

Comments on “Joint Bayesian Model Selection and Estimation of Noisy Sinusoids via Reversible Jump MCMC”

Alireza Roodaki, Julien Bect and Gilles Fleury

Abstract—Reversible jump MCMC (RJ-MCMC) sampling techniques, which allow to jointly tackle model selection and parameter estimation problems in a coherent Bayesian framework, have become increasingly popular in the signal processing literature since the seminal paper of Andrieu and Doucet (*IEEE Trans. Signal Process.*, 47(10), 1999). Crucial to the implementation of any RJ-MCMC sampler is the computation of the so-called Metropolis-Hastings-Green (MHG) ratio, which determines the acceptance probability for the proposed moves.

It turns out that the expression of the MHG ratio that was given in the paper of Andrieu and Doucet for “Birth-or-Death” moves is erroneous, and has been reproduced in many subsequent papers dealing with RJ-MCMC sampling in the signal processing literature. This note fixes the erroneous expression and briefly discusses its cause and consequences.

Index Terms—Signal decomposition; Bayesian inference; Markov Chain Monte Carlo methods; Trans-dimensional problems.

I. INTRODUCTION

Model selection and parameter estimation are fundamental tasks arising in many (if not all) signal processing problems, when parametric models are employed. Andrieu and Doucet [1] pioneered the use of Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) [2] sampling in “signal decomposition” problems, by tackling joint model selection and parameter estimation for an unknown number of sinusoidal signals observed in white Gaussian noise. This seminal paper was followed by many others in the signal processing literature [3–11], relying systematically on the original paper [1] for the computation of the acceptance ratio of “Birth-or-Death” moves—the most elementary type of trans-dimensional move, which either adds or removes a component from the signal decomposition. However, the expression of the acceptance ratio for Birth-or-Death moves provided by [1, Equation (20)] turns out to be erroneous, and the same error has been reproduced in many subsequent papers dealing with RJ-MCMC sampling in the signal processing literature (referred to above).

In this note, we briefly justify the computation of the acceptance ratio for the Birth-or-Death moves and provide the correct acceptance ratio for the problem considered in [1], which can be used in the problems addressed in the papers mentioned above. For a complete justification and mathematical results, we refer the reader to [12, 13] and references therein. Section II recalls, very quickly, the hierarchical model and RJ-MCMC sampler developed by [1] for the problem of Bayesian detection and estimation of sinusoids in white Gaussian noise. Section III provides results for computing the acceptance ratio for the Birth-or-Death moves and gives the correct expression in the considered problem. Finally, Section IV discusses the effect of using the erroneous ratio and explains the cause of the error. The good news is that published results obtained with the erroneous ratio are still valid, if interpreted as coming from a different prior distribution on the number of components.

This is an author-generated version. The final publication is available from the IEEE, under DOI 10.1109/TSP.2013.2261992.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

E3S—SUPELEC Systems Sciences Dept. of Signal Processing and Electronic Systems, SUPELEC, Gif-sur-Yvette, France. Contact authors: al.roodaki@gmail.com, julien.bect@supelec.fr.

II. MODEL

In this section, we follow closely the model and notations of [1]; the reader is referred to the original paper for more details.

Let $\mathbf{y} = (y_1, \dots, y_N)^t$ be a vector of N observations of an observed signal. We consider the finite family of nested models $\mathcal{M}_0 \subset \mathcal{M}_1 \subset \dots \subset \mathcal{M}_{k_{\max}}$, where \mathcal{M}_k assumes that \mathbf{y} is composed of k sinusoids observed in white Gaussian noise. Let $\boldsymbol{\omega}_k = (\omega_{1,k}, \dots, \omega_{k,k})$ and $\mathbf{a}_k = (a_{c_{1,k}}, a_{s_{1,k}}, \dots, a_{c_{k,k}}, a_{s_{k,k}})$ be the vectors of radial frequencies and cosine/sine amplitudes under model \mathcal{M}_k , respectively; moreover, let \mathbf{D}_k be the corresponding $N \times 2k$ design matrix. Then, the observed signal \mathbf{y} follows under \mathcal{M}_k a normal linear regression model:

$$\mathbf{y} = \mathbf{D}_k \cdot \mathbf{a}_k + \mathbf{n},$$

where \mathbf{n} is a white Gaussian noise with variance σ^2 . The unknown parameters are, then, assumed to be the number of components k , the component-specific parameters $\boldsymbol{\theta}_k = (\mathbf{a}_k, \boldsymbol{\omega}_k)$ and the noise variance σ^2 which is common to all models. The joint prior distribution is chosen to have the following hierarchical structure:

$$p(k, \boldsymbol{\theta}_k, \sigma^2) = p(\mathbf{a}_k | k, \boldsymbol{\omega}_k, \sigma^2) p(\boldsymbol{\omega}_k | k) p(k) p(\sigma^2),$$

where the prior over \mathbf{a}_k is the conventional g -prior distribution, which is a zero mean Gaussian with $\sigma^2 \delta^2 (\mathbf{D}_k^t \mathbf{D}_k)^{-1}$ as its covariance matrix. Conditional on k , the radial frequencies are independent and identically distributed, with a uniform distribution on $(0, \pi)$. The noise variance σ^2 is endowed with Jeffreys improper prior, i.e. $p(\sigma^2) \propto 1/\sigma^2$. The number of components k is given a Poisson distribution with mean Λ , truncated to $\{0, 1, \dots, k_{\max}\}$. The parameters \mathbf{a}_k and σ^2 can be integrated out analytically, and the resulting marginal posterior becomes

$$p(k, \boldsymbol{\omega}_k | \mathbf{y}) \propto (\mathbf{y}^t \mathbf{P}_k \mathbf{y})^{-N/2} \frac{\Lambda^k \pi^{-k}}{k! (\delta^2 + 1)^k} \mathbb{1}_{(0, \pi)^k}(\boldsymbol{\omega}_k), \quad (1)$$

with

$$\mathbf{P}_k = \mathbf{I}_N - \frac{\delta^2}{1 + \delta^2} \mathbf{D}_k (\mathbf{D}_k^t \mathbf{D}_k)^{-1} \mathbf{D}_k^t$$

when $k \geq 1$ and $\mathbf{P}_0 = \mathbf{I}_N$.

III. BIRTH-OR-DEATH KERNELS

Inference under this hierarchical Bayesian model is carried out in [1] using an RJ-MCMC sampler on $\mathbb{X} = \bigcup_{k=0}^{k_{\max}} \{k\} \times (0, \pi)^k$ with target density (1). We only focus here on the birth-or-death moves which propose a jump by one component between models. More precisely, assuming that the vector $\boldsymbol{\omega}_k$ of radial frequencies is *not* sorted, a birth move inserts a new component with radial frequency $\omega^* \in (0, \pi)$, generated according to some proposal distribution $q(\omega)$, at a *randomly* selected location. A death move, on the contrary, removes a *randomly* selected component from the current state. We assume that a discrete uniform distribution is used in both cases.

Proposition 1: Let us denote by $\boldsymbol{\omega}_k \oplus_i \omega^*$ the vector obtained by inserting ω^* at location i in $\boldsymbol{\omega}_k$. Then, the MHG ratio for a birth move from $\mathbf{x} = (k, \boldsymbol{\omega}_k)$ to $\mathbf{x}' = (k+1, \boldsymbol{\omega}_k \oplus_i \omega^*)$ is

$$r(\mathbf{x}, \mathbf{x}') = \frac{p(k+1, \boldsymbol{\omega}_k \oplus_i \omega^* | \mathbf{y})}{p(k, \boldsymbol{\omega}_k | \mathbf{y})} \times \frac{p_d(\mathbf{x}')}{p_b(\mathbf{x})} \times \frac{1}{q(\omega^*)}. \quad (2)$$

See [12, Proposition 1.11] or [13, Proposition 2] for a proof. Now, setting q to a uniform distribution on $(0, \pi)$ and using for probabilities of selecting birth and death moves

$$\frac{p_d(\mathbf{x}')}{p_b(\mathbf{x})} = \frac{p_0(k)}{p_0(k+1)} = \frac{k+1}{\Lambda}$$

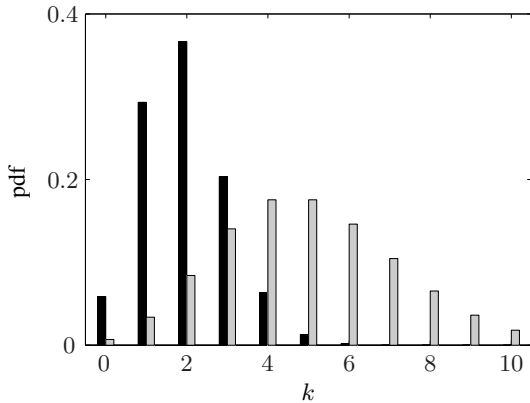


Figure 1. Probability distribution functions of the Poisson (gray) and accelerated Poisson (black) distributions with mean $\Lambda = 5$.

as in [1], with p_0 standing for the (truncated Poisson) prior distribution of k , we finally find

$$\begin{aligned} r(\mathbf{x}, \mathbf{x}') &= \left(\frac{\mathbf{y}^t \mathbf{P}_{k+1} \mathbf{y}}{\mathbf{y}^t \mathbf{P}_k \mathbf{y}} \right)^{-N/2} \frac{\Lambda \pi^{-1}}{(1+k)(1+\delta^2)} \cdot \frac{k+1}{\Lambda} \cdot \frac{1}{\pi^{-1}} \\ &= \left(\frac{\mathbf{y}^t \mathbf{P}_{k+1} \mathbf{y}}{\mathbf{y}^t \mathbf{P}_k \mathbf{y}} \right)^{-N/2} \frac{1}{1+\delta^2}. \end{aligned} \quad (3)$$

IV. THE ERROR AND ITS IMPACT

Note that the expression of the ratio proposed in [1, Equation (20)] differs from the one we find here (3) by a factor $1/(k+1)$. A similar error in computing RJ-MCMC ratios has been reported in the field of genetics [14, 15]. In fact, using the expression of the birth ratio with an additional factor of $1/(k+1)$, as in [1], amounts to assigning a different prior distribution over k called “accelerated Poisson distribution” [15] which reads

$$p_2(k) \propto \frac{e^{-\Lambda} \Lambda^k}{(k!)^2} \mathbb{1}_{\mathbb{N}}(k). \quad (4)$$

Figure 1 illustrates the difference between both the accelerated (black) and the usual (gray) Poisson distributions when mean $\Lambda = 5$. It can be observed that the accelerated Poisson distribution (4) puts a stronger emphasis on “sparse” models, i.e., models with a small number of components.

To show the impact of using the erroneous MHG ratio, let us consider an experiment in which the observed signal of length $N = 64$ consists of $k = 3$ sinusoidal components with the radial frequencies $\omega_k = (0.63, 0.68, 0.73)^t$ and amplitudes $a_{c_{i,k}}^2 + a_{s_{i,k}}^2 = (20, 6.32, 20)^t$, $1 \leq i \leq k$. The signal to noise ratio, defined as $\text{SNR} \triangleq \|\mathbf{D}_k \mathbf{a}_k\|^2 / (N\sigma^2)$, is set to a moderate value of 7dB. Samples from the posterior distribution of k are obtained using the RJ-MCMC sampler of [1], with an inverse Gamma prior $\mathcal{IG}(2, 100)$ on δ^2 and a Gamma prior $\mathcal{G}(1, 10^{-3})$ on Λ . For each observed signal in 100 replications of the experiment, the sampler was run twice: once with the correct expression of the ratio, given by (3), and once with the erroneous expression from [1]. Figure 2 shows the frequency of selection of each model under both the Poisson and the accelerated Poisson distribution as a prior for k . It appears that the (unintended) use of the accelerated Poisson distribution, induced by the erroneous expression of the MHG ratio, can result in a significant shift to the left of the posterior distribution of k .

The reason why the MHG ratio in [1] is wrong can be understood from a subsequent paper [16], where the same computation is explained in greater detail. There, we can see that the authors consider that the new component in a birth move is *inserted at the end*. The

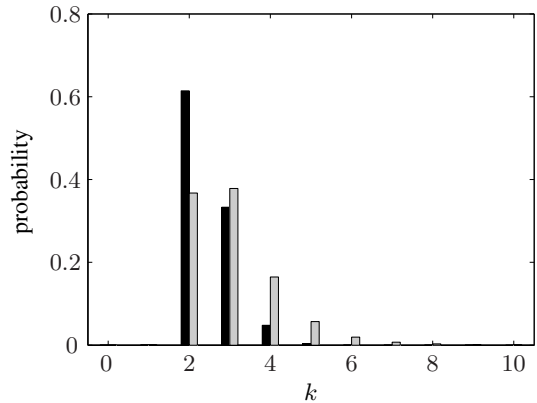


Figure 2. Frequency of selection for each model \mathcal{M}_k for 100 replications of the experiment described in Section IV, using the expression of the ratio given in [1, Equation (20)] (black) and the corrected ratio (3) (gray).

death move, however, is defined as in the present paper: a sinusoid to be removed is *selected randomly* among the existing components. Here is the error: if the new component is inserted at the end during a birth move, then any attempt at removing a component which is not the last one should be rejected during a death move. In other words, the acceptance probability should be zero when any component but the last one is picked to be removed during a death move.

REFERENCES

- [1] C. Andrieu and A. Doucet, “Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC,” *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2667–2676, 1999.
- [2] P.J. Green, “Reversible jump MCMC computation and Bayesian model determination,” *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.
- [3] C. Andrieu, N. Freitas, and A. Doucet, “Robust full Bayesian learning for radial basis networks,” *Neural Computation*, vol. 13, no. 10, pp. 2359–2407, 2001.
- [4] C. Andrieu, E. Barat, and A. Doucet, “Bayesian deconvolution of noisy filtered point processes,” *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 134–146, 2002.
- [5] J.R. Larocque and J.P. Reilly, “Reversible jump MCMC for joint detection and estimation of sources in coloured noise,” *IEEE Transactions on Signal Processing*, vol. 50, pp. 231–240, 2002.
- [6] J.R. Larocque, J.P. Reilly, and W. Ng, “Particle filters for tracking an unknown number of sources,” *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 2926–2937, 2002.
- [7] W. Ng, J.P. Reilly, T. Kirubarajan, and J.R. Larocque, “Wideband array signal processing using MCMC methods,” *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 411–426, 2005.
- [8] M. Davy, S. J. Godsill, and J. Idier, “Bayesian analysis of polyphonic western tonal music,” *Journal of the Acoustical Society of America*, vol. 119, pp. 2498–2517, 2006.
- [9] M. Hong, M.F. Bugallo, and P.M. Djuric, “Joint Model Selection and Parameter Estimation by Population Monte Carlo Simulation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 3, pp. 526–539, 2010.
- [10] M. N. Schmidt and M. Mørup, “Infinite non-negative matrix factorization,” in *Proceedings of the 18th European Signal Processing Conference (EUSIPCO)*, Aug 2010.
- [11] D.V. Rubtsov and J.L. Griffin, “Time-domain Bayesian detection and estimation of noisy damped sinusoidal signals applied to NMR spectroscopy,” *Journal of Magnetic Resonance*, vol. 188, no. 2, pp. 367–379, 2007.
- [12] A. Roodaki, *Signal Decompositions using Trans-dimensional Bayesian Methods*, Ph.D. thesis, École Supérieure d’Électricité (Supélec), Gif-sur-Yvette, France, 2012.
- [13] A. Roodaki, J. Bect, and G. Fleury, “Note on the computation of the Metropolis-Hastings ratio for Birth-or-Death moves in trans-dimensional MCMC algorithms for signal decomposition problems,” Technical report (arXiv:1111.6245v2), École Supérieure d’Électricité (Supélec), Gif-sur-Yvette, France, 2012.
- [14] J. L. Jannink and R. L. Fernando, “On the Metropolis-Hastings acceptance probability to add or drop a quantitative trait locus in Markov chain Monte Carlo-based Bayesian analyses,” *Genetics*, vol. 166, pp. 641–643, 2004.
- [15] M.J. Sillanpaa, D. Gasbarra, and E. Arjas, “Comment on “On the Metropolis-Hastings Acceptance Probability to Add or Drop a Quantitative Trait Locus in Markov Chain Monte Carlo-Based Bayesian Analyses,”” *Genetics*, vol. 167, no. 2, pp. 1037, 2004.
- [16] C. Andrieu, P.M. Djuric, and A. Doucet, “Model selection by MCMC computation,” *Signal Processing*, vol. 81, no. 1, pp. 19–37, 2001.