Institute of Statistical Mathematics
Tokyo, December 2001

Bayesian variable selection and the Swendsen-Wang algorithm

by David Nott (University of New South Wales) and Peter Green (University of Bristol, P.J.Green@bristol.ac.uk).

- Bayesian variable selection
- Auxiliary variables in MCMC
- Swendsen-Wang and the Ising model
- Partial decoupling
- New algorithm for Bayesian variable selection

1. Bayesian variable selection

Let $\mathbf{y} = (y_1, ..., y_n)^T$ be a vector of responses, $\mathbf{X}$ be an $n \times p$ design matrix and consider a linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is a vector of parameters and $\boldsymbol{\epsilon} \sim N(0, \sigma^2\mathbf{I})$ is a vector of zero mean errors.

The problem of variable selection is to decide which predictors should be included in the model for the mean of the responses; the remainder are excluded (or equivalently, the corresponding β_j set to 0).

We take a Bayesian view, and make such inference simultaneously with inference on $\boldsymbol{\beta}$ and σ^2.

Let $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_p)^T$ be a binary vector, and write $q_\gamma = \sum_j \gamma_j$ for the number of nonzero elements of $\boldsymbol{\gamma}$. Let $\mathbf{X}_\gamma$ be the $n \times q_\gamma$ design matrix obtained by removing those columns j from $\mathbf{X}$ for which $\gamma_j = 0$. Similarly let $\boldsymbol{\beta}_\gamma$ be the subvector of $\boldsymbol{\beta}$ obtained by removing components β_j of $\boldsymbol{\beta}$ for which $\gamma_j = 0$.

We assume that

$$\mathbf{y}|\boldsymbol{\gamma}, \mathbf{X}_\gamma, \boldsymbol{\beta}_\gamma, \sigma^2 \sim N(\mathbf{X}\boldsymbol{\gamma}\boldsymbol{\beta}_\gamma, \sigma^2\mathbf{I}).$$

For Bayesian inference on the model parameters we use a hierarchical prior. The prior for $\boldsymbol{\beta}_\gamma$ given $\boldsymbol{\gamma}$ and σ^2 is normal,

$$p(\boldsymbol{\beta}_\gamma|\boldsymbol{\gamma}, \sigma^2) \sim N(0, n\sigma^2(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)^{-1}).$$

The prior on σ^2 is $p(\sigma^2) \propto \sigma^{-2}$, and for our prior on $\boldsymbol{\gamma}$ we use $p(\boldsymbol{\gamma}) = 2^{-p}$, so that all models have equal prior probability. For alternative prior specifications on γ that encourage model parsimony see Denison *et al.* (1998) and Kohn *et al.* (2001).

We are interested in the posterior distribution on $\boldsymbol{\gamma}$ with $\boldsymbol{\beta}_\gamma$ and σ^2 integrated out,

$$p(\boldsymbol{\gamma}|\mathbf{y}) \propto p(\boldsymbol{\gamma})p(\mathbf{y}|\boldsymbol{\gamma}).$$

It can be shown that

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\gamma}) &\propto (1+n)^{-q_\gamma/2} \\ &\times \left(\mathbf{y}^T\mathbf{y} - \frac{n}{n+1}\mathbf{y}^T\mathbf{X}_\gamma(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)^{-1}\mathbf{X}_\gamma^T\mathbf{y}\right)^{-n/2} \end{aligned}$$

For p relatively small, we can compute the posterior $p(\boldsymbol{\gamma}|\mathbf{y})$ exactly, obtaining the normalising constant by summing over all possible values of $\boldsymbol{\gamma}$. For large p, this is not feasible due to the number of terms in the sum, and we use Markov chain Monte Carlo algorithms to identify high posterior probability models.

MCMC: the basic idea

Given a distribution of interest $\pi(\boldsymbol{\theta})$

(*in statistical analysis, more often than not a Bayesian posterior*)

- construct a Markov chain $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots$, whose invariant distribution is π,
- simulate it, and
- treat the realisation as a sample from π.

For example, estimate

$$P_\pi(\boldsymbol{\theta} \in A) \quad \text{by} \quad \frac{\#\{t \leq N : \boldsymbol{\theta}^{(t)} \in A\}}{N}$$

or

$$E_\pi(g(\boldsymbol{\theta})) \quad \text{by} \quad \frac{\sum_{t=1}^{N} g(\boldsymbol{\theta}^{(t)})}{N}$$

The usual MCMC approach to variable selection

Typically, we update one component of $\boldsymbol{\gamma}$ at a time (that is, add or delete single terms from the model), by a Metropolis method.

When columns of **X** are highly collinear, this sampler mixes very slowly.

2. Improving performance of MCMC by augmenting the state space

Perhaps counter-intuitively, it is sometimes possible to improve MCMC performance by augmenting the state vector to include additional components. Two successful recipes are those in which the original model appears as a *conditional* distribution in an augmented model (simulated tempering) and in which it appears as a *marginal* (auxiliary variables).

2a. Simulated tempering

Combat slow mixing by embedding desired model in a family of models, indexed say by α, and treat α now as an additional dynamic variable. Design the family so that for some α, the chain mixes much better.

$$\pi(\boldsymbol{x}) \Rightarrow \pi^\star(\boldsymbol{x}, \alpha_0)$$

Run MCMC on $\pi^\star(\boldsymbol{x}, \alpha)$, and condition on $\alpha = \alpha_0$ by selecting from the output.

Simulated tempering, by changing the temperature

This was the original Marinari / Parisi idea; we set

$$\pi^\star(\theta, \alpha) \propto \{\pi(\theta)\}^\alpha$$

where $\alpha = \alpha_0 = 1$ corresponds to the original model, and $\alpha \to 0$ makes the probability surface 'flatter', or in physical terms, 'warmer'.

The full conditionals change in the same way:

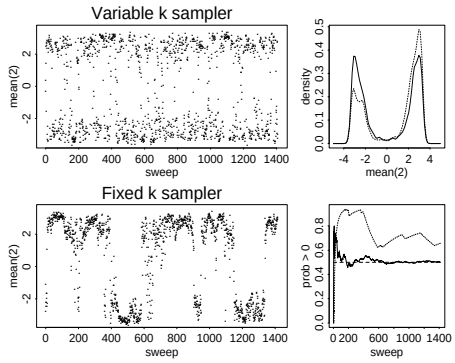
$$\pi^\star(\theta_i|\theta_{-i}, \alpha) \propto \{\pi(\theta_i|\theta_{-i})\}^\alpha$$

so implementation is very easy.

We place a (discrete) artificial prior on α so that the *marginal* for α is approximately uniform.

Simulated tempering, by inventing models

An example from mixture analysis: allowing the number of components to vary gives much better mixing.



2b. Auxiliary variables

Edwards and Sokal (1988) proposed a way to improve mixing by augmenting the state space so that the original target appears as the *marginal* equilibrium distribution. The following interpretation of their approach in statistical language can be found in Besag and Green (*JRSS(B)*, 1993).

Starting from $\pi(\boldsymbol{x})$, take some additional variables u, with $\pi(u|\boldsymbol{x})$ arbitrarily chosen. Then the joint is $\pi(\boldsymbol{x}, u) = \pi(\boldsymbol{x})\pi(u|\boldsymbol{x})$, for which $\pi(\boldsymbol{x})$ is certainly the marginal for $\boldsymbol{x}$.

We could now run a MCMC method for the *joint* target $\pi(\boldsymbol{x}, u)$ (usually a method that updates $\boldsymbol{x}$ and u alternately), and simple ignore the u variable in extracting information from the simulation.

When might this idea be useful? Suppose $\pi(\boldsymbol{x})$ factorises as:

$$\pi(\boldsymbol{x}) = \pi_0(\boldsymbol{x})b(\boldsymbol{x})$$

where $\pi_0(\boldsymbol{x})$ is a (possibly unnormalised) distribution that is easy to simulate from, and $b(\boldsymbol{x})$ is the awkward part, often representing the 'interactions' between variables that are slowing down the chain.

Then take a one-dimensional u with $u|\boldsymbol{x} \sim U[0, b(\boldsymbol{x})]$: we find

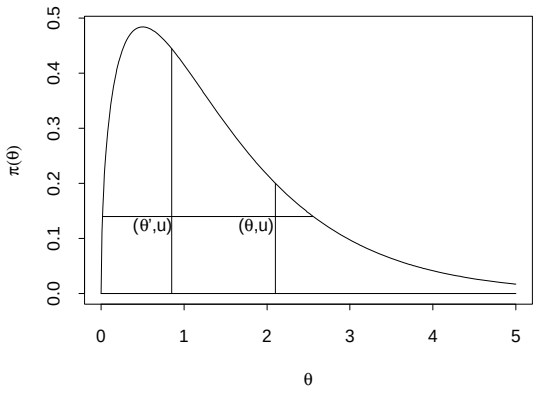
$$\pi(\boldsymbol{x}, u) = \pi(\boldsymbol{x})\pi(u|\boldsymbol{x}) = \pi_0(\boldsymbol{x})b(\boldsymbol{x})\frac{I[0 \le u \le b(\boldsymbol{x})]}{b(\boldsymbol{x})}$$

so that

$$\pi(\boldsymbol{x}|u) \propto \pi_0(\boldsymbol{x})$$

restricted to (conditional on) the event $\{\boldsymbol{x} : b(\boldsymbol{x}) \ge u\}$. At least when this $\pi(\boldsymbol{x}|u)$ can be sampled without rejection, we can easily implement a Gibbs sampler, drawing u and $\boldsymbol{x}$ in turn.

This method has recently been popularised under the name of the 'slice sampler', reflecting the fact that if $\pi_0(\boldsymbol{x}) =$ constant, not only $\pi(u|\boldsymbol{x})$ but also $\pi(\boldsymbol{x}|u)$ are uniform distributions; the latter corresponding to a horizontal slice through the graph of $\pi(\boldsymbol{x})$.



For statistical applications of the idea, see Neal (1997) and Damien, Wakefield and Walker (1999), and for a detailed analysis of the method, see Roberts and Rosenthal (1999).

3. Swendsen-Wang and the Potts and Ising models

The (symmetric) Potts model on an arbitrary graph (V, E) is the probability distribution for a random vector $\boldsymbol{x}$ with components indexed by V and values in a finite set of *colours* $\mathcal{C}$, given by

$$\pi(\boldsymbol{x}) \propto \exp\left\{\beta \sum_{(v,w)\in E} I[\boldsymbol{x}_v = \boldsymbol{x}_w]\right\} = \prod_{e\in E} b_e(\boldsymbol{x}).$$

The Ising model is the special case where $\mathcal{C}$ has just 2 elements, say $\{0, 1\}$ or {white,black}. For now, we assume $\beta > 0$.

There are also asymmetric versions, with an additional multiplicative term $\exp(\sum_{v\in V} \alpha(\boldsymbol{x}_v))$, and inhomogeneous relatives, where $\alpha(\cdot)$ and / or β can vary with v or (v, w).

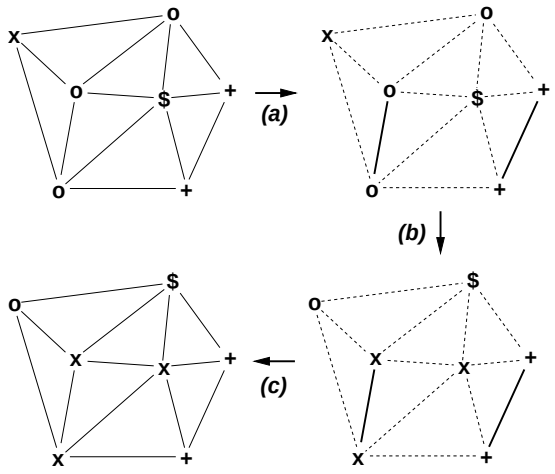
The Swendsen-Wang method is an auxiliary variable MCMC method for these models.

We define one auxiliary variable u_e for each edge e, conditionally independent given $\boldsymbol{x}$, with $u_e|\boldsymbol{x} \sim U(0, b_e(\boldsymbol{x}))$. If $u_e > e^{-\beta}$ we say the edge e is 'on', otherwise 'off'.

It is easy to see that in drawing $\boldsymbol{u}$ given $\boldsymbol{x}$, edges are on with probability $1 - e^{-\beta}$ if $\boldsymbol{x}_v = \boldsymbol{x}_w$, always off if $\boldsymbol{x}_v \neq \boldsymbol{x}_w$.

Simple manipulation shows that $\pi(\boldsymbol{x}|\boldsymbol{u})$ is a random uniform colouring on the clusters determined by the on bonds.

Illustrating the Swendsen-Wang algorithm for a small graph



Illustrating the Swendsen-Wang algorithm: (a) bond variables between like-coloured nodes are 'on' with probability $1 - e^{-\beta}$, always 'off' between unlike-coloured ones; (b) clusters formed by 'on' bonds are re-coloured uniformly at random; (c) the new colouring.

The original applications of auxiliary variable methods were to statistical physics problems, where in particular the Swendsen-Wang method (Swendsen and Wang, 1987) has had a profound influence; see also Edwards and Sokal (1988) and Sokal (1989).

4a. Partial decoupling

Attempts to exploit the idea of auxiliary variables in other statistical models have been less successful, chiefly because $\pi(\boldsymbol{x}|\boldsymbol{u})$ often cannot be directly simulated, so that computational effort is not reduced overall.

The idea of partial decoupling is to choose auxiliary variables that while not completing 'killing' interactions, nevertheless do *reduce* their effect, without making the computation more complicated.

Thus, in general, we might have

$$\pi(\boldsymbol{x}) = \pi_0(\boldsymbol{x}) b(\boldsymbol{x})$$

and use

$$u_e|\boldsymbol{x} \sim U(0, c_e(\boldsymbol{x}))$$

independently, where we no longer require $b(\boldsymbol{x}) = \prod_e c_e(\boldsymbol{x})$.

The resulting conditional distribution is

$$\pi(\boldsymbol{x}|\boldsymbol{u}) \propto \pi_0(\boldsymbol{x}) \frac{b(\boldsymbol{x})}{\prod_e c_e(\boldsymbol{x})}.$$

With a suitable choice of $\{c_e(\boldsymbol{x})\}$, MCMC updates in detailed balance with this conditional distribution will mix quickly yet be computationally inexpensive.

4b. Ising models with negative interactions

If $\mathcal{C} = \{0, 1\}$, and

$$\pi(\boldsymbol{x}) \propto \exp\left\{ \sum_{(v,w)\in E} \beta_{vw} I[\boldsymbol{x}_v = \boldsymbol{x}_w] \right\}$$

where some of the β_{vw} are negative, the Swendsen-Wang idea still applies, except that the direction of the association is reversed, and hence a cluster is interpreted differently.

If $\beta_{vw} < 0$ and $u_{vw} > \exp(\beta_{vw})$, then in updating $\boldsymbol{x}$ given $\boldsymbol{u}$, $\boldsymbol{x}_v$ and $\boldsymbol{x}_w$ are constrained to be *different*.

5. Application to variable selection

We apply the ideas of auxiliary variables with partial decoupling, allowing negative interactions, to the variable selection problem by identifying

$$\pi(\boldsymbol{x}) \quad \text{with} \quad p(\boldsymbol{\gamma}|\mathbf{y})$$

Thus we treat the indices of the regression variables as 'spatial' locations, and try to approximate the posterior distribution of the indicator variables γ_j by an (inhomogeneous) Ising model.

The auxiliary variables u_{ij} define clusters among the components of $\boldsymbol{\gamma}$ as in the Swendsen-Wang algorithm. Components of $\boldsymbol{\gamma}$ within the same cluster satisfy a set of constraints of the form $\gamma_i = \gamma_j$ or $\gamma_i \neq \gamma_j$. (The constraints always have at least one feasible solution, since they are created based on the current value for $\boldsymbol{\gamma}$.)

Let $C = C(\mathbf{u})$ be one cluster defined by the set of auxiliary variables $\mathbf{u}$, and let $\overline{C}$ denote the set of variables not in C. Let $\boldsymbol{\gamma}(C)$ be the subset of $\boldsymbol{\gamma}$ corresponding to the variables in C, and $\boldsymbol{\gamma}(\overline{C})$ denote the remaining components of $\boldsymbol{\gamma}$. Given the constraints, note that there are only two possible values for the vector $\boldsymbol{\gamma}(C)$: from one possible value we can obtain the other by "flipping" the ones to zeros and zeros to ones within the cluster C.

In general, we can update $\boldsymbol{\gamma}(C)$ by a Metropolis-Hastings step. Write $\boldsymbol{\gamma}^{new}$ for a proposed value of $\boldsymbol{\gamma}$ in which $\boldsymbol{\gamma}^{new}(C)$ is generated from the proposal distribution $q(\boldsymbol{\gamma}(C)|\boldsymbol{\gamma}, \mathbf{u}, \mathbf{y})$ and $\boldsymbol{\gamma}^{new}(\overline{C}) = \boldsymbol{\gamma}(\overline{C})$. The Metropolis-Hastings acceptance probability is

$$\min \quad \left\{ 1, \frac{q(\boldsymbol{\gamma}(C)|\boldsymbol{\gamma}, \mathbf{u}, \mathbf{y}) p(\boldsymbol{\gamma}^{new}|\mathbf{y})}{q(\boldsymbol{\gamma}^{new}(C)|\boldsymbol{\gamma}, \mathbf{u}, \mathbf{y}) p(\boldsymbol{\gamma}|\mathbf{y})} \right.$$
$$\left. \times \quad \exp\left(\sum_{i<j} \beta_{ij} (I(\gamma_i = \gamma_j) - I(\gamma_i^{new} = \gamma_j^{new}) \right) \right\}.$$

Since $\boldsymbol{\gamma}^{new}(\overline{C}) = \boldsymbol{\gamma}(\overline{C})$ we can simplify this expression by noting that

$$\exp\left(\sum_{i<j} \beta_{ij} (I(\gamma_i = \gamma_j) - I(\gamma_i^{new} = \gamma_j^{new})) \right)$$
$$= \quad \exp\left(\sum_{(i,j)\in\partial C} \beta_{ij} (I(\gamma_i = \gamma_j) - I(\gamma_i^{new} = \gamma_j^{new})) \right)$$

where

$$\partial C = \{(i, j) : i < j \text{ and exactly one of } i, j \in C\}.$$

This is inexpensive to compute provided that β_{ij} is nonzero only for a fairly small number of pairs $(i, j) \in \partial C$.

Baseline methods for comparison

A special case: fix $\beta_{ij} \equiv 0$, and then draw from the full conditional for $\boldsymbol{\gamma}(C)|\boldsymbol{\gamma}(\overline{C}), \mathbf{u}, \mathbf{y}$ – that is, the Gibbs sampler.

In fact, this is easily improved. Let

$$q = Pr(\boldsymbol{\gamma}^c(C)|\boldsymbol{\gamma}^c(\overline{C}), u, y)$$

be the probability that $\boldsymbol{\gamma}(C)$ stays at its current value (given the current $\boldsymbol{\gamma}(\overline{C})$ and $\mathbf{u}$) with this Gibbs move.

We will flip to the opposite state for $\boldsymbol{\gamma}(C)$ with probability

$$\min\left(1, \frac{1-q}{q}\right),$$

instead of $1 - q$.

It is easy to show that in this case detailed balance is maintained. This changes state more often than Gibbs, and so by Peskun (1973) is superior in terms of asymptotic variance. We call this the 'antithetic' proposal.

Obtaining the interaction parameters

Compromises must be struck in designing the approximating Ising field (that is, choosing the β_{ij}).

If many $\beta_{ij} \neq 0$, then the acceptance probabilities will be expensive to compute; it is therefore better to truncate small $|\beta_{ij}|$ to 0.

Further, following Higdon (*JASA*, 1998), the nonzero $|\beta_{ij}|$ should not be too large, since this will mean that clusters are typically large, so that strong likelihood ratios will inhibit 'flipping'.

Our starting point for choosing β_{ij} is to note that if we have a binary random field with

$$\begin{aligned} \pi(\boldsymbol{\gamma}) &\propto \exp(\sum_{i \in V} \alpha(\boldsymbol{\gamma}_i)) \\ &\times \exp\left\{\sum_{(i,j) \in E} \beta_{ij} I[\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_j]\right\} \end{aligned}$$

(an inhomogeneous Ising model), then for any fixed configuration $\boldsymbol{\gamma}^\star$,

$$\beta_{ij} = 0.5\left(\sum_{=} \log \pi(\boldsymbol{\gamma}) - \sum_{\neq} \log \pi(\boldsymbol{\gamma})\right) \qquad (1)$$

where $\sum_{=}$ (respectively, $\sum_{\neq}$) are sums over $\boldsymbol{\gamma}$ such that $\boldsymbol{\gamma}_k = \boldsymbol{\gamma}_k^\star$ for all $k \neq i, j$, and $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_j$ (resp., $\boldsymbol{\gamma}_i \neq \boldsymbol{\gamma}_j$).

It is also true that for such Ising models on p variables,

$$\beta_{ij} = 0.5^{p-1}\left(\sum_{\boldsymbol{\gamma}:\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_j} \log \pi(\boldsymbol{\gamma}) - \sum_{\boldsymbol{\gamma}:\boldsymbol{\gamma}_i \neq \boldsymbol{\gamma}_j} \log \pi(\boldsymbol{\gamma})\right). \qquad (2)$$

Both of these identities seem to provide plausible approximating β_{ij} for arbitrary binary random fields, although of course the latter involves unfeasibly large summations in the case of interest where p is large.

4 methods

In each case we use the 'antithetic' proposal.

A $\beta_{ij} \equiv 0$

B Use (2), scale to interval [-1,1], then truncate if $|\beta_{ij}| < 0.1$

C Use (1) with $\gamma_j^\star \equiv 1$, scale to interval [-1,1], then truncate if $|\beta_{ij}| < 0.1$. (This baseline ensures that all predictors are included in capturing correlation between the indicators).

D A more complicated method based on an eigenvalue decomposition of the scaled sum-of-products matrix, followed again by scaling and truncation.

Example 1: George and McCulloch simulated data set

They simulated a data set with 15 predictor variables as follows. Let $\mathbf{Z}_1, ..., \mathbf{Z}_{15}, \mathbf{Z} \sim N_{180}(0, \mathbf{I})$. Then let $\mathbf{X}_i = \mathbf{Z}_i + 2\mathbf{Z}$, $i = 1, 3, 5, 9, 10, 12, 13, 14, 15$ and set $\mathbf{X}_2 = \mathbf{X}_1 + 0.15\mathbf{Z}_2$, $\mathbf{X}_4 = \mathbf{X}_3 + 0.15\mathbf{Z}_4$, $\mathbf{X}_6 = \mathbf{X}_5 + 0.15\mathbf{Z}_6$, $\mathbf{X}_7 = \mathbf{X}_8 + \mathbf{X}_9 - \mathbf{X}_{10} + 0.15\mathbf{Z}_7$ and $\mathbf{X}_{11} = \mathbf{X}_{14} + \mathbf{X}_{15} - \mathbf{X}_{12} - \mathbf{X}_{13} + 0.15\mathbf{Z}_{11}$.

This construction results in severe and complicated multicollinearity: there is a correlation of about 0.998 between $\mathbf{X}_i$ and $\mathbf{X}_{i+1}$, $i = 1, 3, 5$ and strong linear dependencies among $(\mathbf{X}_7, \mathbf{X}_8, \mathbf{X}_9, \mathbf{X}_{10})$ and $(\mathbf{X}_{11}, \mathbf{X}_{12}, \mathbf{X}_{13}, \mathbf{X}_{14}, \mathbf{X}_{15})$. Let $\mathbf{X}$ be the design matrix with columns $\mathbf{X}_i$, $i = 1, ..., 15$. Let

$$\boldsymbol{\beta} = (1.5, 0, 1.5, 0, 1.5, 0, 1.5, -1.5, 0, 0, 1.5, 1.5, 1.5, 0, 0)^T$$

and generate the responses $\mathbf{Y}$ as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim N_{180}(0, 2.5^2\mathbf{I})$.

Relative improvements in Monte Carlo standard errors for $\overline{\gamma}_i$ for simulated example for methods B, C and D, relative to method A. Estimates are based on 50,000 iterations from two different starting points for each sampling scheme with $1,000$ iterations burn in.

Predictor	B	C	D
X_2	0.9244	1.1691	1.2619
X_3	3.0625	3.6750	3.2667
X_4	2.7170	3.4286	3.7403
X_6	0.7750	0.8424	1.2500
X_7	1.7966	10.6600	8.8833
X_8	1.7418	13.2667	10.3377
X_9	1.7841	12.4609	10.9247
X_{10}	1.7171	12.4286	9.7875
X_{11}	2.7088	10.8769	8.9494
X_{12}	2.5316	9.0187	9.0187
X_{13}	2.5647	9.5067	11.3175
X_{14}	2.6346	13.9796	13.1731
X_{15}	2.6766	15.4111	14.7553

Example 2: US crime rates

Data set discussed by Ehrlich (1973) and Raftery (1995).

47 states, response = crime rate.

15 predictors, with several strong collinearities among them.

Predictors for US crime data set

Predictor	Description
M	percentage of males aged 14-24
So	indicator variable for a southern state
Ed	mean years of schooling
Po1	police expenditure in 1960
Po2	police expenditure in 1959
LF	labour force participation rate
M.F	number of males per 1000 females
Pop	state population
NW	number of nonwhites per 1000 people
U1	unemployment rate of urban males 14–24
U2	unemployment rate of urban males 35–39
GDP	gross domestic product per head
Ineq	income inequality
Prob	probability of imprisonment
Time	average time served in state prisons

Relative improvements in Monte Carlo standard errors for $\overline{\gamma}_i$ for US crime example for methods B, C and D, relative to method A.

Predictor	B	C	D
M	1.1895	1.6259	1.4967
So	1.1535	1.6871	1.7714
Ed	1.1397	0.9761	1.1027
Po1	1.0389	1.2511	1.3868
Po2	1.7127	2.0000	1.9745
LF	1.4508	2.3289	1.9239
M.F	1.1236	1.2422	1.0471
Pop	0.9274	1.0550	1.1443
NW	0.8587	0.8827	1.0327
U1	1.4809	1.3759	1.3288
U2	1.0645	1.1250	1.2857
GDP	1.0820	1.7069	1.1061
Ineq	1.0886	0.8643	0.9053
Prob	1.3725	1.3642	1.2139
Time	1.0571	1.7619	1.3136

Example 3: Statistical correction of a numerical weather prediction model

Responses: daily maximum temperatures on 369 dayes in August–October, 1993–1996.

62 predictors: averages of 24hr and 36hr meteorological fields from a numerical weather prediction model.

Monte Carlo standard errors for $\overline{\gamma}_i$ for numerical weather prediction model data for methods A, C and D. Estimates are based on 200,000 iterations for each method from two different starting methods with $1,000$ iterations burn in for each sequence. The columns labelled “Relative” for methods C and D give relative improvements of the Monte Carlo standard errors for these methods compared to that for method A.

	Method				
Predictor	A	C		D	
	SE($\overline{\gamma}_i$)	SE($\overline{\gamma}_i$)	Relative	SE($\overline{\gamma}_i$)	Relative
X_6	0.0237	0.0234	1.0107	0.0183	1.2951
X_{26}	0.0194	0.0205	0.9463	0.0160	1.2125
X_{31}	0.0153	0.0125	1.2240	0.0178	0.8596
X_{49}	0.0158	0.0138	1.1449	0.0101	1.5644
X_{56}	0.0152	0.0186	0.8172	0.0195	0.7815
X_{57}	0.0128	0.0090	1.4144	0.0118	1.0894

To follow up

This paper:

`http://www.stats.bris.ac.uk/~peter/papers/bvsswang.ps`

This talk:

`http://www.stats.bris.ac.uk/~peter/slides/tokyo.ps`

My web page:

`http://www.stats.bris.ac.uk/~peter`

My email address:

`P.J.Green@bristol.ac.uk`