 (March 1988), 390-397.

Lenk, P. J. (1984), *Bayesian nonparametric predictive distribution*, PhD dissertation, Dept. of Statistics, Univ. of Mich., Ann Arbor, Mich.

__________(1988), "The logistic normal distribution for Bayesian nonparametric predictive densities," *J. Amer. Statist. Assoc.* 83.

Leonard, T. (1973), "A Bayesian method for histograms," *Biometrika* . 60, 297-308.

__________ (1978), "Density estimation, stochastic processes, and prior information (with Discussion)," *J. Roy. Statist. Soc., Ser. B,* 40, 113-146.

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

Olkin, I., and Spiegelman, C. H. (1987), "A semiparametric approach to density estimation, *J. Amer. Statist. Assoc.* 82, 858-865.

Savage, L. J. (1964), Personal communication.

Titterington, D. Michael (1986), "A general view of statistical smoothing," *A.S.A. Proceedings of Business and Economic Statistics Section,* 79-84.

Whittle, P. (1958), "On the smoothing of probability density functions," *J. Roy. Statist. Soc.* 20, 334-343.

Wilks, S. S. (1962), *Mathematical Statistics,* New York: John Wiley.

__________ (1969), "Smoothing by cheating," *Ann. Mathl. Statist.* , Vol 40, No 4, 1477-82.

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

Good, I. J., and Gaskins, R. A. (1971), "Nonparametric roughness penalties for probability densities," *Biometrika* , 58, 255-277.

__________ (1980), "Density estimation and bump-hunting by the penalized likelihood method exemplified by scattering and meteorite data," *J. Amer. Statistical Assoc.* , 75, 42-56.

Grunbaum, Branko (1967), *Convex Polytopes* , New York: John Wiley, Interscience.

Hill, B. M. (1968), "Posterior distribution of percentiles: Bayes' theorem for sampling from a finite population," *J. Amer. Statistical Assoc.* 63, 677-691.

Hill, B. M. (1987a), "De Finetti's theorem, induction, and A_n, or Bayesian nonparametric predictive inference" (With discussion), To appear in *Bayesian Statistics 3,* J. M. Bernardo, M. H. DeGroot, D. V. Lindley, A. F. M. Smith, eds., Oxford: Oxford Univ. Press.

Hill, B. M. (1987b), "Parametric models for A_n: Splitting processes and mixtures," Manuscript, Dept. of Statistics, Univ. of Michigan, Ann Arbor, Mich.

Jiang, Jyh-Ming (1984), *Distributional properties of linear forms in a Dirichlet vector and applications,* PhD dissertation, Res. Rept. 7/84-#14, Dept. of Mathematics and Statistics, State Univ. of N. Y. at Albany.

__________ (1988), "Starlike functions and linear functions of a Dirichlet distributed vector," *SIAM J. Math. Anal.,* Vol 19, No 2 (March 1988), 390-397.

Lenk, P. J. (1984), *Bayesian nonparametric predictive distribution,* PhD dissertation, Dept. of Statistics, Univ. of Mich., Ann Arbor, Mich.

__________(1988), "The logistic normal distribution for Bayesian nonparametric predictive densities," *J. Amer. Statist. Assoc.* 83.

Leonard, T. (1973), "A Bayesian method for histograms," *Biometrika* . 60, 297-308.

__________ (1978), "Density estimation, stochastic processes, and prior information (with Discussion)," *J. Roy. Statist. Soc., Ser. B,* 40, 113-146.

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

Olkin, I., and Spiegelman, C. H. (1987), "A semiparametric approach to density estimation, *J. Amer. Statist. Assoc.* 82, 858-865.

Savage, L. J. (1964), Personal communication.

Titterington, D. Michael (1986), "A general view of statistical smoothing," *A.S.A. Proceedings of Business and Economic Statistics Section,* 79-84.

Whittle, P. (1958), "On the smoothing of probability density functions," *J. Roy. Statist. Soc.* 20, 334-343.

Wilks, S. S. (1962), *Mathematical Statistics,* New York: John Wiley.