A tutorial on Bayesian inference for variable dimension models

Sergio Hernandez

October 12, 2007

1 Introduction

Variable dimension models are problems where the parameter space is not well defined, therefore the sample space is a infinite collection of unrelated subspaces. If the considered statistical model is not defined in concise way, then the dimensionality of the parameter space can also be part of the model uncertainty. These problems have been studied in the context of Bayesian model comparison and model selection, having several applications in statistical signal processing, image analysis, model-based clustering and financial data anaylisis.

A common problem in MCMC methods for the variable dimension case is the lack of a dominating measure for the target distribution. For a dataset d we can consider a finite set of models $\mathcal{M} = \mathcal{M}_1, \mathcal{M}_2, .., \mathcal{M}_k, ..$ indexed by a parameter $k \in \mathcal{K}$. Each model has a parameter vector $\theta_k \in \Theta_k$. If k is the dimensionality of the model, then Bayesian estimate can make use of a prior distribution $p(k)$. However the parameter space θ_k will be fixed to k, and no dominating measure can be defined for a model jump.

Green(1995) proposed a *reversible* transition kernel for the probability $\pi_k(\theta_k)$. The method recasts the time reversibility property of a Markovian transition kernel $\mathcal{X}(s)$, which is a neccesary condition for ensuring convergence to a particular stationary distribution.

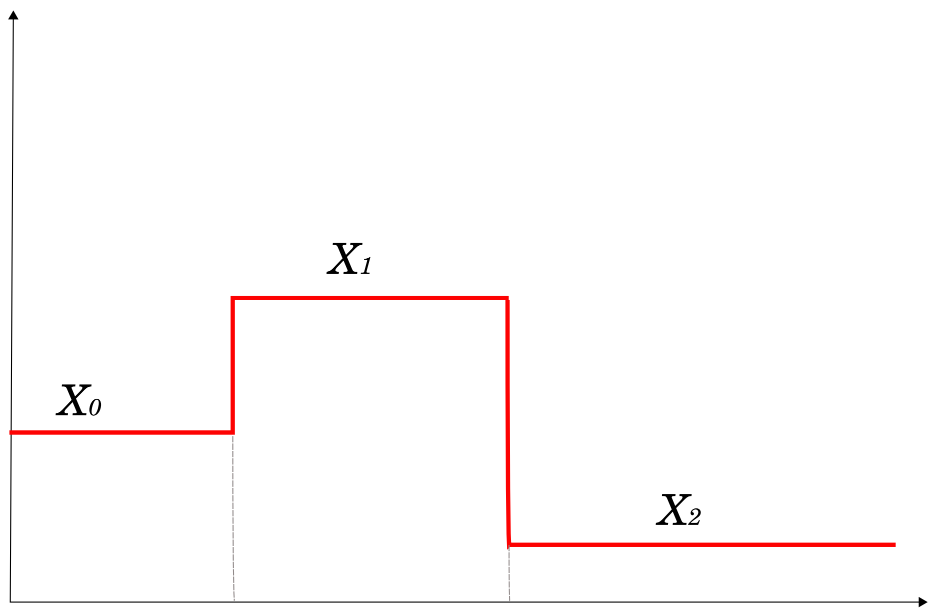

Figure 1: Reversible-jump sampler

Similar approaches were proposed for the variable dimension problem. Geyer and Moller (1994) formulated a Metropolis-Hasting method for simulating spatial point processes and Carlin and Chib (1995) uses pseudo-priors. However, the approach proposed by Green can be thought as a more general formalization. Stephens (2000) proposed a continuous-time simulation method using birth-death processes, and later on Cappe and Robert (2004) demonstrated the covergence of the continuous-time algorithm to its discrete-time counterpart as a special case of the reversible jump algorithm.

2 Reversible-jump algorithm

Consider a finite (or infinite) collection of models $\mathcal{M}_k$ with corresponding sampling distribution $f(\cdot|\theta_k)$, parameters θ_k and a prior distribution over the indices $p(k)$. Posterior inference can then be carried in the same way as the posterior expectation of a random variable $x = (\mathcal{M}_k, \theta_k)$, which is defined across models.

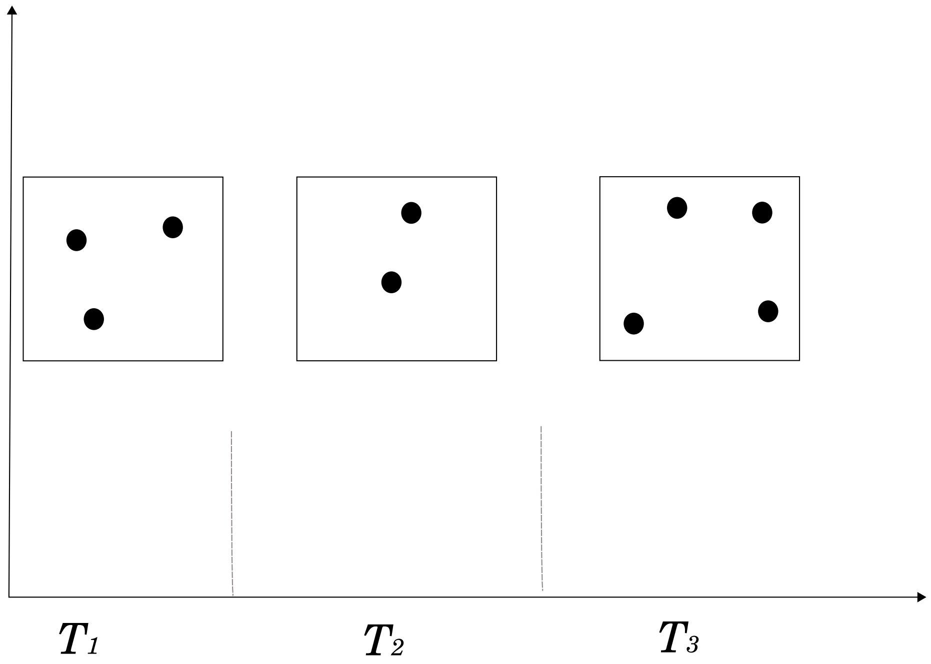

Figure 2: Birth-death sampler

$$\pi(x) = p(k)\pi_k(\theta_k)$$

Which is a density defined with respect to a probability measure on the union of spaces $\Theta = \bigcup_k \{k\} \times \Theta_k$

The the reversible kernel K of invariant distribution π can be written as :

$$\int_A \int_B K(x, dy)\pi(x)dx = \int_B \int_A K(y, dx)\pi(y)dy$$

Where $A, B \in \Theta$, $x \in A$ and $y \in B$.

If the model jumps are decomposed into moves between pairs of models, then an artificial space can act as the supplement between Θ_{k1} and Θ_{k2} in order to create a *bijection* among the two spaces.

The artificial space where both models can be compared is generated by augmenting the state space, imposing a dimension matching condition for the kernel. θ_{k1} is completed by simulating a random variable $u_1 \sim g_1(u_1)$ into (θ_{k1}, u_1) as well as (θ_{k2}, u_2), with $u_2 \sim g_2(u_2)$.

the probability of acceptance for the move $\mathcal{M}_{k1}$ to $\mathcal{M}_{k2}$ is then:

$$\min\left(\frac{\pi(k_2,\theta_{k2})}{\pi(k_1,\theta_{k1})}\,\frac{\pi_{21}g_2(u_2)}{\pi_{12}g_1(u_1)}\left|\frac{\partial T(\theta_{k1},u_1)}{\partial(\theta_{k1},u_1)}\right|,1\right)$$

2.1 Green's algorithm

The pseudo-code of Green's algorithm is then:

Iteration t $(t \geq 1)$**:** if $x^{(t)} = (m, \theta^{(m)})$,

1. Select model $\mathcal{M}_m$ with probability π_{mn}
2. Generate $u_{mn} \sim \varphi_{mn}(u)$ and set $(\theta^{(n)}, v_{nm}) = \Psi_{m\rightarrow n}(\theta^{(m)}, u_{mn})$
3. Take $x^{(t+1)} = (n, \theta^{(n)})$ with probability

$$\min\left(\frac{\pi(n,\theta^{(n)})}{\pi(m,\theta^{(m)})}\,\frac{\pi_{nm}\varphi_{nm}(v_{nm})}{\pi_{mn}\varphi_{mn}(u_{mn})}\left|\frac{\partial_{m\rightarrow n}(\theta^{(m)},u_{mn})}{\partial(\theta^{(m)},u_{mn})}\right|,1\right)$$

and take $x^{(t+1)} = x^{(t)}$ otherwise.

Where $\pi_m n$ represents the probability of a jump between θ_m and θ_n.

3 Example : Mixture modeling

Let's consider the following mixture of Gaussian distributions:

$$\mathcal{M}_k = \sum_{j=1}^{k} p_{jk}\mathcal{N}(\mu_{jk}, \sigma_{jk}^2)$$

The log-likelihood of the data $d = (d_1, .., d_n)$ can be written as:

$$\pi_k(\theta_k) = \sum_{i=1}^{n}\log(\sum_{j=1}^{k}\frac{p_{jk}}{\sqrt{2\pi\sigma jk^2}}\exp\left[\frac{(d_i-\mu_{jk})}{2\sigma_{jk}}\right])$$

The priors are :

1. $k \sim \text{Poisson}(\lambda)$

2. $p_{1:k} \sim \text{Dirichlet}(1, .., 1)$

3. $\mu_{1:k} \sim \text{Normal}(0, \kappa)$ with log-pdf:

$$\log p(\mu_{1:k}|\kappa) = -\frac{k}{2}\, 2\pi\kappa - \frac{1}{2\kappa}\sum_{i=1}^{k}\mu_i^2$$

4. $\sigma_{1:k} \sim \text{Inverse-Gamma}(\alpha, \beta)$ with log-pdf:

$$\log p(\sigma_{1:k}|\alpha, \beta) = k(\alpha \log\beta - \log\Gamma(\alpha)) - \sum_{i=1}^{k}\left[(\alpha+1)\log(\sigma_i) + \frac{\beta}{\sigma_i}\right]$$

The simplest approach implements model jumps of no more than one dimension, using *birth* and *death* steps. The birth step adds a new component drawn from the prior, and in the death step one of the components is killed by random selection.

The birth probability can be written as:

$$p_B = \min\left(\frac{\pi_{(k+1)k}}{\pi_{k(k+1)}}\frac{(k+1)!}{k!}\frac{\pi_{k+1}(\theta_{k+1})}{\pi_k(\theta_k)(k+1)\varphi_{k(k+1)}(u_{k(k+1)})}, 1\right)$$

and the death probability is $p_D = 1 - p_B$.

References

[1] Green, P.J., (2001). *Complex stochastic systems*: Chapman-Hall.

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

[3] Marin, J.M and Robert, C.P., (2007). *Bayesian core : A practical approach to Bayesian computational statistics.* London : Springer.

[4] Cappe O. and Robert C.P. and Ryden T., (2003). "Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers", Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65(22) : 679-700.