JIRSS (2008)
Vol. 7, Nos. 1-2, pp 23-33

Inferences about a Mean Vector under a Mixture of Dirichlet Process Model

Amin Ghalamfarsa Mostofi[1], **Javad Behboodian**[2]

[1]Department of Statistics, Shiraz University, Iran. (ghalamfarsa@susc.ac.ir)
[2]Department of Mathematics, Islamic Azad University, Shiraz, Iran. (behboodian@susc.ac.ir)

Abstract. We study the problem of testing the hypothesis that the mean vector of a random vector belongs to a given set. For this purpose, we consider a semiparametric mixture of Dirichlet process model in which the mean vector has a prior distribution concentrated on the set of interest. A computational method is given to obtain a sample from the posterior distribution of the mean vector. On the basis of this sample, we can obtain the Bayes estimate and the posterior probability that the hypothesis is true. We give a numerical example to demonstrate the application of this method.

1 Introduction

In recent years, the mixture of Dirichlet process (MDP) models has played an important role in Bayesian nonparametric and semiparametric inference. In a MDP model, the probability density function is written as a mixture of known nonnegative measurable functions

Key words and phrases: 0-1 loss function, Bayesian inference, Dirichlet process, Gibbs sampler, mean vector.

and then a Dirichlet process prior is assigned to the mixing distribution. Unlike the Dirichlet process prior, these models provide a class of priors that selects a continuous distribution with probability 1. Lo (1984) obtained the Bayes estimator of the density function and its functionals in a MDP model. Then, Brunner and Lo (1989) applied the results to an unknown unimodal density (by the Khintchine and Shepp theorem (Feller, 1971), a unimodal density can be written as a scale mixture of uniform densities). Brunner (1995) then gave a generalization to a linear regression model with symmetric unimodal error terms. A Dirichlet mixture of normals is another example of the MDP model, which has a normal density function as its kernel, (e.g., Ferguson (1983), West (1990), Escobar (1994)), and Escobar and West (1995)) Escobar (1994) applied a computational approach, based on a Gibbs sampler algorithm, to find the Bayes estimate of normal means in this model. He also considered the case when a prior distribution is assigned to the parameters of the Dirichlet process. Escobar and West (1995) studied Bayesian density estimation using a Dirichlet mixture of normal densities. Consistency of the posterior of such models is discussed by Ghosal, et al. (1999) and Lijoi et al. (2005). Using a semiparametric MDP prior, Merrick et al. (2003) proposed a proportional hazards model to describe the relationship between machine tool life and operational variables. Further applications can be found in Mukhopadhyay and Gelfand (1997 and references therein), Kottas and Gelfand (2001), Hoff (2004), Kottas et al. (2004) and Lijoi, et al. (2005). See also Walker et al. (1999) and Muller and Quintana (2004) for a review of Bayesian nonparametric inference, and MacEachern and Muller (2000) for other surveys of MDP models.

The main contribution of this article is to review Dirichlet Process mixing and show its application in testing hypothesis about a mean vector. Let μ be the mean vector of a random vector $\mathrm{Y} = (Y_1, ..., Y_m)$ and let A denote some subset of R^m . We wish to test whether or not the sample is compatible with the hypothesis $H_0 : \mu \in A$.

Here we review Dirichlet process mixing and study a related computational procedure. Let $k(.,.) : R^m \times R^1 \to R^1$ be a known nonnegative valued and measurable function with the following property:

$$\int_{R^m} k(\mathrm{y}, \theta)\, d\mathrm{y} = 1 \qquad \forall\theta \in R \, . \tag{1.1}$$

Given the distribution function $G(.)$, we suppose that the random vector Y has the probability density function:

$$f_{\mathrm{Y}}(\mathrm{y}|G) = \int_R k(\mathrm{y}, \theta)\, G(d\theta). \tag{1.2}$$

We further assume a Dirichlet process prior for $G(\,.)$, $G \in DP(cG_0)$, where G_0 is a specified distribution function and $c > 0$ is a precision parameter (see Ferguson, 1973). In fact, the distribution $G(\,.)$ is the infinite dimensional parameter of the model. The Bayes estimate of (1-2) for the no sample problem is

$$E(f_{\mathrm{Y}}(\mathrm{y}|G)) = f_{\mathrm{Y}}(\mathrm{y}|G_0). \tag{1.3}$$

For a random sample $D = (\mathrm{Y}_1, ..., \mathrm{Y}_n)$ from $f_{\mathrm{Y}}(\,.\,|G)$, we can apply the results given by Lo (1984) to obtain the Bayes estimator of a functional $H(f_{\mathrm{Y}}(.|G))$. Since Lo's expressions are intractable, some authors provide approximation methods to compute posterior moments of functionals. See e.g., Escobar (1994), Gelfand and Mukhopadhyay (1995), Escobar and West (1995), Ishwaran and Zarepour (2000), Gelfand and Kottas (2002) and Ishwaran and Zarepour (2002).

Here, we briefly review the work of Gelfand and Kottas (2002). For this purpose, we restate the MDP model as follows

$$\begin{cases} \mathrm{Y}_i|\, G, \theta_1, ..., \theta_n \;\sim\; k(., \theta_i), \;\; i = 1, ..., n \;\; (indep.) \\ \theta_1, ..., \theta_n|G \;\sim\; G(.) \qquad\qquad (i.i.d.) \\ G \;\in\; D\,P(cG_0) \end{cases} \tag{1.4}$$

Then, $\mathrm{Y}_1, ..., \mathrm{Y}_n$ are $i.i.d.$ random vectors from $f_{\mathrm{Y}}(\,.\,|G)$. In what follows, we use the bracket notation of Gelfand and Smith (1990) to write the distribution of random variables. For example the distribution of the random variable Z at z is denoted by $[z]$ and given the value w of a random variable W, the conditional distribution of Z is denoted by $[z|w]$. Then,

$$[\theta_i|\theta_j; j \neq i, D] \propto \prod_{i=1}^{n} k(\mathrm{y}_i, \theta_i)\, [\theta_i|\theta_j; j \neq i], \tag{1.5}$$

where,

$$[\theta_i|\theta_j, j \neq i] = \frac{1}{c+n-1}(cG_0 + \sum_{j \neq i}^{n} \delta_{\theta_j}), \tag{1.6}$$

for $i = 1, ..., n$ (δ_θ is assumed to be a degenerate distribution at θ). Using these n conditional distributions, we can implement a Gibbs sampler to obtain a sample point from the posterior distribution of $\theta_1, ..., \theta_n$.

Let $\theta_b^* = (\theta_{b1}^*, ..., \theta_{bn}^*)$; $b = 1, ..., B$; be a sample from $[\theta_1, ..., \theta_n|D]$ and define

$$G_{0b}^*(.) = (c+n)^{-1}\{cG_0(.) + \sum_{i=1}^{n} \delta_{\theta_{bi}^*}(.)\}. \tag{1.7}$$

It can be shown that $G_b^* = \sum_{j=1}^{\infty} w_j \delta_{\theta_j}$ is a realization from $[G|\theta_b^*]$, where $w_1 = z_1$, $w_j = z_j(1 - z_{j-1})...(1 - z_1)$, $j = 2, 3, ...; z_j \sim Beta(1, c+n)(i.i.d)$ and $\theta_j \sim G_{0b}^*$ $(i.i.d.)$ (Sethuraman (1994)).

For sufficiently large J, we can use a partial sum approximation, $\sum_{j=1}^{J} w_j \delta_{\theta_j}$. If $\theta_1^{'}, ..., \theta_l^{'}$ is a sample from G_b^*, then the Monte Carlo integration

$$H_b^* = \frac{1}{l} \sum_{j=1}^{l} H(k(., \theta_j^{'})) \tag{1.8}$$

is a realization from $[H(f_{\mathrm{Y}}(.|G))|D]$, where H is a linear functional,

$$H(f_{\mathrm{Y}}(.|G)) = \int H(k(., \theta))G(d\theta). \tag{1.9}$$

The sample H_b^*; $b = 1, ..., B$; can be used to approximate the posterior moments of $H(f_{\mathrm{Y}}(.|G))$. For the linear functionals $H_1(f(.|G))$ and $H_2(f(.|G))$ and an arbitrary function $T(., .)$, one may use $T_b^* = T(H_{1b}^*, H_{2b}^*)$ as a realization from $[T(H_1, H_2)|D]$ (See Gelfand and Kottas (2002) for further details). For example, suppose that we want to estimate the conditional probability density $f(\mathrm{y}^{(1)}|\mathrm{y}^{(2)}, G)$, where the random vector Y is partitioned as $\mathrm{Y} = (\mathrm{Y}^{(1)}, \mathrm{Y}^{(2)})$. We consider the "p.d.f.-at-a-point" functionals

$$H_1(f_{\mathrm{Y}}(.|G)) = f(\mathrm{y}^{(1)}, \mathrm{y}^{(2)}|G), \tag{1.10}$$

and

$$H_2(f_{\mathrm{Y}}(.|G)) = f(\mathrm{y}^{(2)}|G) = \int k(\mathrm{y}^{(2)}|\theta)\, G(d\theta). \tag{1.11}$$

If we define

$$T(H_1, H_2) = \frac{H_1}{H_2}, \tag{1.12}$$

then, $T(H_{1b}^*, H_{2b}^*)$; $b = 1, ..., B$; is a sample from the conditional predictive density $E(f(\mathrm{y}^{(1)}|\mathrm{y}^{(2)}, G)|D)$.

The rest of this article is organized as follows. Section 2 is devoted to an approach for testing hypotheses about the mean vector of a random vector under a MDP prior. We consider appropriate prior distributions concentrated on the sets of interest.

In Section 3, we give an example to illustrate the approach.

2 Inference about the mean vector

In this section, we propose a semiparametric procedure to test H_0 : $\mu \in A$ versus $H_1 : \mu \in A^c$, where μ denotes the mean vector of Y, A is some subset of R^m and A^c is its complement. For this purpose, we consider the function $k(.,.)$ defined in Section 1 and suppose that it satisfies

$$\int_{R^m} \mathrm{y}\, k(\mathrm{y},\theta)\, d\mathrm{y} = 0 \qquad \forall\, \theta \in R. \tag{2.1}$$

The random vector Y is assumed to have the following probability density function

$$f_{\mathrm{Y}}(\mathrm{y}|\mu\,, G) = \int_R k(\mathrm{y}-\mu,\theta)\, G(d\theta), \tag{2.2}$$

where $\mu = (\mu_1, ..., \mu_m)$ and $G(\,.)$ is a distribution function. It can be shown that the mean vector of Y is μ. (Given $G(\,.)$ and μ, Y $-\,\mu$ has the density.) Consider the 0-1 loss function $\{\, L(a_i,\mu), i = 0, 1\}$, where a_i denotes the decision to accept that the hypothesis H_i is true; $i = 0, 1$,

$$L(a_0,\mu) = \begin{cases} 0 & \mu \in A \\ 1 & \mu \in A^c \end{cases}, \tag{2.3}$$

and

$$L(a_1,\mu) = 1 - L(a_0,\mu). \tag{2.4}$$

To minimize the expected posterior loss, we make the decision a_0 if (and only if)

$$P(\mu \in A|data) \geq \frac{1}{2}. \tag{2.5}$$

Note that the hypothesis H_0 must have a positive probability under the prior distribution of μ. (For example, the Bayes rule always rejects the null hypothesis $H_0 : \mu = \mu_0$ if we assign an absolutely continuous prior distribution to μ.)

We suppose that $G \in DP(cG_0)$ and consider the following prior distribution for μ:

$$\begin{cases} \mu|\mu \in A \sim \pi_0(\,.\,) \\ \mu|\mu \in A^c \sim \pi_1(\,.\,) \\ P(\mu \in A) = p_0 \end{cases}, \tag{2.6}$$

where $\pi_0(\,.\,)$ and $\pi_1(\,.\,)$ are specified density functions and $0 < p_0 < 1$. Our prior opinion about the hypothesis H_0 is expressed by p_0. The densities $\pi_0(\,.\,)$ and $\pi_1(\,.\,)$ are selected by the statistician, but otherwise arbitrary non-informative uniform densities can be chosen. Let $D = (\mathrm{Y}_1, ..., \mathrm{Y}_n)$ be a random sample from $f_{\mathrm{Y}}(.|\mu\,, G)$. Following Mukhopadhyay and Gelfand (1997), we consider latent θ_i associated with Y_i and suppose that

$$\mathrm{Y}_i|\, G, \theta_1, ..., \theta_n, \mu \;\sim\; k(\,y - \mu\,, \theta_i)\,;\quad i = 1, ..., n\;\; (indep.). \tag{2.7}$$

Given $G(\,.)$ and μ, $\theta_i^{,}s$ are assumed to be *i.i.d.* random variables with the distribution $G(\,.)$. Then,

$$[\mu_i|\theta_1, ..., \theta_n, \mu_j; j = 1, ..., m\,; j \neq i\,; D] \propto \prod_{h=1}^{n} k(\mathrm{y}_h - \mu\,, \theta_h)[\mu], \quad i = 1, ..., m \tag{2.8}$$

and

$$[\theta_i|\mu\;, \theta_j;\; j = 1, ..., n\,;\; j \neq i; D] \;\propto\; q_{i0}g_0(.)\,k(\mathrm{y}_i - \mu,\,.\,) + \sum_{j=1,\, j\neq i}^{n} q_{ij} I_{\theta_j}(.)\;; \quad i = 1, ..., n \tag{2.9}$$

where $g_0(.)$ is the density associated with $G_0(.)$,

$$q_{i0} \propto c(\int g_0(\theta)k(\mathrm{y}_i - \mu, \theta)d\theta)^{-1}, \tag{2.10}$$

and

$$q_{ij} \propto k(\mathrm{y}_i - \mu\,, \theta_j), \tag{2.11}$$

subject to $\sum_{j\neq i} q_{ij} = 1$. Using the conditional distributions (2-8) and (2-9), a Gibbs sampler can be implemented to provide a sample, $\mu_1, ..., \mu_l$; from the posterior, $[\mu|D]$. The posterior probability $P(\mu \in A|D)$ and the Bayes estimate of μ are respectively approximated by $\frac{1}{l}\sum_{j=1}^{l} I_A(\mu_j)$ and $\bar{\mu} = \frac{1}{l}\sum_{j=1}^{l}\mu_j$. The Bayes rule rejects H_0 when the posterior probability of $\{\mu \in A\}$ is less than 0.5.

An extension of this method considers more than two hypotheses, i.e.

$$H_i : \mu \in A_i\,;\; i = 1, ..., t, \tag{2.12}$$

where $A_1, ..., A_t$ form a partition for R^m. Then, we accept the hypothesis which has the highest posterior probability.

3 Numerical example

In this section, we used the lumber data (Johnson and Wichern (1988), exercise 5.14) to illustrate our approach. The variables are Y_1 =stiffness and Y_2 =bending strength and the units are pounds / $(\text{inches})^2$. In this exercise, it is assumed that $\mu_{10} = 2000$ and $\mu_{20} = 10000$ represent " typical " values for stiffness and bending strength respectively. The aim is to verify whether the data are consistent with these values. Using normal probability plots, we conclude that the assumption of bivariate normal distribution is not reasonable. Instead, we consider the density function (2-2), where $k(.,.,\theta)$ is the joint density function of two jointly normal distributed variables with zero means, unit variances and correlation coefficient θ. The random distribution function $G(\,.)$ is assumed to be a Dirichlet process, where $c = 1$ and $G_0(\,.)$ is a uniform distribution function on the interval (-1,1). Here, we wish to test $H_0 : (\mu_1, \mu_2) \in (c_1, d_1) \times (c_2, d_2)$. We assume that μ_i has the following prior distribution

$$\begin{cases} \mu_i | \mu_i \in (c_i, d_i) \sim U(\,c_i, d_i) \\ \mu_i | \mu_i \in (-a_i, c_i] \bigcup [d_i, a_i) \sim U(-a_i, c_i] \bigcup [d_i, a_i) \\ P(\mu_i \in (c_i, d_i)) = 0.5 \\ P(\mu_i \notin (-a_i, a_i)) = 0 \end{cases} , \quad i = 1, 2, \tag{3.1}$$

where $-a_i < c_i < d_i < a_i$ and a_i is sufficiently large, for $i = 1, 2$. In other words, $A_i = (c_i, d_i)$ and $A_i^c = (-a_i, c_i] \bigcup [d_i, a_i)$, for $i = 1, 2$ and $\pi_0(\,.\,)$ and $\pi_1(\,.\,)$ are non-informative uniform densities. We also take $a_1 = a_2 = 100000$.

It can be shown that the conditional distribution (2-8) is proportional to

$$\phi(\mu_i | \frac{\sum_{h=1}^{n} (1-\theta_h^2)^{-1} (y_{1h} - \theta_h (y_{2h} - \mu_j))}{\sum_{h=1}^{n} (1-\theta_h^2)^{-1}}, (\sum_{h=1}^{n} (1-\theta_h^2)^{-1})^{-1}) [\mu_i] \tag{3.2}$$

where $\phi(.|\mu, \sigma^2)$ denotes the probability density function of a normal distribution with mean μ and variance σ^2 and $[\mu_i]$ is the density function associated with (3-1).

In order to generate a sample point from (3-2), we first observe the result of some Bernoulli trial that has probability of success 0.5. If the Bernoulli trial results in a success, then we generate a sample point from the density function that is proportional to

$$\phi(\mu_i|\frac{\sum_{h=1}^{n}(1-\theta_h^2)^{-1}(y_{1h}-\theta_h(y_{2h}-\mu_j))}{\sum_{h=1}^{n}(1-\theta_h^2)^{-1}},(\sum_{h=1}^{n}(1-\theta_h^2)^{-1})^{-1})I_{(c_i,d_i)}(\mu_i) \quad (3.3)$$

Otherwise, the sample point is generated from the density function that is proportional to

$$\phi(\mu_i|\frac{\sum_{h=1}^{n}(1-\theta_h^2)^{-1}(y_{1h}-\theta_h(y_{2h}-\mu_j))}{\sum_{h=1}^{n}(1-\theta_h^2)^{-1}},(\sum_{h=1}^{n}(1-\theta_h^2)^{-1})^{-1})I_{(-a_i,c_i]\bigcup[d_i,a_i)}(\mu_i)$$

Table 3.1 shows the results for some different values of each of c_1, d_1, c_2 and d_2. For example, the Bayes rule accepts $H_0 : (\mu_1, \mu_2) \in$ (1500, 2500)×(4000, 13000) and rejects$H_0 : (\mu_1, \mu_2) \in$ (1700, 2300)× (6000, 11000). The typical values μ_{10} and μ_{20} belong to (c_1, d_1) and (c_2, d_2) respectively if we consider each of the first three hypotheses. The last column shows the Bayes estimates of μ_1 and μ_2 under the corresponding prior distributions.

Table 3.1: Posterior probabilities and Bayes estimates

(c_1, d_1)	(c_2, d_2)	$P((\mu_1, \mu_2) \in (c_1, d_1)\times (c_2, d_2)\|D)$	$(\hat{\mu}_1, \hat{\mu}_2)$
(1500,2500)	(4000,13000)	0.64	(2065.44,8266.68)
(1600,2400)	(5000,12000)	0.50	(1939.34,8370.71)
(1700,2300)	(6000,11000)	0.25	(2018.53,8373.34)
(1900,2100)	(8000,9000)	0.10	(2094.53,8461.79)
(1800,1900)	(8200,8400)	0.01	(1973.72,8400.90)
(1500,2500)	(8000,9000)	0.55	(1921.18,8259.15)

Acknowledgment

The authors thank an anonymous referee for helpful comments on an earlier version of the paper, and the editor for his kind cooperation.

References

Brunner, L. J. (1995), Bayesian linear regression with error terms that have symmetric unimodal densities. Journal of nonparametric Statistics, **4**, 335-348.

Brunner, L. J. and Lo, A. Y. (1989), Bayes methods for a symmetric unimodal density and its mode. Annals of Statist., **17**(4), 1550-1566.

Escobar, M. D. (1994), Estimating normal means with a dirichlet process prior. Journal of the American Statistical Association, **89**, 268-277.

Escobar, M. D. and West, M. (1995), Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association, **90**(430), 577-588.

Feller, W. (1971), An Introduction to Probability theory and its applications. Vol.II, New York: Wiley.

Ferguson, T. S. (1973), A Bayesian analysis of Some nonparametric problems. Annals of Statistics, **1**, 209-230.

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

Gelfand, A. E. and Kottas, A. (2002), A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. Journal of Computational and Graphical Statistics, **11**, 284-305.

Gelfand, A. E. and Mukhopadhyay, S. (1995), On nonparametric Bayesian inference for the distribution of a random sample. The Canadian Journal of Statistics, **23**, 411-420.

Gelfand, A. E. and Smith, A. F. M. (1990), Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association, **85**, 398-409.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. V. (1999), Posterior consistency of Dirichlet mixtures in density estimation. Annals of Statistics, **27**, 143-158.

Hoff, P. D. (2004), Clustering Based on Dirichlet Mixtures of Attribute Ensembles. Technical Report, No. 448, University of Washington.

Ishwaran, H. and Zarepour, M. (2000), Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. Biometrika, **87**, 371-390.

Ishwaran, H. and Zarepour, M. (2002), Exact and approximate sum-representations for the Dirichlet process, unpublished.

Johnson, R. A. and Wichern, D. W. (1988), Applied Multivariate Statistical Analysis. 2nd edition, Prentice Hall International Inc.

Kottas, A. and Gelfand, A. E. (2001), Modeling variability order: A semiparametric Bayesian approach. Methodology and Computing in Applied Probability, **3**, 427-442.

Kottas, A., Muller, P., and Quintana, F. (2004), Nonparametric Bayesian modelling for multivariate ordinal data. Journal of Computational and Graphical Statistics, **14**(3), 610-625.

Lijoi, A., Prunster, placeI., and Walker, S. G. (2005), On consistency of nonparametric normal mixtures for Bayesian density estimation. Journal of the American Statistical Association, **100**(472), 1292-1296.

Lijoi, A., Mena, R. H., and Prunster, placeI. (2005), Hierarchical mixture modeling with normalized inverse-Gaussian priors. Journal of the American Statistical Association, **100**(472), 1278-1291.

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

MacEachern, S. N. and Muller, P. (2000), Efficient MCMC schemes for robust model extensions using encompassing Dirichlet process mixture models. in F.Ruggeri and D. Rios-Insua (eds), Robust Bayesian Analysis, State New York: Springer-Verlag, 295-316.

Merrick, J. R. W., Soyer, R., and Mazzuchi, T. A. (2003), A Bayesian semiparametric analysis of the reliability and maintenance of machine tools. Technometrics, **45**(1), 58-69.

Mukhopadhyay, S. and Gelfand, A. E. (1997), Dirichlet process mixed generalized linear models. Journal of the American Statistical Association, **92**(438), 633-639.

Muller, P. and Quintana, F. A. (2004), Nonparametric Bayesian data analysis. Statistical Science, **19**(1), 95-110.

Sethuraman, J. (1994), A constructive definition of Dirichlet priors. Statistica Sinica, **4**,639-650.

Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999), Bayesian nonparametric inference for random distributions and related functions. J. R. Statist. Soc., B, **61**, part 3, 485-527.

West, M. (1990), Bayesian Kernel Density Estimation. Discussion paper, 90- A02, Duke University, Institute of Statistics and Decision Sciences.