ISBA 2000, Proceedings, pp. 000–000

Towards Perfect Sampling for Bayesian Mixture Priors

DUNCAN J. MURDOCH and XIAO-LI MENG
University of Western Ontario, Canada and The University of Chicago, U.S.A.

SUMMARY

Perfect sampling using the coupling from the past (CFTP) algorithm was introduced by Propp and Wilson in 1996. In much the way rejection sampling allows one to convert samplers from one distribution into samplers from another, CFTP allows one to convert Markov chain Monte Carlo algorithms from approximate samplers of the steady-state distribution into perfect ones. Since 1996 CFTP has been applied to many different Markov chains. However, its use in routine Bayesian computation is still in the early stages of development. This paper provides a couple of building blocks for its potentially routine application in Bayesian mixture priors, including a t mixture coupler, and demonstrates the types of difficulties that currently prevent CFTP from being applied routinely in Bayesian computation.

Keywords: COUPLING FROM THE PAST, EXACT SAMPLING, MCMC, MIXTURES.

1. INTRODUCTION

Markov chain Monte Carlo (MCMC) is commonly used to explore posterior distributions. In this technique, a Markov chain X_t is constructed so that the marginal distribution of sampled values converges to the posterior of interest; X_0 is set to some arbitrary value and then the chain is updated to produce $X_1, X_2, \ldots$. After skipping an initial segment ("burning in"), sampled values X_t are used to estimate properties of the posterior such as moments, modes, quantiles, etc.

A worry for practitioners is knowing how long the chain needs to be run before the marginal distribution of X_t is close enough to the posterior. The *coupling from the past* algorithm (CFTP; described below) formulated by Propp and Wilson (1996,1998) does not face this problem: it allows a Markov chain simulation method to be used to draw *exactly* from the limiting distribution of the Markov chain.

Propp and Wilson's work has stimulated a great amount of theoretical and methodological work in the emerging MCMC area of *perfect sampling* or *exact sampling*. David Wilson maintains an annotated bibliography covering much of this work; see `http://dimacs.rutgers.edu/~dbwilson/exact`; he has also written a primer on perfect sampling (Wilson, 2000b) in which CFTP and other perfect sampling algorithms are discussed. However, the use of perfect

sampling in routine applications of Bayesian computation is still in its infancy (see Green and Murdoch, 1999, Møller and Nicholls, 1999 and Murdoch, 2000). In this paper we explore the use of CFTP for Bayesian mixture priors; the case of mixture likelihoods was considered in Hobert *et al.* (1999) and will be further explored in future work. We present several ideas that are potentially fruitful and show the types of difficulties one currently faces in attempted routine use of perfect sampling for Bayesian analysis.

2. COUPLING FROM THE PAST: A BRIEF REVIEW

We start with the observation that if a positive recurrent Markov chain X_t had been run from the indefinite past ($t = -\infty$), then by time $t = 0$ it would have reached steady-state, and X_0 would be a draw from the limiting distribution. What CFTP does is to calculate the value of X_0 while only carrying out a finite amount of computation.

To illustrate, we note that the usual computer simulation of a Markov chain can be written in stochastic recursive sequence (SRS) form as

$$X_{t+1} = \phi(X_t, U_{t+1}), \tag{1}$$

where U_t is an i.i.d. sequence of random values from an easily simulated distribution, and ϕ is a deterministic function. Calculation of X_0 appears at first to be impossible without going back to $t = -\infty$: it depends on X_{-1}, which depends on X_{-2}, and so on. However, CFTP works because we do not always need to know X_{-1} exactly in order to calculate X_0, and we need to know even less about sampled values further in the past. If we go far enough back, it is possible that no matter what value the chain took, it will evolve to the same X_0.

This is implemented by imagining the entire sequence U_t to have been sampled once and then fixed. (In practice we normally draw U_t values only when we need them.) We then consider a collection of chains started at some time $-T < 0$ from every possible state of X_{-T}. These are all updated using the same updating function ϕ and the same sequence of U_t values. If the resulting values of X_0 vary from chain to chain, then a larger value of T is used, and the same calculation is repeated. If the resulting values of X_0 all agree, then increasing T would have no effect, since any path has to pass through *some* state at time $-T$. The common value X_0 is clearly a draw from the limiting distribution.

The CFTP algorithm may or may not terminate, depending on the choice of updating function. It is essential that ϕ be chosen so that it causes states to *coalesce*, i.e., so that for at least some values of U and $Y \neq Z$, $\phi(Y, U) = \phi(Z, U)$. (This condition is not sufficient, but we will not give exact sufficient conditions: in practice a satisfactory ϕ function is demonstrated by the success of the CFTP algorithm!) Furthermore, even when a CFTP algorithm terminates with probability one in theory, it may take too long or too much memory to be practical. This is the key difficulty we need to overcome in order to make CFTP a routine tool in Bayesian computation. There have been a variety of clever methods developed in the literature, including the "bracketing" method of Propp and Wilson (1996) which can take advantage of a monotonicity property of

ϕ. However, since these methods require additional structures of the underlying Markov chain, their applicability depends critically on whether our chain meets these specific requirements. A challenge in routine applications of CFTP is either to identify such structures in a problem at hand (e.g., identify a useful partial ordering for the "bracketing" method), or to construct a new method, perhaps by tailoring an existing method for a different problem, to make the coalescence fast enough to be practical. The work reported in this paper is an example of such an effort.

An additional issue that can be problematic in routine practical implementation is that, because the U_t values may be re-used many times as T is increased, programming CFTP can be tricky. Wilson (2000a) has devised a variation called "read-once CFTP" that avoids this problem, allowing U_t values to be sampled only forward in time. Rather than search for a time $-T$ from which coalescence has occurred, read-once CFTP looks through fixed size blocks (of B steps, say), until it finds one for which coalescence occurs entirely within the block; see Figure 2 of Section 7 for an illustration. Taking advantage of the fact that the occurrence of this event is independent from block to block, Wilson shows that the last observation in a single path from the end of one coalescent block to the start of the next one is equivalent to a perfect CFTP sample.

3. MIXTURE PRIORS

The simplest way for a mixture to enter a Bayesian model is in the prior distribution. We may have a k component mixture prior, $\pi(\theta) \propto \sum_{i=1}^{k} p_i \pi_i(\theta)$. This could arise if the investigator was unsure which of several reasonably sharp priors should apply, or if s/he wanted to mix a small proportion of a diffuse prior with a relatively sharp one, in case the knowledge that led to the sharp prior turned out to be inapplicable, or if a complex prior was built up of components from relatively simple distributions.

Suppose we have data y, leading to likelihood $\mathcal{L}(\theta \,|\, y)$. The posterior is then $\pi(\theta \,|\, y) \propto \sum_i p_i \pi_i(\theta) \mathcal{L}(\theta)$. In order to construct a Markov chain, we follow the standard approach and augment the parameter to (θ, z) where $z \in \{1, \ldots, k\}$ indicates the mixture component. The joint density for (θ, z) is proportional to $p_z \pi_z(\theta) \mathcal{L}(\theta)$, whose marginal for θ is our target posterior density. With this augmentation, we can implement the Gibbs sampler by alternating between the two conditional distributions

$$\pi(z \,|\, \theta, y) \;=\; \frac{p_z \pi_z(\theta) \mathcal{L}(\theta)}{\sum_i p_i \pi_i(\theta) \mathcal{L}(\theta)} = \frac{p_z \pi_z(\theta)}{\sum_i p_i \pi_i(\theta)}, \tag{2}$$

$$\pi(\theta \,|\, z, y) \;\propto\; \pi_z(\theta) \mathcal{L}(\theta). \tag{3}$$

Sampling from $\pi(z \,|\, \theta, y)$, a discrete distribution on k values, is straightforward. Sampling from $\pi(\theta \,|\, z, y)$ may be less so, but it is typically easier than sampling directly from the desired mixture distribution, $\pi(\theta \,|\, y)$.

We may write this Markov chain in SRS form in many ways; a simple one is as follows, for the case where θ is one dimensional. We let $U_t = (U_t^{(1)}, U_t^{(2)})$,

where $U_t^{(i)}$ are i.i.d. $U(0,1)$ random variables. Then both θ_{t+1} and z_{t+1} are selected via the inverse CDF method from their conditional distributions using $U_t^{(1)}$ and $U_t^{(2)}$, respectively. The usefulness of this method depends on how easy it is to compute the inverse CDF function of $\pi(\theta \,|\, z, y)$, especially when θ is continuous and unbounded.

Does this chain converge in distribution to $\pi(\theta, z \,|\, y)$? Typically it does, though in cases where the conditional distributions $\pi(\theta \,|\, z, y)$ do not have areas of common support it will not. Does the CFTP algorithm work, at least in theory? Again under the conditions where the chain has a unique limiting distribution it will: there will be a possibility that two different chains will be updated with the same z_{t+1} value, and from then on all future updates of both will be identical. The following normal example illustrates this point as well as the most basic steps of our mixture coupler.

4. A NORMAL MIXTURE PRIOR EXAMPLE

To introduce the basic ideas of this paper, we start with an artificial $k = 3$ component mixture prior for a one-dimensional location parameter θ:

1. 0.1% from $N(0, 10)$ (a slightly diffuse prior in case our guesses are wrong);
2. 49.9% from $N(1, 1)$ (our first guess about θ);
3. 50% from $N(20, 1)$ (our second guess about θ).

Suppose we have one observation, $y = 12.1$, from $N(\theta, \tau^2)$, where $\tau^2 = 1$. The one observation setting is not restrictive under the normality assumption with known variance, as we can replace y by $\bar{y}$ and modify τ^2 accordingly.

Because this is a mixture of normal distributions, we can calculate the posterior analytically. In particular, with prior $\pi(\theta) = \sum_{z=1}^{k} p_z N(\mu_z, \sigma_z^2)$ and $y \sim N(\theta, \tau^2)$, we have

$$\theta \,|\, y \sim \sum_{z=1}^{k} w_z(y) N\left(\mu_z(y), \sigma_z^2(y)\right), \tag{4}$$

where $\mu_z(y) = (\sigma_z^{-2}\mu_z + \tau^{-2}y)/(\sigma_z^{-2} + \tau^{-2})$, $\sigma_z^2(y) = \left(\sigma_z^{-2} + \tau^{-2}\right)^{-1}$, and

$$w_z(y) \propto \frac{p_z}{\sqrt{\sigma_z^2 + \tau^2}} \exp\left\{-\frac{(\mu_z - y)^2}{2(\sigma_z^2 + \tau^2)}\right\}. \tag{5}$$

For our example, the components of the posterior turn out to be (1) 86.8% of $N(11, 10/11)$, (2) 0.0% of $N(6.55, 1/2)$, and (3) 13.2% of $N(16.05, 1/2)$.

To illustrate our basic mixture coupler, we pretend that the weight calculations above were not feasible, and try to determine the weights by simulation. We will simulate directly from the component distributions; for many other problems this might not be feasible, but it would still often be easier to simulate from the components than from the mixture as a whole, especially when the weights are

not known. There are also other ways of getting around this problem (see Section 6). Following the general CFTP strategy described in Section 2, we started our coupler from all three possible states of z, and then proceeded as follows for the three chains:

1. For sampling θ given z_t, we simulated a common $X \sim N(0,1)$ and then applied a linear transformation depending on the value of z_t to obtain θ_{t+1} as a draw from $\pi(\theta|z_t, y)$. Therefore, if two chains have the same z_t value, they will also produce the same θ_{t+1}.

2. To sample z given θ_{t+1}, we calculated the multinomial distributions $\pi(z|\theta_{t+1}, y)$ for the three (or fewer) values of θ_{t+1}, and sampled from them using the inverse CDF method with a common $U \sim U(0,1)$. Thus, if two chains have the same θ_{t+1}, they will produce the same z_{t+1}.

With this basic coupler, coalescence of CFTP is fairly slow. The median value of T for coalescence was 128. However, the "perfect" nature of the sampler is indicated in Figure 1, which plots the simulated values of θ (left panel) and the posterior density of θ using the estimated weights from the proportions of the simulated z's (right panel), both against the true density curve.

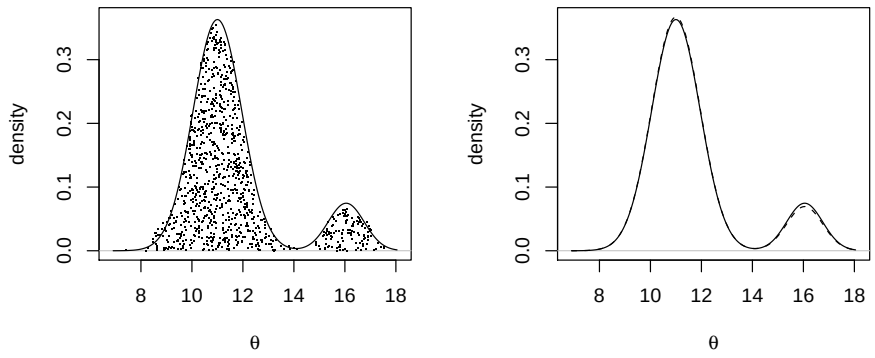

Figure 1. *Plot of posterior density for θ. In both plots, the solid line is the true density; on the left, 1,000 simulated values are shown using the variable width jittered plotting method of Lee and Tu (1997). On the right, the density estimate obtained by estimating the weights from the proportions of the 1,000 simulated z values, is shown dashed.*

5. TRANSFORMING THE COMPONENTS

In the example above, we observed that coalescence was quite slow because the Gibbs sampler did not mix well: it was quite difficult to make a jump from one mode to the other. A well-known and often effective technique for speeding up a Gibbs sampler is to reduce the dependence among its components via transformation (i.e., reparametrization), as we demonstrate below.

First, we can shift each component so that the modes are aligned. That is, we change variables from (θ, z) to $(\phi, z) = (\theta - m_z, z)$, where m_z is the mode of the conditional distribution of $\theta \,|\, z$. The Jacobian is 1, so the joint density of (ϕ, z) is proportional to $p_z \pi_z(\phi + m_z)\mathcal{L}(\phi + m_z)$ and the Gibbs conditionals are

$$\pi(z \,|\, \phi, y) = \frac{p_z \pi_z(\phi + m_z)\mathcal{L}(\phi + m_z)}{\sum_i p_i \pi_i(\phi + m_i)\mathcal{L}(\phi + m_i)}, \tag{6}$$

$$\pi(\phi \,|\, z, y) \propto \pi_z(\phi + m_z)\mathcal{L}(\phi + m_z). \tag{7}$$

This produces a vast improvement in the speed for our normal example, with the median value of T needed for coalescence reduced from 128 to 1.

Second, we could go further and also rescale each component so that the dispersions of the conditional densities of $\theta \,|\, z$ match each other more closely. Specifically, we can transform (θ, z) to $(\phi, z) = (A_z^{-1}(\theta - m_z), z)$, where the conditional covariance of θ is $\Sigma_z = A_z A_z^T$. With this transformation, the Jacobian is $|\, A_z \,|$, so the the Gibbs conditionals are

$$\pi(z \,|\, \phi, y) = \frac{|\, A_z \,| \, p_z \pi_z(A_z\phi + m_z)\mathcal{L}(A_z\phi + m_z)}{\sum_i |\, A_i \,| \, p_i \pi_i(A_i\phi + m_i)\mathcal{L}(A_i\phi + m_i)}, \tag{8}$$

$$\pi(\phi \,|\, z, y) \propto \pi_z(A_z\phi + m_z)\mathcal{L}(A_z\phi + m_z). \tag{9}$$

For our normal example, this location-scale transformation makes ϕ independent of z, so our mixture coupler would always coalesce in one step. In other words, in this special case, this transformed CFTP algorithm is the same as directly drawing ϕ and z from their joint density, which is the product of two marginal densities, and thus directly obtaining independent draws of θ via $\theta = A_z\phi + m_z$. In problems where CFTP is needed, this simplification does not happen. However, this is not necessarily a problem because our goal is to find a suitable transformation that will render fast coalescence, not necessarily in one step. In other words, we can choose some approximate m_z and A_z to achieve both simplicity and fast mixing rate, as we show in the next section.

6. A T MIXTURE COUPLER VIA FURTHER AUGMENTATION

In the previous section we avoided the difficult construction of a CFTP coupler for the continuous and typically unbounded θ by augmenting θ with z, which only has k possible states, and using the two-step Gibbs sampler. (In practice, k is typically small; for large k, the multi-stage CFTP of Meng (2000) may help to reduce the computational load for starting the chain at every possible state.) A key for this method to work is that the resulting conditional draw of θ (or ϕ with the transformation) given z is easy to couple with different values of z.

When this direct conversion does not lead to easily coupled draws for θ or ϕ given z, further augmentation could be helpful. Consider the same mixture problem as in Section 4 except the normal likelihood is replaced by a d-dimensional t likelihood with ν degrees of freedom. We retain the normal prior; it is quite common and convenient to use normal distributions to represent prior information regardless of the form of the likelihood. Specifically, suppose we have i.i.d samples $Y = \{Y_1, \ldots, Y_n\}$ from a d-dimensional location-scale t distribution $t_d(\theta, \Sigma_0; \nu)$, where the degrees of freedom ν and scale parameter Σ_0 are assumed to be known. Our prior for θ is given by $\pi(\theta) = \sum_{z=1}^{k} p_z N(\mu_z, \Sigma_z)$. (Note that $k = 1$ gives the usual posterior from a t-likelihood with normal prior.) In this case $\pi(\theta \,|\, z, Y)$ does not correspond to any convenient multivariate distribution and it is not easy to couple directly. However, using the decomposition,

$$Y = \theta + \frac{\Sigma_0^{1/2} Z}{\sqrt{q}}, \quad Z \sim N_d(0, I), \quad q \sim \chi^2_\nu/\nu, \quad Z \perp q, \tag{10}$$

we can augment Y into $\{Y, q\} = \{(Y_i, q_i), i = 1, \ldots, n\}$. It is easy to check (as in van Dyk and Meng, 2001) that given z and q (and Y)

$$\theta \,|\, Y, q, z \;\sim\; N_d\left(\mu_{z,q}, \Sigma_{z,q}\right), \tag{11}$$

where

$$\mu_{z,q} = \Sigma_{z,q}\left[\Sigma_0^{-1} \sum_i q_i Y_i + \Sigma_z^{-1} \mu_z\right], \quad \Sigma_{z,q}^{-1} = (\sum_i q_i)\Sigma_0^{-1} + \Sigma_z^{-1}. \tag{12}$$

Given θ, $\{z, q_1, \ldots, q_n\}$ are mutually independent with

$$q_i \,|\, \theta, Y \;\sim\; \frac{\chi^2_{\nu+d}}{(Y_i - \theta)^T \Sigma_0^{-1} (Y_i - \theta) + \nu}, \quad i = 1, \ldots, n, \tag{13}$$

and $\pi(z \,|\, \theta, Y)$ is a multinomial distribution calculated in the same way as in (2) with $\mathcal{L}(\theta)$ being the t-likelihood. We thus have a two-step Gibbs sampler alternating between drawing from θ_{t+1} given $\{z_t, q_t\}$ using (11)-(12), and $\{z_{t+1}, q_{t+1}\}$ given θ_{t+1} using (13) and the multinomial (2).

The addition of the continuous component q_t to the Gibbs sampler makes the construction of a coupler trickier. Murdoch and Green (1998) showed that coalescence of a coupled Gibbs sampler follows from coalescence of all but one component; here we need either θ_t or *both* z_t and q_t to coalesce. Fortunately, the simple univariate scaled χ^2 conditionals given in (13) allow us to directly apply Wilson's (2000b, Section 2.4.8) layered multiscale gamma coupler to sample q_i following (13). This coupler gives a discrete set of updates to the uncountable set of scaled $\chi^2_{\nu+d}$ distributions. When the scale factor is in an interval (a, b), the set is bounded above if $b < \infty$ and finite if, additionally, $a > 0$. We start with $a = 0$ and $b = 1/\nu$, from (13). When θ_{t+1} is calculated, the upper bound on q_t is sufficient to give an upper bound on $|\theta_{t+1}|$, which in turn gives $a > 0$ when we come to the q_{t+1} calculation. With a finite set of possible values of q_{t+1}, we then

pair each of them to the k possible values of z_{t+1}, and thereby determine a finite number of paths to follow. In a practical implementation this number may still be too large to handle, and we may still only keep track of the bounds; eventually, it will by chance take on as small a multiple of k as desired. There is a tradeoff between waiting a long time for a small multiple versus using a lot of computer memory for a large one.

The underlying Gibbs sampler here suffers the same slow mixing problem as the one in Section 4. Indeed, the two Gibbs samplers become the same when $\nu \to \infty$ (the d-dimensional version of the Gibbs sampler of Section 4 is obvious). We again can apply the transformation idea in Section 5 to speed up the mixing. However, we do not want to use the conditional mode $\mu_{z,q}$ and "standard deviation" matrix $\Sigma_{z,q}^{1/2}$ from (11) for the transformation. This is because although such a transformation will make ϕ independent of $\{q, z\}$, it will upset the important conditional independence structure in (13). To get around this problem, we can use the conditional mode m_z and "standard deviation" matrix A_z calculated from *pretending* the sampling density were normal $N_d(\theta, \Sigma_0)$ instead of $t_d(\theta, \Sigma_0; \nu)$; that is, we can use the same transformation $\phi = A_z^{-1}(\theta - m_z)$ as in Section 5. The resulting conditional distribution for ϕ is obviously, by (11),

$$\phi \,|\, Y, q, z \;\sim\; N_d\left(A_z^{-1}(\mu_{z,q} - m_z),\; A_z^{-1}\Sigma_{z,q}(A_z^{-1})^T\right). \tag{14}$$

Conditional on ϕ, z and $q = \{q_1, \ldots, q_n\}$ are no longer independent, but conditional on both ϕ and z, we still have, independently for $i = 1, \ldots, n$,

$$q_i \,|\, \phi, Y, z \;\sim\; \frac{\chi^2_{\nu+d}}{(Y_i - A_z\phi - m_z)^T\Sigma_0^{-1}(Y_i - A_z\phi - m_z) + \nu}. \tag{15}$$

Thus the Gibbs step for drawing $\{z_{t+1}, q_{t+1}\}$ given ϕ_{t+1} can be easily accomplished by first drawing z_{t+1} from z given ϕ_{t+1}, which is the same multinomial as in (8) with $\mathcal{L}$ being the t likelihood, and then drawing q_{t+1} using (15) with $z = z_{t+1}$ and $\phi = \phi_{t+1}$. The multiscale gamma coupler can be implemented in the same way with appropriately modified bounds.

7. NUMERICAL ILLUSTRATION AND CONCLUDING REMARKS

To illustrate the couplers of Section 6, we start with the example of Section 4, but using $t_1(\theta, 1, \nu)$ in place of the $N(\theta, 1)$ likelihood. The prior for θ is the same normal mixture. With large degrees of freedom ($\nu = 50$), this is very similar to the previous example, and coalescence using the transformed coupler is fast, with $T < 10$ typically (see left panel of Figure 2). The untransformed coupler is slower, with $T > 200$ in most simulations. Using read-once CFTP to generate an i.i.d. sample of 1000 observations from the posterior required a total of about 8000 time steps with the transformed coupler, and about 670,000 time steps with the untransformed one. The block size B for the read-once CFTP was 6, the median of nine forward coalescence times we ran in advance.

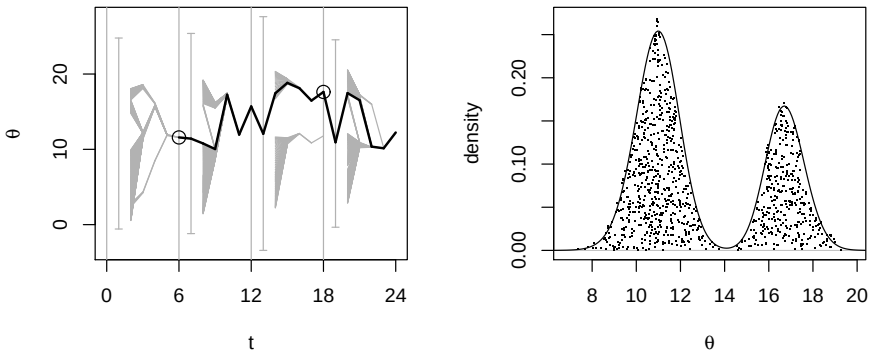

Figure 2. *An illustration of read-once CFTP for the t mixture with $\nu = 50$ and block size 6. On the left the steps required to obtain the first two draws are shown: the vertical gray lines represent the unbounded set of θ values at the start of each block; the vertical gray brackets show the bounds on θ from the multiscale gamma coupler; the other gray lines are the collection of updates of all of the finite states. The solid black line is the single path started from the end of the first coalescent block; the circles indicate the two draws. On the right 1000 draws are shown using the method of Lee and Tu (1997) together with the true posterior density function (normalized by numerical integration).*

On the other hand, when $\nu = 1$, coalescence was much slower. For the transformed coupler, T was in the neighborhood of 6200. We believe it took this long because the transformation was optimized for a normal likelihood, but with $\nu = 1$ the actual likelihood often is far from that. For the untransformed coupler, we gave up after 10^6 steps without observing any coalescence. Convergence may also be slower with $d > 1$. While working on a single observation with $d = 2$ and $\nu = 50$, the chains took about the same time to coalesce as with $d = 1$. However, the slow convergence with small ν is exacerbated in high dimensions: for example, the transformed algorithm failed to coalesce in 10^6 steps with $\nu = 1$ and $d = 2$.

With the t distribution, a sample size of 1 is hardly an interesting problem. We have used our coupler successfully on larger ($n = 50$) randomly generated datasets with large ν; it works well there provided the data are centered at a good estimate $\hat{\theta}$, presumably because the likelihood is so dominant in the posterior. In other words, it is better to bound $|\theta - \hat{\theta}|$ than to bound $|\theta|$ in the multiscale gamma coupler. Interestingly, very large n can also help to overcome the small ν problem when n is large enough so that the normal transformation is again approximately optimal. For other cases, the coupler fails because there are too many (e.g., 10^6) possibilities for the q vector: we did not have enough computer resources to follow all of them. Obtaining better transformations for such cases appears to be a key to make our coupler work in general.

The difficulties we encountered here, namely, (1) too slow to coalesce (due to slow mixing of the forward chain) and (2) too many states to trace, are typical

in current attempts to implement CFTP for Bayesian problems with unbounded and continuous state spaces. For models where the k-component mixture appears in the likelihood, the latter problem could be especially severe, as there will be k^n terms in the posterior, in contrast to k terms as in mixture-prior models. Overcoming these difficulties requires better constructions (e.g., better transformations, better bounds), which in turn require a level of analytic effort greater than is needed to find the corresponding forward MCMC algorithm. However, such efforts are worthwhile, especially for those of us whose research goal is to construct effective and reliable algorithms for general users.

REFERENCES

van Dyk, D. A. and Meng, X. L. (2001). The Art of Data Augmentation (with discussion). *The Journal of Computational and Graphical Statistics*, 10, to appear.

Green, P. J. and Murdoch, D. J. (1999) Exact sampling for Bayesian inference: towards general purpose algorithms. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, 301-321. Oxford University Press.

Hobert, J. P., Robert, C. P. and Titterington, D. M. (1999). On perfect simulation for some mixtures of distribution. *Statistics and Computing*, to appear.

Lee, J. J. and Tu, Z. N. (1997). A versatile one-dimensional distribution plot: The BLiP plot. *The American Statistician*, 51:353–358.

Meng, X. L. (2000). Towards a more general Propp-Wilson algorithm: Multistage Backward Coupling. In *Monte Carlo Methods – Fields Institute Communications Vol. 26* (Editor: Neal Madras), 85-93. AMS.

Murdoch, D. J. (2000). Exact sampling for Bayesian inference: Unbounded state spaces. In *Monte Carlo Methods – Fields Institute Communications Vol. 26* (Editor: Neal Madras), 111-121. AMS.

Murdoch, D. J. and Green, P. J. (1998). Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, 25:483–502.

Møller, J. and Nicholls, G. K. (1999). Perfect simulation for sample-based inference. Preprint.

Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252.

Propp, J. G. and Wilson, D. B. (1998). How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *Journal of Algorithms*, 27:170–217.

Wilson, D. B. (2000a). How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms* 16: 85-113.

Wilson, D. B. (2000b). Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP). In *Monte Carlo Methods – Fields Institute Communications Vol. 26* (Editor: Neal Madras), 143-179. AMS.