Advanced Simulation Methods

Chapter 7 - Reversible Jump MCMC

In this chapter, we review some computational questions arising in model choice. We then introduce the reversible jump algorithm, a generic MCMC algorithm to sample from distributions which are defined on a union of subspaces of different dimensions. This methodology has had a huge impact in statistics since its introduction by Peter Green in 1995.

1 Bayesian model selection

Assume you have a countable set $\{\mathcal{M}_1, \mathcal{M}_2, ...\}$ of Bayesian models to describe some data y. To each Bayesian model $\mathcal{M}_k$ is associated a random parameter θ_k of prior density $p(\theta_k \mid \mathcal{M}_k)$ on the parameter space Θ_k and a likelihood function $\mathcal{L}(y; \theta_k, \mathcal{M}_k)$. Hence the posterior density on the parameters associated to model $\mathcal{M}_k$ is defined on Θ_k and has density

$$\pi(\theta_k \mid y, \mathcal{M}_k) = \frac{\mathcal{L}(y; \theta_k, \mathcal{M}_k)\, p(\theta_k \mid \mathcal{M}_k)}{p(y \mid \mathcal{M}_k)}$$

where

$$p(y \mid \mathcal{M}_k) = \int_{\Theta_k} \mathcal{L}(y; \theta_k, \mathcal{M}_k)\, p(\theta_k \mid \mathcal{M}_k) d\theta_k$$

is usually called the marginal likelihood or “evidence”.

Here we are interested in performing Bayesian inference about the model. To achieve this, we need to specify a prior distribution on $\{\mathcal{M}_1, \mathcal{M}_2, ...\}$; i.e.

$$p(\mathcal{M}_k) = \mathbb{P}(\mathcal{M} = \mathcal{M}_k).$$

Now Bayesian inference on the model and parameter relies on the joint posterior

$$\pi(\mathcal{M}_k, \theta_k \mid y) = \frac{\mathcal{L}(y; \theta_k, \mathcal{M}_k)\, p(\theta_k \mid \mathcal{M}_k) p(\mathcal{M}_k)}{\sum_j \left(\int_{\Theta_j} \mathcal{L}(y; \theta_j, \mathcal{M}_j)\, p(\theta_j \mid \mathcal{M}_j) d\theta_j \right) p(\mathcal{M}_j)}$$

which is defined on the space

$$\Theta = \bigcup_k \{k\} \times \Theta_k.$$

From this joint posterior, we can also obtain the marginal on the models by integrating out the parameters,

$$p(\mathcal{M}_k \mid y) = \frac{p(y \mid \mathcal{M}_k) p(\mathcal{M}_k)}{\sum_j p(y \mid \mathcal{M}_j) p(\mathcal{M}_j)}$$

and the comparison between two models $\mathcal{M}_i$ and $\mathcal{M}_j$ can be summarised by standard posterior odds

$$\frac{p(\mathcal{M}_i \mid y)}{p(\mathcal{M}_j \mid y)} = \frac{p(y \mid \mathcal{M}_i) p(\mathcal{M}_i)}{p(y \mid \mathcal{M}_j) p(\mathcal{M}_j)}.$$

This ratio is usually called the Bayes factor for models $\mathcal{M}_i$ and $\mathcal{M}_j$.

2 Model evidence estimation using standard Monte Carlo techniques

We could try to use standard Monte Carlo to estimate $p(y \mid \mathcal{M}_k)$ for each model $\mathcal{M}_k$. Note that we can write

$$p(y \mid \mathcal{M}_k) = \int_{\Theta_k} \mathcal{L}(y; \theta_k, \mathcal{M}_k)\, p(\theta_k \mid \mathcal{M}_k) d\theta_k$$

and hence we can sample independently $\theta^{(1)}, \ldots \theta^{(n)}$ from the prior distribution $p(d\theta_k \mid \mathcal{M}_k)$ and approximate the evidence by

$$\frac{1}{n}\sum_{t=1}^{n} \mathcal{L}\left(y; \theta^{(t)}, \mathcal{M}_k\right).$$

This yields a consistent Monte Carlo estimator (as $n \to \infty$) but might have a large variance if the likelihood is peaked compared to the prior distribution, because then most sampled points will not contribute much to the estimate. Alternatively we can also write

$$\begin{aligned} p(\theta_k \mid y, \mathcal{M}_k) &= \frac{\mathcal{L}(y \mid \theta_k, \mathcal{M}_k) p(\theta \mid \mathcal{M}_k)}{p(y \mid \mathcal{M}_k)} \\ \Leftrightarrow (\mathcal{L}(y \mid \theta_k, \mathcal{M}_k))^{-1} p(\theta_k \mid y, \mathcal{M}_k) &= \frac{p(\theta \mid \mathcal{M}_k)}{p(y \mid \mathcal{M}_k)} \end{aligned}$$

so that, integrating θ_k on both sides we obtain

$$\int (\mathcal{L}(y \mid \theta_k, \mathcal{M}_k))^{-1} p(\theta_k \mid y, \mathcal{M}_k) d\theta_k = \frac{1}{p(y \mid \mathcal{M}_k)},$$

from which you can obtain the “harmonic mean” estimator (“harmonic” is to stay betwen quotes as its variance is typically infinite!). The harmonic estimator is

$$\left(\frac{1}{n}\sum_{t=1}^{n} \mathcal{L}\left(y; \theta^{(t)}, \mathcal{M}_k\right)^{-1}\right)^{-1}$$

where $\theta^{(1)}, \ldots, \theta^{(n)}$ is approximately from the posterior distribution $p(\theta_k \mid y, \mathcal{M}_k)$. For instance you can run a Metropolis-Hastings algorithm to obtain the sample, and if you store the likelihood evaluations $\mathcal{L}\left(y; \theta^{(t)}, \mathcal{M}_k\right)$ along the way, you obtain the harmonic estimator for free.

In general standard tools perform poorly at estimating the model evidence $p(y \mid \mathcal{M}_k)$, which is currently considered a very challenging problem. Moreover if the collection of models is countably infinite, then clearly MCMC cannot be performed for each model separately. Alternatively, we could envision an MCMC method that samples directly from the joint distribution $\pi(\mathcal{M}_k, \theta_k \mid y)$. However as Θ is a union of spaces of different dimensions, there are measure-theoretic subtleties that prevent the direct use of the previously described algorithms such as Metropolis-Hastings. Reversible Jump MCMC is a methodology that allows to sample from $\pi(\mathcal{M}_k, \theta_k \mid y)$, by extending Metropolis-Hastings; see [1] for the original article and Chapter 11 in [2].

3 Reversible Jump Markov chain Monte Carlo

The reversible jump algorithm will alternatively use “within model” kernels and “between-models” kernels. Starting from model index $k^{(0)}$ and parameter $\theta^{(0)} \in \Theta_{k^{(0)}}$, where do we go? Either we propose a “within model” move or we propose a move to another model. For within model moves we can use a standard Metropolis-Hastings algorithm, so the question now is about “between-models” moves. We are interested in designing a Markov kernel moving (k, θ_k) to $(k', \theta_{k'})$.

3.1 Reversible Markov kernel across dimensions

We will show how to design a reversible Markov kernel P allowing moves “across models” and leaving the joint distribution $\pi(dk, d\theta_k)$ invariant. We will actually construct P so that it satisfies the *reversibility* condition, in the sense that for all $A = \cup_{k \in K_A} \{k\} \times A_k$ and $B = \cup_{k' \in K_B} \{k'\} \times B_{k'}$

$$\int_{((k,\theta_k),(k',\theta'_{k'}))\in A\times B} \pi(dk, d\theta_k) P((k,\theta_k), d(k', \theta'_{k'})) = \int_{((k,\theta_k),(k',\theta'_{k'}))\in A\times B} \pi(dk', d\theta'_{k'}) P((k',\theta'_{k'}), d(k, \theta_k)). \tag{1}$$

Note also that both integrals in Eq. 1 are on a common space of dimension $2 + \dim(\theta_k) + \dim(\theta'_{k'})$, and that $\dim(\theta_k)$ could be different from $\dim(\theta'_{k'})$, which is the whole point. Reversibility will imply that P leaves $\pi(dk, d\theta_k)$ invariant. For simplicity, let us write $x = (k, \theta_k)$, living in some space $\mathcal{X}$; we will return to (k, θ) later.

A quick remark on the measure-theoretic notation: here we note $\pi(dx)P(x, dx')$ instead of $\pi(x)P(x, x')dxdx'$, for the good reason that the measure $\pi(dx)P(x, dx')$ might not admit a density with respect to a dominating measure. We will go back to that point at the end of the section. It is enough to understand (or admit!) that the argument x of π might be of varying dimension, and that P is a Markov kernel across spaces of (potentially) varying dimensions.

Mimicking the Metropolis-Hastings algorithm, let us look for a Markov kernel P such that 1) first a candidate x' is drawn from a distribution $q(x \to dx')$, 2) x' is accepted with some probability $a(x \to x')$; otherwise the previous state x is taken as the new state.

Thus the kernel takes the form

$$P(x, dx') = q(x \to dx')a(x \to x') + (1 - a(x))\delta_x(dx'),$$

where $1 - a(x)$ denotes the probability of rejecting a candidate given a current point x, i.e.

$$a(x) = \int_{\mathcal{X}} q(x \to dx')a(x \to x').$$

Hence we can rewrite Eq. 1 as

$$\int_{\mathcal{A}\times\mathcal{B}} \pi(dx)q(x \to dx')a(x \to x') \tag{2}$$

$$= \int_{\mathcal{A}\times\mathcal{B}} \pi(dx')q(x' \to dx')a(x' \to x). \tag{3}$$

Indeed from Eq. 1 we can subtract the term corresponding to

$$\int_{\mathcal{A}\cap\mathcal{B}} \pi(dx)P(x, \{x\}) = \int_{\mathcal{A}\times\mathcal{B}} \pi(dx)(1 - a(x))\delta_x(dx')$$

i.e. the case $x = x'$, corresponding to rejection of the proposal, since that same term appearing on both sides.

Now the integrals in Eq. 2 and Eq. 3 are nicer than the ones in Eq. 1: indeed we can assume that $\pi(dx)q(x \to dx')$, seen as a measure on $\mathcal{X}^2$, admits a dominating measure, written $dx\,dx'$. Thus we can simply rewrite Eq. 2-3 as

$$\int_{\mathcal{A}\times\mathcal{B}} \pi(x)q(x \to x')a(x \to x')dxdx'$$

$$= \int_{\mathcal{A}\times\mathcal{B}} \pi(x')q(x' \to x)a(x' \to x)dxdx'.$$

Without going into measure theoretical details, the reason why we can assume this, and why we couldn't assume the existence of a dominating measure on $\pi(x)P(x, dx')$, is because of the point mass in P; we can assume that q does not have such a point mass.

Now that our measure theoretical difficulties are over, we can look for a way to construct $q(x \to x')$ and $a(x \to x')$ such that the terms in Eq. 2 and Eq. 3 are equal, when x and x' are of different dimensions.

3.2 Constructing q and a using dimension matching and deterministic mappings

For clarity, we now come back to the original notation $(k, \theta) = x$. We write $\pi(dx) = \pi(k, \theta)d\theta$ and $q(x \to dx') = q(k \to k')q_{k\to k'}(\theta \to \theta')d\theta'$. The idea of dimension matching is to extend θ and θ' with auxiliary variables u and u' such that the extended variables are of common dimension. Hence introduce u drawn from (an arbitrary) $\varphi_{k\to k'}$ and u' drawn from (an arbitrary) $\varphi_{k'\to k}$ such that

$$\dim(\theta) + \dim(u) = \dim(\theta') + \dim(u').$$

To construct the proposal with density $q_{k\to k'}(\theta \to \theta')$, we first draw $u \sim \varphi_{k\to k'}$. We then introduce a diffeomorphism $G_{k\to k'}$ taking a couple (θ, u) and returning $(\theta', u') = G_{k\to k'}(\theta, u)$. A diffeomorphism is simply a differentiable function, with an inverse that is also differentiable. We will also write θ' as $\theta'(\theta, u)$ and u' as $u'(\theta, u)$, to emphasize that it is a deterministic mapping. Then we can write for any sets A, B:

$$\begin{aligned}
&\sum_{k,k' \in K_A \times K_B} \int_{(\theta,\theta') \in A_k \times B_{k'}} \pi(k,\theta) q(k \to k') q_{k \to k'}(\theta \to \theta') a(\theta \to \theta') d\theta d\theta' \\
&= \sum_{k,k' \in K_A \times K_B} \int_{(\theta,\theta'(\theta,u)) \in A_k \times B_{k'}} \pi(k,\theta) q(k \to k') \varphi_{k \to k'}(u) a(\theta \to \theta'(\theta,u)) d\theta du,
\end{aligned}$$

where note that we keep writing the constraint $\theta'(\theta, u) \in B_{k'}$ in terms of θ' and not in terms of u; it is equivalent to a constraint on u because $G_{k\to k'}$ is a diffeomorphism. We could instead write $(\theta, u) \in A_k \times U_{k'}$ where $U_{k'}$ is such that $(\theta, \theta'(\theta, u)) \in A_k \times B_{k'} \Leftrightarrow (\theta, u) \in A_k \times U_{k'}$.

On the other hand, we can derive a similar equality for the term in Eq. 3. Given θ' we propose to first sample $u' \sim \varphi_{k'\to k}$ and apply the inverse of $G_{k\to k'}$, denoted by $G_{k'\to k}$, to obtain $(\theta, u) = G_{k'\to k}(\theta', u')$. Hence we obtain the term

$$\sum_{k,k' \in K_A \times K_B} \int_{(\theta(\theta',u'),\theta') \in A_k \times B_{k'}} \pi(k',\theta') q(k' \to k) \varphi_{k' \to k}(u') a(\theta' \to \theta(\theta',u')) d\theta' du'.$$

To equate both integrands in Eq. 2 and Eq. 3, we perform a change of variables for the latter integral. We will transform $d\theta' du'$ to $d\theta du$, and the Jacobian of the transformation $G_{k\to k'}$ appears:

$$\begin{aligned}
&\sum_{k,k' \in K_A \times K_B} \int_{(\theta(\theta',u'),\theta') \in A_k \times B_{k'}} \pi(k',\theta') q(k' \to k) \varphi_{k' \to k}(u') a(\theta' \to \theta(\theta',u')) d\theta' du' \\
&= \sum_{k,k' \in K_A \times K_B} \int_{(\theta,\theta'(\theta,u)) \in A_k \times B_{k'}} \pi(k',\theta'(\theta,u)) q(k' \to k) \varphi_{k' \to k}(u'(\theta,u)) a(\theta'(\theta,u) \to \theta) \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right| d\theta du
\end{aligned}$$

Now that we have rewritten both sides of Eq. 1 in the form of integrals with respect to $d\theta du$, we see that both integrals would be equal (for any choice of A, B) if we find an expression for $a(\theta \to \theta')$ such that the integrands are equal pointwise, i.e. for all (k, k', θ, u)

$$\begin{aligned}
&\pi(k,\theta) q(k \to k') \varphi_{k \to k'}(u) a(\theta \to \theta'(\theta,u)) \\
&= \pi(k',\theta'(\theta,u)) q(k' \to k) \varphi_{k' \to k}(u'(\theta,u)) a(\theta'(\theta,u) \to \theta) \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right|. \qquad (4)
\end{aligned}$$

Consider the following acceptance probability:

$$a(\theta \to \theta') = \min\left(1, \frac{\pi(k',\theta')\varphi_{k'\to k}(u') q(k' \to k)}{\pi(k,\theta)\varphi_{k \to k'}(u) q(k \to k')} \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right| \right).$$

We just check that it satisfies Eq. 4. We have indeed

$$\begin{aligned}
&\pi(k,\theta) q(k \to k') \varphi_{k \to k'}(u) a(\theta \to \theta') \\
&= \pi(k,\theta) q(k \to k') \varphi_{k \to k'}(u) \min\left(1, \frac{\pi(k',\theta')\varphi_{k'\to k}(u') q(k' \to k)}{\pi(k,\theta)\varphi_{k \to k'}(u) q(k \to k')} \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right| \right) \\
&= \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right| \min\left(\pi(k,\theta) q(k \to k') \varphi_{k \to k'}(u) \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right|^{-1}, \pi(k',\theta')\varphi_{k'\to k}(u') q(k' \to k) \right) \\
&= \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right| \pi(k',\theta')\varphi_{k'\to k}(u') q(k' \to k) \min\left(\frac{\pi(k,\theta) q(k \to k') \varphi_{k \to k'}(\mathrm{u})}{\pi(k',\theta')\varphi_{k'\to k}(u') q(k' \to k)} \left| \frac{\partial G_{k'\to k}(\theta,u)}{\partial(\theta,u)} \right|, 1 \right) \\
&= \pi(k',\theta')\varphi_{k'\to k}(u') q(k' \to k) a(\theta' \to \theta) \left| \frac{\partial G_{k\to k'}(\theta,u)}{\partial(\theta,u)} \right|.
\end{aligned}$$

We have used that the Jacobian associated to $G_{k\to k'}$ is the inverse of the Jacobian associated to $G_{k'\to k}$. To summarize, we have found a proposal mechanism, consisting of sampling k', then u from $\varphi_{k\to k'}$ and then deterministically mapping (θ, u) to $(\theta', u') = G_{k\to k'}(\theta, u)$, and an acceptance probability $a(\theta \to \theta')$, such that the resulting Markov kernel P satisfies Eq. 1. This will constitute our "between-models" move.

3.3 Another representation for the random walk Metropolis-Hastings

The presentation of "between-models" moves also works when θ and θ' are in fact of the same dimension. A natural question is then: what does it do? Is it the same as standard Metropolis-Hastings? Consider the vanilla Metropolis-Hastings algorithm, where a proposal X' is made according to

$$X' = X + W$$

with $W \sim g$. Assume g is a symmetric distribution, i.e. its density satisfies $g(w) = g(-w)$. We can also write the proposal as

$$(X', W') = T(X, W)$$

where $W \sim g$ and $T : (x, w) \mapsto (x + w, w)$. For the backward move we generate $W' \sim g$ and apply the inverse transformation $T^{-1} : (x, w) \mapsto (x - w, w)$. We indeed have $T \circ T^{-1}(x, w) = (x, w)$. Moreover the Jacobian of T (and of T^{-1}) is equal to 1. Hence the acceptance probability obtained above can be written

$$a(X \to X') = \min\left(1, \frac{\pi(X'(X, W))g(W'(X, W))}{\pi(X)g(W)} \times \left|\frac{\partial T(X, W)}{\partial(X, W)}\right|\right) = \min\left(1, \frac{\pi(X + W)}{\pi(X)}\right)$$

using $X' = X'(X, W) = X + W$ and $W' = W'(X, W) = W$. Indeed we fall back to the usual random walk Metropolis-Hastings acceptance, so that in the "between-models" move can be seen as a (admittedly complicated) generalization of a standard move.

4 Algorithmic description

Having described the between-model moves, by specifying the Markov kernel P, we are now in position to describe the full algorithm, allowing to sample from the transdimensional posterior distribution defined on $\cup_{k\in\mathcal{K}}\{k\} \times \Theta_k$. For each model $\mathcal{M}_k$ we introduce a "within-model" standard Metropolis-Hastings kernel S_k leaving $\pi(\theta_k \mid y, \mathcal{M}_k)$ invariant. We thus have a collection of "within-model" moves, S_k for each $\mathcal{M}_k$, and a "between-model" Markov kernel P. We introduce the probability β of performing a "within-model" move, which we might take close to 1 so that most moves are performed within models, with the occasional "between-models" attempt with probability $1 - \beta$. Let us start from model index $k^{(0)}$ and parameter $\theta^{(0)} \in \Theta_{k^{(0)}}$.

Algorithm. (Reversible jump Markov chain Monte Carlo). Starting with $\left(k^{(0)}, \theta^{(0)}\right)$ iterate for $t = 1, 2, 3, ...$

1. With probability β, set $k^{(t)} = k^{(t-1)}$ and perform one step of $S_{k^{(t)}}$ leaving $\pi(\theta_{k^{(t)}} \mid y, \mathcal{M}_{k^{(t)}})$ invariant.
2. With probability 1-β, propose another model $k' \sim q(k' \mid k^{(t-1)})$. Then draw a random variable $u_{k^{(t-1)}\to k'} \sim \varphi_{k^{(t-1)}\to k'}$ and apply the deterministic mapping $G_{k^{(t-1)}\to k'}$ to obtain a proposal $\theta' \in \Theta_{k'}$ and $u_{k'\to k^{(t-1)}}$. With probability

$$a(\theta^{(t-1)} \to \theta') = \min\left(1, \frac{\pi(\theta')\varphi_{k'\to k^{(t-1)}}(u_{k'\to k^{(t-1)}})q(k^{(t-1)} \mid k')}{\pi(\theta^{(t-1)})\varphi_{k^{(t-1)}\to k'}(u_{k^{(t-1)}\to k'})q(k' \mid k^{(t-1)})}\left|\frac{\partial G_{k^{(t-1)}\to k'}(\theta^{(t-1)}, u_{k^{(t-1)}\to k'})}{\partial(\theta, u)}\right|\right)$$

accept, i.e. set $\theta^{(t)} = \theta', k^{(t)} = k'$. Otherwise reject, i.e. set $\theta^{(t)} = \theta^{(t-1)}, k^{(t)} = k^{(t-1)}$.

Proposition 1 *The Markov kernel associated to the reversible jump algorithm admits $\pi\left(\mathcal{M}_k, \theta_k \mid y\right)$ as invariant density.*

Note that we identify a model $\mathcal{M}_k$ with its index k, so we write equivalently $\pi\left(\mathcal{M}_k, \theta_k \mid y\right)$ or $\pi\left(k, \theta_k \mid y\right)$. The proof is straightforward since we already showed that the "between-models" move is reversible with respect to π so that it leaves it invariant. We just check that each Markov kernel S_k also leaves π invariant, which is true by assumption. Indeed it does not change k and then given k it leaves $\pi(\theta_k \mid y, \mathcal{M}_k)$.

5 Toy Example

Consider a problem with two possible models $\mathcal{M}_1$ and $\mathcal{M}_2$. Model $\mathcal{M}_1$ has a single parameter $\theta \in \Theta_1$. Model $\mathcal{M}_2$ has two parameters $(\theta_1, \theta_2) \in \Theta_2$. The joint posterior is defined on

$$\Theta = \{1\} \times \Theta_1 \cup \{2\} \times \Theta_2.$$

We need to propose moves $G_{1\to 2}$ and $G_{2\to 1}$ such that $G_{2\to 1} = G_{1\to 2}^{-1}$. Assume we move from $\mathcal{M}_1$ to $\mathcal{M}_2$ using

$$(\theta_1, \theta_2) = G_{1\to 2}(\theta, u) = (\theta - u, \theta + u)$$

where u is some auxiliary variable from distribution φ, so that the associated reverse move from $\mathcal{M}_2$ to $\mathcal{M}_1$ is simply

$$G_{2\to 1}(\theta_1, \theta_2) = \left(\frac{\theta_1 + \theta_2}{2}, \frac{\theta_1 - \theta_2}{2}\right).$$

We have $G_{2\to 1} \circ G_{1\to 2}(\theta, u) = G_{2\to 1}(\theta - u, \theta + u) = (\theta, u)$. We have $|\frac{\partial G_{1\to 2}(\theta,u)}{\partial(\theta,u)}| = 2$ and $|\frac{\partial G_{2\to 1}(\theta,u)}{\partial(\theta,u)}| = \frac{1}{2}$. If we propose a move from $\mathcal{M}_1$ to $\mathcal{M}_2$ with probability q_{12} and a move from $\mathcal{M}_2$ to $\mathcal{M}_1$with probability q_{21}, then the acceptance rate of a move from $\mathcal{M}_1$ to $\mathcal{M}_2$ is given by

$$\min\left(1, \frac{\pi(2, \theta_1, \theta_2)}{\pi(1, \theta)} \frac{1}{\varphi(u)} \frac{q_{21}}{q_{12}} \times 2\right).$$

References

[1] P.J. Green, Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, vol. 82, no. 4, pp. 711-732, 1995.

[2] C.P. Robert & G. Casella, *Monte Carlo Statistical Methods*, Springer, 2004.