UNIVERSIDADE DE SÃO PAULO

DIAGNOSTIC OF CONVERGENCE OF A RAO-BLACKWELLISED ESTIMATE OF THE MARGINAL DENSITY VIA CALIBRATED DIVERGENCE MEASURES

JOSEMAR RODRIGUES

Nº 41

NOTAS

Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas de São Carlos

ISSN - 0103-2577

DIAGNOSTIC OF CONVERGENCE OF A RAO-BLACKWELLISED ESTIMATE OF THE MARGINAL DENSITY VIA CALIBRATED DIVERGENCE MEASURES

JOSEMAR RODRIGUES

Nº 41

NOTAS DO ICMSC
Série Estatística

São Carlos
Out./1997

Diagnostic of Convergence of a Rao-Blackwellised estimate of the marginal density via Calibrated Divergence Measures

Josemar Rodrigues
ICMSC-USP-13560-970-São Carlos-SP-Brazil

Keys words: Computational Statistics; Gibbs sampler, $\phi-$ divergence.

Abstract

This paper presents a diagnostic procedure using calibrated divergence measures to study the performance of a Rao-Blackwellised estimate (RB) of the marginal density. The Gibbs sampling approach is used to compute the divergence measures. A new formulation of the Anchored Ratio Convergence Criterion (ARC^2) and the Difference Convergence Criterion (DC^2) introduced by Zellner and Min (1995) is considered in terms of these calibrated divergence measures. Some examples are analyzed to illustrate the utility of these procedures.

1 Introduction

The Rao-Blackwell estimative of the marginal density via Gibbs sampler (GS) was introduced by Gelfand and Smith (1990) and discussed analytically by Liu, Wong, and Kong (1991). Also, it was used by Zellner and Min (1995) to formulate the Anchored Ratio Convergence Criterion (ARC^2) to determine whether the GS has converged to a correct distribution. The ARC^2 is used when the models parameters may be divided into two (vector or scalar) α and β and explicit conditional densities $p(\alpha \mid \beta)$ and $p(\beta \mid \alpha)$ from the joint density $p(\alpha, \beta)$ are given.

The ARC^2 is computed by using the a Rao-Blackwellised estimate of the marginal densities $p(\alpha)$ and $p(\beta)$ given by:

$$\hat{p}(\alpha) = \frac{\sum_{j=1}^{N} p(\alpha \mid \beta^{(j)})}{N}$$
$$\hat{p}(\beta) = \frac{\sum_{j=1}^{N} p(\beta \mid \alpha^{(j)})}{N}, \qquad (1)$$

where $(\alpha^{(j)}, \beta^{(j)}), \quad j = 1, \ldots, N$ denotes the sequence of GS draws from the conditional densities, $p(\alpha \mid \beta)$ and $p(\beta \mid \alpha)$. The motivation to formulate $\hat{p}(\alpha)$ and $\hat{p}(\beta)$ in (1) is the Rao-Blackwell theorem which was originally illustrated by Gelfand and Smith (1990) and analytically by Liu, Wong and Kong (1991).

In this paper we have the following two main purposes:

- To apply the calibrated divergence measures introduced by Peng and Dey (1995) to measure the perfomance of the Rao-Blackwell estimatives given in (1) for any particular Gibbs sequence.
- There exist various methods to carry out some form of priori statistical analysis of run lengths to assess convergence in the GS algorithms. These methods are called convergence diagnostics. Two of these methods, the ARC^2 and the DC^2, are reformed in terms of the calibrated divergence measures to diagnostic the convergence of a long Gibbs sequence. To diagnostic convergence, we first consider the ARC^2 with a long sequence where we extract every Kth observation up to complete a Gibbs sequence $(\alpha^{(j)}, \beta^{(j)})$ for $j = 1, \ldots, N$. Some illustrative examples are considered in the last section. The DC^2 is modified to detect convergence by minitoring importance weights calculated by splitting a single long run of a Markov chain, generated by the GS. In Example 3, special attention will be given to the application of this new DC^2 to fit a two-component normal mixture model.

2 The calibrated ϕ-divergence measures and Monte Carlo Estimates (Peng/Dey, 1995)

First we consider a general divergence measure as formulated in Peng/Dey (1995) to measure the discrepancy between the RB estimate defined in (1) and the marginal density $p(\alpha)$. This ϕ-measure between the RB estimate $\hat{p}(\alpha)$ and $p(\alpha)$ is defined as

$$D_{\phi}(\alpha) = D(\hat{p}(\alpha), p(\alpha)) = \int \phi\left(\frac{\hat{p}(\alpha)}{p(\alpha)}\right) p(\alpha) d\alpha, \tag{2}$$

where ϕ is a convex function with $\phi(1) = 0$. Dey and Birmwal (1993) considered several choices of ϕ . For example,

$$\begin{aligned}
&\text{Kullback-Leibler Divergence:} \quad \phi(x) = -log(x),\\
&\text{Hellinger distance:} \quad \phi(x) = (\sqrt{x} - 1)^2/2,\\
&\text{Variation distance:} \quad \phi(x) = \frac{1}{2} \mid x - 1 \mid,\\
&\text{Chi-square:} \quad \phi(x) = (x - 1)^2.
\end{aligned} \tag{3}$$

As we do not know the marginal density, $p(\alpha)$, next result gives a way to compute the ϕ-divergence measure without worrying about the normalization of the joint distributions involved.

Theorem 1:

The ϕ-divergence measure between the RB estimate, $\hat{p}(\alpha)$, and the marginal density $p(\alpha)$ is given by

$$D_{\phi}(\alpha) = \int \phi\left(\frac{\delta_{\alpha}(\alpha, \beta)}{\int \delta_{\alpha}(\alpha, \beta) p(\alpha, \beta) d\alpha d\beta}\right) p(\alpha, \beta) d\alpha d\beta, \tag{4}$$

where

$$\delta_{\alpha}(\alpha, \beta) = \frac{\hat{k}_{\alpha}(\alpha, \beta)}{k(\alpha, \beta)}$$

and $k(\alpha, \beta)$ and $\hat{k}_{\alpha}(\alpha, \beta)$ are the kernels of the joint probabilities $p(\alpha, \beta)$ and

$$\hat{p}_{\alpha}(\alpha, \beta) = \hat{p}(\alpha) p(\beta \mid \alpha), \tag{5}$$

respectively.

Proof:

The proof this result is similar to Peng and Dey's proof so it was omitted. The above result suggests the following way to compute the ϕ-divergence measure via Monte Carlo: Suppose we have the Gibbs sample, $(\alpha^{(j)}, \beta^{(j)})$, for $j =$

$1, \ldots, N$ from the joint distribution $p(\alpha, \beta)$. Then the Monte Carlo estimate of $D_\phi(\alpha)$ is

$$\hat{D}_\phi(\alpha) = \frac{1}{N} \sum_{j=1}^{N} \phi \left(\frac{\delta_\alpha(\alpha^j, \beta^j)}{N^{-1} \sum_{j=1}^{N} \delta_\alpha(\alpha^{(j)}, \beta^{(j)})} \right) \tag{6}$$

Given $D_\phi(\alpha) = d$, it would be useful to the pratictioner to have some scheme in terms of d to study the performance of the RB estimate or the convergence of the Gibbs sampling. The idea is to calibrate the ϕ-divergence measure which is similar to that of Peng and Dey (1995). Let $B(p)$ denote the Bernoulli distribution with probability p to an event and we find a function $g(d)$ such that $D_\phi(\alpha) = D(B(\frac{1}{2}), B(g(d)) = d$ The number $g(d)$ is our calibration of d. The calibration tells us that, as measured by (2), the difference d, between the RB estimative and $p(\alpha)$, is the same as that between $B(\frac{1}{2})$ and $B(g(d))$. This latter difference is one that we can readily appreciate. For example, if $g(d) = 0.99$, then RB estimative and $p(\alpha)$ are quite different. If $g(d) = 0.501$, then they are similar. Given $D_\phi(\alpha) = d$, it is not difficult to check that $g(d)$ satisfies the following equation (see, Peng and Dey, 1995):

$$d = \frac{\phi(2g(d)) + \phi(2(1 - g(d))}{2} \tag{7}$$

For example, for the variation distance we have that

$$d = \frac{1}{2} \mid 1 - 2g(d) \mid \tag{8}$$

Thus if $g(d) = 0.6$ it corresponds to $d = 0.1$, similarly, $g(d) = 0.75$ corresponds to $d = 0.25$. Based on Peng and Dey (1995) we suggest the following scale to study the performance of the RB estimative:

$$\begin{aligned} d &\in [0.1, 0.25] : \text{ a mild convergence} \\ d &> 0.25 : \text{a weak convergence} \\ d &< 0.1 : \text{a strong convergence} \end{aligned} \tag{9}$$

In similar manner we can obtain a scale for other choices of ϕ.

3 $\phi-$Convergence Diagnostics

The GS assumes that the conditional distributions $p(\alpha \mid \beta)$ and $p(\beta \mid \alpha)$ are known from which draws can be made. Thus GS can be used to compute normalized constants, posteriors moments and other quantities associated with $p(\alpha, \beta)$. In this section, we are interested to determine whether the GS not only has converged ,but also has converged to correct distribution. With this purpose in mind, Zellner and Min (1995) introduced the Anchored Ratio Criterion (ARC^2) in the following way: For any two points (α_1, β_1) and (α_2, β_2) they computed the following components of the ARC^2:

$$
\begin{aligned}
\hat{\theta}_\alpha &= \frac{\hat{p}(\alpha_1)p(\beta_1 \mid \alpha_1)}{\hat{p}(\alpha_2)p(\beta_2 \mid \alpha_2)}, \\
\hat{\theta}_\beta &= \frac{\hat{p}(\beta_1)p(\alpha_1 \mid \beta_1)}{\hat{p}(\beta_2)p(\alpha_2 \mid \beta_2)}, \quad \text{and} \\
\theta &= \frac{k(\alpha_1, \beta_1)}{k(\alpha_2, \beta_2)}.
\end{aligned}
\tag{10}
$$

The value of θ is known exactly and can be used as an " anchor" to check the convergence of the GS. Then if the GS has converged $\hat{\theta}_\alpha \approx \hat{\theta}_\beta \approx \theta$.

The DC^2 introduced by Zellner and Min is as follows: Given the parameters α and β, we define

$$\eta = p(\alpha)p(\beta \mid \alpha) - p(\beta)p(\alpha \mid \beta) = 0.$$

Now if we use the Rao-Blackwellised estimates given by (1), we define a sample measure given by

$$\hat{\eta} = \hat{p}(\alpha)p(\beta \mid \alpha) - \hat{p}(\beta)p(\alpha \mid \beta).$$

If $\hat{\eta}$ is very small, then the estimated marginals have converged to the correct values while if $\hat{\eta}$ is large has not ocurred. In addition to the problems raised by many authors about the ARC^2 and DC^2 we can see the following three difficulties:

- How do we choose the points (α_1, β_1) and (α_2, β_2)?
- How do we decide whether $\hat{\theta}_\alpha$ or $\hat{\theta}_\beta$ is close to θ?
- How do we choose the points (α, β) and to decide whether $\hat{\eta}$ is close to zero?
- Large sample theory procedures are used to perform ARC^2 and DC^2.

In order to avoid the above questions we suggest a new formulation of the ARC^2 and DC^2 in terms of the ϕ-divergence measure, called " ϕ-Convergence Criterion " (ϕ-CC) and the " ϕ- Difference Criterion (ϕ-DC) , respectively.

The ϕ-CC rule is motivated by the following result:

Theorem 2:

The GS converged only and only if

$$D_\phi(\alpha) = D_\phi(\beta) = 0$$

Proof: It is very trivial from the definition (2). Motivated by the above result and assuming that ϕ is the variation distance we introduce based on (9) the ϕ-CC as follows:

$$\begin{aligned} \max(\hat{D}_\alpha, \hat{D}_\beta) &< 0.1: \quad \text{a strong convergence} \\ \max(\hat{D}_\alpha, \hat{D}_\beta) &\in [0.1, 0.25]: \quad \text{a mild convergence} \\ \max(\hat{D}_\alpha, \hat{D}_\beta) &> 0.25: \quad \text{a weak convergence.} \end{aligned} \tag{11}$$

It is interesting to observe from (4) that the kernel $k(\alpha, \beta)$ used in the ϕ-CC corresponds in a similar way to the quantity θ in the ARC^2. The ϕ-DC diagnostic can be constructed as follows: We begin by running a chain which is split into blocks of N observations. Then for the l^{th} block of the chain, we define

$$D_l = D(\hat{p}_{\alpha,l}(\alpha, \beta), \hat{p}_{\beta,l}(\alpha, \beta)) \tag{12}$$

where $\hat{p}_{\alpha,l}(\alpha, \beta)$ and $\hat{p}_{\beta,l}(\alpha, \beta)$ are the joint density estimatives of $p(\alpha, \beta)$ via the l^{th} block. Thus, we can define an estimative for the variation distance between these joint density estimatives as follows:

$$\hat{D}_l = \max\{\frac{1}{2N}\sum_k \mid \frac{\hat{p}_{\alpha,l}(\alpha^{(k)}, \beta^{(k)})}{\hat{p}_{\beta,l}(\alpha^{(k)}, \beta^{(k)})} - 1 \mid, \frac{1}{2N}\sum_k \mid \frac{\hat{p}_{\beta,l}(\alpha^{(k)}, \beta^{(k)})}{\hat{p}_{\alpha,l}(\alpha^{(k)}, \beta^{(k)}} - 1 \mid\} \tag{13}$$

where $(\alpha^{(k)}, \beta^{(k)})$, $k = 1, \ldots, N$, are the GS draws of the l^{th} block. As in (11), we can monitor $\hat{D}_l$ over blocks $l = 1, 2, \ldots,$ to assess convergence via the following calibrated divergence measure:

$$\begin{aligned} \hat{D}_l &< 0.1: \quad \text{a strong convergence} \\ \hat{D}_l &\in [0.1, 0.25]: \quad \text{a mild convergence} \\ \hat{D}_l &> 0.25: \quad \text{a weak convergence.} \end{aligned}$$

This diagnostic eliminates the need for multiple replications and greatly reduces computational expenses. Also, we do not need to choose points in the parametric space as in Zellner and Min's paper. This is very similar to Ritter and Tanner's diagnostic (1992). In the next section, we discuss three illustrative examples concerned to ϕ-CC and ϕ-DC where ϕ is the variation distance.

4 Some illustrative Examples

Example 1

Suppose that the kernel of the joint distribution of (α, β) is

$$k(\alpha,\beta) = \binom{n}{\alpha} \beta^{\alpha+a-1}(1-\beta)^{n-\alpha+b-1}, \qquad (14)$$
$$\alpha = 0,1,\ldots,n \qquad 0 \leq \beta \leq 1,$$

from which the following results can be derived:

(a) The marginal density of α is the beta-binomial distribution with density

$$p(\alpha) = \binom{n}{\alpha} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(\alpha+a)\Gamma(n-\alpha+b)}{\Gamma(a+b+n)}, \qquad (15)$$
$$\alpha = 0,1,\ldots,n.$$

(b)

$$p(\alpha \mid \beta) \quad \text{is Binomial} \quad (n,\beta). \qquad (16)$$

(c)

$$p(\beta \mid \alpha) \quad \text{is Beta} \quad (\alpha+a, n-\alpha+b). \qquad (17)$$

(d)

$$p(\beta) \quad \text{is Beta} \quad (a,b). \qquad (18)$$

In this example, using a starting value $\beta = 5$ and the GS to generate a long run size M we pick off every k^{th} value to obtain the Gibbs sample $(\alpha^{(j)}, \beta^{(j)})$ $j = 1,\ldots,N$ of the joint distribution $p(\alpha,\beta)$ such that $M = KN$. We use this sample to obtain the RB estimates of the marginal densities and the ϕ-CC in (11) to check for convergence. Also, because the marginal densities are known we compare the RB estimates with them.

Table 1:GS estimates and calibration of the Variation Distance Divergence for $n = 16, a = 2, b = 4$

M	K	N	$\hat{D}_\phi(\alpha)$	$\hat{D}_\phi(\beta)$	$g(d)$
1000	2	500	0.0286	0.0194	0.5286
2000	4	500	0.0338	0.0255	0.5388
3000	6	500	0.0326	0.0221	0.5326
4000	8	500	0.0326	0.0221	0.5326
5000	10	500	0.0250	0.0294	0.5294
7500	15	500	0.0216	0.0197	0.5216
10000	20	500	0.0188	0.0098	0.5188

In Table 1 we present the GS estimates of the variation distance of the marginal densities and the caliblration of the $max(\hat{D}_\phi(\alpha), \hat{D}_\phi(\beta))$ for different values of M and K. From (11) we see an strong convergence of the GS algorithm even for $M = 1000$ and $K = 2$

Since, we know the form of the marginal density $p(\beta \mid \alpha)$ we compare it,in Table 2, to its respective RB estimate for different values of β. We see that the RB estimates are very close to the actual marginal $p(\beta)$. Last, Figure 1 and 2 present the histograms of the GS observations of $p(\alpha)$ and $p(\beta)$, respectively.

Table 2:. RB Estimates of the Marginal density for β and Exact density for $M = 1000, \quad K = 2$ and $N = 500$.

β	$\hat{p}(\beta)$	$p(\beta)$
0.04	0.7572	0.8235
0.13	1.7435	1.7710
0.17	1.9988	1.9839
0.21	2.1093	2.0760
0.27	2.1209	2.0932
0.31	2.0411	2.0245
0.36	1.8805	1.8669
0.40	1.7321	1.7104
0.47	1.4335	1.3971
0.58	0.8437	0.8376
0.63	0.6234	0.6367
0.79	0.1322	0.1371

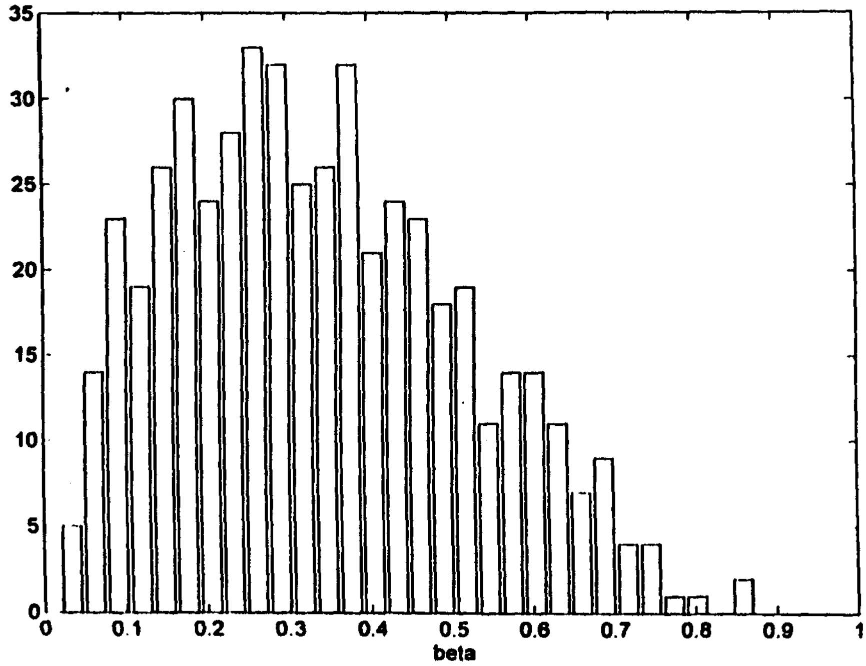

Figure 1: Histogram of Samples of size 500 from the Beta-Binomial with $n = 16$. $a = 2$ and $b = 4$.

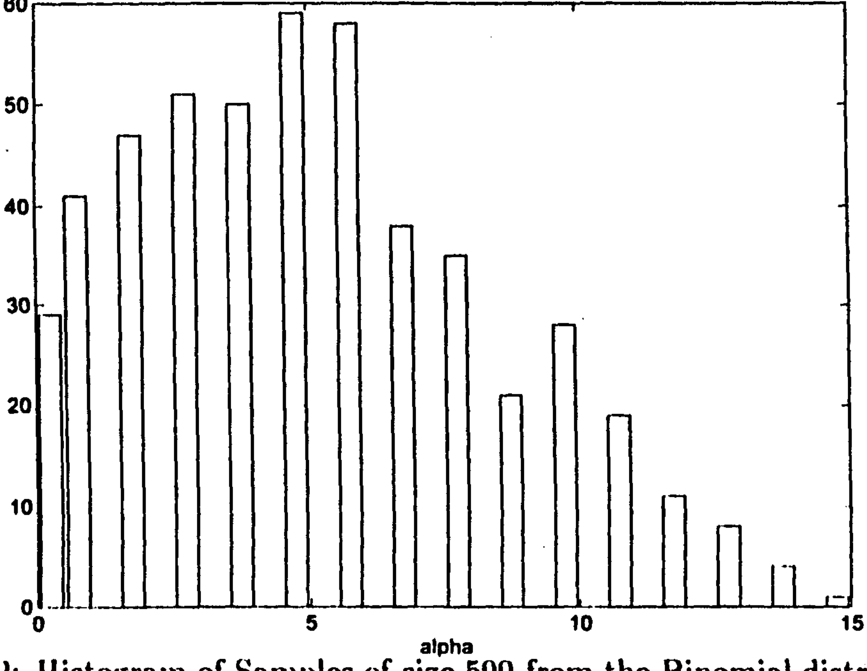

Figure 2: Histogram of Samples of size 500 from the Binomial distribution with $a = 2$ and $b = 4$.

Example 2

Suppose that the kernel of the joint distribution of (α,β) is

$$k(\alpha,\beta) \quad = \quad \beta^{-(T+1)}\exp\{-[\nu s^2 + m_{xx}(\alpha-\hat{\alpha})^2]/2\beta^2\} \qquad (19)$$

with $\nu = T-1$, $\nu s^2 = \sum_{i=1}^{T}(y_i - x_i\hat{\alpha})^2$, $\hat{\alpha} = \sum_{i=1}^{T} x_i y_i / m_{xx}$ and $m_{xx} = \sum_{i=1}^{T} x_i^2$. This kernel is associated with the regression model $y_i = x_i\alpha + e_i$, $i = 1,\ldots,T$, with the e_i's independently draw from a zero mean normal distibution with variance β^2 and a diffuse prior for (α,β).

To illustrate the perfomance of the RB estimates and the use of the $\phi - CC$ to check for convergence of the GS, we generate the data with $T = 20$, $\alpha = 1$ and $\beta^2 = 4$. This model was also used by Zellner and Min (1995) to illustrate the application of the Difference Convergence Criterion DC^2 introduced by them. If we assume a diffuse prior for (α,β), the following results can be derived:

- $p(\beta)$ is an Inverted Gamma (s^2,ν);
- $p(\alpha) \mid \beta$ is normal $(\hat{\alpha},\beta^2/m_{xx})$;
- $p(\beta \mid \alpha)$ is an Inverted Gamma $(\sum_{i=1}^{T}(y_i - x_i\alpha))^2/T, T)$;
- $p(\alpha)$ is univariate Student t with $(\hat{\alpha}, m_{xx}/s^2, \nu)$

As before, Table 3 and 4 show the performace of the RB estimates via variation distance and a comparison beteween the RB estimative, $\hat{p}(\beta)$ and $p(\beta)$, respectively.

Table 3: GS estimates and Variation Distance for Zellner and Min's generated data

M	K	N	$\hat{D}_\phi(\alpha)$	$\hat{D}_\phi(\beta)$	$g(d)$
1000	2	500	0.0156	0.0016	0.5156
2000	4	500	0.0125	0.0047	0.5125
3000	6	500	0.0110	0.0011	0.5110
4000	8	500	0.0078	0.0030	0.5078
5000	10	500	0.0151	0.0067	0.5151
7500	15	500	0.0166	0.0017	0.5140
10000	20	500	0.0166	0.0017	0.5166

Table 4:RB estimates of the Marginal density for β and Exact density for $M = 1000, \quad K = 2$ and $N = 500$.

β	$\hat{p}(\beta)$	$p(\beta)$
1.68	1.1419	1.1483
1.84	1.3069	1.3122
1.96	1.1817	1.1898
2.00	1.1309	1.1321
2.23	0.6497	0.6461
2.25	0.6269	0.6232
2.32	0.4987	0.4946
2.53	0.2343	0.2309
2.89	0.0577	0.0565
4.49	0.0000	0.0000

Example 3

In this example we illustrate the $\phi-$DC for a two-component normal mixture model introduced by Brooks et al. (1996). This model is

$$X \sim pN(\alpha, \sigma_1^2) + (1-p)N(\beta, \sigma_2^2) \quad \alpha < \beta, \tag{20}$$

where $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (σ^2 : known), so that we have three parameters; α, β, p. Following Diebolt and Robert (1994), we introduce the hyper-parameters $z_1, z_2, \ldots, z_n$ where n is the number of observations in the sample, and z_i is an indicator variable taking the value 1 if observation x_i is assigned to the component with mean α and zero otherwise.

Assuming the following priors,

$$\begin{array}{c} p \sim U(0,1) \\ \alpha \sim N(\nu_1, \sigma_\nu^2) \quad \text{and} \quad \beta \sim N(\nu_2, \sigma_\nu^2), \end{array}$$

we obtain the following posterior conditionals via the likelihood;

$$\begin{aligned}
p \mid \{z_i\} &\sim Beta(1 + \sum z_i, n + 1 - \sum z_i) \\
\alpha \mid \{z_i\}, \beta &\sim N(\frac{\sigma_\nu^2 \sum x_i z_i + \sigma^2 \nu_1}{\sigma_\nu^2 \sum z_i + \sigma^2}, \frac{\sigma^2 \sigma_\nu^2}{\sigma_\nu^2 \sum z_i + \sigma^2}) I_{(\alpha<\beta)}(\alpha) \\
\beta \mid \{z_i\}, \alpha &\sim N(\frac{\sigma_\nu^2 \sum x_i (1 - z_i) + \sigma^2 \nu_1}{\sigma_\nu^2 \sum (1 - z_i) + \sigma^2}, \frac{\sigma^2 \sigma_\nu^2}{\sigma_\nu^2 \sum (1 - z_i) + \sigma^2}) I_{(\alpha<\beta)}(\beta) \\
z_i \mid \alpha, \beta, p &\sim Bernoulli(p_i)
\end{aligned}$$

where

$$p_i = \frac{pf(x_i \mid \mu = \alpha)}{pf(x_i \mid \mu = \alpha) + (1-p)f(x_i \mid \mu = \beta)}.$$

Note that since $\sigma_1^2 = \sigma_2^2$, we have that

$$f(x \mid \mu) = \exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2]$$

Let us take the following data in order to fit the model via Gibbs Sampler:

$$\sigma^2 = 1.0, \quad \sigma_\nu^2 = 20.0, \quad \nu_1 = 5, \quad \nu_2 = 20,$$

, and nine observations,

2.3,3.7,4.1,10.9,11.6,12.8,20.1,21.4,22.3

The existence of full conditionals for the parameters α and β allows us to suggest the following Rao-Blackwellised estimates for the joint distribution of (α, β) :

$$\begin{aligned} \hat{p}_\alpha(\alpha,\beta) &= \frac{\sum_j p(\alpha \mid \beta^{(j)}, z^{(j)})}{N} \frac{\sum_k p(\beta \mid \alpha, z^{(k)}, p^{(k)})}{N} \\ \hat{p}_\beta(\alpha,\beta) &= \frac{\sum_j p(\beta \mid \alpha^{(j)}, z^{(j)})}{N} \frac{\sum_k p(\alpha \mid \beta, z^{(k)}, p^{(k)})}{N}. \end{aligned} \tag{21}$$

where $(\alpha^{(j)}, \beta^{(j)}, z^{(j)}, p^{(j)})$, $\quad j = 1, \ldots, N$ are the GS draws and $(z^{(k)}, p^{(k)})$, $\quad k = 1, \ldots, N$, are the GS draws given α and β.

Figure 2 displays the variation distance between $\hat{p}_{\alpha,l}(\alpha,\beta)$ and $\hat{p}_{\beta,l}(\alpha,\beta)$ given by the Monte Carlo estimate

$$\hat{D}_l = \text{maxa}\{\frac{1}{2N}\sum_j \mid \frac{\hat{p}_{\alpha,l}(\alpha^{(j)},\beta^{(j)})}{\hat{p}_{\beta,l}(\alpha^{(j)},\beta^{(j)})} - 1 \mid, \frac{1}{2N}\sum_j \mid \frac{\hat{p}_{\beta,l}(\alpha^{(j)},\beta^{(j)})}{\hat{p}_{\alpha,l}(\alpha^{(j)},\beta^{(j)})} - 1 \mid\}. \tag{22}$$

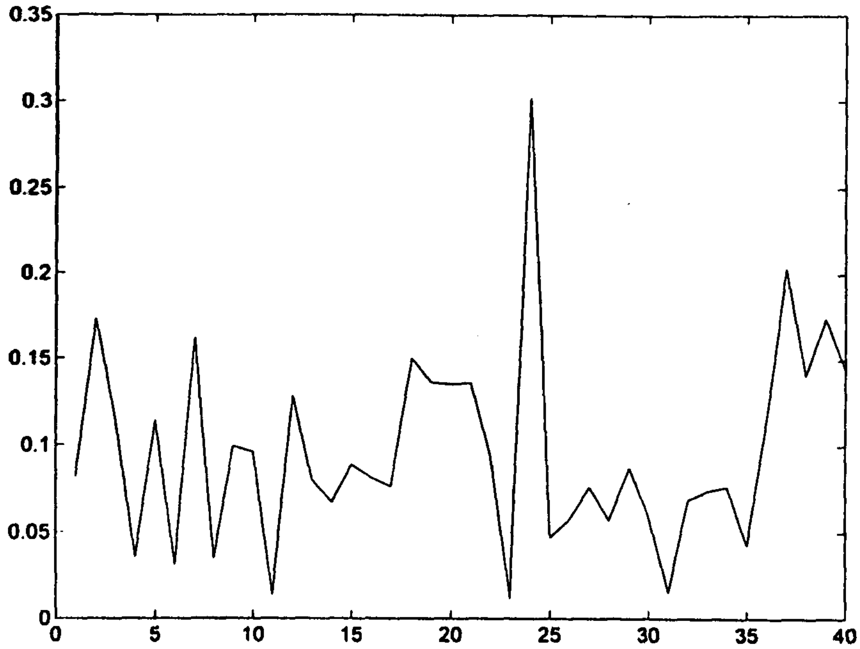

Figure 3: The variation distance for $N = 10$ of a long chain separated into 40 blocks of size 50.

From the calibrated estimative $\hat{D}_l$ we see a clear indication that a strong and mild convergence have been achieved by $l \neq 25$. This result agrees with the diagnostics of Brooks et al. (1996) which the variation distance was computed in a non trivial way and suggesting that convergence was achieved at around 125 iteractions.

References

Brooks, S.P., Dellaportas, P, and Roberts, G.O. (1996). An approach to Diagnosing total variation convergence of MCMC algorithms, Technical Report, Statistical Laboratory, University of Cambridge.

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.

Liu, J,, Wong W.H., and Kong, A. (1994). Covariance structure of the Gibbs sampler eith applications to the comparisons of estimators and augmentation shemes, Biometrika, 81,1,27-40.

Peng, F., and Dey, D. (1995). Bayesian analysis of outliers problems using divergence measures, The Canadian Journal of Statistics, v. 23, no. 2,

199-213.
Ritter, C. and Tanner, M. (1992). Facilitating the Gibbs Sampler: The Gibbs Stopper and Griddy-Gibbs Sampler, Journal of the American Satatistical Association , 87,861-868.
Zellner, A., and Min, C.k. (1995). Gibbs Sampler Convergence Criteria, Journal of American Statistical Association, 90,921-927.

NOTAS DO ICMSC

SÉRIE ESTATÍSTICA

040/97 BARATELA, D.S.; RODRIGUES, J. - Uma caracterização da existência da posteriori marginal do parâmetro N do modelo de Jelinski-Moranda.

039/97 ACHCAR, J.A.; BRASSOLATI, D. - Use of markov chain Monte Carlo methods for a bayesian analysis of software reliability models.

038/97 FRANCELIN, R.A.; BALLINI, R.; ANDARDE, M.G. - Back-propagation vs. Box and Jenkins model to streamflow forescasting.

037/97 ACHCAR, J.A.; LOIBEL, S. - Constant hazard function models with a change-point: a bayesian analysis using markov chain Monte Carlo methods

036 /97 MOALA, F..A.; RODRIGUES, J. - Bayesian inference of the Weibull reliability function via Laplace approximation

.035/96 ACHCAR, J.A.; STORANI, K. - Nonhomogeneous poisson processes assuming a inverse Gaussian order statistics model for software reliability data: a bayesian approach.

034/96 ACHCAR, J.A.; LEANDRO, R.A. - Regression models for bivariate survival data: a bayesian approach.

033/96 ACHCAR, J.A. - Use of gibbs-with-metropolis-hastings algorithms for a bayesian analysis of complex network reliability systems.

032/96 ACHCAR, J.A.; LEANDRO, R.A. - Use of markov chain Monte Carlo methods in a bayesian analysis of the Block and Basu bivariate exponential distribution.

031/96 CEREGATO, S.A.; RODRIGUES, J. - Utilização da inferência bayesiana em experimentos de captura-recaptura.