DETECTION AND ESTIMATION OF SIGNALS BY REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO COMPUTATIONS

Petar M. Djurić [†], *Simon J. Godsill* [‡], *William J. Fitzgerald* [‡], *and Peter J.W. Rayner* [‡]

[†]Department of Electrical Engineering
State University of New York at Stony Brook
Stony Brook, NY 11794-2350

[‡]Engineering Department
University of Cambridge
Cambridge, CB2 1PZ

ABSTRACT

Markov Chain Monte Carlo (MCMC) samplers have been a very powerful methodology for estimating signal parameters. With the introduction of the reversible jump MCMC sampler, which is a Metropolis-Hastings method adapted to general state spaces, the potential of the MCMC methods has risen to a new level. Consequently, the MCMC methods currently play a major role in many research activities. In this paper we propose a reversible jump MCMC sampler based on predictive distributions obtained by integrating out unwanted parameters. The proposal distributions are approximations of the posterior distributions of the remaining parameters and are computed by sampling importance resampling (SIR). We apply the method to the problem of signal detection and parameter estimation of signals. To illustrate the proposed procedure, we present an example of sinusoids embedded in noise.

1. INTRODUCTION

Standard signal processing problems are signal detection and estimation of signal parameters. The list of detection and estimation methods is large, and the method that is applied to observed data normally depends on the nature of the observed data. Recently, the statistical and signal processing community have given considerable attention to parameter estimation methods known as Markov Chain Monte Carlo (MCMC) [1], [5], [9]. MCMC methods are in essence numerical integrations implemented by Markov chains. In many estimation problems one needs to integrate over the posterior distribution of the model parameters or data sample spaces given the model parameters, and therefore it is not surprising that MCMC is attracting so much attention.

The principal idea of MCMC methods is to generate samples from a multivariate distribution by constructing a Markov chain of simpler distributions from which it is easy to sample and whose equilibrium is the original multivariate distribution. There are many ways for constructing such chains, but two have been particularly convenient. One is the Gibbs sampler [3], and the other is the Metropolis-Hastings algorithm [7], [8], of which the Gibbs sampler is only a special case.

This work was supported by the National Science Foundation under Award No. MIP-9506743

Recently, an important and qualitatively new method has been added to the family of MCMC samplers [6]. It is more general and is applicable to problems where the Markov chain sampler can jump between parameter subspaces corresponding to different models, which are typically of different dimensions. As such, this method, also known as *reversible jump* MCMC, finds immediate application in many difficult signal and image processing problems where not only the parameters of the signal model are unknown, but also the dimensionality of the parameter vector. This includes detection of signals, mixture deconvolution [10], multiple change point problems [6], time series modeling [11], object recognition [5], [6], and variable selection [4].

In our paper, we exploit a reversible jump MCMC sampler on problems where the number of signals is unknown and some of their parameters are nuisance parameters. Our objective is to obtain a sampler that converges rapidly to the equilibrium distribution. The approach is based on *predictive* distributions of the considered models obtained by integrating out the nuisance parameters. More specifically, first we determine the predictive densities, which are functions of the parameters of interest, and given the observed data, act as likelihoods. Then we use them to construct proposals for making the transitions from one model parameter space to another, or for moving within the same model parameter space. This is accomplished by approximating the likelihood via the sampling importance resampling technique (SIR). Once the reversible jump MCMC sampling is completed, we obtain the joint posterior distribution of the number of signals and their parameters. Then the samples obtained can be used for inference about the models and their parameters.

We provide an example which shows the steps in implementing the proposed procedure. We assume that the observed data represent closely spaced sinusoids embedded in noise. The number of sinusoids and their parameters are unknown. A reversible jump MCMC sampler is constructed which produces the joint distribution of frequencies and number of sinusoids. The sampler constantly attempts to jump from a current to adjacent models, and the probabilities for such moves are calculated for each attempt. From the resulting joint distribution, we easily deduce the most likely number of sinusoids and the estimates of their frequencies. A numerical example is also presented.

2. GENERAL PROBLEM STATEMENT

We assume that there is a set of models which represent candidates for the generating mechanism of some observed data $\mathbf{y}$. The models are parametric with some of their parameters being of interest, and the remaining being nuisance parameters. We would like to develop a procedure that provides the joint a posteriori distribution of the models and their parameters, from which we can make inference about many quantities. For example, the estimates of the a posteriori probabilities of the models would certainly be important, as well as various estimates of the parameters of interest. In brief, if the models are denoted by $\mathcal{M}_0$, $\mathcal{M}_1$, $\mathcal{M}_2$, ..., $\mathcal{M}_k$, ..., and the associated parameters $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, ..., $\boldsymbol{\theta}_k$, ..., then the objective is to estimate $p(\mathcal{M}_i, \boldsymbol{\theta}_i|\mathbf{y})$.

3. REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO SAMPLING STRATEGY

The reversible jump MCMC sampler is an extension of the standard MCMC method in that it allows for jumps between models and their parameter spaces of different dimensions as the sampling proceeds. It is based on the Metropolis-Hastings method and it involves *moves* that represent changes of models or updating of the current model parameters. The move that requires updating of parameters is implemented in the usual way as there is no change of dimensions and/or change of parameter spaces. The proposal of new model and the evaluation of the probability of its acceptance is done, however, by imposing dimension matching and retaining detailed balance. Overall, of course, the Markov chain has to be irreducible and aperiodic.

Suppose that the current model is $\mathcal{M}_k$ and its parameters $\boldsymbol{\theta}_k$, and that there is a proposal for a move m to the model $\mathcal{M}_j$. One way to construct the reversible jump sampler is to allow for a transition which is accepted with probability α, where

$$\alpha = \min(1, \alpha_0) \tag{1}$$

and

$$\alpha_0 = \frac{p(\mathbf{y}|\boldsymbol{\theta}_j, \mathcal{M}_j)p(\boldsymbol{\theta}_j, \mathcal{M}_j)g(\boldsymbol{\theta}_k, \mathbf{y}, \mathcal{M}_k)j(\mathcal{M}_k, \mathcal{M}_j)}{p(\mathbf{y}|\boldsymbol{\theta}_k, \mathcal{M}_k)p(\boldsymbol{\theta}_k, \mathcal{M}_k)g(\boldsymbol{\theta}_j, \mathbf{y}, \mathcal{M}_j)j(\mathcal{M}_j, \mathcal{M}_k)} \tag{2}$$

where the $p(\mathbf{y}|\cdot)$'s are the likelihood functions of the models, $p(\boldsymbol{\theta}_j, \mathcal{M}_j)$ and $p(\boldsymbol{\theta}_j, \mathcal{M}_j)$ their priors, the $g(\cdot)$'s the proposal distributions for all the parameters of the appropriate models, and $j(\mathcal{M}_k, \mathcal{M}_j)$ and $j(\mathcal{M}_j, \mathcal{M}_k)$ the probabilities of proposing transitions from $\mathcal{M}_j$ to $\mathcal{M}_k$, and from $\mathcal{M}_k$ to $\mathcal{M}_j$, respectively .

The reversible jump sampler has two general types of moves. One is the updating of the current model's parameters, and the other, transition to a different model. Once the sampler is in the parameter space of model $\mathcal{M}_k$, a move is chosen whether to stay with the same model, or to move to another model, say $\mathcal{M}_j$, which is reachable from $\mathcal{M}_k$ in one step. If such an attempt is made, then a proposal for the parameter values of the destination model has to be provided, which is obtained from the proposal distribution $g(\boldsymbol{\theta}_j, \mathbf{y}, \mathcal{M}_j)$. A new set of parameters is sampled from the parameter space of $\mathcal{M}_k$ using $g(\boldsymbol{\theta}_k, \mathbf{y}, \mathcal{M}_k)$, and the probability of acceptance α is calculated from (1). Then a random number from the uniform distribution on (0,1) is drawn, and if it is less than α, the transition to $\mathcal{M}_j$ is accepted. Otherwise, the model remains as $\mathcal{M}_k$. The procedure continues in the same fashion long enough to allow for convergence to the equilibrium distribution.

4. IMPLEMENTATION ISSUES

The convergence of the reversible jump MCMC sampler is an important issue, and to improve it, it is desirable to decrease the dimension of the parameter space over which we draw the samples. This is more so when some of the parameters are nuisance parameters. Also, in many signal processing problems some parameters can be estimated much more easily than others, and once the model is known they can readily be obtained. So, it is preferable to avoid sampling from the spaces of these parameters. This is accomplished by integrating the parameters out from the model. For example, if there are linear parameters in a model, we could remove them from the list of parameters for sampling, thereby save time for the sampler and improve convergence. Similarly, in the usual case of signal in white Gaussian noise with unknown variance, the variance can also be straightforwardly integrated out.

Recall that with the reversible jump MCMC method there are frequent comparisons of models with different dimensions. Then it seems that our definition of the priors may be critical for the rapid convergence of the posterior to the equilibrium distribution, especially in cases when we want to be as uninformative about the parameters as possible. If the priors are noninformative, the unknown proportionality constants will cause problems due to the dimension mismatch in the comparison of the models.

One approach that avoids these difficulties is based on predictive densities. Predictive densities have been successfully applied to model selections before (see for example [2] and its references). The underlying concept for using them is to partition the data into an estimation and validation sets, that is $\mathbf{y}$ is split into $\mathbf{y}_e$ and $\mathbf{y}_v$. The estimation set is used to obtain proper priors for the nuisance parameters, and the validation set for obtaining the likelihoods for the remaining unknown parameters.

Thus, we propose elimination of the nuisance parameters by analytical integration using the concept of predictive densities. In many problems of interest in signal processing, this is easily accomplished. Then we proceed with the reversible jump MCMC sampler where the predictive densities, which are functions of the unknown parameters, act as likelihood functions.

When we make the moves with the sampler, we have to provide proposals for new values of the parameters. We do it by uniformly sampling from the parameter space of the parameter, then we reweight the so obtained samples by the likelihood, and finally resample from the reweighted discrete distribution, i.e., apply the SIR procedure [1].

5. AN EXAMPLE OF SINUSOIDS EMBEDDED IN NOISE

Here we present an example that is of frequent interest in the signal processing community. It involves the detection of sinusoids in noise and the estimation of their parameters. Let the model with k sinusoids be represented by

$$\mathbf{y} = \mathbf{H}_k \mathbf{a}_k + \mathbf{w} \tag{3}$$

where $\mathbf{H}_k$ is an $N \times 2k$ matrix, $\mathbf{a}_k$ a $2k \times 1$ vector of amplitudes, and $\mathbf{w}$ a zero mean Gaussian vector with covariance matrix $\sigma^2 \mathbf{I}$. Without loss of generality it is assumed that the unknown frequencies of the sinusoids, $\mathbf{f}_k$, take values from the interval $(0, 1/2)$. In $\mathbf{H}_k$ every two columns span the signal space of a different sinusoid. For example, the first and second columns of $\mathbf{H}_k$ are defined by

$$\mathbf{h}_1^T = [1 \ \cos(2\pi f_1) \ \cos(2\pi f_1 2) \cdots \cos(2\pi f_1 (n-1))] \tag{4}$$

and

$$\mathbf{h}_2^T = [0 \ \sin(2\pi f_1) \ \sin(2\pi f_1 2) \cdots \sin(2\pi f_1 (n-1))]. \tag{5}$$

The variance and the amplitudes are nuisance parameters, so we would like to integrate them out. As prescribed in the previous section, since we work with improper priors for the amplitudes and variance, i.e.,

$$p(\mathbf{a}_k, \sigma^2) \propto \frac{1}{\sigma^2} \tag{6}$$

we use predictive densities. If m is the number of samples of $\mathbf{y}_e$ used for obtaining proper priors for the nuisance parameters, and $\mathbf{y}_v$ is the remaining portion of the data, we obtain

$$p(\mathbf{y}_v | \mathbf{y}_e, \mathbf{f}_k, \mathcal{M}_k) = \frac{1}{(2\pi)^{\frac{n-m}{2}}} \frac{\Gamma(\frac{n-2k}{2})}{\Gamma(\frac{m-2k}{2})} \times \frac{\left(\frac{\mathbf{y}_e^T \mathbf{P}_e^\perp \mathbf{y}_e}{2}\right)^{\frac{m-2k}{2}}}{\left(\frac{\mathbf{y}^T \mathbf{P}^\perp \mathbf{y}}{2}\right)^{\frac{n-2k}{2}}} \frac{|\mathbf{H}_e^T \mathbf{H}_e|^{\frac{1}{2}}}{|\mathbf{H}^T \mathbf{H}|^{\frac{1}{2}}} \tag{7}$$

where $\mathbf{P}_e^\perp$ is a projection matrix formed from $\mathbf{H}_e$, that is $\mathbf{P}_e^\perp = \mathbf{I} - \mathbf{H}_e(\mathbf{H}_e^T \mathbf{H}_e)^{-1} \mathbf{H}_e^T$, and $\mathbf{P}^\perp$ is similarly formed from $\mathbf{H}$. Note that we have dropped the indices k from the projection matrices, but it should be clear from the context that they have rank equal to $2k$. Also, it should be kept in mind that the dependence of $p(\mathbf{y}_v | \mathbf{y}_e, \mathbf{f}_k, \mathcal{M}_k)$ on $\mathbf{f}_k$ is not emphasized in (7), but it should be obvious that the projection matrices and the determinants are functions of $\mathbf{f}_k$.

In the case of no sinusoids, the predictive densities become

$$p(\mathbf{y}_v | \mathbf{y}_e, \mathcal{M}_0) = \frac{1}{(2\pi)^{\frac{n-m}{2}}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{m}{2})} \frac{\left(\frac{\mathbf{y}_e^T \mathbf{y}_e}{2}\right)^{\frac{m}{2}}}{\left(\frac{\mathbf{y}^T \mathbf{y}}{2}\right)^{\frac{n}{2}}}. \tag{8}$$

With (7) and (8) on hand, we can proceed with the sampling as prescribed in the previous section.

6. NUMERICAL RESULTS

To demonstrate the approach with a numerical example, we have created 50 data samples according to

$$y[n] = a_1 \cos(2\pi f_1 n + \phi_1) + a_2 \cos(2\pi f_2 n + \phi_2) + w[n] \tag{9}$$

where $n = 0, 1, 2, \cdots, N-1$, $f_1 = 0.215$, $f_2 = 0.225$, $N = 50$, $\phi_1 = 1$ rad, $\phi_2 = 1.5$ rad, $w[n]$ was a white Gaussian noise with variance $\sigma^2 = 1$, and the amplitudes were $a_1 = a_2 = \sqrt{2}$, which corresponds to signal-to-noise ratios of 0 dB. Note that the separation of the sinusoids is two times smaller than the Rayleigh resolution.

The sampler could jump only to adjacent models, which were defined as models with one more or one less sinusoid than the current model. In other words, if the current model was $\mathcal{M}_k$, $k > 0$, the sampler could jump to $\mathcal{M}_{k-1}$, $\mathcal{M}_{k+1}$, or stay with $\mathcal{M}_k$. Each of these moves had probability of 1/3. If the current model was $\mathcal{M}_0$, there were only two possibilities, one to jump to $\mathcal{M}_1$, and the other to stay with $\mathcal{M}_0$. The probability of these moves was 1/2. It is important to note that the maximum number of sinusoids was not limited. The region from which the frequencies were sampled was defined by $\mathcal{F} = (0.20, 0.24)$. To get the proposal for the sampled frequency, 20 samples were generated from a uniform distribution defined on $\mathcal{F}$. They were reweighted according to the appropriate predictive density and the frequency was then resampled. For the priors of the frequencies we used the uniform discrete prior, in this case 1/20. Finally, in each cycle the estimation data set was randomly selected, and the data samples in that set were not necessarily successive.

We let the reversible jump MCMC sampler go through 10,000 iterations. The first 200 hundred iterations were not used in our summary computations. In Figure 1 we show the moves of the sampler between iterations 4001 and 6000. We observe that the sampler had occasional visits to models 1, 3, and 4, but overall it spent most of the time with the model with two sinusoids. In Figure 2 the posterior probabilities of the various models are plotted, and there we notice that the posterior probability of model 2 is greater than 0.9. In Figure 3, the histograms of the sampled frequencies from model $\mathcal{M}_2$ are shown. Their means are $\hat{f}_1 = 0.2143$ and $\hat{f}_2 = 0.2275$. For higher signal-to noise ratios, the sampler would stay even longer with the second model, whereas with the decrease of the signal-to-noise ratio, the more and more preferred model is the one with single sinusoid. If the signal-to-noise ratio continues to decrease, visits to model $\mathcal{M}_0$ become increasingly common.

Again, we emphasize that the number of hypothesized sinusoids is not limited, provided we have enough data to obtain proper predictive densities. In general, for as long as there is a scheme which allows reachability of any model from any point in the state space, the reversible jump MCMC sampler will move through it and spend most of the time with the model or set of models which are the most likely.

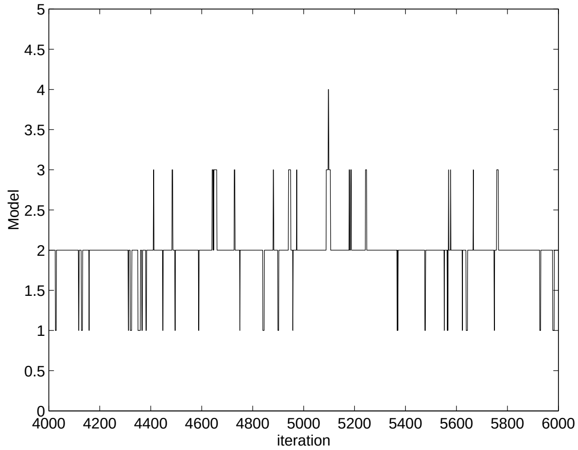

Figure 1: Moves of the sampler between iterations 4001 and 6000.

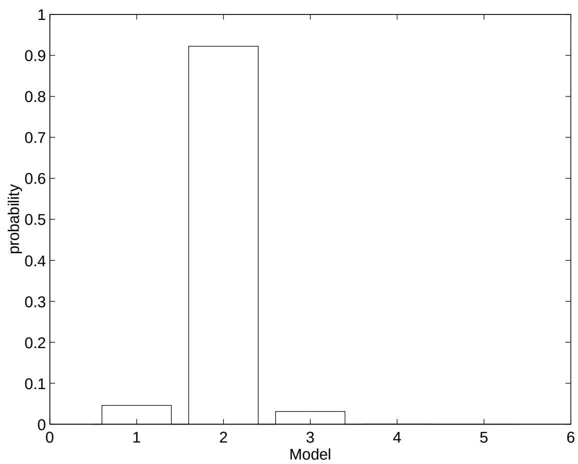

Figure 2: Posterior probabilities of the models.

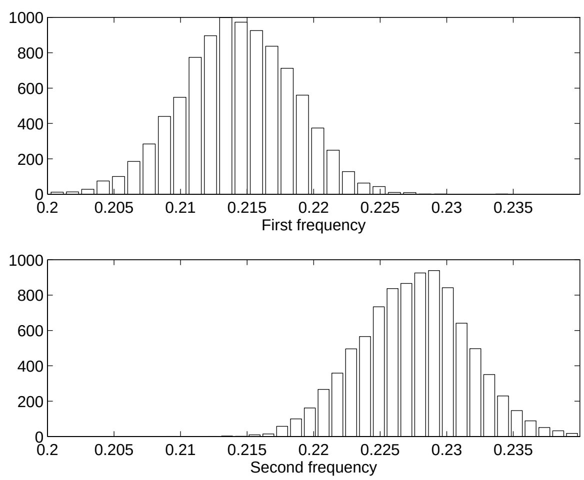

Figure 3: Histograms of the sampled frequencies from $\mathcal{M}_2$.

7. CONCLUSIONS

A reversible jump MCMC sampler has been proposed for problems with nuisance parameters that can readily be integrated out. Since the priors for these parameters are assumed improper, we use predictive densities to construct the likelihoods for the remaining parameters. An example is provided that shows the implementation of the proposal along with some numerical results. The procedure is rather general and can be applied to many other problems with models whose parameters are analytically marginalizable.

8. REFERENCES

[1] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*, New York: John Wiley & Sons, 1994.

[2] P.M. Djurić and S. M. Kay, "Model selection based on Bayesian predictive densities," IEEE Transactions on Signal Processing, vol. 42, pp. 1685-1697, 1994.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE Transactions on Pattern Analysis, and Machine Intelligence, vol. 6, pp. 721-741, 1984.

[4] E. I. George and R. E. McCulloch, "Variable selection via Gibbs sampling," Journal of the American Statistical Association, vol. 85, pp. 398-409, 1993.

[5] W. R. Gilks, S. Richardson, and D. J. Spieglehalter, *Markov Chain Monte Carlo in Practice*, New York: Chapman Hall, 1996.

[6] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," Biometrika, vol. 82, pp. 711-732, 1995.

[7] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," Biometrika, vol. 57, pp. 97-109, 1970.

[8] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller," J. Chem. Phys., vol. 21, pp. 1087-1091, 1953.

[9] J. J. K. O'Ruanaidh and W. J. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, New York: Springer Verlag, 1996.

[10] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," Journal of the Royal Statistical Society B, vol. 59, pp. 731-792, 1997.

[11] P. T. Troughton and S. J. Godsill, "A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves," Technical Report, CUED/F - INFENG/TR.304, University of Cambridge, 1997.