Bayesian Analysis (2010) **5**, Number 2, pp. 263–282

A General Purpose Sampling Algorithm for Continuous Distributions (the t-walk)

J. Andrés Christen[*] and Colin Fox[†]

Abstract. We develop a new general purpose MCMC sampler for arbitrary continuous distributions that requires no tuning. 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 is provably convergent under the usual mild requirements. We restrict proposal distributions, or 'moves', to those that produce an algorithm that is invariant to scale, and approximately invariant to affine transformations of the state space. Hence scaling of proposals, and effectively also coordinate transformations, that might be used to increase efficiency of the sampler, are not needed since the t-walk's operation is identical on any scaled version of the target distribution. Four moves are given that result in an effective sampling algorithm.

We use the simple device of updating only a random subset of coordinates at each step to allow application of the t-walk to high-dimensional problems. In a series of test problems across dimensions we find that the t-walk is only a small factor less efficient than optimally tuned algorithms, but significantly outperforms general random-walk M-H samplers that are not tuned for specific problems. Further, the t-walk remains effective for target distributions for which no optimal affine transformation exists such as those where correlation structure is very different in differing regions of state space.

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. The t-walk is available for R, Python, MatLab and C++ at http://www.cimat.mx/~jac/twalk/.

Keywords: MCMC, Bayesian inference, simulation, t-walk

1 Introduction

We develop a new MCMC sampling algorithm that contains neither adaptivity nor tuning parameters yet that can sample from target distributions with arbitrary scale and correlation structure. We dub this algorithm the "t-walk" (for "traverse" or "thoughtful" walk, as opposed to a random-walk MCMC). Unlike adaptive algorithms that attempt to *learn* the scale and correlation structure of complex target distributions (Andrieu and Thoms 2008), the t-walk is designed to be *invariant* to this structure. Because the t-walk is constructed as a Metropolis-Hastings algorithm on the product space it is

[*] Centro de Investigación en Matemáticas (CIMAT), Guanajuato, Mexico, http://www.cimat.mx/~jac

[†] Department of Physics, University of Otago, New Zealand, mailto:fox@physics.otago.ac.nz

DOI:10.1214/10-BA603

provably convergent under the usual mild conditions.

Application areas are in sampling continuous densities with unknown scale and correlation structure. In applications where a change of variables could 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 adaptive (e.g. in the sense of Bai et al. 2008, in our case the t-walk is a homogeneous Markov chain), 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 adjust 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 applying 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 (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; Laine and Tamminen 2008; Watzenig and Fox 2009, just to mention some recent examples). Therefore, developing a generic and easy to use MCMC method like the t-walk will help non-statisticians who are looking to use Bayesian inferential methods in their research fields.

Because the t-walk is useful as a black-box sampling algorithm it 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 significantly reduces the 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 1 to 200. Good results are obtained, always simulating from the objective function successfully for all examples that range across different scales and dimensions. We also report on several other examples where the t-walk has been used successfully, with dimensions of up to 576.

Even though the t-walk is not adaptive, it is useful to compare it to existing adaptive algorithms, since the purpose is similar. 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 et al. 1998; Brockwell and Kadane 2005; Haario et al. 2001; Andrieu and Thoms 2008), and the adaptive direction samplers (ADS) that maintain several points in the state space (see, for example, Gilks et al. 1994; Gilks and Roberts 1996; Eidsvik and Tjelmeland 2004) including the "evolutionary Monte Carlo" that combines ADS with moves from genetic algorithms to speed up a Metropolis coupled MC (Liang and Wong 2001). The t-walk may be viewed as an adaptive sampler of the second type, since it maintains a set of two points in the state space and moves them with some structure. However, several important differences should be noted. Firstly, computational effort in the ADS scales poorly in problem dimension since the number of points maintained scales super-linearly (Gilks, Roberts, and George 1994). Consequently, even if an ADS achieves a constant integrated autocorrelation time (IAT, see Geyer 1992) per dimension (which is optimal, see, e.g. Roberts and Rosenthal 2001) the computation time and storage scales super-quadratically. In contrast the t-walk maintains two points, independent of dimension, and achieves constant IAT per dimension in standard examples (see Section 4). Hence the computational effort required to achieve a given variance reduction in estimates scales linearly in problem dimension. Secondly, by focusing on *invariance* of the sampler, rather than adaptivity, we have proposed a sampler that operates efficiently across a wide range of target distributions, without further problem-specific work. The invariance property means that demonstrating the effectiveness of the t-walk for a suite of canonical stylized test problems (as we do in Sections 3 and 4) demonstrates effectiveness for all problems that differ by a coordinate transformation. We are unaware of successful applications of adaptive MCMC schemes to a comprehensive suite of objective functions, and note that ADS has been shown to be inefficient in many cases (see, for example, Gilks, Roberts, and George 1994). Further, the ADS requires 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. In Section 4 we compare the t-walk with optimally-tuned M-H MCMC algorithms in a suit of standard examples. Finally a discussion of the paper is given 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 $\mathbb{R}^n$), 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

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

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. Note, however, that we are not considering two independent parallel chains in each $\mathcal{X}$; instead the whole process lies in $\mathcal{X} \times \mathcal{X}$. We will randomly choose from four different proposals, to be defined below, each characterized by a particular function $h(\cdot, \cdot)$. We will first choose an option in (1) and second create the proposal (y, y') simulating from the corresponding h function.

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. Note that restriction to proposal 1 implies that only a single evaluation of the target density is required, in either case.

It is straightforward to show that if the random variable h is invariant to affine transformations, i.e. $h(\phi x, \phi x') = \phi h(x, x')$ for any affine transformation ϕ, then so are the proposals 1 and the resulting MCMC sampler. We formalize this in Theorem 1. Design of an invariant sampling algorithm then rests on the question of whether it is possible to find one or more random variables h that give an effective sampling algorithm. We have found that the four choices for h, given below, give adequate mixing across a wide range of target distributions of moderate dimension.

For high dimensional problems we select a random subset of coordinates to be updated at each step, as follows. In each of the four moves below we simulate a Bernoullian sequence of independent indicator variables $I_j \sim \text{Be}(p), j = 1, 2, \ldots, n$. If $I_j = 0$ coordinate j is not updated. The probability p of updating a given coordinate is chosen so that $np = \min(n, n_1)$ and we set $n_I = \sum_{j=1}^{n} I_j$. That is, the expected number of parameters to be moved at each iteration is n_1 for $n \geq n_1$, while for $n \leq n_1$ all coordinates are used in each move (we use $n_1 = 4$, see Section 2.5).

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

$$h_{\mathrm{w}}(x,x')_j = \begin{cases} x_j + \left(x_j - x'_j\right)\alpha_j & I_j = 1 \\ x_j & I_j = 0, \end{cases}$$

for $j = 1, 2, \ldots, n$, where $\alpha_j \in \mathbb{R}$ are i.i.d. r.v. with density $\psi_{\mathrm{w}}(\cdot)$. Considering the second case in (1), $g(y|x,x') = \prod_{I_j=1} g_j(y_j|x_j,x'_j)$, where $g_j(y_j|x_j,x'_j) = \psi_{\mathrm{w}}\left(\frac{y_j-x_j}{x_j-x'_j}\right)/|x_j-x'_j|$. It is straightforward to verify that if $\alpha = \frac{y_j-x_j}{x_j-x'_j}$, then $\frac{g_j(x_j|y_j,x'_j)}{g_j(y_j|x_j,x'_j)} = \frac{\psi_{\mathrm{w}}\left(\frac{x_j-y_j}{y_j-x'_j}\right)}{\psi_{\mathrm{w}}\left(\frac{y_j-x_j}{x_j-x'_j}\right)}\left|\frac{x_j-x'_j}{y_j-x'_j}\right| = \frac{\psi_{\mathrm{w}}\left(\frac{-\alpha}{1+\alpha}\right)}{\psi_{\mathrm{w}}(\alpha)}\left|\frac{1}{1+\alpha}\right|$.

If $\alpha > -1$ then $\left|\frac{1}{1+\alpha}\right| = \frac{1}{1+\alpha}$ and this proposal is symmetric ($\frac{g_j(x_j|y_j,x'_j)}{g_j(y_j|x_j,x'_j)} = 1$) when

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

We achieve this by setting

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

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

$$\alpha = \frac{a_{\mathrm{w}}}{1+a_{\mathrm{w}}}\left(-1+2u+a_{\mathrm{w}}u^2\right)$$

where $u \sim U(0,1)$. 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 (we set $a_{\mathrm{w}} = 1.5$, see Section 2.5).

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 and scale 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

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

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

The case $\beta \equiv 0$ is similar to 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 required in the leap-frog move. As noted by MacKay (2003, p. 394), this move has similarities to the 'snooker' move used in ADS. 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 it is not possible to make both the proposal and the acceptance ratio independent of the dimension of state space, n, except for the case $\beta \equiv 1$. However, by setting $\psi_{\mathrm{t}}(1/\beta) = \psi_{\mathrm{t}}(\beta)$, for all $\beta > 0$, the ratio of proposals is simplified to $\beta^{n_I - 2}$ (see below). By direct calculation it is easy to see that a density of this kind may be obtained by using a density $\nu(\cdot)$ on $\mathbb{R}^+$ and defining $\psi_{\mathrm{t}}(\beta) = C\{\nu(\beta^{-1} - 1)I_{(0,1]}(\beta) + \nu(\beta - 1)I_{(1,\infty)}(\beta)\}$, for a normalizing constant C (assuming $\int_0^1 \nu(\beta^{-1} - 1)d\beta < \infty$). A simple and convenient result is obtained with $\nu(y) = (a_{\mathrm{t}} - 1)(y + 1)^{-a_{\mathrm{t}}}$, for any $a_{\mathrm{t}} > 1$, in which case

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

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

$$x\beta = \begin{cases} u^{\frac{1}{a_{\mathrm{t}} + 1}}, & \text{with prob. } \dfrac{a_{\mathrm{t}} - 1}{2a_{\mathrm{t}}} \\ u^{\frac{1}{1 - a_{\mathrm{t}}}}, & \text{with prob. } \dfrac{a_{\mathrm{t}} + 1}{2a_{\mathrm{t}}}, \end{cases} \tag{2}$$

where $u \sim U(0, 1)$. We want steps to be taken around the length of $||x - x'||$, thus a good idea is that $P(\beta \leq 2) > 0.9$. We set $a_{\mathrm{t}} = 6$ giving $P(\beta < 2) \approx 0.98$, see Section 2.5. A plot of $\psi_{\mathrm{t}}(\beta)$ with $a_{\mathrm{t}} = 6$ is presented in Figure 1.

Following the above transformation, it is clear that

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

A note of caution is prudent here, regarding calculation of the acceptance probability for this move. Since the range of h_{t} is a subspace of $\mathcal{X}$ it is most convenient to use the reversible jump MCMC formalism (Green and Mira 2001) for evaluating the acceptance ratio. The corresponding Jacobian determinant equals $\beta^{n_I - 2}$, and since $\psi_{\mathrm{t}}(1/\beta) =$

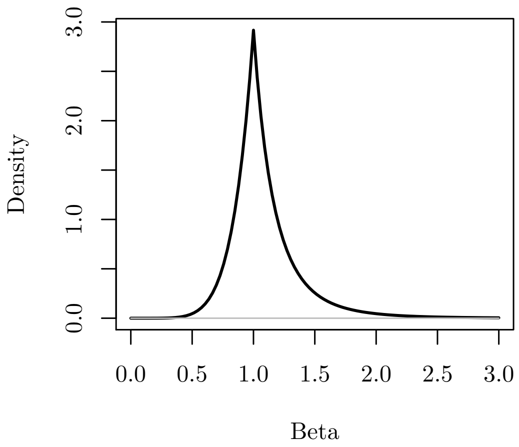


Figure 1: $\psi_{\mathrm{t}}(\beta)$ with $a_{\mathrm{t}} = 6$ giving $P(\beta < 2) \approx 0.98$.

$\psi_{\mathrm{t}}(\beta)$ the acceptance ratio is $\frac{\pi(y')}{\pi(x')}\beta^{n_I - 2}$ or $\frac{\pi(y)}{\pi(x)}\beta^{n_I - 2}$, for the first and second cases in (1), 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 when applied to a large number of uncorrelated variables. In examples with correlation as high as $1 - 10^{-7}$, or higher, (typical of examples from inverse problems) the traverse move is effective in mixing along the principal axis of the distribution, but is very slow in mixing in directions perpendicular to this 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 somehow cured if we try to avoid x and x' collapsing to each other. Note that the walk and traverse moves simply do not work if $x = x'$. We employ two further moves that make bold proposals, precisely for avoiding $x \approx x'$, 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 function

$$h_{\mathrm{h}}(x,x')_j = \begin{cases} x_j + \dfrac{\sigma(x,x')}{3} z_j & I_j = 1 \\ x_j & I_j = 0, \end{cases}$$

with $z_j \sim N(0,1)$, where $\sigma(x,x') = \max_{I_j=1} |x_j - x'_j|$. For this proposal

$$g_{\mathrm{h}}(y \mid x,x') = \frac{(2\pi)^{-n_I/2} 3^{n_I}}{\sigma(x,x')^{n_I}} \exp\left\{ -\frac{9}{2\sigma(x,x')^2} \sum_{I_j=1} (y_j - x_j)^2 \right\} \prod_{I_j=0} \delta_{x_j}(y_j).$$

Note that this move is centred at x.

Finally we consider the blow move defined by

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

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

$$g_{\mathrm{b}}(y \mid x,x') = \frac{(2\pi)^{-n_I/2}}{\sigma(x,x')^{n_I}} \exp\left\{ -\frac{1}{2\sigma(x,x')^2} \sum_{I_j=1} (y_j - x'_j)^2 \right\} \prod_{I_j=0} \delta_{x_j}(y_j).$$

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

2.4 Convergence

Let $K_{\mathrm{m}}(\cdot,\cdot)$ be the corresponding M-H transition kernel for proposal q_{m}, where $\mathrm{m} \in \{\mathrm{w},\mathrm{t},\mathrm{h},\mathrm{b}\}$. Strong aperiodicity is ensured by the positive probability of rejection in the M-H scheme. (For example, when $n \neq 2$, in the traverse move there is always a positive probability that $n_I \neq 2$ and $\beta < \left(\frac{\pi(x)}{\pi(y)}\right)^{\frac{1}{n_I - 2}}$, resulting in an acceptance probability of less than 1. If $n = 2$, and if $\pi(\cdot)$ is locally constant so that the walk and traverse moves produce acceptance probabilities equal to 1, in such case there is a positive probability that either the blow or hop moves will produce an acceptance probability less than 1.) It may be seen, using the properties of the M-H method, that each K_{m} satisfies detailed balance with $f(x,x')$. We form the transition kernel

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

where $\sum_{\mathrm{m}} w_{\mathrm{m}} = 1$, 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).

2.5 Parameter settings

In our implementation of the t-walk we set the move probabilities $w_{\mathrm{w}}, w_{\mathrm{t}}, w_{\mathrm{h}}, w_{\mathrm{b}} = 0.4918, 0.4918, 0.0082, 0.0082$ (move ratios of 60:60:1:1). These values were chosen to give the minimum integrated autocorrelation time (IAT, see Section 4), 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.

The other three important parameter settings required are n_1, a_w and a_t, the expected number of parameters to be moved and the Walk and Traverse moves proposal parameters, respectively. Based on many examples, those shown here and many more, we established reasonable test ranges for each parameter, namely $n_1 \in [2, 20], a_w \in [0.3, 2]$ and $a_t \in [2, 10]$. We performed an optimization of these parameters by calculating the IAT's for many examples across dimensions from $n = 2$ to $n = 150$ (we utilized the examples presented in Section 4 running the t-walk in 500 runs of a Latin Hypercube Design within the parameter ranges). The rounded optimal results are $n_1 = 4, a_w = 1.5$ and $a_t = 6$.

Indeed, we cannot consider *all* possible objective functions. However, we have seen the t-walk to succeed in sampling in many examples already (see the Discussion Section 5). Our intention is that the above parameter settings are left as default and the user does not need to alter them to achieve a reasonable performance.

2.6 Properties

We consider that the most important property of the t-walk is that it is invariant to affine transformations. All moves were developed with that in mind. Given a transformation of the space $\mathcal{X}$, $\phi(z) = az + b$, where $a \in \mathbb{R}, a \neq 0$ and $b \in \mathbb{R}^n$, 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 $(x_0, x'_0) \in \phi(\mathcal{X}) \times \phi(\mathcal{X})$, or by applying the t-walk kernel to π with starting values $(\phi^{-1}(x_0), \phi^{-1}(x'_0))$, and then transforming the resulting chain with ϕ. The following Theorem states that the t-walk is invariant to changes in scale and reference point.

Theorem 1. *Let $V \in \mathcal{X} \times \mathcal{X}$ and $A \subset \mathcal{X} \times \mathcal{X}$ (a measurable set). The t-walk transition kernel using objective $\lambda(z) = |a^{-n}|\pi(\phi^{-1}(z))$, K_λ, and the t-walk kernel using objective π, K_π, have the invariance property*

$$K_\lambda(\phi(V), \phi(A)) = K_\pi(V, A). \tag{3}$$

Proof: Let $W' = (y, y') \in \phi(\mathcal{X}) \times \phi(\mathcal{X})$. Elementary calculations show that $|a^{-n_I}|q_{\mathrm{m}}($

$\phi^{-1}(W) \mid V) = q_{\mathrm{m}}(W \mid \phi(V))$ for $\mathrm{m} = \mathrm{w}, \mathrm{b}, \mathrm{h}$ and $|a^{-1}| q_{\mathrm{t}}(\phi^{-1}(W) \mid V) = q_{\mathrm{t}}(W \mid \phi(V))$ (in this case the transformation is univariate, that is, along the line $x_j - x'_j$, and therefore the necessary Jacobian is simply $|a^{-1}|$; in what follows we will take in this case $n'_I = 1$ and $n'_I = n_I$ for the other moves). It is clear that the M-H acceptance probability considering proposal m for λ is (let $f'(W') = \lambda(y)\lambda(y')$)

$$\rho^{\lambda}_{\mathrm{m}}(\phi(V), W') = \min\left\{1, \frac{f'(W')}{f'(\phi(V))} \frac{q_{\mathrm{m}}(\phi(V)|W')}{q_{\mathrm{m}}(W'|\phi(V))}\right\}.$$

Equivalently for π

$$\rho^{\pi}_{\mathrm{m}}(V, \phi^{-1}(W')) = \min\left\{1, \frac{f(\phi^{-1}(W'))}{f(V)} \frac{q_{\mathrm{m}}(V|\phi^{-1}(W'))}{q_{\mathrm{m}}(\phi^{-1}(W')|V)}\right\}.$$

Since $f(\phi^{-1}(W')) = |a^{2n}| f'(W')$ and $f(V) = |a^{2n}| f'(\phi(V))$, and the above relation on the q_m's, we see that $\rho^{\lambda}_{m}(\phi(V), W') = \rho^{\pi}_{m}(V, \phi^{-1}(W'))$. It is clear then that the probabilities of accepting a jump have the property

$$\begin{aligned} r^{\lambda}_{\mathrm{m}}(\phi(V), \phi(A)) = \textstyle\int_{\phi(A)} \rho^{\lambda}_{\mathrm{m}}(\phi(V), W') q_{\mathrm{m}}(W' \mid \phi(V)) dW' = \\ \textstyle\int_{\phi(A)} \rho^{\pi}_{\mathrm{m}}(V, \phi^{-1}(W')) |a^{-n'_I}| q_{\mathrm{m}}(\phi^{-1}(W') \mid V) dW' = \\ \textstyle\int_{A} \rho^{\pi}_{\mathrm{m}}(V, W) |a^{-n'_I}| q_{\mathrm{m}}(W \mid V) |a^{n'_I}| dW = r^{\pi}_{\mathrm{m}}(V, A). \end{aligned}$$

Since the t-walk kernel is a mixture of the individual kernels for $\mathrm{m} = \mathrm{w}, \mathrm{t}, \mathrm{b}, \mathrm{h}$, and together with the fact that the probability of not jumping in either case is the same $(1 - r^{\lambda}_{\mathrm{m}}(\phi(V), \phi(\mathcal{X} \times \mathcal{X})) = 1 - r^{\pi}_{\mathrm{m}}(V, \mathcal{X} \times \mathcal{X}))$, establishes the result. ■

It is immediate that the above result also holds for n steps into the t-walk and therefore $K^n_{\pi}(V, A) = K^n_{\lambda}(\phi(V), \phi(A))$ and since also $f(A) = f'(\phi(A))$ we have that

$$||K^n_{\pi}(V, \cdot) - f(\cdot)||_{TV} = ||K^n_{\lambda}(\phi(V), \cdot) - f'(\cdot)||_{TV}.$$

The above establishes the following 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. More importantly, for many applications, the t-walk is effectively invariant for more general classes of transformations. When all components are selected, i.e. $n_I = n$, Theorem 1 remains valid for the t-walk limited to the traverse and walk moves with the more general change in scale using $a = \mathrm{diag}(a_j)$, a diagonal matrix, with $a_j \in \mathbb{R} \setminus \{0\}$. While the hop and blow moves are not invariant under this transformation, the operation of those moves is effectively unchanged. Further, when a represents a rotation, the traverse move is invariant while the remaining moves are effectively invariant. In particular, the *density* over the walk move is invariant. When a is allowed to be any nonsingular matrix, i.e. ϕ is a general invertible affine transformation, the traverse move remains invariant. The simple form of coordinate selection that we employ for high dimensional problems means that these further invariances do not actually hold in general, however the operation of the sampler does not suffer and it appears that the beneficial consequences of the invariances are preserved.

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 $||x^{(t)} - x'^{(t)}||$ 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

$$\pi(x) = C \exp\left\{ -\tau \left(\sum_{i=1}^{2} (x_i - m_{1,i})^2 \right) \left(\sum_{i=1}^{2} (x_i - m_{2,i})^2 \right) \right\}, \tag{4}$$

for some $(m_{1,1}, m_{1,2})$ and $(m_{2,1}, m_{2,2})$ that approximately locate two modes (on $\mathbb{R}^2$), and scale parameter τ (C 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 istribution.

We present two further 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 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

In this section we demonstrate the usefulness of the t-walk in a high dimension example, that arises as the posterior distribution over a semiparametric age model for radiocarbon calibration in paleoecology. In that work, cores taken from peat bogs are sectioned and then subject to radiocarbon dating, from which we wish to build a model for age as a function of depth. See Blaauw and Christen (2005) for details.

For a single core we have a series of radiocarbon determinations with standard errors $y_j \pm \sigma_j$ taken at depth d_j for $j = 1, 2, \ldots, m$. Hence $\{y_j, \sigma_j, d_j\}_{j=1,2,\ldots,m}$ are the (known) measured data. We follow Blaauw and Christen (2005), who take the measured values of σ_j and d_j to be the true values, and use a piecewise linear model for the relation

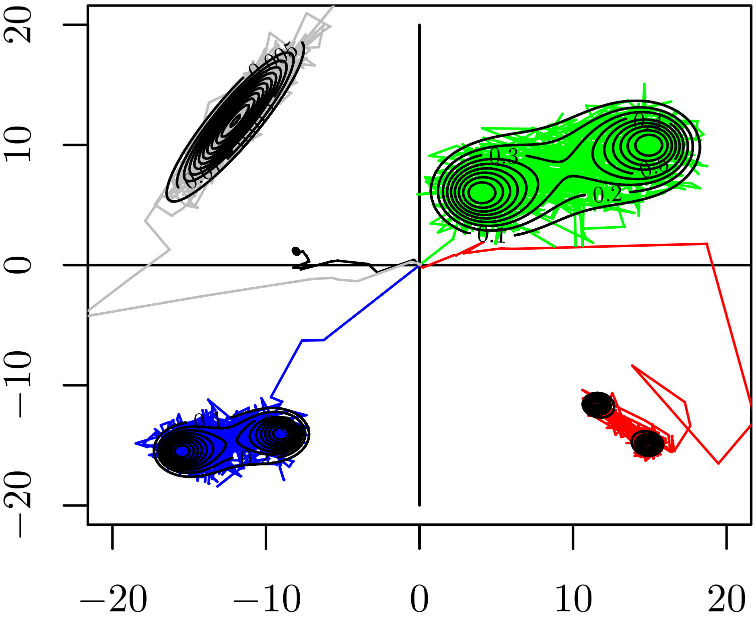
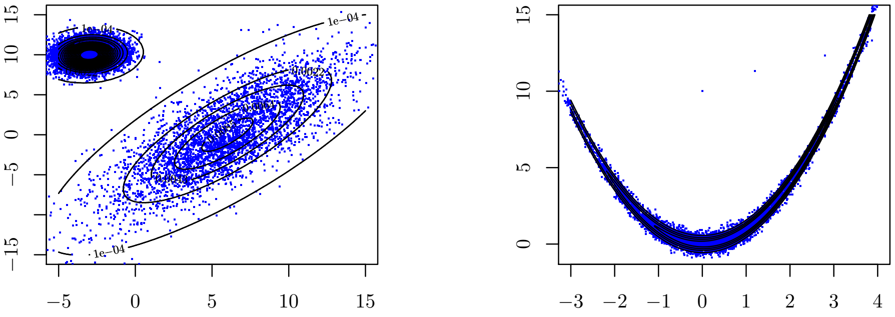


Figure 2: Sample paths for one component in the t-walk. Upper left quadrant: Bivariate normal distribution with correlation 0.95. Other quadrants, counterclockwise from the lower left: distribution in (4) with $\tau = 0.01, 0.1, 0.001, 1000$ (for $\tau = 1000$ the scale is such that the distribution 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.

between the (unknown) true age of peat and depth, d,

$$G(d; \mathbf{x}) = x_1 + \sum_{j=2}^{i} x_j \Delta c + x_{i+1}(d - c_i) \qquad \text{if } c_i \leq d < c_{i+1}$$

where the uniformly spaced depths $c_i = c_1 + (i - 1)\Delta c$, $i = 1, 2, \ldots, n - 1$ are fixed, given (known) c_1, Δc, and n, while x_1 (the age-depth model abscissa) and $x_2, \ldots, x_{n-1}$ (the age-depth model accumulation rates) are parameters to be inferred. The usual normal likelihood model is assumed, $y_j \mid \sigma_j, d_j, \mathbf{x} \sim N(\mu(G(d_j; \mathbf{x})), \sigma_j)$, where $\mu(\cdot)$ is



(a) (b)

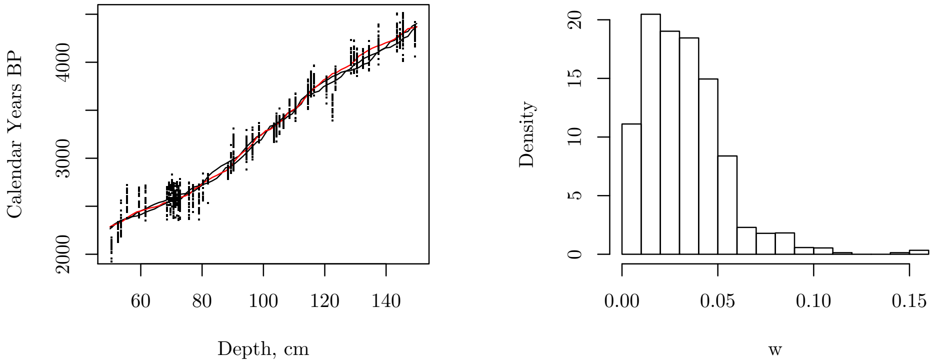
Figure 3: Sample points for one component in the t-walk (a) mixture of bivariate normals, low mode with weight 0.7, $\mu_1 = 6, \sigma_1 = 4, \mu_2 = 0, \sigma_2 = 5, \rho = 0.8$, high mode with weight 0.3, $\mu_1 = -3, \sigma_1 = 1, \mu_2 = 10, \sigma_2 = 1, \rho = 0.1$. We took 100,000 iterations with an acceptance rate of around 45%. (b) "Rosenbrock" (Rosenbrock 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 3.1.

the radiocarbon calibration curve, see Blaauw and Christen (2005) for details.

Additionally, a (prior) model is proposed for the (peat accumulation) rates $x_j = wx_{j+1} + (1 - w)z_j$, where $w \sim \text{Beta}(\alpha_w, \beta_w)$ and $z_j \sim \text{Gamma}(\alpha_z, \beta_z)$. Here α_w, β_w, α_z and β_z are known (representing the prior information available on accumulation rates, see Blaauw and Christen 2005). Therefore, the unknown parameters to be sampled by the t-walk are $x_1, \ldots, x_{n-1}$, and w that we denote x_n thus the unknown set of paremeters may be written as the n-vector $\mathbf{x}$.

A simple program (in C++) is used to calculate $-\log f(\mathbf{x}|\{y_j, \sigma_j, d_j\}_{j=1,2,\ldots,m})$. Restricted support of parameters (eg. $w \in [0, 1]$) is enabled in most implementations of the t-walk by providing a 'Supp' function that returns True or False according to whether or not the input $\mathbf{x}$ is in the support of the objective function (in the MatLab implementation the function providing the log target density returns the value `-Inf` for arguments outside its support). The other inputs required are the starting values for $\mathbf{x}$ and $\mathbf{x}'$.

In this example we set $x_n = w = 0.4$, $x'_n = w' = 0.1$, draw $x_{n-1}, x'_{n-1} \sim \text{Gamma}(\alpha_z, \beta_z)$, and take $x_j = wx_{j+1} + (1 - w)z_j$ and $x'_j = w'x'_{j+1} + (1 - w')z'_j$ for



(a) (b)

Figure 4: (a) MAP estimator (red) and two sample age-depth models for core EngXV. For each of the $m = 57$ radiocarbon determinations a sample of 25 calendar ages were simulated and plotted (small dots; calendar ages measured in 'years Before Present' (BP), where 'present' is AD 1950). (b) Histogram for the marginal posterior distribution of w.

$j = n-2, \ldots, 2$ drawing $z_j, z'_j \sim \text{Gamma}(\alpha_z, \beta_z)$. Finally, we draw $x_1, x'_1 \sim N(y_1, \sigma_1)$. This provides initial, random, values for $\mathbf{x}$ and $\mathbf{x}'$ that are in the support of the objective function.

We used the data set called "EngXV" with $m = 57$ determinations (Blaauw, van Geel, Mauquoy, and van der Plicht 2004), using $n = 71$ (70 parameters for the age-depth model plus w) and ran 300,000 iterations of the t-walk (taking 1 minute on a MacBook Pro lap top). Two sample age-depth models and the MAP estimator are presented in Figure 4(a), and a histogram approximating the marginal distribution of w is presented in Figure 4(b).

4 Comparisons with optimally-tuned M-H MCMC

Roberts and Rosenthal (2001) present a review of optimal scaling for a random-walk Metropolis Hastings (M-H) algorithm applied to some simple models. For minimum integrated autocorrelation time (IAT), the proposal window must be tuned to give 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, with the values $C_j = 1$ (model 1), $C_1 = 2$ and $C_j = 1; j = 2, 3, \ldots, n$ (model 2), and $C_1 = 1$ and $C_j \sim \text{Exp}(1); j = 2, 3, \ldots, n$ (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 or equally efficient than *any* generic method, including the t-walk. However, fine tuning a M-H MCMC constitutes significant effort in 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. Avoiding this difficulty is the idea behind adaptive methods (see Andrieu and Thoms, 2008, for a recent review), and the t-walk.

Roberts and Rosenthal (2001) argue that IAT divided by the dimension of the parameter space is a good measure for comparing convergence rates, or efficiency, among MCMC samplers across space dimension. In all cases we calculate the IAT for x_1, as done in Roberts and Rosenthal (2001). We fine tune a random walk M-H algorithm for model 1 for $n = 10$ to an acceptance rate of 0.234. In this case, the variance for the normal distribution in the random walk should be $2.38/\sqrt{10}$ as explained by Gelman, Roberts, and Gilks (1996). We use that same sampler in the four models above at dimensions $n = 2, 5, 10, 25, 50, 75, 100, 125, 150$, multiplying the corresponding jumping variance by $\sqrt{10}/\sqrt{n}$ to obtain the expected correct rescaling (Gelman et al. 1996). The results are presented as dashed lines in Figure 5. Moreover, the random walk Metropolis Hastings algorithm was also run with a slightly suboptimal variance of 2.0 instead of $0.75 = 2.38/\sqrt{10}$, with rescaling as before for models 1, 2 and 3. Also, no rescaling was applied, using just the optimal scaling for $n = 10$, with results shown by dotted lines in Figure 5. We also ran the t-walk for all models with the results shown in Figure 5 as solid lines.

For the random walk M-H with the correct scaling, IAT$/n$ remains low, as expected by the theoretical results of Gelman et al. (1996), at least for models 1, 2 and 3, and once it is tuned at the optimal scaling. However, exporting this sampler to model 0 does not work as well, particularly for low dimensional cases. Indeed, without the correct scaling the sampler fails radically, as seen in the dotted lines in Figure 5; this failing is a very well known fact by any dedicated practitioner of MCMC. Note for example, for n=100, the correct scaling is $0.238 = 2.38/\sqrt{100}$ whereas we are using 0.75 ($= 2.38/\sqrt{10}$) and the resulting IAT/n is already around ten times the optimal. Even more interesting is to see that a slight (within the same order of magnitude) change in the variance of the random walk M-H algorithm leads to a clear under performance, that gets worse with increasing dimension. This occurs even when using the correct rescaling $\sqrt{10}/\sqrt{n}$, and leads to an IAT$/n$ well above 20 in most cases (results shown only for model 3, dotted and dashed line in Figure 5, where it fails radically).

Note that in the case of the t-walk, for *all* models the IAT increases more or less linearly with n and thus IAT$/n$ remains in most cases below 15. (It is the case that IAT$/n$ remains bounded by 30 for all the examples presented in the previous sections, including the high dimension $n = 71$ in Section 3.2.) Only in the very ideal, and artificial, case of knowing the *exact* optimal scaling does the random walk M-H outperform the t-walk for models 1 and 2, and even in those cases the t-walk remains very competitive. It is worth mentioning that the *same* t-walk algorithm is used in all examples (in this Section and elsewhere in the paper), requiring no tuning parameters, no rescaling of any sort, and only needing as input the (log of the) objective function and two initial points in the sample space.

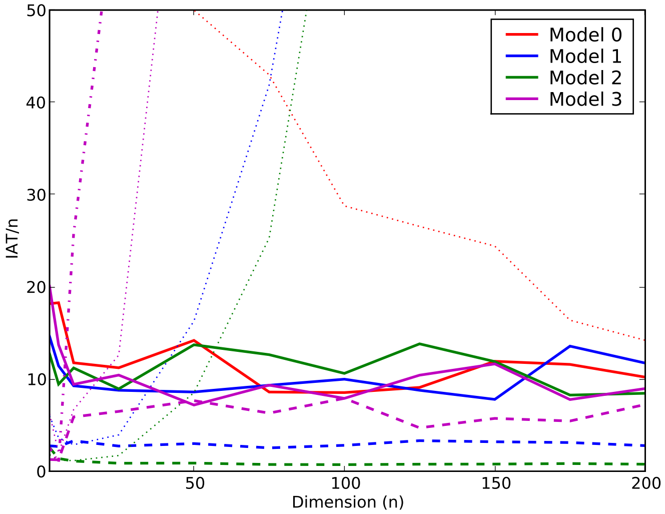


Figure 5: Integrated Autocorrelation Times dived by the dimension, over various dimensions for the t-walk, solid lines. A random walk Metropolis Hastings was used, with the correct rescaling, dashed lines and with a fix scaling (optimal at $n = 10$), dotted lines. The optimal scaling performs slightly better than the t-walk for models 1 and 2 and is already comparable in the case of model 3. A random walk M-H with a slightly wrong rescaling was also run, obtaining IAT/n well above 20 in most cases. This rescaling completely fails for model 3, see dotted and dashed line. The models are taken from Roberts and Rosenthal (2001), see text.

5 Discussion

The t-walk has unique performing characteristics, sampling efficiently from target distributions with 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, such as those employed by ter Braak (2006). 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. nor some prior knowledge of the number of modes, tails etc. As presented in the numerical examples, we have evidence that the t-walk will perform satisfactorily with common densities (posterior distributions in common Bayesian statistical analyses). 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.

Besides the examples we have already mentioned, we and other colleagues have implemented the t-walk in a series problems. Indeed, we now treat the t-walk as our sampler of first choice and have been pleasantly surprised to find that it always provides useful output, and avoids several of the difficulties seen in standard MCMC. These examples include a reliability example ($n = 2$), reservoir effects in radiocarbon dating problems ($n = 3 - 6$), a bacterial horizontal gene transfer model ($n > 50$, Zenil-López 2008), electrical capacitance tomography using polygonal representations ($n = 32-128$, where not having to calculate Jacobians for subspace moves was very liberating), pixel-based impedance imaging ($n = 576$), and in fitting analytic models to groundwater pump tests ($n = 2 - 15$). We have also combined the t-walk kernel with Gibbs kernels, when the full conditionals for some blocks of parameters have known distributions and the rest have full conditionals that are difficult to sample from. Such examples arose in fitting spatial Gaussian ($n = 2$) processes and other in an Econometric time series model ($n = 9$, Lence, Hart, and Hayes 2009).

The t-walk is available as an R (R Development Core Team 2008) and Python `http://www.python.org/` packages (and is also available in MatLab and C++) at `http://www.cimat.mx/~jac/twalk/` and will soon be included in the PyMC package at `http://code.google.com/p/pymc/`.

References

Andrieu, C. and Thoms, J. (2008). "A tutorial on adaptive MCMC." *Statistics and Computing*, 18(4): 343–373. 263, 264

Bai, Y., Roberts, G., and Rosenthal, J. (2008). "On the Containment Condition for Adaptive Markov Chain Monte Carlo Algorithms." Technical report, No. 0806, Department of Statistics, University of Toronto. 264

Bavencoff, F., Vanpeperstraete, J., and Le Cadre, J. (2006). "Constrained bearings-only target motion analysis via Markov chain Monte Carlo methods." *IEEE Transactions on Aerospace and Electrical Systems*, 42(4): 1240–1263. 264

Blaauw, M. and Christen, J. (2005). "Radiocarbon peat chronologies and environmental change." *Applied Statistics*, 54(4): 805–816. 273, 275

Blaauw, M., van Geel, B., Mauquoy, D., and van der Plicht, J. (2004). "Radiocarbon wiggle-match dating of peat deposits: advantages and limitations." *Journal of Quaternary Science*, 19: 177–181. 276

Brockwell, A. and Kadane, J. (2005). "Identification of Regeneration Times in MCMC Simulation, With Application to Adaptive Schemes." *Journal of Computational and Graphical Statistics*, 14(2): 436–458. 264

Eidsvik, J. and Tjelmeland, H. (2004). "On directional Metropolis-Hastings algorithms." *Statistics and Computing*, 16(1): 93–106. 265

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. 264

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: 203–216. 264

Gelman, A., Roberts, G., and Gilks, W. (1996). "Efficient Metropolis jumping rules." In Bernardo J.M, D. A., Berger J.O. and A.M.F., S. (eds.), *Bayesian Statistics V*, 599–608. Oxford: Oxford University Press. 277

Geyer, C. (1992). "Practical Markov Chain Monte Carlo." *Statistical Science*, 7(4): 473–511. 265

Gilks, W. and Roberts, G. (1996). "Strategies for improving MCMC." In Gilks, W., Richardson, S., and Spiegelhalter, D. (eds.), *Markov Chain Monte Carlo in Practice*, 89–114. London: Chapman and Hall. 265

Gilks, W., Roberts, G., and George, E. (1994). "Adaptive direction sampling." *The Statistician*, 43: 179–189. 265

Gilks, W., Roberts, G., and Sahu, S. (1998). "Adaptive Markov Chain Monte Carlo Through Regeneration." *Journal of the American Statistical Association*, 93: 1045–1054. 264

Green, P. and Mira, A. (2001). "Delayed Rejection in Reversible Jump Metropolis–Hastings." *Biometrika*, 88: 1035–1053. 268

Haario, H., Saksman, E., and Tamminen, J. (2001). "An adaptive Metropolis algorithm." *Bernoulli*, 7: 223–242. 264

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. 264

Laine, M. and Tamminen, J. (2008). "Aerosol model selection and uncertainty modelling by adaptive MCMC technique." *Atmospheric Chemistry and Physics*, 8(24): 7697–7707. 264

Lence, S. C., Hart, C., and Hayes, D. (2009). "An Econometric Analysis of the Structure of Commodity Futures Prices (poster paper)." In *Agricultural & Applied Economics Association & American Council on Consumer Interests 2009 Joint Annual Meeting, 26-28 July*. Milwaukee, Wisconsin, US. 279

Liang, F. and Wong, W. (2001). "Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models." *Journal of the American Statistical Association*, 96(454): 653–666. 265

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambrige, UK: Cambridge University Press. 268

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL http://www.R-project.org/ 279

Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. New York: Springer. 271

Roberts, G. O. and Rosenthal, J. S. (2001). "Optimal scaling for various Metropolis-Hastings algorithms." *Statistical Science*, 16(4): 351–367. 265, 276, 277, 278

Rosenbrock, H. (1960). "An automatic method for finding the greatest or least value of a function." *The Computer Journal*, 3: 175–184. 275

Symonds, J. P. R., Reavell, K., Olfert, J., Campbell, B., and Swift, S. (2007). "Diesel soot mass calculation in real-time with a differential mobility spectrometer." *Journal of Aerosol Science*, 38(1): 52–68. 264

ter Braak, C. (2006). "A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces." *Statistics and Computing*, 16: 239–249. 278

Warnes, G. (2000). "The Normal Kernel Coupler: An adaptive Markov Chain Monte Carlo method for efficiently sampling from multi-modal distributions." Ph.D. thesis, University of Washington. 264

Watzenig, D. and Fox, C. (2009). "A review of statistical modelling and inference for electrical capacitance tomography." *Measurement Science and Technology*, 20(5): (052002) 1–22. 264

Zenil-López, R. (2008). "A Hidden Markov Model of Horizontal Gene Transfer in Bacteria." Master's thesis, Centro de Investigación en Matemáticas, Guanajuato, Mexico. 279

Acknowledgments

J Andrés Christen was partially founded by grant SEMARNAT-2004-C01-0007 (Mexico). We thank Sergio Lence, Rosana Zenil, Patricia Bautista, José Miguel Ponciano and Sergio Pérez E. for enthusiastic feedback on implementing and using the t-walk.