EXACT MULTIVARIATE BAYESIAN BOOTSTRAP DISTRIBUTIONS OF MOMENTS

by

Mauro Gasparini
Purdue University

Technical Report #94-6

Department of Statistics
Purdue University

March 1994

Exact Multivariate Bayesian Bootstrap Distributions of Moments

Mauro Gasparini[1]

Purdue University

Abstract

The common unknown probability law P of a random sample $Y_1, \ldots, Y_n$ is assigned a Dirichlet process prior with index α (Ferguson, 1973). It is shown that the posterior joint density of several moments of P converges, as $\alpha(\mathbb{R}) \to 0$, to a multivariate B-spline, which is, therefore, the Bayesian Bootstrap joint density (Rubin, 1981) of the moments.

The result provides the basis for possible default nonparametric Bayesian inference on unknown moments.

[1] Research partially supported by the National Science Foundation, Grant DMS-9303556.

[0] *Key words and phrases.* Dirichlet priors, multivariate B-splines, Bayesian bootstrap.

[0] *AMS 1980 subject classifications.* Primary 62G05; secondary 62P99.

1 Introduction

The need for a default prior to represent vague initial information in nonparametric Bayesian statistics is generally recognized. For this purpose, in the presence of a random sample $Y_1, \ldots, Y_n$, with unknown common probability law P, some researchers would consider using a Dirichlet prior on P (Ferguson 1973) with a very small total mass $\alpha(\mathbb{R})$ of the index α, or a limit as $\alpha(\mathbb{R}) \to 0$. The reason is that, in such a limit case, the posterior law of the infinite dimensional parameter P is centered around the empirical measure. The results obtained are then comparable to standard frequentist results, as illustrated by the applications in Section 5 of Ferguson (1973) and much of the following literature on Dirichlet priors. This reconciliation between frequentist and Bayesian approach is appealing to the scientist who feels opposed to the use of prior information for philosophical reasons. It also provides a possible default choice for the compilation of Bayesian software.

Limiting results from Dirichlet priors of the above nature have been referred to as the Bayesian Bootstrap (BB) by Rubin (1981) and Lo (1987), among others. Such a convention is followed in the present paper, where a few more applications of the BB are illustrated. The focus is on the posterior limiting distribution - the BB distribution - of the vector-valued functional of the parameter P composed of the first s moments: $\boldsymbol{\mu}(P) = (\mu_1, \ldots, \mu_s)(P)$, where $\mu_j(P) := \int y^j P(dy), j = 1, \ldots, s$.

Cifarelli and Regazzini (1990) obtain the proper distribution of the mean $\mu_1(P)$ when P is chosen accordingly to a Dirichlet process satisfying minimal conditions. In particular, their results, applied to the posterior process on P, provide the researcher with a *bona fide* nonparametric posterior distribution on $\mu_1(P)$. Letting $\alpha(\mathbb{R}) \to 0$, they obtain, as a by-product, an expression for the BB density of $\mu_1(P)$, namely

$$M(\mu; y_{(1)}, \ldots, y_{(n)}) = (n-1) \sum_{i=1}^{k} \frac{(\mu - y_{(i)})^{n-2}}{\prod_{j \neq i}(y_{(j)} - y_{(i)})} \quad \text{if } y_{(i)} \leq \mu < y_{(i+1)} \tag{1}$$

for $i = 1, \ldots, n-1$ and 0 otherwise, where $y_{(1)} < y_{(2)} < \ldots < y_{(n)}$ are the order statistics, supposed distinct for sake of simplicity. Although the interest of the authors is not on its applications to statistical practice, such a density turns out to be $n-3$ times continuously differentiable, bell shaped, log-concave, centered around $\bar{Y} = \sum Y_i/n$ and with variance $S^2/(n+1)$, where $S^2 = \sum(Y_i - \bar{Y})^2/n$ is the sample variance. These results are comparable to those derived by a classical sampling theoretical approach or by bootstrap based inference on $\mu_1(P)$.

Generalizing the limit results of Cifarelli and Regazzini to more than one dimension is of some relevance because of the possibility constructing, for example, joint BB regions for unknown moments. Multidimensional results of this sort are actually already available in the literature. Density (1) is in fact a well known classical univariate B-spline, introduced first by Curry and Schoenberg (1966). Its natural extension to higher dimensions, the *multivariate B-splines*, is precisely the BB density of the functional $\boldsymbol{\mu}(P)$. Multivariate B-splines are nowadays a well understood object, see for example Dahmen and Micchelli (1983).

Sections 2 and 3 contain a summary of the relevant definitions and results about Dirichlet priors and multivariate B-splines. Section 4 is an application of these results to the Bayesian nonparametric problem. Section 5 contains asymptotic results.

2 The Bayesian Bootstrap as a limit of Dirichlet posterior processes

Let P, the unknown probability measure of real observanda $Y_1, \ldots, Y_n$, be distributed *a priori* as a Dirichlet process with index α, as in Ferguson (1973). Write $P \sim \mathcal{D}(\alpha)$. A fundamental property of Dirichlet process priors is that they are conjugate to random sampling, in the sense expressed by the following:

Theorem 1 (Ferguson, 1973) *If $P \sim \mathcal{D}(\alpha)$ and if, given P, $Y_1, \ldots, Y_n$ are i.i.d. P, then the posterior law of P is again Dirichlet, namely*

$$P|Y_1 = y_1, \ldots, Y_n = y_n \sim \mathcal{D}(\alpha + \sum_{i=1}^{n} \delta_{y_i}) \tag{2}$$

where δ_x represents the unit mass measure at x.

The Bayesian Bootstrap describes the limit of the posterior law of P as $\alpha(\mathbb{R}) \to 0$. More formally, the Bayesian Bootstrap may be understood in terms of weak convergence of probability laws if P is viewed as a random element taking values on the space $\mathcal{P}$ of all probability measures on the real numbers, endowed with the topology of weak convergence. We then have

Theorem 2 *Under the same conditions of Theorem 1, $\mathcal{D}(\alpha + \sum_{i=1}^{n} \delta_{y_i})$, the posterior law of P, converges weakly to $\mathcal{D}(\sum_{i=1}^{n} \delta_{y_i})$, as $\alpha(\mathbb{R}) \to 0$.*

Proof. This is a corollary of Theorem 3.2 in Sethuraman and Tiwari (1982). ∎

$\mathcal{D}(\sum_{i=1}^{n} \delta_{y_i})$ may be called the BB law of P. The result is equivalent to stating that P is, in the limit, a random distribution with finite support $\{y_1, \ldots, y_n\}$ and masses $(\Pi_1, \ldots, \Pi_n)$ distributed according to a Dirichlet distribution with parameter $(1, 1, \ldots, 1)$, as defined, for example, in Wilks (1962).

The BB behavior of the random vector $\boldsymbol{\mu}(P)$ may also be related to Dirichlet posterior processes in terms of weak convergence, as in the following:

Theorem 3 *Under the same conditions of Theorem 1, if $\alpha = \alpha(\mathbb{R})Q$ and Q is a probability measure such that $\int |y^{2s}| Q(dy) < \infty$, then, as $\alpha(\mathbb{R}) \to 0$, $\boldsymbol{\mu}(P)$ converges in distribution to $\boldsymbol{\mu} := (\sum y_i \Pi_i, \sum y_i^2 \Pi_i, \ldots, \sum y_i^s \Pi_i)'$, where $(\Pi_1, \ldots, \Pi_n)'$ is a random vector having a Dirichlet distribution with parameter $(1, 1, \ldots, 1)$.*

Proof. By application of the Cramer-Wold device, consider a linear combination of the components of $\boldsymbol{\mu}(P)$, say $C = \int \sum a_j y^j P(dy)$. Then C satisfies the conditions of

Corollary 2.7 of Hannum, Hollander and Langberg (1981) and converges in distribution to the corresponding linear combination of the components of $\boldsymbol{\mu}$. ∎

A direct analysis of the convergence of the multivariate densities, possibly leading to stronger results than Proposition 3, would require consideration of coalescent knots of the corresponding B-spline and is avoided here, for sake of simplicity. Also, notice that $\boldsymbol{\mu}(P)$ is not an a.s. weakly continuous functional, so its weak convergence is not a direct corollary of Theorem 2.

3 Multivariate B-splines.

Let the random vector $\mathbf{\Pi} = (\Pi_1, \ldots, \Pi_n)'$ have a Dirichlet distribution with parameter (1,1, ..., 1). This is equivalent to saying that $(\Pi_1, \ldots, \Pi_{n-1})$ has constant Lebesgue density - equal to (n-1)! - over the simplex $\{(\pi_1, \ldots, \pi_{n-1}); \pi_j \geq 0, \sum_1^{n-1} \pi_j \leq 1\}$ and 0 otherwise. The mean vector of $\mathbf{\Pi}$ is $E(\mathbf{\Pi}) = n^{-1}(1, \ldots, 1)$ and its variance-covariance matrix is $V(\mathbf{\Pi}) = n^{-2}(n+1)^{-1}\mathbf{A}$ where the matrix $\mathbf{A}$ has diagonal elements equal to $(n-1)$ and off-diagonal elements equal to -1.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^s$ be distinct and not restricted to a hyperplane.

Definition 1 *The density of the random vector*

$$\boldsymbol{\mu} := \Pi_1\mathbf{x}_1 + \ldots + \Pi_n\mathbf{x}_n \tag{3}$$

is called the s-variate B-spline, or simplex spline, with knots $\mathbf{x}_1, \ldots, \mathbf{x}_n$ *and is denoted by* $M(\boldsymbol{\mu}; \mathbf{x}_1, \ldots, \mathbf{x}_n)$.

Such fundamental objects like densities of linear combinations of uniform variates made various appearances in the statistical literature (see for example Watson (1956) and related works) and were called multivariate B-splines in de Boor (1976), where a geometrical interpretation, equivalent to the definition above, is given. This geometrical interpretation is a multivariate extension of the work by Curry and Schoenberg (1966), where an explicit formula for the univariate B-spline is derived through a Peano representation of the divided differences and a classical formula due to Hermite and Genocchi. For $s = 1$,

$$M(\mu; x_1, \ldots, x_n) = (n-1) \sum_{i=1}^{n} \frac{(x_i - \mu)_+^{n-2}}{\prod_{j \neq i}(x_i - x_j)} \tag{4}$$

where $(.)_+$ denotes positive part.

In the multivariate case, recursive formulae due to Dahmen and Micchelli (see the bibliography in Dahmen and Micchelli (1983)), together with the computing power attained in recent years, provide viable alternatives to cumbersome explicit formulae and approximations of the earlier statistical literature. For an account of these formulae and for other applications of multivariate B-splines in statistics and probability, see Dahmen and Micchelli (1986) and Karlin, Micchelli and Rinolt (1986). For the present purposes, the following result is sufficient:

Theorem 4 (Micchelli, 1980) For any $\boldsymbol{\mu} \in \mathbb{R}^s$ and $\lambda_i \in \mathbb{R}$ such that $\sum_{i=1}^n \lambda_i = 1$ and $\boldsymbol{\mu} = \sum_{i=1}^n \lambda_i \mathbf{x}_i$, we have for $n > s + 1$

$$M(\boldsymbol{\mu}; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{n-1}{n-1-s} \sum_{i=1}^{n} \lambda_i M(\boldsymbol{\mu}; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n). \tag{5}$$

From a practical point of view, in order to calculate a multivariate B-spline at a specific point $\boldsymbol{\mu}$, the recursion in (5) above is iterated down to $n = s + 1$, for which

$$M(\boldsymbol{\mu}; \mathbf{x}_1, \ldots, \mathbf{x}_{s+1}) = \frac{s!}{\left| \det \begin{bmatrix} 1 & \ldots & 1 \\ \mathbf{x}_1 & \ldots & \mathbf{x}_s \end{bmatrix} \right|} \tag{6}$$

for $\boldsymbol{\mu}$ in the interior of the convex hull of $\mathbf{x}_1, \ldots, \mathbf{x}_s$ and 0 otherwise. See Micchelli (1979) and Grandine (1988) for further discussion on evaluation problems.

4 Exact Bayesian Bootstrap densities of moments

The results described in Sections 2 and 3 can now be combined by considering $\mathbf{x}_i = (y_i, y_i^2, \ldots, y_i^s)'$, for $i = 1, \ldots, n$, where y_i's are distinct observed values. We conclude that the BB density of $\boldsymbol{\mu}(P)$, is a multivariate B-spline with knots $\mathbf{x}_1, \ldots, \mathbf{x}_n$. [2]

For example, for $s = 1$, it is easy to secure the equivalence of formulae (4) and (1), with $x_i = y_{(i)}$ - the order of the x_i's does not matter - by noticing that their difference can be written as the divided difference of the polynomial $(n-1)(.-\mu)^{n-2}$ at $x_1, \ldots, x_n$ and therefore equals 0.

The multivariate BB density of $\boldsymbol{\mu}(P)$ is logconcave and of global continuity class $\mathcal{C}^{n-s-2}$ if every $s+1$ knots span a convex hull of positive volume in $\mathbb{R}^s$ (see references in Section 3). Multivariate B-splines are therefore very smooth even for a small sample size. This is particularly relevant, since critics of the use of Dirichlet priors have often emphasized their essentially discrete character (and consequently their inappropriateness) for the analysis of continuous data.

Let $m_j := \sum_{i=1}^n y_i^j/n$ be the j-th sample moment, $j = 1, \ldots, s$. Then, since $\boldsymbol{\mu} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]\mathbf{\Pi}$, its mean is given

$$E(\boldsymbol{\mu}) = [\mathbf{x}_1, \ldots, \mathbf{x}_n]E(\mathbf{\Pi}) = (m_1, \ldots, m_s) =: \mathbf{m} \tag{7}$$

and its variance-covariance matrix is

$$V(\boldsymbol{\mu}) = [\mathbf{x}_1, \ldots, \mathbf{x}_n]V(\mathbf{\Pi})[\mathbf{x}_1, \ldots, \mathbf{x}_n]' = \frac{1}{n+1}[m_{j+k} - m_j m_k]_{s \times s}. \tag{8}$$

Figure 1 contains contour and perspective plots of BB joint densities of first and second moments $(\mu_1, \mu_2)(P)$ for two samples of size $n = 5$ and $n = 10$ simulated from a Normal(0,1) distribution. Figure 2 contains contour plots for analogous samples from an exponential distribution with mean 1. The convex hull of points (y_i, y_i^2), $i = 1, \ldots, n$

[2] The theory could be extended to the case of coincident observations by defining B-splines with coalescent knots appropriately, but it is not done here, for sake of simplicity.

supports the BB density and is drawn on the contour plots. Clearly, edge effects are attenuated as the sample size increases and the posterior density approaches normality. The exact computations described so far are in fact applicable to small sample size problems, whereas for large n normal asymptotic approximations described in the next Section are appropriate, under the usual moment conditions.

It is clear how to use computations of this sort to obtain numerically HPD regions and decision theoretical quantities for unknown moments and smooth transformations of them, like the variance $\mu_2 - \mu_1^2$.

The methodology illustrated here extends to the joint posterior density of functionals of the form $\int \psi_i(y)P(dy)$, for smooth real measurable functions $\psi_i, i = 1, \ldots, s$, as in Corollary 1 of Cifarelli and Regazzini (1990).

5 Asymptotic Bayesian Bootstrap densities of moments

A frequentist asymptotic analysis of the results obtained in the previous sections may be carried out by supposing $Y_1, \ldots, Y_n$ are independent and identically distributed random variables with unknown "true" distribution F_0, unknown "true" mean $\mu_{0,1}$ etc... From this point of view, the BB density of several moments is to be viewed as a *random* multivariate B-spline, since its knots are random.

Theorem 5 *If F_0 possesses finite moments $\mu_{0,1}, \ldots, \mu_{0,2s}$, then the BB density of $\boldsymbol{\mu}^* := \sqrt{n}(\boldsymbol{\mu} - \mathbf{m})$ converges weakly, a.s.-$F_0 \times F_0 \times \ldots$, to an s-variate normal with mean $(0, \ldots, 0)$ and variance-covariance matrix $[\mu_{0,j+k} - \mu_{0,j}\mu_{0,k}]_{s\times s}$.*

Proof. The standardized vector $\boldsymbol{\mu}^*$ has a multivariate B-spline density, since it can be written as $\boldsymbol{\mu}^* = \sum \Pi_i \mathbf{x}_{i,n}^*$, with $\mathbf{x}_{i,n}^* := \sqrt{n}(\mathbf{x}_i - \mathbf{m})$. It suffices to show that conditions of Corollary 4 of Dahmen and Micchelli (1981) hold almost surely.

Condition (a) holds a.s. since

$$(\frac{1}{n}\max_{1\leq i\leq n} \| \mathbf{x}^*_{i,n} \|)^2 = \frac{1}{n}\max_{1\leq i\leq n} \| \mathbf{x}_i - \mathbf{m} \|^2 \rightarrow 0$$

and $\mu_{0,2s} < \infty$ implies $\max_{0\leq i\leq n} \| \mathbf{y}_i \|^{2s} /n \rightarrow 0$ a.s., by the strong law of large numbers.

Condition (b) holds a.s., trivially, with $\mathbf{y} = (0, \ldots 0)$.

Condition (c) holds a.s. with A equal to half of the variance-covariance matrix in the statement of the Theorem, since $\mu_{0,2s} < \infty$, by the strong law of large numbers. ∎

Theorem 5, parallel to standard normal asymptotic theory (see for example Serfling (1980), page 68), is also a generalization of Theorem 4.1 of Lo (1987).

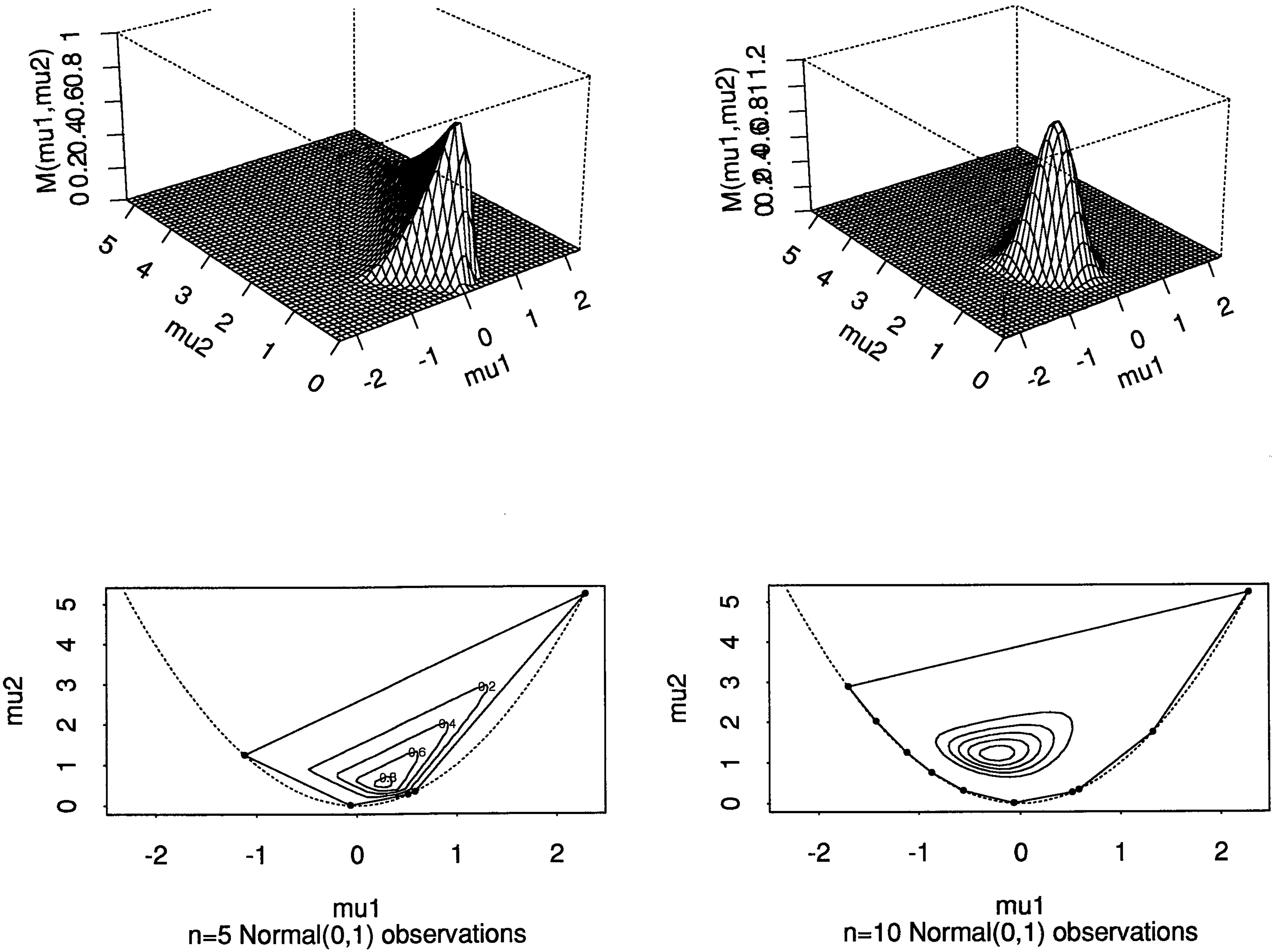
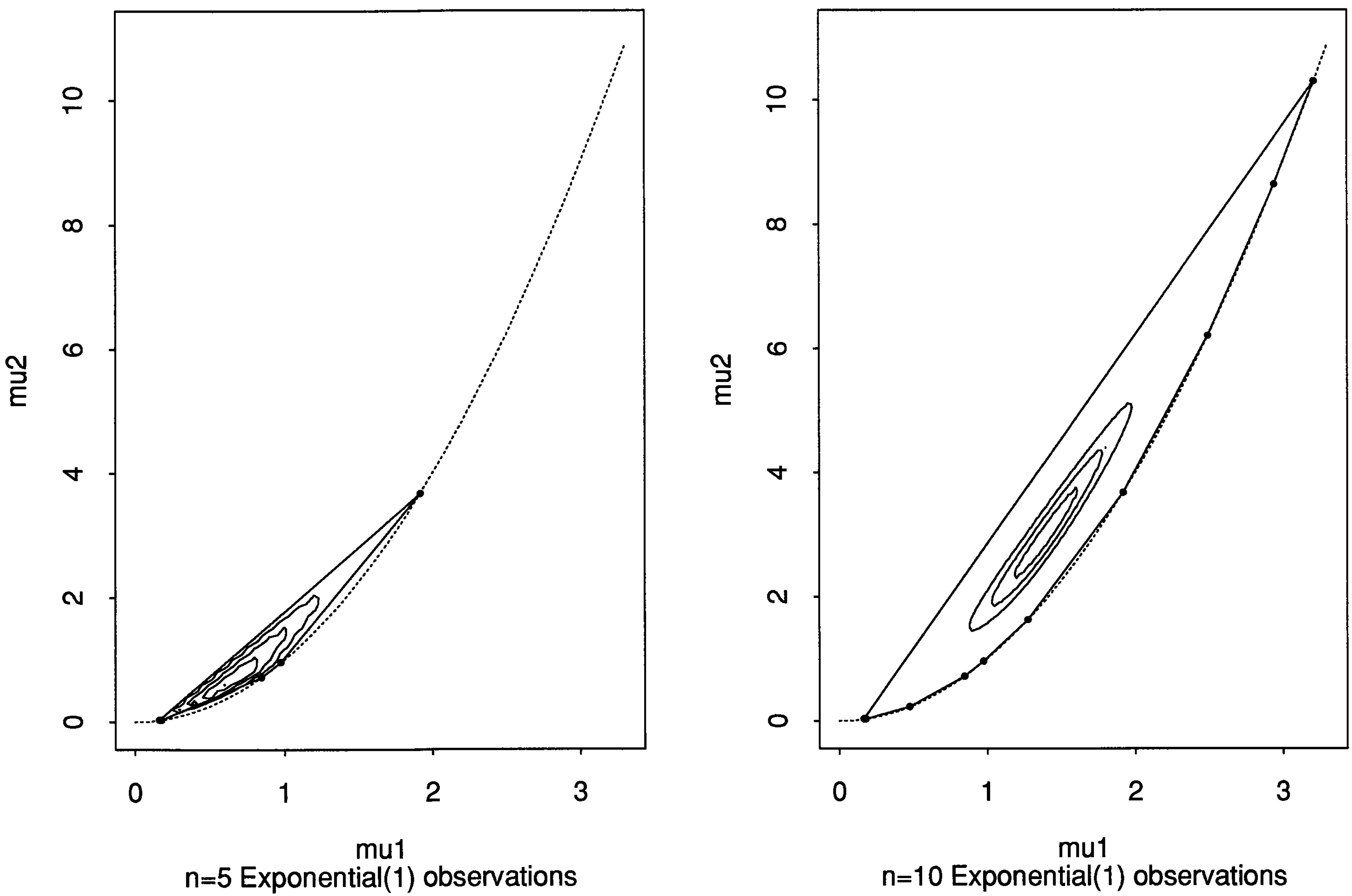

M(mu1,mu2)
0 0.2 0.4 0.6 0.8 1
mu2
mu1
M(mu1,mu2)
mu2
mu1
mu2
mu1
n=5 Normal(0,1) observations
0.2
0.4
0.6
0.8
mu2
mu1
n=10 Normal(0,1) observations

Fig. 1



mu2
10
8
6
4
2
0
0
1
2
3
mu1
n=5 Exponential(1) observations
mu2
10
8
6
4
2
0
0
1
2
3
mu1
n=10 Exponential(1) observations

Fig. 2

References

Cifarelli, D.M. and Regazzini, E. (1990). Distribution functions of means of a Dirichlet process. *Ann. Statist.* **18** 429-442.

Curry, H.B. and Schoenberg, I.J. (1966). On Pólya frequency functions IV: the fundamental spline functions and their limits. *J. d'Analyse Math.* **17** 71-107.

Dahmen, W. and Micchelli, C.A. (1981). On limits of multivariate B-splines. *J. d'Analyse Math.* **39** 256-278.

Dahmen, W. and Micchelli, C.A. (1983). Recent progress in multivariate splines. In *Approximation theory IV,* Chui, C.K., Schumaker, L.L. and Ward, J.D. eds. Academic Press, New York, 17-121.

Dahmen, W. and Micchelli, C.A. (1986). Statistical encounters with B-splines. *Contemporary Mathematics* **59** 17-48.

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

Ferguson, T.S., Phadia E.G. and Tiwari, R.C. (1992). Bayesian nonparametric inference. In *Current Issues in Statistical Inference: Essays in honor of D.Basu* (M. Ghosh and P.K.Patak eds.) IMS Lecture Notes-Monograph series vol.17 127-150.

Grandine, A. T. (1988). The stable evaluation of multivariate simplex splines. *Math. Comp.* **50** 197-205.

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

Karlin, S., Micchelli, C.A. and Rinott, Y. (1986). Multivariate B-splines: A Probabilistic Perspective. *J. Multivariate Anal.* **20** 69-90.

Lo, A.Y. (1987). A large sample study of the Bayesian bootstrap. *Ann. Statist.* **15** 360-375.

Micchelli, C.A. (1979). On a numerically efficient method for computing multivariate B-splines. In *Multivariate approximation theory* Schempp, W. and Zeller, K. eds. Birkäuser. 211-248.

Micchelli, C.A. (1980). A constructive approach to Kergin interpolation in $\mathbb{R}^k$: Multivariate B-splines and Lagrange interpolation. *Rocky Mountain J. Math.* **10** 485-497.

Rubin, D.B. (1981). The Bayesian bootstrap. *Ann. Statist.* **9** 130-134.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* Wiley, New York.

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, vol.2* (S.S.Gupta and J.O.Berger, Eds.) Academic Press, New York, 305-315.

Watson, G.S. (1956). On the joint distribution of circular serial correlation coefficients. *Biometrika* **4** 161-168.

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