View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Antonietta Mira e Geoff Nicholls

# Bridge estimation of the probability density at a point

2001/7

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell’Università dell’Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

© Copyright Antonietta Mira and G. Nicholls
Printed in Italy in May 2001
Università degli Studi dell'Insubria
Via Ravasi 2, 21100 Varese, Italy

All rights reserved. No part of this paper may be reproduced in any form without permission of the Author.

# Bridge estimation of the probability density at a point*

Antonietta Mira †and Geoff Nicholls‡

May 7, 2001

**Abstract**

Bridge estimation, as described by Meng and Wong in 1996, is used to estimate the value taken by a probability density at a point in the state space. When the normalisation of the prior density is known, this value may be used to estimate a Bayes factor. It is shown that the multi-block Metropolis-Hastings estimators of Chib and Jeliazkov (2001) are bridge estimators. This identification leads to more efficient estimators for the quantity of interest.

Keywords: Bayes factor, Bridge estimators, Marginal likelihood, Markov chain Monte Carlo, Metropolis-Hastings algorithms.

## 1 Motivation

In their recent paper, Chib and Jeliazkov (2001) treat the problem of estimating the value taken by the normalised probability density at a point. Their interest is motivated by the Bayesian model choice problem, and the estimation of Bayes factors. It turns out the estimators given in Chib and Jeliazkov (2001) belong to the class of bridge estimators considered in Meng and Wong (1996). This result may seem surprising, given that bridge estimators deal with ratios of normalising constants. However, as we will see, the normalised density at a point is a ratio of normalising constants in which the trivial normalising constant of an atom at the point in question plays the part of the second constant. The bridge in Chib and Jeliazkov (2001) proves to be a sequence of densities, defined on spaces of increasing dimension, running from this atom, and finishing at the target density. The value of making an identification of this kind, is in bringing into play the body of theory

*This work has been supportted by the F.A.R. 2000, of the University of Insubria.

†Università dell'Insubria, Facoltà di Economia, Via Ravasi 2, 21100, Varese, Italy. Email: amira@eco.uninsubria.it

‡Auckland University, Department of Mathematics, Private Bag 92019, Auckland, New Zealand. Email: nicholls@math.auckland.ac.nz

and experience developed in Meng and Wong (1996) and by subsequent authors. In particular, efficiency gains, over the scheme of Chib and Jeliazkov (2001), are immediate, as we will see.

Bayesian inference expresses conflicting beliefs as distinct prior distributions. We can sometimes quantify the relative support, given by the data to each of two prior models, using their Bayes factor. We compute the average value of the likelihood under each of the two priors; the Bayes factor is the ratio of these two average values. Let $f_{Y|X}(y|x)$ denote the likelihood of data $y$ given parameters $x$. Let $f_X(x)$ be an unnormalised prior density, defined for $x$ in some parameter space $\mathcal{X}$, and let $Z_X = \int_{\mathcal{X}} f_X(x)dx$. Let $f_{X|Y}(x|y) = f_{Y|X}(y|x)f_X(x)$ denote the unnormalised posterior density for $x \in \mathcal{X}$ and let $Z_{X|Y} = \int_{\mathcal{X}} f_{X|Y}(x|y)dx$. The average likelihood under the prior is

$$\begin{aligned} \mathsf{E}_{f_X}[f_{Y|X}(y|x)] &= \int_{\mathcal{X}} f_{Y|X}(y|x)\frac{f_X(x)}{Z_X}dx && (1) \\ &= Z_{X|Y}/Z_X. && (2) \end{aligned}$$

The problem of estimating a Bayes factor therefore reduces to the problem of estimating the ratio of the normalising constants of the posterior and prior distributions.

Estimates of ratios of normalising constants are sometimes required in classical inference also, for example when dealing with missing data in likelihood ratios for hypothesis testing. Finally the problem is of great interest in statistical physics. We refer the reader to the introduction of the paper by Meng and Wong (1996) for an extensive treatment of non-Bayesian and non-statistical applications of the ideas presented in the present manuscript.

In Section 2 we review the Meng-Wong family of bridge estimators for the ratio of normalizing constants. We then introduce (Section 3) the estimators proposed by Chib and Jeliazkov (2001) used within both for the single and multiple blocks Metropolis-Hastings algorithm and show that they are bridge estimators. This identification allows us to improve the Chib and Jeliazkov estimators in terms of their efficiency. We end with some concluding remarks.

# 2 The Meng-Wong family of estimators

## 2.1 Bridge estimators

Meng and Wong (1996) give a formula which may be used to generate estimators for ratios of normalising constants. Gelman and Meng (1998) develop these ideas very neatly. For $i = 1, 2$, let $f^{(i)}(x)$ be an unnormalised density with state space $\mathcal{X}^{(i)}$ and normalising constant $Z^{(i)}$. Let an *admissible* weight function $h(x)$ be an arbitrary function, defined on $\mathcal{X}^{(1)} \cap \mathcal{X}^{(2)}$, satisfying

$$0 < \left| \int_{\mathcal{X}^{(1)} \cap \mathcal{X}^{(2)}} h(x) f^{(1)}(x) f^{(2)}(x) dx \right| < \infty. \qquad (3)$$

The ratio of the normalising constants of $f^{(1)}$ and $f^{(2)}$ may be written

$$r_h^{2,1} = \frac{Z^{(1)}}{Z^{(2)}} = \frac{\mathsf{E}_{f^{(2)}}[h(x)f^{(1)}(x)]}{\mathsf{E}_{f^{(1)}}[h(x)f^{(2)}(x)]}. \tag{4}$$

The above ratio leads to the natural Monte Carlo estimator $\hat{r}_h^{2,1}$ for $r_h^{2,1}$. Suppose that, for $i = 1, 2$, sequences $S_i = \{x^{(i),j}\}_{j=1}^{N^{(i)}}$ of $N^{(i)}$ samples $x^{(i),j} \sim f_i$ are available. The samples may be correlated, for example, they may be generated by MCMC. Let $S = \{S_1, S_2\}$ and $\hat{r}_h^{2,1}(S)$ be an estimate of $r_h^{2,1}$ based on $S$. Let

$$RE^2(\hat{r}_h^{2,1}) = \frac{\mathsf{E}_S[(\hat{r}_h^{2,1}(S) - r_h^{2,1})^2]}{(r_h^{2,1})^2}$$

denote the relative mean square error for a particular estimator and given stochastic processes realising $S$. Suppose $h(x) = h_O(x)$ minimises $RE^2(\hat{r}_h^{2,1})$ over all admissible $h$. Meng and Wong (1996) show that $h_O$ is

$$h_O(x) = \left[N^{(1)} f^{(1)}(x) + r_h^{2,1} N^{(2)} f^{(2)}(x)\right]^{-1}.$$

As Meng and Wong note, the presence of $r_h^{2,1}$, the quantity we wish to estimate, in the proposed estimator is not a significant obstacle. The iteration defined by

$$\hat{r}_{h_O}^{2,1}(t+1; S) = \frac{\frac{1}{N^{(2)}} \sum_{j=1}^{N^{(2)}} \frac{f^{(1)}(x^{(2),j})}{N^{(1)} f^{(1)}(x^{(2),j}) + N^{(2)} \hat{r}_{h_O}^{2,1}(t;S) f^{(2)}(x^{(2),j})}}{\frac{1}{N^{(1)}} \sum_{j=1}^{N^{(1)}} \frac{f^{(2)}(x^{(1),j})}{N^{(1)} f^{(1)}(x^{(1),j}) + N^{(2)} \hat{r}_{h_O}^{2,1}(t;S) f^{(2)}(x^{(1),j})}} \tag{5}$$

converges to $\hat{r}_{h_O}^{2,1}(S)$ (usually very rapidly).

Note that, when the samples in $S_i$ are not iid, $N^{(i)}$ is to be replaced, wherever it appears in Eqn. (5), by the "effective sample size" parameter of the set $S_i$. This is difficult as effective sample size is not even uniquely defined in the present setting since it is not clear on which paremeter to base the estimation of the autocorrelation time. However the problem does not seem so great in practice. First, as Meng and Wong note, any reasonable estimate for the effective sample size is adequate, since the relative mean square error $RE^2(\hat{r}_h^{2,1})$ is typically insensitive to the value chosen for this quantity. Secondly, if $f^{(1)}$ is the posterior, $f^{(2)}$ the prior, and if $\tau_{f_{Y|X}}^{(i)}$ is the integrated autocorrelation time of the likelihood in the sequence $S_i$, then replacing the sample size with the estimate $N^{(i)}/\tau_{f_{Y|X}}^{(i)}$ leads to near optimal estimators for $\hat{r}_h^{2,1}$. This second observation is supported by our own experimentation, which we do not report.

## 2.2 Bridging spaces of unequal dimension

There are extensions of the identity in Eqn. (4). Take a sequence, index $i = 0, 1, 2 \ldots B$, of $B+1$ densities $f^{(i)}$, associated normalising constants $Z^{(i)}$, and state spaces $\mathcal{X}^{(i)}$. For each of the $B$ ratios $Z^{(i-1)}/Z^{(i)}, i = 1, 2 \dots B$ fix some weight function $h^{(i)}(\psi_i), \psi_i \in \mathcal{X}^{(i)}$. Consider now the identity

$$\frac{Z^{(0)}}{Z^{(B)}} = \frac{Z^{(0)}}{Z^{(1)}} \times \frac{Z^{(1)}}{Z^{(2)}} \times \dots \frac{Z^{(B-1)}}{Z^{(B)}}. \tag{6}$$

If, for each $i = 1, 2 \dots B$, we can find $h = h^{(i)}$ satisfying Eqn. (3) (with $f^{(1)}$ replaced by $f^{(i-1)}$ and $f^{(2)}$ by $f^{(i)}$) it is straightforward to estimate each ratio in Eqn. (6) as before, and thereby estimate $Z^{(0)}/Z^{(B)}$. However, if the bridging densities $...f^{(i-1)}, f^{(i)}...$ involved are defined on spaces of unequal dimension, Eqn. (3) cannot be satisfied. Chen and Shao (1996) remove this restriction, as we now explain.

Let $x = (x_1, x_2 \dots x_B)$ denote a parameter vector $x \in \mathcal{X}$ with components divided into $B$ blocks of parameters. For $i = 0, 1, ..., B$, let $\psi_i = (x_1, x_2 \dots x_i)$ denote a composite variable built from the first $i$ blocks. In this setting $\mathcal{X}^{(0)}, \mathcal{X}^{(1)} \dots \mathcal{X}^{(B)}$ is a sequence of spaces of increasing dimension, running from a possible atom at $\mathcal{X}^{(0)}$ to the full space $\mathcal{X}^{(B)} = \mathcal{X}$. Following Chen and Shao (1996), introduce a sequence of $B$ explicitly normalised conditional densities $w_{i,i-1}(x_i|\psi_{i-1})$ indexed by $i = 1, 2 \dots B$, with support $x_i \in \mathcal{X}_w^{(i,i-1)}(\psi_{i-1})$.

Let $f^{(i,i-1)}(\psi_i) = f^{(i-1)}(\psi_{i-1}) w_{i,i-1}(x_i|\psi_{i-1})$ and

$$\mathcal{X}^{(i,i-1)} = \{(\psi_{i-1}, x_i); \psi_{i-1} \in \mathcal{X}^{(i-1)}, x_i \in \mathcal{X}_w^{(i,i-1)}(\psi_{i-1})\},$$

so that $\int_{\mathcal{X}^{(i,i-1)}} f^{(i,i-1)}(\psi_i)\, d\psi_{i-1} dx_i = Z^{(i-1)}$. Admissible $h^{(i)}(\psi_i)$ now satisfy a variant of Eqn. (3),

$$0 < \left| \int_{\mathcal{X}^{(i,i-1)} \cap \mathcal{X}^{(i)}} f^{(i,i-1)}(\psi_i) f^{(i)}(\psi_i) h^{(i)}(\psi_i)\, d\psi_i \right| < \infty, \tag{7}$$

and the identity

$$r_h^{i,i-1} = \frac{Z^{(i-1)}}{Z^{(i)}} = \frac{\mathsf{E}_{f^{(i)}}[f^{(i,i-1)} h^{(i)}]}{\mathsf{E}_{f^{(i,i-1)}}[f^{(i)} h^{(i)}]} \tag{8}$$

replaces Eqn. (4) whenever two bridging densities are defined on spaces of unequal dimension. Following Meng and Wong (1996), Chen and Shao (1996) give $h_O^{(i)}$ minimising the MSE of the natural estimator for $r_{h_O}^{i,i-1}$. Note that the more general reversible jump setting of Green (1995) is allowed. One may bridge $f^{(0)}$ to $f^{(B)}$ using an *arbitrary* sequence of distributions. It is not necessary that each distribution adds a factor to a product measure as implied in this exposition.

# 3 The estimators of Chib and Jeliazkov

Returning to estimation of $Z_{X|Y}/Z_X$, it is is natural to choose $f^{(1)} = f_{X|Y}$ and $f^{(2)} = f_X$ in Eqn. (4) and thereby obtain an estimator $\hat{r}_{h_{X|Y,X}}$ for the expected likelihood under prior $f_X$. However, it is quite often the case that the normalising constant of the prior can be calculated in closed form. In this case it may be possible to find a comparison density $f^{(2)}$ with advantages over the prior. For example, the estimator $\hat{r}_{h_{X|Y,2}} Z^{(2)}/Z_X$ formed using some alternative comparison function $f^{(2)}$ may be more easily computed, or may have a smaller sampling variance, than $\hat{r}_{h_{X|Y,X}}$.

In this section we analyse a class of estimators, proposed in Chib and Jeliazkov (2001), for which the prior normalising constant must be known or separately estimated. For ease of notation, let $f(x) = f_{X|Y}(x|y)$ be an unnormalised density on $\mathcal{X}$ and let $\tilde{f}(x) = f(x)/Z$ with $Z = \int_{\mathcal{X}} f(x)dx$. Let a fixed state $x^* \in \mathcal{X}$ be given. Typically $x^*$ is an estimate of the posterior mode or mean obtained from MCMC simulation of the posterior. We will consider the problem of estimating $\tilde{f}(x^*)$, which is equivalent to determining $Z_{X|Y}/Z_X$ when the prior normalisation is known.

We start with "single block" problems: a Metropolis-Hastings Markov chain with equilibrium $f$ is given; in it, all the parameters of $f$ are updated in a single Metropolis-Hastings step. Chib and Jeliazkov (2001) give an estimator for $\tilde{f}(x^*)$ in terms of expectations in the equilibrium and Hastings-proposal densities. We show, in Section 3.1, that their single-block estimator is in the Meng-Wong family, and can be improved by replacing the Chib and Jeliazkov choice of $h$ with the Meng-Wong optimal choice.

It is not always feasible to update all parameters of a posterior density in a single Metropolis-Hastings update. One may then identify *blocks* of variables which may conveniently be updated as a group. Chib and Jeliazkov show how normalising constants may be estimated in multiple block problems. We show, in Section 3.2, that their multi-block estimator is a bridge estimator. Combining the bridging densities of Chib and Jeliazkov (2001) and the optimal weight functions from Meng and Wong (1996), we write down an improved estimator, no less tractable than that of Chib and Jeliazkov.

## 3.1 Single block densities

Let $q(x, x')$ be a normalised proposal density for $x'$ conditional on $x$, and let $\alpha(x, x')$ be the corresponding probability of accepting $x'$ in a Metropolis-Hastings update:

$$\alpha(x, x') = \min\left[1, \; \frac{f(x')\, q(x', x)}{f(x)\, q(x, x')}\right].$$

This update respects the detailed balance relation,

$$q(x, x')\, \alpha(x, x')\, f(x) = q(x', x)\, \alpha(x', x)\, f(x'). \tag{9}$$

Chib and Jeliazkov base an estimator for $\tilde{f}(x^*)$ on the identity

$$\tilde{f}(x^*) = \frac{f(x^*)}{Z} = \frac{\mathsf{E}_f[\alpha(x, x^*) q(x, x^*)]}{\mathsf{E}_{q(x^*,\cdot)}[\alpha(x^*, x)]}. \tag{10}$$

Consider now the generating identity, Eqn. (4). The choices $f^{(1)}(x) = f(x^*)\, q(x^*, x)$, $f^{(2)}(x) = f(x)$,

$$h(x) = h_{CJ}(x) = \frac{\alpha(x^*, x)}{f(x)},$$

$\mathcal{X}^{(1)} = \{x; x \in \mathcal{X}, q(x^*, x) > 0\}$ and $\mathcal{X}^{(2)} = \mathcal{X}$, lead from Eqn. (4) to Eqn. (10). Notice that $Z^{(1)} = f(x^*)$ since $q(x^*, x)$ is normalised over $x$, so that our choices set up $Z^{(1)}/Z^{(2)} = \tilde{f}(x^*)$. Identity Eqn. (10) is a special case of identity Eqn. (4). The optimal $h$ for this $f^{(1)}$ and $f^{(2)}$ is

$$h_{f,q} = \left[N^{(1)} f(x^*)\, q(x^*, x) + N^{(2)}(f(x^*)/Z) f(x)\right]^{-1} \tag{11}$$

and our simulations confirm this. It is straightforward to convert code computing an estimator based on $h_{CJ}(x)$ to code computing the optimal estimator obtained when $\tilde{f}(x^*)$ is estimated using Eqn. (11). The estimator derived from $h_{CJ}(x)$ may be used to seed the iteration defined in Eqn. (5).

## 3.2 Densities with more than one parameter block

### 3.2.1 Notation

Let $\Psi_{i+1} = (x_{i+1}, x_{i+2} \dots x_B)$ denote the complement of $\psi_i = (x_1, x_2 \dots x_i)$ so that $x = (\psi_{i-1}, x_i, \Psi_{i+1})$. Let $\Psi^*_{i+1} = (x^*_{i+1}, x^*_{i+2} \dots x^*_B)$. Let $f(\psi_{i-1}, x_i | \Psi^*_{i+1}) = f(\psi_{i-1}, x_i, \Psi^*_{i+1})$ and

$$\mathcal{X}^*_i = \{(\psi_{i-1}, x_i); (\psi_{i-1}, x_i, \Psi^*_{i+1}) \in \mathcal{X},\ f(\psi_{i-1}, x_i, \Psi^*_{i+1}) > 0\}.$$

Let $Z_i = \int_{\mathcal{X}^*_i} f(\psi_{i-1}, x_i | \Psi^*_{i+1})\, d\psi_{i-1} dx_i$ and $\tilde{f}(\psi_{i-1}, x_i | \Psi^*_{i+1}) = f(\psi_{i-1}, x_i | \Psi^*_{i+1})/Z_i$. Let $\tilde{f}(x^*_i | \Psi^*_{i+1})$ equal the normalised marginal density

$$\tilde{f}(x^*_i | \Psi^*_{i+1}) = \int_{\mathcal{X}^*_{i-1}} \tilde{f}(\psi_{i-1}, x^*_i | \Psi^*_{i+1})\, d\psi_{i-1}.$$

For the special case $i = B$, $\mathcal{X}_B = \mathcal{X}$ is the full space, $Z_B = Z$, and by $\tilde{f}(x^*_B)$ we mean the normalised marginal density

$$\tilde{f}(x^*_B) = \frac{1}{Z} \int_{\mathcal{X}^*_{B-1}} f(\psi_{B-1}, x^*_B)\, d\psi_{B-1}.$$

### 3.2.2 The method of Chib and Jeliazkov

Chib and Jeliazkov use the relation

$$\tilde{f}(x^*) = \tilde{f}(x^*_1 | \Psi^*_2) \times \tilde{f}(x^*_2 | \Psi^*_3) \times \cdots \times \tilde{f}(x^*_{B-1} | \Psi^*_B) \times \tilde{f}(x^*_B), \tag{12}$$

to reduce the problem of estimating $\tilde{f}(x^*)$ to the problem of estimating $\tilde{f}(x^*_i | \Psi^*_{i+1})$ for each $i = 1, \cdots, B$.

Consider now a Metropolis-Hastings Markov chain with equilibrium $\tilde{f}(\psi_{i-1}, x_i | \Psi^*_{i+1})$. Let $q_i(x_i, x'_i | \psi_{i-1}, \Psi^*_{i+1})$ be the normalised proposal density for the Metropolis-Hastings step from $x = (\psi_{i-1}, x_i, \Psi^*_{i+1})$ to $x' = (\psi_{i-1}, x'_i, \Psi^*_{i+1})$. The corresponding Metropolis-Hastings acceptance probability is

$$\alpha(x_i, x'_i|\psi_{i-1}, \Psi^*_{i+1}) = \min\left[1,\ \frac{f(\psi_{i-1}, x'_i|\Psi^*_{i+1})\ q_i(x'_i, x_i|\psi_{i-1}, \Psi^*_{i+1})}{f(\psi_{i-1}, x_i|\Psi^*_{i+1})\ q_i(x_i, x'_i|\psi_{i-1}, \Psi^*_{i+1})}\right].$$

Let $f^*_{i,i-1}(\psi_i) = f(\psi_{i-1}|x^*_i, \Psi^*_{i+1})\, q_i(x^*_i, x_i|\psi_{i-1}, \Psi^*_{i+1})$ and

$$\mathcal{X}^*_{i,i-1} = \{(\psi_{i-1}, x_i); (\psi_{i-1}, x_i, \Psi^*_{i+1}) \in \mathcal{X},\ q_i(x^*_i, x_i|\psi_{i-1}, \Psi^*_{i+1}) > 0\}.$$

Chib and Jeliazkov show that, for $i = 1, \cdots, B$

$$f(x^*_i|\Psi^*_{i+1}) = \frac{\mathsf{E}_{f^*_i}[\alpha(x_i, x^*_i|\psi_{i-1}, \Psi^*_{i+1})\, q_i(x_i, x^*_i|\psi_{i-1}, \Psi^*_{i+1})]}{\mathsf{E}_{f^*_{i,i-1}}[\alpha(x^*_i, x_i|\psi_{i-1}, \Psi^*_{i+1})]}. \tag{13}$$

In this formula "$\mathsf{E}_{f^*_i}$" is an expectation over $\tilde{f}(\psi_{i-1}, x_i|\Psi^*_{i+1})$ and state space $\mathcal{X}^*_i$ and "$\mathsf{E}_{f^*_{i,i-1}}$" is an expectation over $\tilde{f}^*_{i,i-1}(\psi_i)$ and state space $\mathcal{X}^*_{i,i-1}$. These expectations are estimated from samples in the obvious way.

### 3.2.3 Bridge estimation for spaces of unequal dimension

The relation in Eqn. (12) can be written in a suggestive way. Since

$$\tilde{f}(\psi_{i-1}, x^*_i|\Psi^*_{i+1}) = \tilde{f}(x^*_i|\Psi^*_{i+1})\tilde{f}(\psi_{i-1}|x^*_i, \Psi^*_{i+1})$$

we have $\tilde{f}(x^*_i|\Psi^*_{i+1}) = Z_{i-1}/Z_i$ and it follows that Eqn. (12) is equivalent to the identity

$$\tilde{f}(x^*) = \frac{f(x^*)}{Z_1} \times \frac{Z_1}{Z_2} \times \cdots \times \frac{Z_{B-2}}{Z_{B-1}} \times \frac{Z_{B-1}}{Z}. \tag{14}$$

Consider deriving an estimator for $Z_{i-1}/Z_i$ (*ie* $\tilde{f}(x^*_i|\Psi^*_{i+1})$) using Eqn. (8). Referring to Eqn. (8), set

$$\begin{aligned} f^{(i)}(\psi_i) &= f(\psi_{i-1}, x_i|\Psi^*_{i+1}), \\ w_{i,i-1}(x_i|\psi_{i-1}) &= q_i(x^*_i, x_i|\psi_{i-1}, \Psi^*_{i+1}) \end{aligned}$$

so that we identify $f^{(i,i-1)}(\psi_i)$ in Sec. (2.2) with $f^*_{i,i-1}(\psi_i)$, in Sec (3.2.2), and

$$h^{(i)}(\psi_i) = h^{(i)}_{CJ}(\psi_i) = \alpha(x^*_i, x_i|\psi_{i-1}, \Psi^*_{i+1})/f(\psi_{i-1}, x_i|\Psi^*_{i+1}),$$

$\mathcal{X}^{(i)} = \mathcal{X}^*_i$ and $\mathcal{X}^{(i,i-1)} = \mathcal{X}^*_{i,i-1}$. Note that $\mathcal{X}^{(0)} = \{x^*\}$ is an atom with $f^{(0)} = f(x^*)$, and $w_{1,0}(x_1) = q(x^*_1, x_1|\Psi^*_2)$, so that $Z^{(0)} = f(x^*)$. Substituting these functions in Eqn. (8) and using the detailed balance relations for $q(x_i, x'_i|\psi_{i-1}, \Psi^*_{i+1})$, $\alpha(x_i, x'_i|\psi_{i-1}, \Psi^*_{i+1})$ and $f(\psi_{i-1}, x_i|\Psi^*_{i+1})$ we arrive precisely at Eqn. (13). The bridge estimation formulae Eqn. (6) and Eqn. (8) are identical to the estimation relations Eqn. (12) and Eqn. (13). It follows that the block update estimators of Chib and Jeliazkov (2001) are bridge estimators.

Chib and Jeliazkov build their bridge from a neatly chosen sequence of distributions. However, their choice of $h^{(i)} = h^{(i)}_{CJ}$ at each span of the bridge is not optimal, whilst the optimal $h^{(i)} = h^{(i)}_{O}$ is

$$h^{(i)}_{O} = \left[ f^*_{i,i-1}(\psi_i) N^{(i,i-1)} + (Z_{i-1}/Z_i) f(\psi_{i-1}, x_i | \Psi^*_{i+1}) N^{(i)} \right]^{-1}.$$

Here $N^{(i,i-1)}$ is the effective sample size of a set of samples from $f^*_{i,i-1}(\psi_i)$. The estimator derived from $h^{(i)}_{O}$ is not substantially more difficult to compute than that derived from $h^{(i)}_{CJ}(\psi_i)$.

# 4 Conclusions

We have shown that the multi-block Metropolis-Hastings estimators of Chib and Jeliazkov (2001) are bridge estimators. By choosing the free functions in those general bridge estimation identities appropriately we arrive at the identity Chib and Jeliazkov use to define their estimators. Results in Meng and Wong (1996) and Chen and Shao (1997) lead to efficiency gains in estimation. Simulation results on very simple single block problems, which we do not report, show that efficiency gains do indeed result. For multi-block problems these gains are likely to be more strongly marked, since those bridge estimates involve products. Small efficiency gains accumulate from one factor to the next.

Clearly one need not use, in the simulations generating the samples $S_1$ and $S_2$, the same $q(x^*, x)$ one uses in the final estimators. One must have the normalisation of $q$ for the estimators, whereas this is not always needed in the simulation itself. It would be interesting to know to what extent good simulation-$q$'s make good estimation-$q$'s. This point is considered by Chib and Jeliazkov (2001).

Chib and Jeliazkov (2001) assume that the prior normalising constant is known or can be estimated separately. The alternatives we give above make the same assumption. In fact we regard this as undesirable compared to the simplest Meng-Wong estimator $\hat{r}_{h_{X|Y,X}}$ which takes $f^{(1)} = f_{X|Y}$ and $f^{(2)} = f_X$, and does not require an independent value for the prior normalising constant. When that information is available, $\hat{r}_{h_{X|Y,X}}$ can be beaten for efficiency. However, one would almost always compute $\hat{r}_{h_{X|Y,X}}$ in order to check the estimators we discuss above, since computer code calculating those estimators is exposed to a wider range of bugs.

# References


[1] Meng, X-L. and Wong, W.H. (1996), "Simulating ratios of normalising constants via a simple identity: a theoretical exploration," *Statistica Sinica*, 6, 831–860.

[2] Gelman, A. and Meng, X.L. (1998), "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling," *Statistical Science*, 13, 163–185.

[3] Chen,M-H. and Shao, Q-M. (1997), "Estimating ratios of normalizing constants for densities with different dimensions," *Statistica Sinica*, 7, 607–630.

[4] Green, P. J. (1995), "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711–732.

[5] Chib, S. and Jeliazkov, I. (2001), "Marginal likelihood from the Metropolis-Hastings output," *Journal of the American Statistical Association*. To appear.