empec, Vol. 13, 1988, page 209–222

Bayes Prediction Density and Regression Estimation – A Semiparametric Approach

By R. C. Tiwari[1], S. R. Jammalamadaka[2] and S. Chib[3]

Abstract: This paper is concerned with the Bayes estimation of an arbitrary multivariate density, $f(x)$, $x \in R^k$. Such an $f(x)$ may be represented as a mixture of a given parametric family of densities $\{h(x|\theta)\}$ with support in R^k, where θ (in R^d) is chosen according to a mixing distribution G. We consider the semiparametric Bayes approach in which G, in turn, is chosen according to a Dirichlet process prior with given parameter α. We then specialize these results when f is expressed as a mixture of multivariate normal densities $\phi(x|\mu, \Lambda)$ where μ is the mean vector and Λ is the precision matrix. The results are finally applied to estimating a regression parameter.

1 Introduction

In a recent paper, Ferguson (1983) presents a nonparametric Bayes procedure for estimating an arbitrary density $f(x)$ on the real line. This paper extends the results of Ferguson to the multivariate case and considers the estimation of the predictive density as well as the regression parameter. Consider a $k \times 1$ random vector $X = (Y, X_2, \ldots, X_k)'$ where, in the regression context, Y may be regarded as the dependent variable and $(X_2, \ldots, X_k)$ as the set of independent variables. We assume that X has an *unknown* density $f(x)$, $x \in R^k$. Such an $f(x)$ may be represented as a mixture of a multivariate normal densities $\{\phi(x|\mu, \Lambda)\}$, i.e.,

$$f(x) = \int \phi(x|\mu, \Lambda) dG(\mu, \Lambda), \tag{1}$$

where μ, the mean vector and Λ, the precision (or the inverse of the variance) matrix of the normal density, are chosen according to a mixing distribution G. Note that any

[1] R C. Tiwari, University of North Carolina, Charlotte, USA.
[2] S. R. Jammalamadaka, University of California, Santa Barbara, USA.
[3] S. Chib, University of Missouri, Columbia, USA.

0377–7332/88/3–4/209–222 $2.50 © 1988 Physica-Verlag, Heidelberg

distribution on R^k can be approximated by such a mixture to any preassigned accuracy in the Levy metric, and any density on R^k can be approximated similarly in the L_1 norm (cf. Ferguson 1983).

We consider the semiparametric Bayes approach, in which the unknown mixing distribution G, is chosen according to a Dirichlet prior with parameter α, say $D(\alpha)$. Our objective is to find the Bayes estimate $f_n(x)$, of the density of a future observation X_{n+1} given a random sample $x_1, \ldots, x_n$ from $f(x)$; that is, to find the semiparametric Bayes prediction density

$$f_n(x) = E[f(x) | x_1, \ldots, x_n], \tag{2}$$

where, in view of representation (1), the expectation in (2) is with respect to the posterior distribution of G given the sample $(x_1, \ldots, x_n)$ which is a mixture of Dirichlet processes (see, equation (11)).

We also consider the problem of finding the Bayes estimate of the parameter β that minimizes the mean-squared prediction error

$$E\left[Y_{n+1} - \sum_{i=2}^{k} \beta_i X_{i,n+1}\right]^2 \tag{3}$$

over all $\beta = (\beta_2, \ldots, \beta_k) \in R^{k-1}$. Note that (3) is minimized when β is

$$\beta^* = D^{-1} a, \tag{4}$$

where

$$D_{(k-1)\times(k-1)} = ((d_{ij})) = ((E(X_{i,n+1} \cdot X_{j,n+1})))$$

and

$$a_{(k-1)\times 1} = (a_2, \ldots, a_k), \quad \text{with } a_i = E(X_{i,n+1} Y_{n+1}).$$

As stated in Poli (1985) (see also Tiwari, Chib and Jammalamadaka 1988), "the achieved estimate β^* provides the best linear prediction of Y_{n+1} in terms of $(X_{2,n+1}, \ldots, X_{k,n+1})$".

This paper is organized as follows. Section 2 contains preliminaries and some general results which are then specialized to the case of normal mixtures in Section 3. Section 4 contains the Bayes estimate of β^*, while the last section includes some comments on the computational aspects.

2 Preliminaries and General Results

This section provides the basic definitions and results that will be used in the sequel. Let X be a $k \times 1$ random vector. Then X has a k-variate normal distribution with mean vector μ and precision (the inverse of variance) matrix Λ, denoted by $X \sim N_k(\mu, \Lambda)$, if its pdf is given by

$$\phi(x|\mu, \Lambda) = (2\pi)^{-k/2} |\Lambda|^{1/2} \exp\{-(1/2)(x-\mu)'\Lambda(x-\mu)\}, \tag{5}$$

where $\mu \in R^k$, and Λ is a symmetric positive definite (s.p.d.) matrix of order k.

A $k \times k$ random matrix Λ has a Wishart distribution if its pdf is given by

$$f(\Lambda|\Lambda^*, \nu) = c \cdot |\Lambda^*|^{-\nu/2} |\Lambda|^{(\nu-k-1)/2} \cdot \exp\{-(1/2)\ \mathrm{tr}\,(\Lambda\Lambda^{*-1})\}, \tag{6}$$

where Λ^* is a scale matrix of order k, $\nu \geqslant k$ is the degrees of freedom, and c is the normalizing constant given by

$$c^{-1} = 2^{k\nu/2}\pi^{k(k-1)/4} \sum_{j=1}^{k} \Gamma(\nu + 1 - j/2).$$

We shall use the notation $\Lambda \sim W_k(\Lambda^*, \nu)$ to denote that Λ has the pdf given by (6).

The $k \times 1$ vector X has a multivariate Student's t-distribution if its density function is given by

$$f(x|\mu, \Lambda, \nu) = (\nu\pi)^{-k/2} \cdot \frac{\Gamma((\nu+k)/2)}{\Gamma(\nu/2)} \cdot |\Lambda|^{1/2}[1 + \nu^{-1}(x-\mu)'\Lambda(x-\mu)]^{-(k+\nu)/2} \tag{7}$$

where $\mu \in R^k$, $\nu > 0$ is the degrees of freedom and Λ is a s.p.d. matrix of order k. We use the notation $X \sim MVt_k(\mu, \Lambda, \nu)$ to denote that X has the pdf given by (7).

Let $\alpha(\cdot)$ be a finite non-null finitely additive measure on (R^d, R^d). A random probability measure P on (R^d, R^d) is a Dirichlet process with parameter α, and write $P \in D(\alpha)$, if for every finite s and every measurable partition $A_1, \ldots, A_s$ of R^d, the random variables $(P(A_1, \ldots, P(A_s))$ have the Dirichlet distribution with parameters $(\alpha(A_1), \ldots, \alpha(A_s))$ (cf. Ferguson 1973).

Let δ_θ represent the degenerate probability measure at a single point θ. Let G be the distribution function associated with the random probability measure P. Then, under $D(\alpha)$, G can expressed as (cf. Sethuraman and Tiwari 1982):

$$G = \sum_{i=1}^{\infty} p_i \delta_{\theta_i}, \tag{8}$$

where

(i) $\theta_1, \theta_2, \ldots$, are iid on (R^d, R^d) with the common distribution $G_0 = \alpha(\cdot)/M$,

(ii) $(p_1, p_2, \ldots)$ and $(\theta_1, \theta_2, \ldots)$ are independent,

(iii) $q_1 = p_1, q_2 = p_2/(1-p_1), q_3 = p_3/(1-p_1-p_2), \ldots$ are iid Beta $(1, M)$, $M = \alpha(R^d)$.

More generally than (1), one may assume that the unknown density $\psi(x)$ is expressed as a mixture of a family of k-variate densities $\{h(x|\theta)\}$, with the mixing distribution G (on θ) in R^d, i.e.,

$$\psi(x) = \int h(x|\theta) dG(\theta). \tag{9}$$

If this mixing distribution G, is assumed to have a Dirichlet prior, then from (8)

$$\psi(x) = \sum_{i=1}^{\infty} p_i h(x|\theta_i). \tag{10}$$

Let $x_1, \ldots, x_n$ be a random sample from $\psi(x)$ given by (9). This is equivalent to first choosing $\theta_1, \ldots, \theta_n$ i.i.d. from $G_0(\theta)$, and then x_i from $h(x|\theta_i)$, $i = 1, \ldots, n$ independently. Then the posterior distribution of G given $x_1, \ldots, x_n$ is a mixture of Dirichlet processes (cf. Antoniak 1974)

$$G|x_1, \ldots x_n \in \int D\left(\alpha + \sum_{i=1}^{n} \delta_{\theta_i}\right) dH(\theta_1, \ldots, \theta_n | x_1, \ldots, x_n), \tag{11}$$

where $dH(\theta_1, \ldots, \theta_n | x_1, \ldots, x_n)$, the posterior density of $\theta_1, \ldots, \theta_n$ given $x_1, \ldots, x_n$, is

$$dH(\theta_1, \ldots, \theta_n | x_1, \ldots, x_n) \propto \left(\prod_{i=1}^{n} h(x_i|\theta_i) \right) \prod_{i=1}^{n} d\left(\alpha + \sum_{j=1}^{i-1} \delta_{\theta_j} \right)(\theta_i)/M^{(n)}$$

with the notation $M^{(n)} = M(M+1) \ldots (M+n-1)$. From (11) we have

$$E(G(\theta)|x_1, \ldots, x_n) = \frac{M}{M+n} G_0(\theta) + \frac{n}{M+n} \int \hat{G}_n(\theta) dH(\theta_1, \ldots, \theta_n | x_1, \ldots, x_n) \tag{12}$$

where $\hat{G}_n(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta_i}$, is the empirical measure of the observed $\theta_1, \ldots, \theta_n$ and G_0 corresponds to the normalized α-measure. Consequently we have the following:

Theorem 1: The Bayes estimator of $\psi(x)$ under squared error loss function, $\psi_n(x) = E[\psi(x)|x_1, \ldots, x_n]$, is given by

$$\psi_n(x) = \frac{M}{M+n} \psi_0(x) + \frac{n}{M+n} \hat{\psi}_n(x) \tag{13}$$

where $\psi_0(x)$, the estimate of $\psi(x)$ for no sample size, is given by

$$\psi_0(x) = E\psi(x) = \int h(x|\theta) dG_0(\theta) \tag{14}$$

and

$$\hat{\psi}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \int \ldots \int h(x|\theta_i) dH(\theta_1, \ldots, \theta_n | x_1, \ldots, x_n). \tag{15}$$

The nonparametric Bayes estimate of $\psi(x)$ is, therefore, seen to be a weighted average of the prior guess $\psi_0(x)$ given in (14), and $\hat{\psi}_n(x)$ given in (15). Two special cases of interest as M, the strength in the prior goes to zero and infinity, may be considered as in Ferguson (1983).

In particular, as $M \to \infty$, the θ_i's are all distinct and the prediction density is given by

$$\psi^*(x) = \frac{1}{n} \sum_{i=1}^{n} \psi(x|x_i)\,, \tag{16}$$

where

$$\psi(x|x_i) = \frac{\int h(x|\theta)h(x_i|\theta)dG_0(\theta)}{\int h(x_i|\theta)dG_0(\theta)}$$

is the Bayes prediction density of x given the one observation x_i.

3 Results for Normal Mixtures

In this section we specialize the results of the previous section by letting $\theta = (\mu, \Lambda)$ and $h(x|\theta)$ be a multivariate normal density with mean μ and the precision matrix Λ. We also let G_0 be the joint prior distribution of (μ, Λ) given by

$$\mu|\Lambda \sim N_k(\mu^*, \Lambda b^*), \quad b^* > 0 \tag{17}$$

and

$$\Lambda \sim W_k(\Lambda^*, \nu^*). \tag{18}$$

Now, let the unknown density $f(x)$, be a random mixture of a multivariate normal densities as in (1) i.e.,

$$f(x) = \int \phi(x|\mu, \Lambda)dG(\mu, \Lambda) \tag{19}$$

where $\phi(\cdot|\mu, \Lambda)$ is the pdf in (5).

In (19), consider the special choice of a Dirichlet process prior for G with parameter $\alpha = MG_0$, where G_0 is the natural conjugate prior of (μ, Λ) given by (17) and (18) namely the normal-Wishart. It is important to note that nothing prevents us from choosing an arbitrary measure G_0, as the prior of θ.

Lemma 1: Let (μ, Λ) have a joint normal-Wishart prior given by (17) and (18). Then the prior guess of the prediction density at x, $f_0(x)$, is given by (see equation (14))

$$f_0(x) \alpha \left(1 + \frac{b^*}{b^*+1}(x - \mu^*)' \Lambda^* (x - \mu^*)\right)^{-(k+\nu^* - k^* + 1)/2} \tag{20}$$

a k-variate MVt density with mean vector μ^*, precision matrix $\phi_0 = (v^* - k + 1)\Lambda^* b^*/(b^* + 1)$, and $\nu_0 = (\nu^* - k + 1)$ degrees of freedom.

The proof of Lemma 1 is given in the Appendix.

Using Theorem 1 and Lemma 1, we are now able to provide the Bayes prediction density of a future observation X_{n+1}, given $x_1, \ldots, x_n$.

Theorem 2: Given $f(x) = \int \phi(x|\mu, \Lambda) dG(\mu, \Lambda)$, with $G \in D(\alpha)$, and $\alpha = MG_0$, where G_0 is the normal-Wishart prior specified in (17) and (18), the Bayes prediction density of X_{n+1} given $x_1, \ldots, x_n$ is

$$f_n(x) = \frac{M}{M+n} f_0(x) + \frac{n}{M+n} \hat{f}_n(x), \tag{21}$$

where $f_0(x)$ is the MVt density given in (20) and

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \int \ldots \int \phi(x|\mu_i, \Lambda_i) dH((\mu_1, \Lambda_1), \ldots, (\mu_n, \Lambda_n)|x_1, \ldots, x_n). \tag{22}$$

The Bayes prediction density is therefore a weighted mixture of a multivariate-t density $f_0(x)$ and $\hat{f}_n(x)$, with the weights $M/(M + n)$ and $n/(M + n)$, respectively. The density $\hat{f}_n(x)$ can be evaluated numerically using the results of Section 5.

Two special cases which do not require a numerical evaluation are given next. The following Theorem 3 for $M \to 0$ yields the usual parametric result in which (μ, Λ) has the normal-Wishart prior and $x_1, \ldots, x_n$ is a random sample from $\phi(x|\mu, \Lambda)$.

Theorem 3: Let $x_i|\mu, \Lambda \sim N_k(\mu, \Lambda)$ $i = 1, \ldots, n$, $\mu|\Lambda \sim N_k(\mu^*, \Lambda^*)$, and $\Lambda \sim W_k(\Lambda^*, \nu^*)$. Then as $M \to 0$, then density in (21) becomes

$$f_n^0(x) \propto [1 + [1 + (x - \mu^{**})'\phi^{**}(x - \mu^{**})]^{-(k+n+\nu^*-k+1)/2}, \quad (23)$$

where $\bar{x} = \sum_{i=1}^{n} x_i/n$ and $S = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$. □

A proof of this theorem is included in the Appendix. Note that $f_n^0(x)$ in (23) is a k-variate MVt density with mean vector $\mu^{**} = (b^*\mu^* + n\bar{x})/(b^* + n)$, precision matrix $\phi^{**} = (b^* + n)[nb^*(b^* + n)^{-1}(\bar{x} - \mu^*)(\bar{x} - \mu^*)' + \Lambda^{*-1} + S]^{-1}/(b^* + n + 1)$ and $n + \nu^* - k + 1$ degrees of freedom. Setting $n = 1$ and $\bar{x} = x_i$ in Theorem 3 gives the Bayes prediction density of X_{n+1} given one observation $x_i (i = 1, \ldots, n)$:

Corollary 1: Let $x_i|\mu, \Lambda \sim N_k(\mu, \Lambda)$, $\mu|\Lambda \sim N_k(\mu^*, b^*\Lambda)$ and $\Lambda \sim W_k(\Lambda^*, \nu^*)$. Then as $M \to 0$, the Bayes prediction density of X_{n+1} given x_i is

$$f_{1,i}(x) \propto [1 + (x - \mu_i^\dagger)'\psi_i^\dagger(x - \mu_i^\dagger)]^{-(k+\nu-k+2)/2}$$

a k-variate multivariate-t pdf with mean vector $\mu_i^\dagger = (b^*\mu^* + x_i)/(b^* + 1)$ precision matrix $\psi_i^\dagger = (b^* + 1)[b^*(b^* + 1)^{-1}(x_i - \mu^*)(x_i - \mu^*)' + \Lambda^{*-1}]^{-1}/(b^* + 2)$ and $\nu - k + 2$ degrees of freedom.

The second special case, as $M \to \infty$, is covered by the following result which follows from Corollary 1 and (16).

Corollary 2: Let $x_i|\mu, \Lambda \sim N_k(\mu, \Lambda)$ $i = 1, \ldots, n$, $\mu|\Lambda \sim N_k(\mu^*, b^*\Lambda)$, and $\Lambda \sim W_k(\Lambda^*, \nu^*)$ then as $M \to \infty$ the Bayes prediction density of X_{n+1} given $x_1, \ldots, x_n, f_n^\infty(x)$, is given by a finite mixture of multivariate-t densities i.e.,

$$f_n^\infty(x) \propto \frac{1}{n} \sum_{i=1}^{n} [1 + (x_i - \mu_i^\dagger)'\psi_i^\dagger(x_i - \mu_i)]^{-(k+\nu-k+2)/2}.$$

4 Bayes Estimation of a Regression Parameter

The results obtained in the previous section allow us to find the estimate of β^* given in (4), where β^* minimizes the prediction mean-squared error that is specified in (3). Although the estimate of β^* using the Bayes prediction density in Theorem 2 cannot be computed in a closed form, the general principle can be illustrated with the following cases.

From (3) we have

$$\beta^* = D^{-1}a,$$

where for $1 \leqslant i, j \leqslant k-1$, the typical elements of D and a are

$$d_{ij} = E(X_{i,n+1} \cdot X_{j,n+1})$$

and

$$a_i = E(X_{i,n+1} \cdot Y_{n+1}).$$

For an f which is a normal mixture, for the no-sample case, the estimate of β^* can be computed using the following. Let $\mu^* = (\mu_1^*, \ldots, \mu_k^*)'$ and let $\Lambda^{*-1} = ((\Sigma_{ll'}^*))$, $1 \leqslant l$, $l' \leqslant k$, then the i, j-th element of D is

$$d_{ij} = (b^*(\nu^* - k - 1)/(b^* + 1))^{-1} \Sigma_{i+1,j+1}^* + \mu_{i+1}^* \mu_{j+1}^*, \quad 1 \leqslant i, j \leqslant k-1$$

and the j-th element of a is

$$a_j = (b^*(\nu^* - k - 1)/(b^* + 1)^{-1} \Sigma_{1,j+1}^* + \mu_1^* \mu_{j+1}^*, \quad 1 \leqslant j \leqslant k-1.$$

A similar procedure can be used to compute the estimate of β^* when $M \to 0$. Letting $\Lambda^{**-1} = ((\Sigma_{ll'}^{**}))$, $1 \leqslant l$, $l' \leqslant k$, and μ_l^{**} denote the l-th element of μ^{**} we have

$$d_{ij} = (n + \nu^* - k - 1)^{-1} \Sigma_{i+1,j+1}^{**} + \mu_{i+1}^{**} \mu_{j+1}^{**}$$

and

$$a_j = (n + \nu^* - k - 1)^{-1} \Sigma^{**}_{1,j+1} + \mu_1^{**} \mu_{j+1}^{**}, \quad 1 \leqslant j \leqslant k-1.$$

Finally, for the $M \to \infty$ case, d_{ij}'s and a_j's can again be given explicitly. The details are ommitted.

5 Remarks and Computational Aspects

The usual issues in density estimation regarding the kernel and the window-length could be related to the choice of the prior α in our Bayes set-up, although the specifies need further investigation. In particular, the special form (16) corresponds to a variable kernel estimate, as Ferguson notes. Computation of (13), the Bayes estimator of the density can be done along the lines of Ferguson (1983), which contains an illustration for density on R^1. If we define

$$H(x_1, \ldots, x_n) = \int \ldots \int \left[\prod_{i=1}^{n} h(x_i|\theta_i) \right] \prod_{i=1}^{n} d\left(MG_0 + \prod_{j=1}^{i-1} \delta_{\theta_j} \right)(\theta_i)/M^{(n)}$$

then Lo (1978) provides the following representation of the function H

$$dH(\theta_1, \ldots, \theta_n | x_1, \ldots, x_n) = \frac{\left[\prod_{i=1}^{n} h(x_i|\theta_i) \right] \prod_{i=1}^{n} d\left(MG_0 + \prod_{j=1}^{i-1} \delta_{\theta_j} \right)(\theta_i)}{M^{(n)} \cdot h(x_1, \ldots, x_n)}$$

Using this, one can rewrite the expression for $\psi_{n,\alpha}(x)$ in (13) in terms of the function $h(\cdot)$ as

$$\psi_{n,\alpha}(x) = \frac{h(x, x_1, \ldots, x_n)}{h(x_1, \ldots, x_n)} \tag{24}$$

The computation of $\psi_{n,\alpha}(x)$ clearly depends on the evaluation of the ratio in equation (24).

If we expand the product measure which appears in equation (24), there are $n!$ terms and each term of the expansion determines a partition $Q = \{K_1, \ldots, K_n\}$ of the data set $\{x_1, \ldots, x_n\}$ with the property that $\theta_i = \theta_j$ if and only if $x_i \in K$, $x_j \in K$, for some set K in Q. Hence, we can write $h(x_1, \ldots, x_n)$ as

$$h(x_1, \ldots, x_n) = \sum_{Q} P_M(Q) Z(Q), \tag{25}$$

where

$$Z(Q) = \prod_{K \in Q} \int \prod_{x_i \in K} h(x_i | \theta) dG_0(\theta) \tag{26}$$

and $P_M(Q)$ is the probability of selecting a particular partition, Q. Define

$$Y(Q) = \frac{Z(Q)}{n} \sum_{K \in Q} |K| \frac{\int h(x|\theta) \prod_{x_i \in K} h(x_i|\theta) dG_0(\theta)}{\int \prod_{x_i \in K} h(x_i|\theta) dG_0(\theta)} \tag{27}$$

where $|K|$ is the cardinality of the set K. For the specific choice of G_0 that we use in (17) and (18), we can simplify the expression for $Z(Q)$ and $Y(Q)$ given in (26) and (27), respectively.

Given $x_1, \ldots, x_n$ the Monte Carlo procedure entails the following steps:

(i) *Select a partition:* This is done by using Kuo (1986)'s method. Start the first set of partitions with x_1, say. Then, for $i = 1, 2, \ldots, n-1$, x_{i+1} starts a new set with probability $\frac{M}{M+i}$; otherwise it is placed in an existing set with probability $\left(\frac{r}{M+i}\right)$, where r is the number of elements already in that set. In the computations, we need only to record the number of the sets in a partition, and the indices in each class, and for this partitioning process, one may use the indices 1 through n and not the data themselves.

(ii) *Estimating* $\psi_{n,\alpha}$: Once a particular partition Q_i is randomly chosen, compute $Z(Q_i)$ and $Y(Q_i)$ using the equations (26) and (27). This process is replicated N times to give $Z(Q_i)$ and $Y(Q_i)$, $1 \leqslant i \leqslant N$, and the Monte Carlo estimate of $\hat{\psi}_n(x)$ in (13) is given by

$$\tilde{\hat{\psi}}_n(x) = \sum_{i=1}^{N} Y(Q_i) / \sum_{i=1}^{N} Z(Q_i).$$

The estimate of $\psi_{n,\alpha}(x)$ is then computed using (13). The variance of this estimator can be computed using the asymptotic formula for the variance of the ratio of means (see, Cochran 1977, p. 155):

$$\text{Var}\left(\frac{\Sigma Y_i}{\Sigma Z_i}\right) = \text{Var}\left(\frac{\bar{Y}}{\bar{Z}}\right) = \frac{1}{N\mu_z^2}\left[\sigma_y^2 - 2\sigma_{yz}\frac{\mu_y}{\mu_z} + \sigma_z^2\frac{\mu_y^2}{\mu_z^2}\right]$$

where the estimates

$$\bar{Z}(Q) = \hat{\mu}_z = \sum_{i=1}^{N} Z(Q_i)/N, \quad \bar{Y}(Q) = \hat{\mu}_y = \sum_{i=1}^{N} Y(Q_i)/N,$$

$$\hat{\sigma}_z^2 = [N(N-1)]^{-1} \sum_{i=1}^{N} (Z(Q_i) - \bar{Z}(Q))^2,$$

$$\hat{\sigma}_y^2 = [N(N-1)]^{-1} \sum_{i=1}^{N} (Y(Q_i) - \bar{Y}(Q))^2,$$

and

$$\hat{\sigma}_{yz} = [N(N-1)]^{-1}[\Sigma Z(Q_i)Y(Q_i) - N\bar{Z}(Q) \cdot \bar{Y}(Q)]$$

are used in place of the corresponding parameters.

Appendix

Proof of Lemma 1: By using the definition in (14), we have that

$$f_0(x) = \int \phi(x|\mu, \Lambda) dG_0(\mu, \Lambda)$$

$$\propto \int |\Lambda|^{(\nu^* - k)/2} e^{-1/2} \text{ tr } \Lambda\left[(x - \mu^*)(x - \mu^*)' \frac{b^*}{b^* + 1} + \Lambda^{*-1}\right] d\Lambda$$

$$\propto |(x - \mu^*)(x - \mu^*)' \frac{b^*}{b^* + 1} + \Lambda^{*-1}|^{(\nu^* + 1)/2}$$

On using the normalizing constant of the Wishart distribution. Now using the result that if V is a nonsingular $(p \times p)$ matrix, a and $b \in R^p$ (cf. Press 1982, p. 20), then

$$|V + ab'| = |V|[1 + b'V^{-1}a],$$

we get that

$$f_0(x) \propto \left[1 + \frac{b^*}{b^* + 1}(x - \mu^*)'\Lambda^*(x - \mu^*)\right]^{-(\nu^* + 1)/2},$$

This completes the proof. □

Proof of Theorem 3: Notice that $f_n^0(x)$ is the expectation of $\phi(x|\mu, \Lambda)$ w.r.t. the posterior density of (μ, Λ) given $x_1, \ldots, x_n$. This posterior density is well known (e.g., see Press 1982, p. 187) and is given by

$$dG_0((\mu, \Lambda)|x_1, \ldots, x_n) \propto dG_0(\mu, \Lambda) \cdot f(x_1, \ldots, x_n|\mu, \Lambda)$$

$$\propto |\Lambda|^{1/2} \exp\left\{-\frac{1}{2}(\mu - \mu^*)'\Lambda b^*(\mu - \mu^*\right\} \times$$

$$\exp\left\{-\frac{1}{2}(\mu - \bar{x})'n\Lambda(\mu - \bar{x})\right\} \cdot |\Lambda|^{(n+\nu^*-k-1)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\Lambda[S + \Lambda^{*-1}]\right\}$$

from which we get that

$$\mu|x_1, \ldots, x_n, \Lambda \sim N_k(\mu^{**}, (\Lambda b^* + n\Lambda)), \tag{A.1}$$

and

$$\Lambda|x_1, \ldots, x_n, \Lambda^* \sim W_k((S + \Lambda^{*-1} + nb^*(b^* + n)^{-1}(\bar{x} - \mu^*)(\bar{x} - \mu^*)')^{-1}, n + \nu^*). \tag{A.2}$$

Using (A.1) and (A.2) it follows that

$$X_{n+1}|x_1, \ldots, x_n, \Lambda \sim N_k(\mu^{**}, \Lambda(b^* + n)/(b^* + n + 1)). \tag{A.3}$$

Let the pdf of in (A.3) be denoted by $f_n^0(\cdot|x_1, \ldots, x_n, \Lambda)$. Then, from (A.2) and (A.3) the Bayes prediction density of x_{n+1} is

$$f_n^0(x) = \int f_n^0(x|x_1, \ldots, x_n, \Lambda) \cdot W_k(d\Lambda|x_1, x_2, \ldots, x_n)$$

$$\alpha \int_\Lambda |\Lambda|^{1/2} \exp\left(-\frac{1}{2} \operatorname{tr} \Lambda[(b^* + n)(b^* + n + 1)^{-1}(x - \mu^{**})(x - \mu^{**})'\right)$$

$$\cdot |\Lambda|^{(n+\nu^* - k - 1)/2} \exp\left(-\frac{1}{2} \operatorname{tr} \Lambda[S + \Lambda^{*-1} + nb^*(b^* + n)^{-1}(\bar{x} - \mu^*)(x - \mu^*)']\right) d\Lambda$$

from which the result follows. □

References

Antoniak CE (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann Statist 2:1152–1174

Cochran WG (1977) Sampling techniques. John Wiley & Sons, New York

Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. Ann Statist 1:209–230

Ferguson TS (1983) Bayesian density estimation by mixtures of normal distributions. In: Rizvi MH, Rustagi JS, Siegmund (eds) Recent advances in statistics. Academic Press

Kuo L (1986) Computations of mixture of Dirichlet processes. SIAM J Scientific and Statistical Computing 7:60–71

Lo AY (1978) On a class of Bayesian nonparametric estimates: density estimates. Rutgers University, New Brunswick, NJ

Poli I (1985) A Bayesian nonparametric estimator for multivariate regression. Journal of Econometrics 28:171–182

Prakasa Rao BLS (1983) Nonparametric funvtional estimation. Academic Press

Press SJ (1982) Applied multivariate analysis: using Bayesian and frequentist methods of inference. Robert E. Krieger Publishing Co

Sethuraman J, Tiwari RC (1982) Convergence of Dirichlet measures and the interpretations of their parameter. In: Gupta SS, Berger JO (eds) Statistical decision theory and related topics, III, vol 2. Academic Press

Tiwari RC, Chib S, Jammalamadaka S Rao (1988) Nonparametric Bayes estimate for multivariate regression with a Dirichlet invariant prior. In: Ali Festschrift: Forthcoming

Wilks SS (1962) Mathematical statistics, 2nd ed. John Wiley, New York

Copyright of Empirical Economics is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.