A GENERAL PURPOSE SCALE-INDEPENDENT MCMC ALGORITHM

J. Andrés Christen

Comunicación Técnica No I-07-16/12-12-2007
(PE/CIMAT)

A General Purpose Scale-Independent MCMC Algorithm

J. Andrés Christen*

CIMAT, Guanajuato, Mexico.

Colin Fox

Department of Physics, University of Otago, New Zealand.

December 2007

Abstract

We develop a new effectively adaptive, general purpose MCMC for arbitrary continuous distributions and correlation structure. We call this MCMC the *t-walk*. The t-walk maintains two independent points in the sample space, and all moves are based on proposals that are then accepted with a standard Metropolis-Hastings acceptance probability on the product space. Hence the t-walk may be viewed as an adaptive MCMC sampler that maintains a set of two points in the state space and moves them with some structure. However the t-walk is strictly not adaptive on the product space, but does display beneficial self-adjusting behavior on the original state space. Four proposal distributions, or 'moves', are given resulting in an algorithm which is effective in sampling distributions of arbitrary scale, without the requirement for further tuning of parameters. Several examples are presented showing good mixing and convergence characteristics, varying in dimensions from 1 to 200 and with radically different scale and correlation structure, using exactly the same sampler.

KEYWORDS: Adaptive MCMC; Bayesian Inference; Simulation; t-walk.

*J. Andrs Christen (Corresponding author) is Investigator Titular B, Centro de Investigacin en Matemticas, A. C. (CIMAT), A.P. 402, Guanajuato, Gto. 36000, Mexico (E-mail: jac@cimat.mx). Colin Fox is Proffesor Department of Physics, University of Otago, Dunidin, New Zealand (E-mail: fox@math.auckland.ac.nz).

1 INTRODUCTION

We develop a new MCMC sampling algorithm that contains neither adaptivity nor tuning parameters yet that can adapt to target distributions with arbitrary scale and correlation structure. We dub this algorithm the "t-walk" (since moves are designed to *traverse* the state space, in a *thoughtful* way). Because the t-walk is constructed as a Metropolis-Hastings algorithm on the product space it is provably convergent under the usual mild conditions.

Application areas are in sampling continuous densities with arbitrary scale and correlation structure. In applications where a change of variables will be applied to improve sampling from distributions with correlation, the t-walk will sample with adequate efficiency in most cases. Indeed, because the t-walk is not actually adaptive, it can efficiently sample from distributions that have local correlation structure that differs in different parts of state space. On the original state space the step size and direction appear to adapt continuously to the local structure. Hence the t-walk is excellent for initial exploration as it overcomes the need to tune proposals for scale and correlation, which is typically the first difficulty encountered when learning MCMC methods. We expect that for a large number of problems the t-walk will allow sufficiently efficient sampling of the target distribution that no recourse to further algorithm development is required.

There is an increasing interest in using Bayesian methods in a number of scientific and engineering applications that may require the use of sophisticated sampling methods such as MCMC (see Firmani, Avila-Reese, Ghisellini, and Ghirlanda, 2007, Jeffery, von Hippel, Jefferys, Winget, Stein and DeGennaro, 2007, Bavencoff, Vanpeperstraete and Le Cadre, 2006, Symonds, Reavell, Olfert, Campbell and Swift, 2007, Emery, Valenti and Bardot, 2007, just to mention some very recent examples). Therefore, developing a generic, adaptable and easy to use MCMC methods like the t-walk will help non-statisticians who are looking to use Bayesian inferential methods in their field of work.

Because the t-walk is useful as a black-box sampling algorithm it potentially allows researchers to focus on data analysis rather than MCMC algorithms. Even though it may be not quite as efficient as a well-tuned algorithm, its use reduces time from problem specification to data analysis in one off research jobs, since the only input required is the log of the target distribution and two initial points in the parameter space. Also, the t-walk will prove useful in multiple data analyses where details of the posterior distribution depend sufficiently on a particular data set that adjustment would be required to the proposal in a standard Metropolis-Hastings algorithm, allowing for automatic use of MCMC sampling.

We show that the t-walk performs well with several examples of dimension from one to 200. Good results are obtained, always simulating from the objective function successfully for all examples that range across different scales and dimensions.

A review of adaptive MCMC algorithms was given by Warnes (2000, chap 1) who classified adaptive algorithms under two broad groups as follows. Those MCMC samplers that aim at updating tuning parameters using information of the chain and/or of the objective function (see, for example, Gilks, Roberts and Sahu, 1998, Brockwell and Kadene, 2005, Haario, Saksman and Tamminen, 2001), and the adaptive direction samplers (ADS) that maintain several points in the state space (see, for example, Gilks, Roberts and George, 1994, Gilks and Roberts, 1996, Eidsvik and Tjelmeland, 2004, or the "evolutionary Monte Carlo" that combines ADS with moves from genetic algorithms to speed up a Metropolis coupled MC in Liang and Wong, 2001). However, we have not found successful applications of these MCMC schemes to a comprehensive suite of objective functions. Most require specific mathematical calculations to be made for each objective function and in many cases the regularity conditions for convergence are complex. In contrast, the t-walk has mild convergence requirements since it mixes a set of standard Metropolis-Hastings kernels, and *only* requires evaluation of the target density.

The paper is structured as follows: in Section 2 we explain the t-walk and establish its ergodic properties (based on standard results for M-H algorithms). In Section 3.1 we present several two dimensional examples and in Section 3.2 we present a more complex example involving a mixture of normals. Finally a discussion of the paper may be found in Section 5.

2 THE T-WALK DESIGN

For an objective function (posterior distribution, etc.) $\pi(x)$, $x \in \mathcal{X}$ ($\mathcal{X}$ has dimension n and is a subset of R^n), we form the new objective function $f(x, x') = \pi(x)\pi(x')$ on the corresponding product space $\mathcal{X} \times \mathcal{X}$. While a general proposal has the form

$$q\{(y, y') \mid (x, x')\},$$

we consider the two restricted proposals

$$(y, y') = \begin{cases} (x, h(x', x)), & \text{with prob. } 0.5 \\ (h(x, x'), x'), & \text{with prob. } 0.5 \end{cases} \quad (1)$$

where $h(x, x')$ is a random variable used to form the proposal. That is, we change only one of x or x' in each step.

Within a Metropolis-Hastings scheme, we need to calculate the corresponding acceptance ratio. Denoting the density function of $h(x, x')$ by $g(\cdot \mid x, x')$, the ratio is equal to

$$\frac{\pi(y')}{\pi(x')} \frac{g(x' \mid y', x)}{g(y' \mid x', x)}$$

for the first case in equations 1 and

$$\frac{\pi(y)}{\pi(x)} \frac{g(x \mid y, x')}{g(y \mid x, x')}$$

for the second case. In all four of the moves below we simulate $\phi_j \sim Be(p); j = 1, 2, \ldots, n$ independently. If ϕ_j coordinate j is not updated. p is choosen so as $np \leq 5$. At each iteration we let $n_\phi = \sum_{j=1}^{n} \phi_j$.

We have found that the following four choices for h give adequate mixing across a wide range of target distributions.

2.1 Walk move

In many applications, particularly with weak correlations, we find that mixing of the chain is primarily achieved by a *scaled random walk* that we refer to as the walk move.

The walk move is defined by the function

$$h_{\mathrm{w}}(x,x')_j = x_j + \phi_j\left(x_j - x'_j\right) z_j$$

for $j = 1, 2, \ldots, n$, where $z_j \in \mathbb{R}$ are i.i.d. r.v. with density $\psi_{\mathrm{w}}(\cdot)$. This proposal is symmetric when

$$\psi_{\mathrm{w}}\left(\frac{-z}{1+z}\right) = (1+z)\,\psi_{\mathrm{w}}(z).$$

We achieve this by setting

$$\psi_{\mathrm{w}}(z) = \begin{cases} \dfrac{1}{k\sqrt{1+z}}, & z \in \left[\dfrac{-a}{1+a}, a\right] \\ 0, & \text{otherwise,} \end{cases}$$

for any $a > 0$, with normalizing constant $k = 2\left(\sqrt{1+a} - 1/\sqrt{1+a}\right)$. This density is simple to simulate using the inverse cumulative distribution as

$$z = \frac{a}{1+a}\left(-1 + 2u + au^2\right)$$

where $u \sim U(0,1)$. We set $a = 1/2$ in our implementation of the t-walk. Consequently, the Hastings ratio for the second case is

$$\frac{g_{\mathrm{w}}(x \mid y, x')}{g_{\mathrm{w}}(y \mid x, x')} = 1,$$

and similarly for the first case. Hence the acceptance probability is simply given by the ratio of target densities.

2.2 Traverse move

A typical difficulty experienced by samplers using random walk moves is with densities with strong correlation between a few, or several, variables. A typical solution is to rotate coordinates of the state variables or, equivalently, the proposal distributions. However, that is not feasible with distributions where the correlation structure changes through state space. (An example of such a distribution may be found in Figure 3(b).)

For those applications, efficiency of the sampler is greatly enhanced by the 'traverse move' defined by

$$h_{\mathrm{t}}(x, x') = \begin{cases} x' + \beta(x' - x) & \phi_j = 1 \\ \\ x & \phi_j = 0. \end{cases}$$

where $\beta \in R^+$ is a r.v. with density $\psi_{\mathrm{t}}(\cdot)$.

The case $\beta \equiv 0$ is similar as Skilling's leap-frog move (see MacKay, 2003, sec. 30.4) restricted to two states and a subset of coordinates. Since the t-walk maintains just two states, the traverse move does not have the random selection of states as in the leap-frog move. As noted by MacKay (2003, p. 394), this move has similarities to the 'snooker' move used in ADS's. The traverse move is therefore much simpler than either leap-frog or snooker, and like the leapfrog move, is more widely applicable than the snooker move since calculation of conditional densities is not required.

Since just one random number is used in this proposal, except for the case $\beta \equiv 1$, it is not possible to make both the proposal and the acceptance ratio independent of the dimension of state space, n. However, by setting $\psi_{\mathrm{t}}(1/\beta) = \psi_{\mathrm{t}}(\beta)$, for all $\beta > 0$, the ratio of propolsals is simplified to $\beta^{n_\phi - 2}$, see below. A density of this kind may be obtained by using a density $\phi(\cdot)$ on R^+ and defining $\psi_{\mathrm{t}}(\beta) = K\{\phi(\beta^{-1}-1)I_{(0,1]}(\beta) + \phi(\beta-1)I_{(1,\infty)}(\beta)\}$, for a normalizing constant K (we need $\int_0^1 \phi(\beta^{-1} - 1)d\beta < \infty$). A simple and convenient result is obtained with $\phi(y) = (a-1)(y+1)^{-a}$, for any $a > 1$, in which case

$$\psi_{\mathrm{t}}(\beta) = \frac{a-1}{2a}\{(a+1)\beta^a I_{(0,1]}(\beta)\} + \frac{a+1}{2a}\{(a-1)\beta^{-a} I_{(1,\infty]}(\beta)\},$$

which is a mixture of two distributions and may be easily sampled from with the following algorithm

$$\beta = \begin{cases} u^{\frac{1}{a+1}}, & \text{with prob. } \frac{a-1}{2a} \\ u^{\frac{1}{1-a}}, & \text{with prob. } \frac{a+1}{2a}, \end{cases} \tag{2}$$

where $u \sim U(0,1)$. We want steps to be taken around the length of $||x - x'||$, thus we state that $P(\beta < 2) \approx 0.9$. This is achieved with $a = 4$ giving $P(\beta < 2) \approx 0.92$. A plot of $\psi_{\mathrm{t}}(\beta)$ with $a = 4$ is presented in Figure 1. Following the above transformation, it is clear that

$$g_{\mathrm{t}}(y \mid x, x') = f\left(\frac{||y - x'||}{||x - x'||}\right) ||x - x'||^{-1}.$$

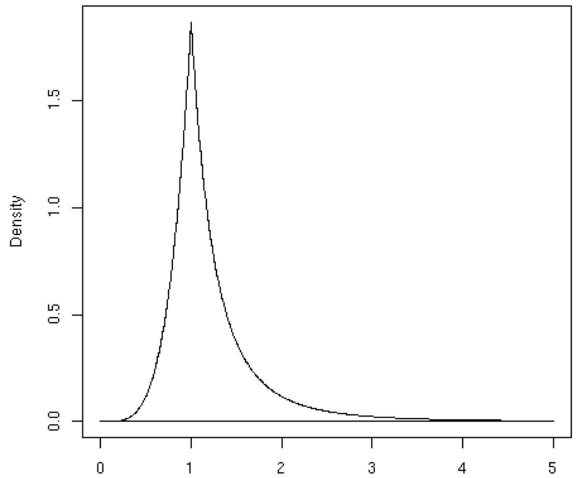


Figure 1: $\psi_t(\beta)$ with $a = 4$ giving $P(\beta < 2) \approx 0.92$.

A note of caution is prudent here, regarding calculation of the acceptance probability for this move. Since the range of h is a subspace of $\mathcal{X}$ it is most convenient to use the reversible jump MCMC formalism (see Green and Mira, 2001) for evaluating the acceptance ratio The corresponding Jacobian determinant equals $\beta^{n_\phi - 2}$, and since $\psi_t(1/\beta) = \psi_t(\beta)$ the acceptance ratio is $\frac{\pi(y')}{\pi(x')}\beta^{n_\phi - 2}$ or $\frac{\pi(y)}{\pi(x)}\beta^{n_\phi - 2}$, for the first and second cases, respectively.

The discussion in MacKay (2003) of why Skilling's leapfrog method works largely applies to the traverse move. In particular example 30.3 of MacKay (2003) and its solution, shows that applying these moves to a Gaussian distribution in n dimensions with covariance matrix proportional to the identity results in an expected acceptance ratio of e^{-2n}. Hence this move has a very low acceptance ratio for large number of uncorrelated variables. However when many variables are strongly correlated the effective value of n decreases, giving good mixing for such distributions, as intended. In examples with correlation as high as $1 - 10^{-7}$ (typical of examples from inverse problems) the traverse move is effective in mixing *along* the long axis of the distribution, but is very slow in mixing in directions perpendicular to the long axis. Then combining the traverse move with the other moves in the t-walk results in an effective sampling algorithm.

2.3 Hop and Blow moves

The walk and traverse moves are not, by themselves, enough to guarantee irreducibility of the chain over arbitrary target distributions. It is therefore necessary to introduce further moves to ensure this. Further, both the walk and traverse moves can lead to extremely slow mixing for distributions with very high correlation (say 0.9999 or higher), as mentioned above. We find that these difficulties are cured by employing two further moves that make bold proposals, but are chosen with relatively low probability (see below). We call these moves the hop and blow moves. We have at least one bimodal example in which switching between modes is improved substantially by choosing the hop and blow moves 10% of the time.

A hop move is defined by the choice for h;

$$h_{\mathrm{h}}(x, x') = (x_j + \phi_j z_j \sigma(x, x')/3)$$

with $z_j \sim N(0,1)$, where $\sigma(x, x') = \max_{j=1,2,\dots,n} \phi_j |x_j - x'_j|$. We call this the *hop* move. For this proposal

$$g_{\mathrm{h}}(y \mid x, x') = \frac{(2\pi)^{-n_\phi/2} 3^{n_\phi}}{\sigma(x, x')^{n_\phi}} \exp\left\{ -\frac{9}{2\sigma(x, x')^2} \sum_{j=1}^{n} (y_j - x_j)^2 \right\}.$$

Note that this move is centred at x.

Finally we consider the blow move defined by

$$h_{\mathrm{b}}(x, x')_j = \begin{cases} x'_j + \sigma(x, x') z_j & \phi_j = 1 \\ x_j & \phi_j = 0, \end{cases}$$

with $z_j \sim N(0,1)$. We thus have

$$g_{\mathrm{b}}(y \mid x, x') = \frac{(2\pi)^{-n_\phi/2}}{\sigma(x, x')^{n_\phi}} \exp\left\{ -\frac{1}{2\sigma(x, x')^2} \sum_{j=1}^{n} (y_j - x'_j)^2 \right\}.$$

Note that, as opposed to the walk and hop moves above, this move is centred at x'.

2.4 Convergence

Let $K_\alpha(\cdot, \cdot)$ be the corresponding M-H transition kernel for proposal q_α, where $\alpha \in \{\mathrm{w}, \mathrm{t}, \mathrm{h}, \mathrm{b}\}$. (There is a positive probability of obtaining n_ϕ and this ensures strong aperiodicity.) It may be seen, using the properties of the M-H method, that each K_α also satisfies detailed balance with $f(x, x')$. We form the transition kernel:

$$K\{(x,x'),(y,y')\} = \sum_{\alpha \in \{\mathrm{w,t,h,b}\}} w_\alpha K_\alpha\{(y,y') \mid (x,x')\},$$

where $\sum_\alpha w_\alpha = 1$, which consequently also satisfies the detailed balance condition with f. Assuming that also K is f-irreducible (note that hop and blow moves ensure f-irreducibility), then f is the limit distribution of K (see Robert and Casella, 1999, chapter 6, for details).

In our implementation of the t-walk we set the move probabilities $w_\mathrm{w}, w_\mathrm{t}, w_\mathrm{h}, w_\mathrm{b} = 0.4918, 0.4918, 0.0082, 0.0082$. These values were chosen to give the minimum integrated autocorrelation time, i.e. roughly the number of iterations per independent sample, across the two-dimensional bi-modal examples presented later in Section 3. Interestingly, these values were close to optimal for each of the example target distributions considered, and little compromise was required.

2.5 Properties

The following theorem states that the t-walk is invariant to changes in scale and reference point.

Theorem 1 *Given a transformation of the space $\mathcal{X}$, $\phi(x) = ax + b$, where $a \in R, a \neq 0$ and $b \in R^n$, that generates the new objective function $\lambda(z) = |a^{-n}|\pi(\phi^{-1}(z))$, one may generate a realization of the t-walk either by applying the t-walk kernel with λ as objective function, with starting values z_0, z_0', or by applying the t-walk kernel to π, with starting values $\phi^{-1}(z_0), \phi^{-1}(z_0')$, and then transforming the resulting chain with ϕ.*

Proof: Let $V_0 = (\phi^{-1}(z_0), \phi^{-1}(z_0'))$ and $W_1 \in \phi(\mathcal{X}) \times \phi(\mathcal{X})$. Elementary calculations lead to the fact that $|a^{-n}|q_\alpha(\phi^{-1}(W_1) \mid V_0) = q_\alpha(W_1 \mid \phi(V_0))$ for $\alpha = w, t, b, h$. Using this it is easy to see for the M-H acceptance probabilities, using π and λ, that $\rho^\pi_{h_j}(V_0, \phi^{-1}(W_1)) = \rho^\lambda_{h_j}(\phi(V_0), W_1)$. It is clear then that $\rho^\pi_{h_j}(V_0, \phi^{-1}(W_1))|a^{-n}|q_{h_j}(\phi^{-1}(W_1) \mid V_0) = \rho^\lambda_{h_j}(\phi(V_0), W_1) q_{h_j}(W_1 \mid \phi(V_0))$. This, together with the fact that the probability of not jumping in either case is the same, $1 - r^\lambda_{h_j}(\phi(V_0)) = 1 - r^\pi_{h_j}(V_0)$, establishes the result (see Robert and Casella, 1999, p. 235 for similar notation). •

What we have just proved is that applying the t-walk kernel with π and transforming it with ϕ has density $|a^{-n}| q_{h_j}(\phi^{-1}(W_1) \mid V_0) + \delta_{V_0}(W_1)\{1 - r^{\lambda}_{h_j}[\phi(V_0)]\}$ which is equal to $K_\lambda(\phi(V_0), W_1)$. It is immediate that this also holds for n steps into the t-walk and therefore, for any set B (of the transformed space) $K^n_\pi(V_0, \phi^{-1}(B)) = K^n_\lambda(\phi(V_0), B)$ and since also $f_\pi(\phi^{-1}(B)) = f_\lambda(B)$ we have that

$$||K^n_\pi(V_0, \cdot) - f_\pi(\cdot)||_{TV} = ||K^n_\lambda(\phi(V_0), \cdot) - f_\lambda(\cdot)||_{TV},$$

where $f_\pi(A) = \int_A \pi(dx)\pi(dx')$ and $f_\lambda(B) = \int_B \lambda(dx)\lambda(dx')$.

The above establishes a powerful characteristic of the t-walk; its performance (speed of convergence, autocorrelations, etc) remain unchanged with a change in scale and position as given by ϕ. Moreover, if the t-walk is limited to moves the traverse and walk moves, the theorem remains valid even if we use a more general change in scale using $a = diag(a_j)$, a diagonal matrix, with $a_j \in R, a_j \neq 0$. Furthermore, the traverse move is invariant when a is any nonsingular matrix.

Although the t-walk contains random walk type updates, it may not be reduced solely to a random walk MCMC in the usual sense. Since x_n *and* x'_n both have $\pi(\cdot)$ as limit distribution, then $||x_n - x'_n||$ is a property of π and in the limit has the distribution of the distance between two points sampled independently from π. Hence the "step size" (in some loose sense) cannot actually be manipulated or designed in any way. However, when viewed as a sampler on the original space, the step size appears to adapt to the characteristics of shape and size of the section of π that is being analyzed.

3 NUMERICAL EXAMPLES

3.1 2-dimensional examples

We present some simple examples with two parameters. First we experiment with a bimodal, correlated, objective function that has the form

$$h(x) = K \exp\left\{-\tau \left(\sum_{i=1}^{2}(x_i - m_{1i})^2\right)\left(\sum_{i=1}^{2}(x_i - m_{2i})^2\right)\right\}, \tag{3}$$

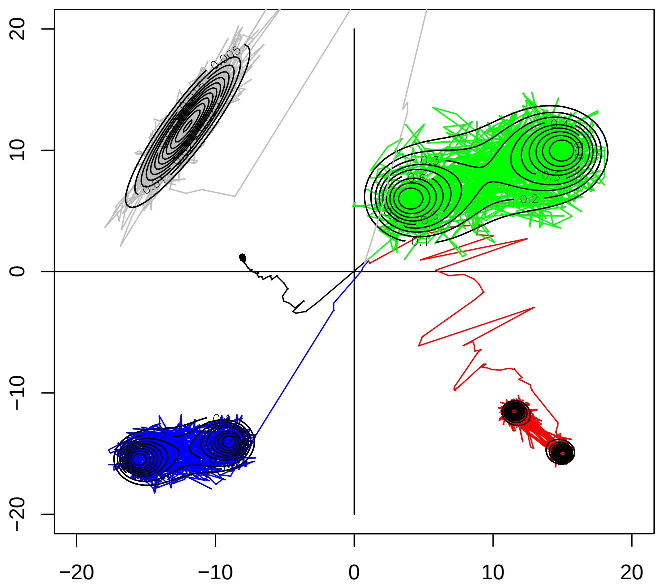
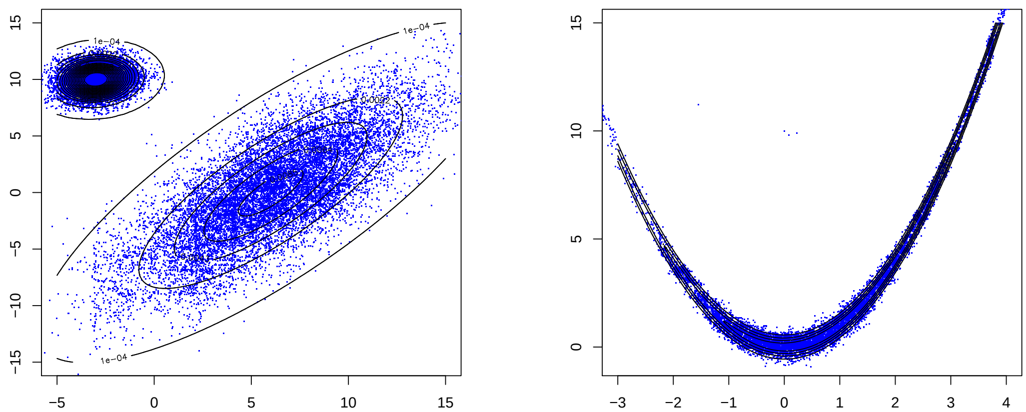


Figure 2: Sample paths for one component in the t-walk. Upper left quadrant: Bivariate normal distribution with correlation 0.95. Other quadrants, counterclockwise from the lower left: distribution in (3) with $\tau = 0.01, 0.1, 0.001, 1000$ (for $\tau = 1000$ the scale is such that the distribution shape can not be distinguished and is reduced to a point). In all cases we had an acceptance ratio of 40 to 50%, the starting points where $x_0 = (0, 0)$ and $x'_0 = (1, 1)$, with a sample of 5000 iterations.

for some m_1 and m_2 that approximately locate modes, and scale parameter $\tau > 0$ (K is a normalization constant). In figure 2 we present an illustration of the t-walk sample paths over quite different choices of the above distribution, and also on a correlated bivariate normal distribution.

We present two quite extreme two dimensional examples. Figure 3(a) shows a mixture of two rather contrasting bivariate normals, one flat, oval highly correlated mode and one peaked low correlated, forming an objective function with two modes. Figure 3(b) shows a strongly correlated hook shape objective function with thin edges and a thicker mid section where the mode is (see figure 3 for more details).

Note that we have run the t-walk over seven quite different objective functions, varying radically



(a) (b)

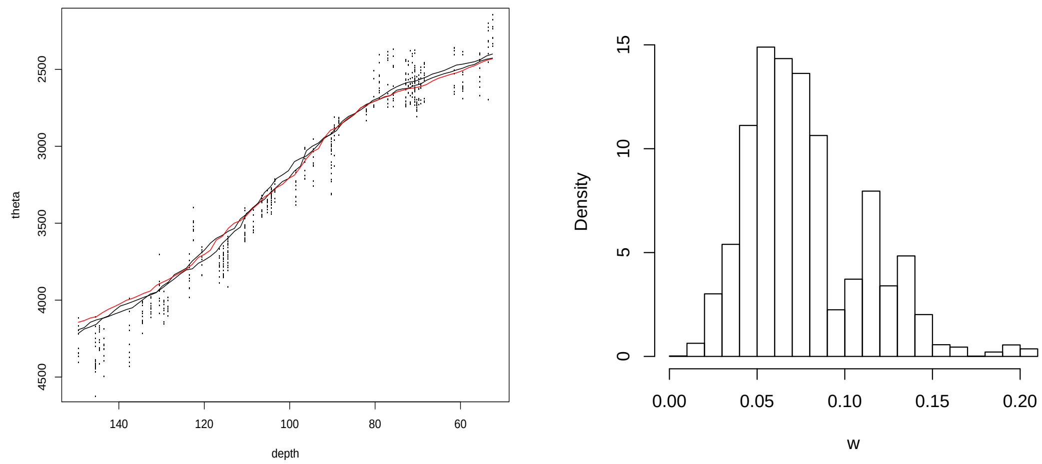
Figure 3: Sample points for one component in the t-walk (a) mixture of bivariate normals, low mode with weight 0.7, $\mu_1 = 6, \sigma_1 = 4, \mu_2 = 0, \sigma_2 = 5, \rho = 0.8$, high mode with weight 0.3, $\mu_1 = -3, \sigma_1 = 1, \mu_2 = 10, \sigma_2 = 1, \rho = 0.1$. We took 100,000 iterations with an acceptance rate of around 45%. (b) "Rosenbrok" (see Rosenbrok, 1960) density equal to $\pi(x, y) = C \exp\left[-k\left\{100(y - x^2)^2 + (1 - x)^2\right\}\right]$ (for some normalizing constant C), with $k = 1/20$. We used 100,000 iterations. This is quite a difficult density to plot and we needed to chop off the two tips of the hook so the corresponding algorithm in R could plot the contours correctly. In this case we obtained an acceptance ratio of about 13% with 100,000 points, lower than all other examples presented in this section.

in scale, correlation, modes, etc. The t-walk performed well and more or less similarly in all cases. Next we present a more complex example of dimension 15.

3.2 Higher dimension example

We study an example of a semiparametric age model for radiocarbon calibration in paleoecology. This problem was studied in Blaauw and Christen (2005) using a picewise linear approach and here we experiment using the t-walk with an age model with several parameters. It is not our intention to justify nor establish the validity of this approach in paleoecology but only to demonstrate the usefulness of the t-walk in an high dimension example. We have a series of radiocarbon determinations $y_j \pm \sigma; j = 1, 2, \ldots, m$ taken along a peat core at depths d_k. A semiparametric model is proposed to establish a relation between the (unknown) age of peat and depth, d, $G(d, x) = x_1 + \sum_{j=2}^{i} x_j \Delta c + x_{i+1}(d - c_i)$; where $c_i \leq d < c_{i+1}$ and the c_i's are depths uniformly spaced along the peat core with difference Δc. The usual normal model is considered, $y_j \mid d_j, x \sim N(\mu(G(d_j, x)), \sigma_j)$, where $\mu(\cdot)$ is the radiocarbon calibration curve, see Blaauw and Christen (2005) for details. Additionally, a model is proposed for the (peat accumulation) rates $x_j = w x_{j-1} + (1 - w) z_j$, where $w = x_n$ and $z_j \sim Gamma(\alpha, \beta)$, with α and β known (representing the information available on accumulation rates, see Blaauw and Christen, 2005).

A simple program (in C++) is used to calculate $-log f(x|y_1, y_2, \ldots, y_m)$. The other input required are the initial points for x and x'. For this we set $w = 0.4$ and $w' = 0.1$ and $x_{n-1} \sim Gamma(\alpha, \beta)$, $x_1 \sim N(\mu(G(d_1, x)), \sigma_1)$. This provides initial ("city park") values for x and x'. The data set use for the example is called "EngXV" with $m = 56$ determinations (see Blaauw, Geel, Mauquoy and Plicht, 2004). We took $n = 71$ (70 parameters for the age-depth model plus w) and ran 200,000 iterations of the t-walk (this took 2 minutes in a Mac G4 iBook machine). Two sample age-depth models and the MAP estimator is presented in Figure 4(a) and a histogram approximating the marginal distribution of w is presented in Figure 4(b).



(a) (b)

Figure 4: (a) MAP estimator (red) and two sample age-depth models for core EngXV. For each of the $m = 56$ determinations a sample of 10 calendar ages were simulated and plotted (small dots). (b) Histogram for the marginal posterior distribution of w.

4 COMPARISONS

Roberts and Rosenthal (2001) present a review of some optimal scaling for a random walk Metropolis Hasting algorithm for some simple models. In terms of Integrated Autocorrelation Time (IAT), the random walk Metropolis hasting most be tune to have an acceptance rate of 0.234, for the type of models considered by them. In particular, they consider the objective $\pi(x) = \prod_{j=1}^{d} C_j g(C_j x_j)$, where g is the standard Normal distribution, and in section 7.1 they take $C_j = 1$, model 1, $C_1 = 2$ and $C_j = 1; j = 2, 3, \ldots, n$, model 2 and $C_1 = 1$ and $C_j \sim Exp(1); j = 2, 3, \ldots, n$, model 3. Also we consider $C_j = 10$, model 0. We have already mentioned that a finely tuned MCMC for a particular objective function should be more of equally efficient that *any* generic method, including the t-walk. But fine tuning a Metropilis Hastings MCMC is the whole issue of applying the method. While very flexible and very general indeed, a M-H MCMC can be extremely ineffective and in high dimensions very difficult to tune. This is the idea behind and motivation for adaptive or self adjusting methods like the t-walk (see Roberts and Rosenthal, 2001, for more examples).

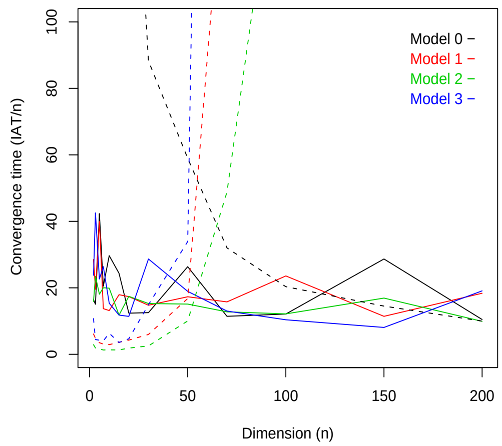


Figure 5: Integrated Autocorrelation Times dived by the dimension, over various dimensions for the t-walk, solid lines, and a random walk Metropolis Hastings, dashed lines. The models are taken from Roberts and Rosenthal (2001), see text.

Roberts and Rosenthal (2001) consider the IAT divided by the dimension of the parameter space as a means to compare convergence rates, or efficiency, among MCMC samples and compensate for space dimension. We fine tune a random walk M-H for model 1 for $n = 10$ (to an acceptance rate near 0.234) and use that same sampler in the rest of the examples. We also ran the t-walk in all examples, using dimensions $n = 2, 3, 5, 7, 10, 15, 20, 30, 50, 70, 100, 150, 200$. The results are shown in Figure 5. Note that in the case of the t-walk, for *all* models the IAT increases linearly with n and thus IAT/n remains in most cases below 30. For the random walk M-H IAT/n is lower, as expected, for n around 10 but soon it exploits making the chain extremely inefficient. The contrary effect is seen for model 0. Moreover, it is also the case that IAT/n remains bounded by 30 for all the examples presented in the previous sections, including the high dimension ($n = 71$) above. We would like to stress the fact that the *same* algorithm was used in all examples, requiring no tuning parameters besides the (log) of the objective function and two initial points in the sample space.

5 DISCUSSION

The t-walk has unique performing characteristics, adapting nicely to radically different scales, correlations, and across several dimensions with no tuning parameters. The very same sampler was used in all the examples shown here considering dimensions from 2 to 200.

However, we have found an example in which extremely high correlations in a high-dimensional problem lead to very slow mixing of the t-walk. Examples of posterior distributions with many highly correlated parameters arise, for example, in the field of inverse problems such as conductivity imaging. We intend to work on this problem to extend the applicability of this approach by developing moves that depend on a few more than two points in state space. We also look to develop a version of the t-walk that may cope with a mixture of discrete and continuous parameters.

We believe that the t-walk is already a useful improvement on existing attempts at creating automatic, generic, self adjusting, MCMC's. The current design results in a simple, mathematically tractable algorithm that lends itself to use as a black-box sampler, since only evaluation of the objective function is needed; there being no need to calculate any conditional distributions, etc. As presented in the numerical examples, we have evidence that the t-walk will perform satisfactory with common densities (posterior distributions in common Bayesian statistical analyses) of dimension up to perhaps 200 or more. For these problems the t-walk can be used as a black-box simulation technique, either for exploratory analysis of the objective density at hand or for final MCMC simulation. The t-walk is available in R (R Development Core Team, 2005) and C++ at `http://www.cimat.mx/~jac/twalk/`.

6 ACKNOWLEDGMENTS

Colin Fox was partially founded by the Department of Mathematics, University of Auckland (New Zealand) and J Andrés Christen was partially founded by grant SEMARNAT-2004-C01-0007 (Mexico).

References

[1] Bavencoff, F., Vanpeperstraete, J.M. and Le Cadre, J.P. (2006), "Constrained bearings-only target motion analysis via Markov chain Monte Carlo methods *IEEE Transactions on Aerospace and Electrical Systems*, **42**(4), 1240-1263.

[2] Blaauw, M., B. van Geel, D. Mauquoy, J. van der Plicht (2004). "Radiocarbon wiggle-match dating of peat deposits: advantages and limitations". *Journal of Quaternary Science*, **19**, 177-181.

[3] Brockwell, A.E. and Kadane, J.B. (2005), "Identification of Regeneration Times in MCMC Simulation, With Application to Adaptive Schemes", *Journal of Computational and Graphical Statistics*, **14**(2), 436–458.

[4] Blaauw, M. and Christen, J.A. (2005), "Radiocarbon peat chronologies and environmental change", *Applied Statistics*, **54**(4), 805-816.

[5] Eidsvik, J. and Tjelmeland, H. (2004), "On directional Metropolis-Hastings algorithms", *Statistics and Computing*, **16**(1), 93–106.

[6] Emery, A. F., Valenti, E. and Bardot, D. (2007), "Using Bayesian inference for parameter estimation when the system response and experimental conditions are measured with error and some variables are considered as nuisance variables, *Measurement Science & Technology*, **18**(1), 19-29.

[7] Firmani, C., Avila-Reese, V., Ghisellini, G. and Ghirlanda, G. (2007), "Long gamma-ray burst prompt emission properties as a cosmological tool, *Revista Mexicana de Astronomia y Astrofísica*, **43**(1), 203-216.

[8] Gilks, W.R. Roberts, G.O. and George, E.I. (1994) "Adaptive direction sampling", *The Statistician*, **43**, 179–189.

[9] Gilks, W.R. and Roberts, G.O. (1996), "Strategies for improving MCMC", in: *Markov Chain Monte Carlo in Practice*, Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. eds. Chapman and Hall: London, 89–114.

[10] Gilks, Roberts and Sahu (1998), "Adaptive Markov Chain Monte Carlo Through Regeneration", *Journal of the American Statistical Association*, **93**, 1045–1054.

[11] Green, P.J. and Mira, A. (2001), "Delayed Rejection in Reversible Jump Metropolis–Hastings", *Biometrika*, **88**, 1035-1053.

[12] Haario, H. Saksman, E. and Tamminen, J. (2001), "An adaptive Metropolis algorithm", *Bernoulli*, **7**, 223-242.

[13] Jeffery, E. J., von Hippel, T., Jefferys, W. H., Winget, D. E., Stein, N. and DeGennaro, S. (2007), "New techniques to determine ages of open clusters using white dwarfs, *Astrophysical Journal*, **658**(1), 391-395.

[14] Liang, F. and Wong, W.H. (2001), "Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models", *Journal of the American Statistical Association*, **96**454, 653-666.

[15] MacKay, David J. C. (2003), "Information Theory, Inference, and Learning Algorithms",Cambridge University Press.

[16] R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing: Vienna.

[17] Robert, C.P. and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer Texts in Statistics, Springer-Verlag: New York.

[18] Roberts, G. O. and Rosenthal, J. S. (2001), "Optimal scaling for various Metropolis-Hastings algorithms", *Statistical Science*, **16**(4), 351-367.

[19] Rosenbrock, H.H. (1960), "An automatic method for finding the greatest or least value of a function", *The Computer Journal*, **3**, 175-184.

[20] Symonds, J. P. R., Reavell, K.S.J., Olfert, J.S., Campbell, B.W. and Swift, S.J. (2007), "Diesel soot mass calculation in real-time with a differential mobility spectrometer, *Journal of Aerosol Science*, **38**(1), 52-68.

[21] Warnes, G.R. (2000), *The Normal Kernel Coupler: An adaptive Markov Chain Monte Carlo method for efficiently sampling from multi-modal distributions*, PhD Thesis: University of Washington.