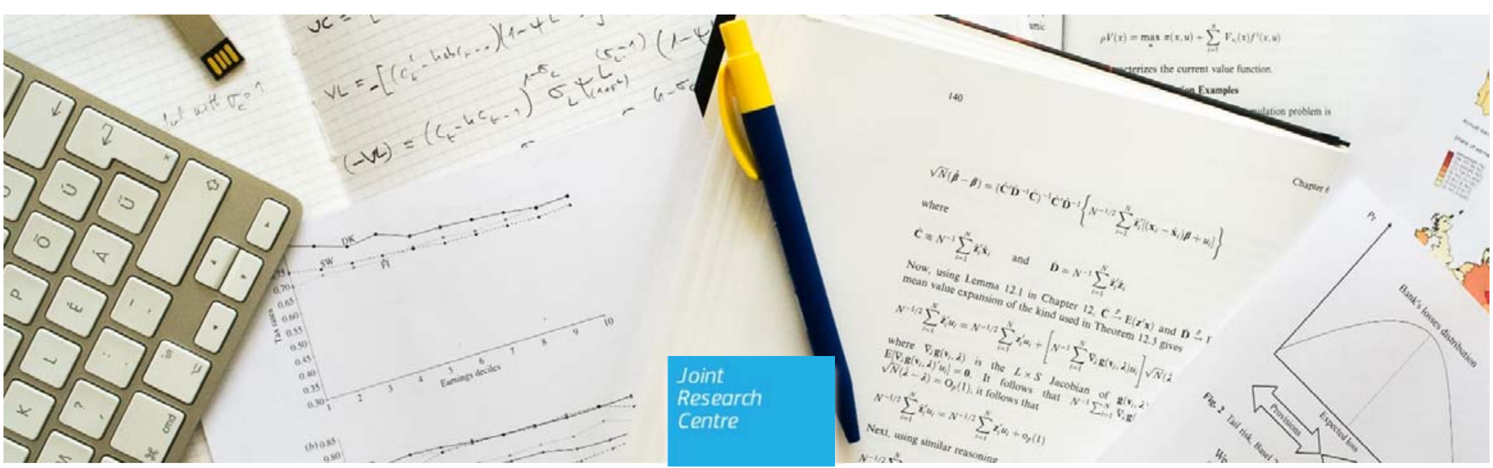
European Commission

JRC TECHNICAL REPORTS

The slice sampler and centrally symmetric distributions

Christophe Planas and Alessandro Rossi

JRC Working Papers in Economics and Finance, 2018/11



Joint Research Centre

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information
Name: Alessandro Rossi
Address: Via E. Fermi, Ispra (VA), Italy
email: alessandro.rossi1@ec.europa.eu
Tel.: +39 0332 789724

JRC Science Hub
https://ec.europa.eu/jrc

JRC113316

PDF ISBN 978-92-79-93405-6 ISSN 2467-2203 doi:10.2760/10835

Luxembourg: Publications Office of the European Union, 2018

How to cite this report: Christophe Planas and Alessandro Rossi, *The slice sampler and centrally symmetric distributions*, JRC Working Papers in Economics and Finance, 2018/11, ISBN 978-92-79-93405-6, doi:10.2760/10835, JRC113316.

The slice sampler and centrally symmetric distributions *

Christophe Planas and Alessandro Rossi

European Commission, Joint Research Centre (JRC), Ispra, Italy

June 2018

Abstract

We point out that the simple slice sampler generates chains that are correlation-free when the target distribution is centrally symmetric. This property explains several results in the literature about the relative performance of the simple and product slice samplers. We exploit it to improve two algorithms often used to circumvent the slice inversion problem, namely stepping out and multivariate sampling with hyperrectangles. In the general asymmetric case, we argue that symmetrizing the target distribution before simulating greatly enhances the efficiency of the simple slice sampler. To achieve symmetry we focus on the Box-Cox transformation with parameters chosen to minimize a measure of skewness. This strategy is illustrated with several sampling problems.

JEL CODE: C11, C15.

KEYWORDS: Box-Cox transformation, Markov Chain Monte Carlo, Multivariate sampling.

*The views expressed in this paper are those of the authors and should not be attributed to the European Commission. Merrill W. Liechty and Jingjing Lu are gratefully acknowledged for making available the code implementing the (d+1) and d(d+1)/2 product slice samplers for truncated normal distributions.

E-Mails: *christophe.planas@ec.europa.eu*, *alessandro.rossi1@ec.europa.eu*.

1 Introduction

The slice sampler is a data augmentation technique to generate Markov chains whose mixing properties generally ensure a rapid convergence to the target distribution. Roberts and Rosenthal (1999) for instance show that for well-behaved distributions, convergence takes place in less than six hundred iterations. This result assumes algebraic knowledge of the region below the slice, a situation which is often difficult to achieve in practice. To circumvent the slice inversion problem, Neal (2003a) has proposed several algorithms known as stepping out in univariate cases and hyperrectangle sampling in multivariate contexts. The few tunings required makes these algorithms attractive for routine use. Since its popularization by Neal, the slice sampler has been applied in a wide variety of disciplines like for instance acoustics (Jasa and Xiang, 2009), climate research (Tarasov et al., 2012), economics (Kline and Tamer, 2016), finance (Li, 2011), genetics (Dunson and Xing, 2009), machine learning (Bishop, 2006), and spatial modelling (Agarwal and Gelfand, 2005).

We argue that the class of centrally symmetric distributions offers new insights into both theoretical and algorithmic properties of the simple slice sampler. The concept of central symmetry is discussed for instance in Serfling (2006); it is sometimes also referred to as 'equal symmetry' (Hollander, 1968) or 'radial symmetry' (Nelsen, 1993). Cases of applied interest include the whole family of elliptically contoured distributions. Besides its relevance in many inferential problems, the class of centrally symmetric distributions deserves a special attention because of a property which has been overlooked in the literature: when applied to target distributions that are centrally symmetric, the simple slice sampler generates chains that are uncorrelated. This optimal feature is a consequence of Lemma 3.2 in Liu, Wong, Kong (LWK, 1994) and it follows from the interleaving Markov property of samplers built by data augmentation. Other data augmentation schemes such as the product slice sampler share this optimal property under the more demanding condition that all the components that factorize the target distribution are symmetric around the central point. This observation explains several results in the literature about the relative performance of the simple and product slice samplers (Neal, 2003b).

When the slice interval cannot be inverted algebraically, we show that stepping out preserves the zero-correlations property when the target distribution is unimodal. Some correlations instead arise in the case of univariate multimodal distributions as well as with hyperrectangle sampling. For such cases we show that important efficiency gains can be obtained from a simple amendment to Neal algorithms that exploits central symmetry.

Next, the optimality of the simple slice sampler for centrally symmetric distributions offers the opportunity to symmetrize the target distribution before applying the slice sampler. To achieve symmetry we focus on the Box-Cox power transformation (Box and Cox, 1964), with parameters chosen to minimize the Mardia measure of skewness (Mardia, 1970). Neal algorithms amended to take symmetry into account can then be implemented on the new scale, the draws being reset to the original scale by inverse transformation. This strategy has the advantage of simplicity, its implementation is almost costless, and the many experiments we report confirm its effectiveness in enhancing efficiency. It provides a valid alternative to overrelaxation and reflective methods for improving the efficiency of the slice sampler.

Section 2 discusses the slice sampler in the central symmetric case. Neal algorithms are reviewed and a simple amendment that takes symmetry into account is presented. The advantage of exploiting symmetry when approximating the slice region is illustrated with several examples taken from the literature. Section 3 considers the general case of asymmetric distributions. Several examples show that the strategy of symmetrizing before implementing the slice sampler with the amended Neal algorithms is greatly beneficial. Section 4 concludes.

2 Efficiency of the simple slice sampler for symmetric distributions

To simplify we first consider the univariate case. Given a density $\pi(x)$ having support $\mathsf{X} \in \Re$, suppose that we wish to calculate the expectation $E_\pi h(x) = \int_{\mathsf{X}} h(x)\pi(x)dx$ of a real-valued and π-integrable function $h(x)$. Suppose further that this integral being intractable we resort to the Markov Chain Monte Carlo simulation method known as the slice sampler. The simple slice sampler generates a homogeneous Markov chain $\{x_n\} \equiv \{x_1, x_2, \cdots\}$ by introducing an auxiliary variable u such that (u, x) are jointly uniformly distributed as in:

$$\pi(u, x) = \mathcal{U}\{0 < u \leq \pi(x)\}, \tag{2.1}$$

from which u and x are iteratively drawn from their conditional distributions $\pi(u|x)$ and $\pi(x|u)$ (see for instance Chapter 8 in Robert and Casella, 2004). The resulting chain $\{x_n\}$ is π-irreducible, aperiodic, and Harris recurrent (Mira and Tierney, 2002), so it converges in distribution to $\pi(x)$ and the average $\overline{h}_n(x) = \frac{1}{n}\sum_i h(x_i)$ converges almost surely to $E_\pi h(x)$. By Lemma 3.1 in LWK the Markov chain $\{x_n\}$ is reversible so provided that $h^2(\cdot)$ is π-integrable, the average $\overline{h}_n(x)$ admits a central limit theorem with associated variance (see e.g. Jones, 2004):

$$V(\overline{h}_n(x)) = \frac{1}{n}V_\pi(h(x))(1 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}\rho_{j-i})$$

where ρ_{j-i} denote the autocorrelation between $h(x_i)$ and $h(x_j)$. For n large enough, the expression above is well approximated by

$$V(\overline{h}_n(x)) \simeq \frac{1}{n}V_\pi(h(x))(1 + 2\sum_{i=1}^{n}\rho_i) = \frac{1}{n}V_\pi(h(x))\, IF_h \qquad (2.2)$$

where the inefficiency factor IF_h – also known as autocorrelation times – summarizes the efficiency of the Markov transition kernel to generate samples of $h(x)$ with $x \sim \pi(x)$. Chains with smaller inefficiency factor produce sample averages $\overline{h}_n(x)$ which estimate $E_\pi h(x)$ more accurately. LWK analyse the correlation structure of the Gibbs sampler within a two-component augmentation scheme like the simple slice sampler. Taking benefit from the interleaving Markov property of the sequences $\{x_n\}$ and $\{u_n\}$ which implies that $E(x_i, x_{i+1}|u_i) = E(x_i|u_i)E(x_{i+1}|u_i)$, LWK show that the first autocovariance of the chain $\{x_n\}$ verifies $Cov(x_i, x_{i+1}) = V_\pi[E_\pi(x|u)]$. This result suggests that asymmetry of the target distribution is a source of inefficiency. A simple example is given by the exponential distribution: the smaller u the larger $E_\pi(x|u)$, giving rise to a non-null first-autocorrelation. Hence we expect skewness of the target distribution to impair the performance of the simple slice sampler. To illustrate we consider an example discussed in Roberts and Rosenthal (2002) and Robert and Casella (2004) as a case where the simple slice sampler appears to be poor.

Example 1: Let us consider the distribution $\pi(x) = e^{-x^{1/d}}/\,d!$ defined over the support $\Re^+$ and indexed by the positive integer d. Closed-form expressions for the slice interval, the skewness, and the first-autocorrelation of the Markov transition kernel implied by the simple slice sampler are given in Appendix A. The skewness values displayed in Table 1 for $d = 1, 2, 5, 10$, and 20 reveal that this distribution becomes increasingly asymmetric as d increases. As can be seen in Table 1, the first-autocorrelation ρ_1 increases together with the skewness of the target distribution up to an upper bound equal to 3/4. Although not displayed in Table 1, all chain autocorrelations increase together with the skewness of the distribution. This example stresses that the simple slice sampler can be inefficient when applied to unimodal distributions that are strongly asymmetric.

Table 1. Skewness of $e^{-x^{1/d}}/\, d!$ and first autocorrelation implied by the slice sampler

	$d = 1$	$d = 2$	$d = 5$	$d = 10$	$d = 20$
Skewness	2	4.30	24.72	4.70×10^2	1.91×10^5
ρ_1	0.50	0.64	0.73	0.74	0.75

Conversely, LWK's result $Cov(x_i, x_{i+1}) = V_\pi[E_\pi(x|u)]$ implies that consecutive draws will be uncorrelated when $E_\pi(x|u)$ is constant. This occurs when the target density is centrally symmetric, i.e. when $x - E_\pi(x)$ is distributed like $E_\pi(x) - x$, since in this case the centre of the slice interval $S(u)$ falls exactly on the mean of x, i.e. $E_\pi(x|u) = E_\pi(x)\ \forall u \in (0, \max \pi(x))$. This feature equally holds for multivariate distributions. We give this special property of the simple slice sampler in Proposition 1.

Proposition 1 *Suppose the target density* $\pi(\mathbf{x})$, $\mathbf{x} \in \Re^d$, *is centrally symmetric, i.e.* $\pi(\mathbf{x} - E_\pi(\mathbf{x})) = \pi(E_\pi(\mathbf{x}) - \mathbf{x})$. *Then the simple slice sampler generates draws which are uncorrelated at all lags.*

Proposition 1 is a consequence of Lemma 3.2 in LWK. It concerns all lags since by Theorem 3.1 in LWK the autocorrelations of a marginal chain obtained by data augmentation are nonnegative and decreasing, so a null autocorrelation at the first lag implies zero autocorrelations at all lags. We illustrate Proposition 1 revisiting the previous example.

Example 1 (cont'd): In Roberts and Rosenthal (2002), the distribution $\pi(x) = e^{-x^{1/d}}/\,d!$, $x \in \Re^+$, arises by applying the transformation $x = \|\mathbf{z}\|^d$, $\mathbf{z} \in \Re^d$, to the distribution $\pi(\mathbf{z}) = e^{-\|\mathbf{z}\|}$. By property of the Euclidian norm, $\pi(\mathbf{z})$ is centrally symmetric about zero, i.e. $\pi(\mathbf{z}) = \pi(-\mathbf{z})$. The slice region $S(u) = \{\mathbf{z} : u \leq \pi(\mathbf{z})\}$ corresponds to the sphere generated by $\{\mathbf{z} : \|\mathbf{z}\| \leq -\log u\}$ from which the vector $\mathbf{z}$ can be uniformly drawn given u. This scheme yields a first correlation equal to zero for all elements of vector $\mathbf{z}$ whatever the dimension d. All inefficiency factors are then equal to one.

To benefit from this optimal property of the simple slice sampler, the slice region $S(u)$ must be known in closed form. Example 1 gives one such case but most often the slice interval cannot be inverted analytically. Neal (2003a) proposes several algorithms to circumvent the inversion problem. Given a current point x_0, a draw u from $\mathcal{U}\{0 < u \leq \pi(x_0)\}$, stepping out proceeds as follows:

Neal stepping out algorithm for univariate distributions:

(i). Random positioning: build a random interval (L, R) of length W around x_0 by setting $L = x_0 - \gamma W$, $\gamma \sim \mathcal{U}(0, 1)$, and $R = L + W$;

(ii). Expanding: check whether the bounds L and R lie outside of the slice interval $S(u)$ and expand the bounds otherwise, i.e. set $L = L - W$ until $\pi(L) < u$ and $R = R + W$ until $\pi(R) < u$;

(iii). Shrinking: draw a candidate $x^c \sim \mathcal{U}(L, R)$. If $x^c \notin S(u)$, shrink the interval (L, R) by setting either $L = x^c$ if $x^c < x_0$ or $R = x^c$ if $x^c \geq x_0$. Repeat until $x^c \in S(u)$ and then set $x_1 = x^c$.

(iv). Set $x_0 = x_1$ and sample a new u from $\mathcal{U}\{0 < u \leq \pi(x_0)\}$.

This algorithm generates draws that are uncorrelated when the target distribution is symmetric and unimodal. Indeed in such cases the interval (L, R) where to draw the candidate x^c always contains $S(u)$, guaranteeing that the equality $E_\pi(x|u) = E_\pi(x)$ holds for all u. This optimal property of stepping out is shared with doubling, another algorithm proposed by Neal for delineating the slice interval. It does not however generalize to distributions which are multimodal. Consider for instance the case of a symmetric bimodal distribution with zero-mean defined over a connected support. Suppose also that the slice interval is made up of two disjoint intervals, say $S(u) = (-S_2, -S_1) \cup (S_1, S_2)$, and that the current state x_0 belongs to (S_1, S_2). When $W < 2S_1$, the expanding step will end up with (L, R) such that the set of admissible draws $(L, R) \cap S(u) = (S_1, S_2)$ does not contain the mean of the distribution. In this case the equality $E_\pi(x|u) = E_\pi(x)$ does not hold giving rise to chain correlations. When $W > 2S_1$ it is the shrinking step that generates autocorrelations even if $(L, R) \cap S(u) = S(u)$. For instance assume that the expanding step has ended up with $(L, R) = (-S_2, S_2)$. The next state x_1 will belong to $(-S_2, -S_1)$ only if the first candidate draw x^c falls into $(-S_2, -S_1)$, an event which occurs with probability $(S_2 - S_1)/(2S_2) < 1/2$. This outcome is not affected by the more general situation where $L < -S_2$ and $R > S_2$. Hence in multimodal cases stepping out favours permanence in the sub-interval which contains the current state: a positive correlation follows.

When applied to multimodal distributions, Neal stepping out algorithm thus breaks the interleaving property. This feature can be exploited to revert the correlation sign by positioning the initial slice interval around the mirror image of x_0 with respect to the distribution centre, i.e. around $2E_\pi(x) - x_0$ instead of around x_0. Such a switch is possible because as long as it is acceptable, the point around which the slice interval is built has no relevance for the convergence of the algorithm. We label antithetic stepping out this simple amendment to Neal algorithm given below:

Antithetic stepping out for univariate symmetric distributions:

(i)-(iii). Like in Neal stepping out for univariate distributions.

(iv). Set $x_0 = 2E_\pi(x) - x_1$ and sample a new u from $\mathcal{U}\{0 < u \leq \pi(x_0)\}$.

This algorithm requires knowledge of the distribution centre $E_\pi(x)$. We discuss the performance of the two samplers in Example 2 where two symmetric normal mixtures are considered, one unimodal to illustrate the optimality of Neal stepping out for such distributions, and a bimodal one to enlighten the advantage of exploiting symmetry. The comparison is made in terms of the relative inefficiency factor defined as:

$$RIF = \frac{IF_A}{IF_B} \times \frac{\text{Time}_A}{\text{Time}_B} \tag{2.3}$$

where A and B refer to the two competing algorithms with inefficiency factor IF_A and IF_B and computational time Time_A and Time_B. The RIF measures the factor by which sampler A's run-time must be increased to achieve the same precision as sampler B; values larger than one point to a greater efficiency of scheme B.

Remark 1: In general, the performance of stepping out depends on the length W chosen for the initial slice interval. Wide intervals imply more mixing but the number of rejections can increase noticeably. On the opposite tiny intervals imply stickiness. All results reported in the sequel are obtained with W set equal to three times the standard deviation of x; this value is larger than the one standard deviation suggested by Neal, but in our experience it often gives a good trade-off between the mixing and the run-time.

Example 2: Let us consider the following two normal mixtures: the unimodal kurtotic distribution $\pi(x) = 2/3\ \phi_1(x) + 1/3\ \phi_{1/10}(x)$ and the separate bimodal distribution $\pi(x) = 1/2\ \phi_{1/2}(x-3/2) + 1/2\ \phi_{1/2}(x+3/2)$, where $\phi_\sigma(x)$ denotes the normal density with zero-mean and standard deviation σ. These two distributions can be visualized in Figure 1 of Marron and Wand (1992). For the two versions of stepping out, Table 2 reports the inefficiency factor, the average number of evaluations of the target density used to simulate one draw, and the relative inefficiency factor.

Table 2. Slice samplers efficiency in Example 2

	Kurtotic	Bimodal
Plain stepping out		
IF	1.00	2.96
#eval	6.42	6.19
Antithetic stepping out		
IF	1.01	0.40
#eval	6.42	6.19
RIF	0.99	7.40

Notes: kurtotic distribution: $\pi(x) = 2/3\ \phi_1(x) + 1/3\ \phi_{1/10}(x)$; bimodal distribution: $\pi(x) = .5\ \phi_{1/2}(x - 3/2) + .5\ \phi_{1/2}(x + 3/2)$; IF refers to the inefficiency factor defined by setting $h(x) = x$ in (2.2); it is calculated using one hundred thousand of simulations after discarding the first one thousand with autocorrelations weighted by a Parzen window of length one hundred; #eval gives the average number of evaluations of $\pi(x)$ used to simulate one draw; the relative inefficiency factor RIF is defined in (2.3) with antithetic stepping out taken as algorithm B.

For the kurtotic distribution, both algorithms yield an inefficiency factor equal to one. This confirms the efficiency of plain stepping out to sample from symmetric unimodal distributions, and also that the antithetic version does not harm in such cases. As expected some correlations arise when applying plain stepping out to the symmetric bimodal distribution: the inefficiency factor rises to 2.96. The antithetic algorithm does not show this feature: the negative correlations induced by the switching mechanism in step (iv) lead to an inefficiency factor equal to 0.40. Overall, the relative inefficiency factor is about 7 in favour of the antithetic version. As the two algorithms perform an equal number of evaluations of the target distribution, the run-time is equivalent and all the improvement pertains to the mixing. Although it stems from a trivial amendment, antithetic stepping out appears to be quite powerful for sampling from symmetric distributions with known centre. We could check that other choices of W such that one and ten times the scale of x does not alter this outcome.

Remark 2: The bimodal distribution of Example 2 is defined over a connected support. Mira and Roberts (2003) point out that, when the support is disconnected and the initial interval length W is too small, stepping out fails to produce an irreducible Markov chain. This failure also occurs with the antithetic version. In such situations increasing W seems to be the only viable route.

Proposition 1 refers specifically to the simple slice sampler. Variants involving further auxiliary variables are however available: in particular, the product slice sampler factorizes the target density like in $\pi(x) \propto f_0(x) \prod_{i=1}^m f_i(x)$, and introduces m auxiliary variables $u_1, \cdots, u_m$ whose joint distribution together with x is proportional to $f_0(x)\mathbf{1}\{x \in \bigcap_{i=1}^m S_i(u_i)\}$, $S_i(u_i)$ denoting the interval $S_i(u_i) = \{x : 0 < u_i \leq f_i(x)\}$ (see Edwards and Sokal, 1988). The aim is to simplify the slice inversion problem by carefully choosing the functions $f_i(x)$. The product slice sampler generates uncorrelated draws when $E_\pi(x|u_1, \cdots, u_m) = E_\pi(x)$ for all $u_1, \cdots, u_m$. Since $E_\pi(x|u_1, \cdots, u_m) = E_{f_0}(x|x \in \bigcap_{i=1}^m S_i(u_i))$, zero correlations will be obtained when the centre of the slice region $\bigcap_{i=1}^m S_i(u_i)$ falls on $E_{f_0}(x)$ whatever $u_1, \cdots, u_m$. This is rather peculiar since it requires that $f_0(x)$ and the factors $f_i(x)$ are all symmetric around $E_\pi(x)$. Applying the product slice sampler to a target distribution which is symmetric but whose components are either asymmetric or with different centres is thus sub-optimal: in such cases the simple slice sampler performs better. This explains the large increase in autocorrelation times, from 1.1 with the simple slice sampler to up to 187 with the product version, that Neal (2003b) reports when simulating from the Bernoulli logistic regression model discussed in Example 5 of Damien, Wakefield, and Walker (1999).

In the multivariate context, delineating the slice region $S(u) = \{\mathbf{x} : u \leq \pi(\mathbf{x})\}$ is more involving. Neal (2003a) proposes the following algorithm:

Neal multivariate slice sampling with hyperrectangles:

(i). Random positioning: given $\mathbf{x}_0 \in \Re^d$, build a random hyperrectangle $\prod_{j=1}^{d}(L_j, R_j)$ by setting for each coordinate $L_j = x_{0j} - \gamma_j W_j$, $\gamma_j \sim \mathcal{U}(0,1)$, and $R_j = L_j + W_j$.

(ii). Shrinking: draw a candidate $\mathbf{x}^c$ by simulating $x_j^c \sim \mathcal{U}(L_j, R_j)$ for $j = 1, 2, \cdots, d$. If $\mathbf{x}^c \notin S(u)$, shrink the hyperrectangle by setting either $L_j = x_j^c$ if $x_j^c < x_{0j}$ or $R_j = x_j^c$ if $x_j^c \geq x_{0j}$. Repeat until $\mathbf{x}^c \in S(u)$ and then set $\mathbf{x}_1 = \mathbf{x}^c$.

(iii). Set $\mathbf{x}_0 = \mathbf{x}_1$ and sample a new u from $\mathcal{U}\{0 < u \leq \pi(\mathbf{x}_0)\}$.

This algorithm keeps the number of evaluations of the target density reasonably low. Contrary to the univariate case however it generates correlations even for symmetric distributions that are unimodal. Indeed each time the candidate $\mathbf{x}^c$ is rejected as falling outside of the slice region, the hyperrectangle is updated by removing layers of $\prod_{j=1}^{d}(L_j, R_j)$, making the new hyperrectangle more concentrated around $\mathbf{x}_0$. This mechanically rises the chain correlations. Removing the shrinking step does not overcome this problem because the number of rejections can become prohibitive. Like in the univariate case, this feature can be exploited to enhance efficiency as detailed below:

Antithetic multivariate slice sampling with hyperrectangles:

(i)-(ii). Like in Neal multivariate slice sampling with hyperrectangles.

(iii-bis). Set $\mathbf{x}_0 = 2E_\pi(\mathbf{x}) - \mathbf{x}_1$ and sample a new u from $\mathcal{U}\{0 < u \leq \pi(\mathbf{x}_0)\}$.

As in the univariate case, approximating the slice interval around $\mathbf{x}_1$ or around its image $2E_\pi(\mathbf{x}) - \mathbf{x}_1$ does not harm convergence. We illustrate the performance of the antithetic multivariate slice sampler with another example taken from the literature.

Example 3: Liechty and Lu (LL, 2010) adapt the product slice sampler to the d-dimensional normal distribution truncated to a region A, say $N(\mathbf{0}, \Sigma)\mathbf{1}_A$, using the factorization:

$$\pi(\mathbf{x}) \propto \prod_{i=1}^{d} \exp\{-\frac{1}{2}a_{ii}\ x_i^2\} \prod_{i=1}^{d} \prod_{j=i+1}^{d} \exp\{-\frac{1}{2}a_{ij}\ x_i x_j\}\mathbf{1}_A \tag{2.4}$$

where a_{ij}, $i, j = 1, \cdots, d$, are the elements of the precision matrix Σ^{-1}. LL propose two product slice samplers which differ in the number of auxiliary variables: the $d+1$-method introduces one variable for each term in the first product plus one further variable for the second product in (2.4); the $d(d+1)/2$-method introduces instead one auxiliary variable for each factor. For the quadrivariate normal distribution with support $\Re^4$, $(-.05, .05)^4$, $(-.0325, .0325)^4$, as considered in the LL's simulation study, Table 3 reports the performance of Neal multivariate algorithm and its antithetic version against the two product slice samplers. The performance is evaluated in terms of IF (2.2), the average number of evaluations of the target density by iteration, plus the RIF (2.3) with the antithetic algorithm taken as benchmark. For the product samplers the average number of evaluations of the target density is measured by the average number of rejections by iteration plus one. We have used the code distributed as supplemental material of LL (2010); it can be checked that the numbers in Table 3 are in broad agreement with those displayed in Table 1 of LL.

For the truncations considered, the two LL samplers give inefficiency factors which are close to one. As each term in the factorization (2.4) is symmetric about the origin, this optimal performance of the two product samplers results from central symmetry. With Neal multivariate algorithm, the inefficiency factors range from 7 to 11: as expected the shrinking step implies some stickiness. In contrast the antithetic algorithm produces inefficiency factors below one-fifth. The $d(d+1)/2$ product sampler necessitates the smallest number of evaluations, about one and a half by iteration; the $(d+1)$ product sampler is more demanding in this respect. With less than three evaluations by iteration the Neal and the antithetic samplers are moderately more costly than the $d(d+1)/2$ algorithm. In terms of RIF the antithetic sampler dominates substantially, the smallest RIF taking value about seventeen.

Table 3. Slice samplers efficiency in Example 3

	Truncation regions		
	$\Re^4$	$(-.05, .05)^4$	$(-.0325, .0325)^4$
Plain multivariate			
max IF	11.07	8.42	7.44
#eval	2.93	2.82	2.75
RIF	88.14	57.66	48.65
$(d+1)$ auxiliary variables			
max IF	1.03	1.02	1.04
#eval	61.49	10.60	3.96
RIF	162.77	39.47	20.91
$d(d+1)/2$ auxiliary variables			
max IF	1.01	1.01	1.05
#eval	1.56	1.41	1.25
RIF	17.32	19.97	21.59
Antithetic multivariate			
max IF	0.13	0.14	0.15
#eval	2.92	2.83	2.76
RIF	1.00	1.00	1.00

Notes: max IF refers to the maximum inefficiency factor of the four-dimensional chain, calculated on one hundred thousand of simulations using a Parzen window of length one hundred; the relative inefficiency factor RIF is calculated as in (2.3) with the antithetic multivariate sampler taken as algorithm B.

Turning to the case of asymmetric distributions, we investigate whether some efficiency gains could be achieved by symmetrizing.

3 Efficiency gain from symmetrizing

We first consider the univariate case. Suppose an invertible transformation indexed by parameter λ, say $y = g(x;\lambda)$, makes the target distribution centrally symmetric on the new scale. Then the slice sampler could be applied to the symmetrized density $\pi(y)$ to simulate a correlation-free chain $\{y_n\}$, the draws being then reset to the original scale using the inverse transformation $x = g^{-1}(y;\lambda)$. To achieve symmetry we focus on the class of power transformations introduced by Box and Cox (1964):

$$y = g(x;\lambda) \quad = \quad \begin{cases} [(x+\lambda_2)^{\lambda_1} - 1]/\lambda_1 & \text{if } \lambda_1 \neq 0 \\ \ln(x+\lambda_2) & \text{if } \lambda_1 = 0 \end{cases} \tag{3.1}$$

where $\lambda = (\lambda_1, \lambda_2) \in \Re \times \Re^+$. While the shift λ_2 can be set equal to any value that guarantees $x + \lambda_2 > 0$, the power parameter λ_1 must instead be estimated. An estimator is built by considering a measure of symmetry say $R(y_1, \cdots, y_n)$ such that $E_\pi(R(y_1, \cdots, y_n)) = 0$ when $\pi(y)$ is symmetric; solving $R(g(x_1;\lambda), \cdots, g(x_n;\lambda)) = 0$ yields the estimate $\hat{\lambda}$. Several measures of symmetry are discussed in Taylor (1985). Focusing on the skewness under the assumption that the first three moments exist, the symmetrization of the target distribution proceeds as follows:

(i) Apply the simple slice sampler to $\pi(x)$ to get a preliminary sample $(x_1, \cdots, x_b)$;

(ii) Set $\lambda_2 = \max(0, -\min(x_1, \cdots, x_b)(1+\delta))$ for some $\delta \geq 0$;

(iii) Find λ_1 that minimizes the square of the skewness calculated on the transformed sample $g(x_1;\lambda), \cdots, g(x_b;\lambda)$.

The slice sampler can then be applied to the transformed variable y with density:

$$\pi(y) \quad = \quad \begin{cases} \pi((1+\lambda_1 y)^{1/\lambda_1} - \lambda_2)\,(1+\lambda_1 y)^{\frac{1-\lambda_1}{\lambda_1}} & \lambda_1 \neq 0 \\ \pi(\exp(y) - \lambda_2)\,\exp(y) & \lambda_1 = 0 \end{cases} \tag{3.2}$$

with $\lambda = \hat{\lambda}$. In general the slice interval will not be invertible on the new scale, and depending on whether $\pi(y)$ is unimodal or multimodal the use of either plain or antithetic stepping out is more appropriate. As the transformation guarantees $1 + \lambda_1 y > 0$, the bound L (R) must be greater (lower) or equal to $-1/\lambda_1$ when λ_1 is positive (negative).

Strictly speaking, zero correlations in the chain $\{y_n\}$ does not imply that the transformed draws $\{h(x_n)\} = \{h(g^{-1}(y_n; \lambda))\}$ that are used to estimate $E_\pi(h(x))$ will be uncorrelated as well. Depending on the shape of $\pi(y)$ some correlations may indeed resurge with the inverse transformation. The ideal situation occurs when the joint distribution of u and y is rectangular, that is when u and y are independent. Being more assertive is however difficult; we can only report that in the many experiments we have made including the examples that follow we have faced no case of strong correlations resurging on the original scale.

We illustrate the performance of the symmetrizing strategy revisiting Example 1.

Example 1 (cont'd): In spite of its strong asymmetry - see Table 1, the distribution $\pi(x) = e^{-x^{1/d}}/\, d!$ is successfully symmetrized by the Box-Cox transformation, the skewness on the new scale not exceeding 0.01 in absolute value even when $d = 20$. Ten thousand simulations have been used to estimate the parameter λ. The transformed distribution remains unimodal so plain stepping out can be safely used. For the two implementations Table 4 reports the inefficiency factor, the average number of evaluations of the target density by iteration, and the relative inefficiency factor (2.3). Without symmetrizing the inefficiency factor increases with d, reaching 16.37 when $d = 20$. Symmetrizing almost annihilates the chain correlations: the inefficiency factor stays constant at around 1.2 regardless of d. Symmetrizing also hastens the algorithm: while on the original scale the average number of evaluations necessary to generate one draw increases with d up to 15.3, this number remains stable about 5 when the Box-Cox transformation is used. Overall, the relative inefficiency factor ranges from 2.3 to 37.3 for d increasing from 1 to 20: hence the more asymmetric the distribution, the more advantageous the symmetrization. In this example the gain of symmetrizing is substantial.

Table 4. Slice samplers efficiency in Example 1

	$d=1$	$d=2$	$d=5$	$d=10$	$d=20$
Plain stepping out					
IF	2.97	4.74	7.24	8.92	16.37
#eval	4.37	4.62	6.01	8.97	15.30
Box-Cox & stepping out					
IF	1.16	1.17	1.23	1.22	1.19
#eval	5.01	5.01	4.85	4.79	4.98
RIF	2.35	2.91	6.53	12.39	37.78

Notes: target distribution $\pi(x) = e^{-x^{1/d}}/\,d!$; IF refers to the inefficiency factor defined by setting $h(x) = x$ in (2.2) and calculated using one million of simulations after discarding the first one thousand; the autocorrelations are weighted by a Parzen window of length one hundred; #eval gives the average number of evaluations of the target density used to simulate one draw; the relative inefficiency factor RIF is defined in (2.3) with the Box-Cox transformation plus stepping out taken as algorithm B.

We report further evidence with the following example.

Example 4: Let us consider the following three univariate distributions: the truncated normal $\pi(x) = \phi_1(x)$ for $x \in \Re^+$, the skewed logistic $\pi(x) = 2e^{-x}/\big((1+e^{-x})^2(1+e^{-10x})\big)$ for $x \in \Re$ proposed by Gupta and Kundu (2010), and the beta distribution $\pi(x) = Beta(0.5, 10)$, $x \in (0,1)$. These distributions have skewness between 1 and 2.31 — see Table 5. Ten thousand points have been used to optimize the λ parameter. In the three cases the Box-Cox transformation corrects the asymmetry, the skewness falling below 0.03 in absolute value, while preserving unimodality. Without symmetrizing stepping out yields inefficiency factors between 1.9 and 4.7; upon symmetrizing, the inefficiency factors never exceed 1.2. Also the average number of evaluations of the target distribution slightly decreases after symmetrizing. Altogether, the relative inefficiency factor (2.3) ranges from 2 to 5 in favour of symmetrizing.

Table 5. Slice samplers efficiency in Example 4

	Truncated normal	Skewed logistic	Beta
Skewness	1.00	1.46	2.31
Stepping out			
IF	1.94	2.01	4.74
#eval	5.14	5.36	5.60
Box-Cox & stepping out			
IF	1.06	1.03	1.17
#eval	4.96	5.14	4.69
RIF	1.90	2.04	4.84

Notes: truncated normal distribution: $\pi(x) = \phi_1(x)$, $x \in \Re^+$; skewed logistic: $\pi(x) = 2e^{-x}/\big((1+e^{-x})^2(1+e^{-10x})\big)$, $x \in \Re$; beta: $\pi(x) = Beta(0.5, 10)$, $x \in (0,1)$; IF refers to the inefficiency factor defined by setting $h(x) = x$ in (2.2) and calculated using one million of simulations after discarding the first one thousand; the autocorrelations are weighted by a Parzen window of length one hundred; #eval gives the average number of evaluations of the target density used to simulate one draw; the relative inefficiency factor RIF is defined in (2.3) with Box-Cox transformation plus stepping out taken as algorithm B.

We turn to the multivariate case. To symmetrize a d-dimensional distribution we follow Andrews, Gnanadesikan, and Warner (1971), applying a Box-Cox transformation to each coordinate separately as described below:

(i) Run the slice sampler to get a preliminary sample of length b, say $(\mathbf{x}_1, \cdots, \mathbf{x}_b)$;

(ii) Apply the Box-Cox transformation (3.1) to each element of $\mathbf{x}$ in turn to get $\mathbf{y} = (y_1, \cdots, y_\ell, \cdots, y_d)$ where $y_\ell = g(x_\ell, \lambda_\ell)$, and $\lambda_\ell = (\lambda_{\ell 1}, \lambda_{\ell 2})$ for $\ell = 1, \cdots, d$; this yields the transformed sample $(\mathbf{y}_1, \cdots, \mathbf{y}_b)$;

(iv) Select $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_d)$ which minimizes the Mardia (1970) measure of multivariate skewness calculated on $(\mathbf{y}_1, \cdots, \mathbf{y}_b)$.

The distribution of the transformed variable verifies:

$$\pi(\mathbf{y}) \quad = \quad \pi(g^{-1}(y_1, \lambda_1), \cdots, g^{-1}(y_d, \lambda_d)) \mid \prod_{\ell=1}^{d} \partial g^{-1}(y_\ell, \lambda_\ell)/\partial y_\ell \mid \qquad (3.3)$$

which is straightforward to evaluate at $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}$, similarly to the univariate case (3.2).

The symmetry of $\pi(\mathbf{y})$ can then be fruited using the antithetic multivariate algorithm described in Section 2. Except in particular situations however, central symmetry will not hold exactly. Also $E_\pi(\mathbf{y})$ will generally be unknown so it must be estimated using the preliminary sample $(\mathbf{y}_1, \cdots, \mathbf{y}_b)$, yielding $\overline{\mathbf{y}} = (1/b)\sum_i \mathbf{y}_i$. Hence the antithetic multivariate algorithm requires a further check since given the current state $\mathbf{y}_0$, the mirror point $2\overline{\mathbf{y}} - \mathbf{y}_0$ must fall within the slice region. Otherwise the switch cannot take place and the slice region must be built around $\mathbf{y}_0$ instead of around its image, as detailed below.

Antithetic slice sampling for symmetrized multivariate distributions:

(i)-(ii). Like in Neal multivariate slice sampling with hyperrectangles; this yields the new draw $\mathbf{y}_1$.

(iii). Sample a new u, say u^*, from $\mathcal{U}\{0 < u \leq \pi(\mathbf{y}_1)\}$.

(iv). If $2\overline{\mathbf{y}} - \mathbf{y}_1 \in S(u^*)$ set $\mathbf{y}_0 = 2\overline{\mathbf{y}} - \mathbf{y}_1$; otherwise set $\mathbf{y}_0 = \mathbf{y}_1$.

We illustrate the performance of this strategy with the following example.

Example 5: We modify the four-variate truncated normal distribution discussed in Example 3 by considering an asymmetric truncation region, namely the positive subspace of $\Re^4$. Table 6 details the performance of the LL product slice samplers, of the Neal multivariate algorithm, and of its antithetic variant implemented on the symmetrized distribution. For this distribution the Mardia measure of skewness amounts to 2.2, and upon symmetrizing it falls to 0.04. On the original scale the Neal algorithm yields a maximum inefficiency factor equal to 11.85. Augmenting the number of auxiliary variables with the $d+1$ and $d(d+1)/2$ product samplers further increases the maximum inefficiency factor by one-third. With the antithetic algorithm, all inefficiency factors remain instead close to 1. This outcome is obtained at the moderate cost of 15% further evaluations of the target distribution compared to the Neal algorithm, due to the need to check that the mirror point falls within the slice region. In contrast the product samplers perform many more evaluations of the target distribution due to a large number of rejections: if augmenting the number of auxiliary variables helps inverting the slice region, it also implies a computational cost. Overall, to attain the precision of the antithetic algorithm, the Neal multivariate sampler must run five times longer, and the LL samplers at least fourty times longer. The antithetic sampler thus yields substantial efficiency gains also in this multivariate example.

Table 6. Slice samplers efficiency in Example 5

	Original scale			After symmetrizing
	Neal	$(d+1)$	$d(d+1)/2$	Antithetic
max IF	11.85	14.21	14.03	0.91
# eval	2.89	25.35	15.05	3.37
RIF	4.97	72.38	40.71	1.00

Notes: truncated normal distribution: $\pi(\mathbf{x}) = N(\mathbf{0}, \Sigma)\mathbf{1}_A$, where A refers to the positive subspace of $\Re^4$ and Σ is taken as in LL (2010); IF is the inefficiency factor defined by setting $h(x) = x$ in (2.2) and calculated using one hundred thousand simulations after discarding the first one thousand; the autocorrelations are weighted by a Parzen window of length one hundred; #eval gives the average number of evaluations of the target density used to simulate one draw; the relative inefficiency factor RIF is defined in (2.3) with Box-Cox transformation plus antithetic sampler taken as algorithm B.

Finally we consider an application to Gaussian spatial models.

Example 6: Gaussian spatial modelling

Suppose we wish to describe the surface temperature $t(s_i)$ recorded at some known locations $s_1, \cdots, s_n$ using a Gaussian spatial model with the latitude and longitude as explanatory variables:

$$t(s_i) = \beta_0 + \beta_1 \,\text{latitude}(s_i) + \beta_2 \,\text{longitude}(s_i) + \epsilon(s_i)$$

Following De Oliveira, Kedem, and Short (1997), we assume that the errors $\epsilon(s_i)$ are related through an exponential covariance function:

$$Cov(\epsilon(s_i), \epsilon(s_j)) = \sigma^2 \exp\{-\frac{\|s_i - s_j\|}{\theta}\}$$

where $\theta > 0$ and $\|s_i - s_j\|$ denotes the geodetic distance between two locations (see Banerjee, 2005). Agarwal and Gelfand (2005) argue in favour of the Bayesian approach to analyze such Gaussian spatial models. Berger, De Oliveira, and Sans (2001) however warn against the use of improper priors in this framework. As prior distributions, we consider the independent and proper distributions $\beta_1 \sim Beta(4,4)\mathbf{1}_{(-100,100)}$, $\beta_2 \sim Beta(4,4)\mathbf{1}_{(-100,100)}$, $\sigma^2 \sim Beta(2,10)\mathbf{1}_{(0,200)}$, and $\theta \sim Beta(3,4)\mathbf{1}_{(0,700)}$. The parameter β_0 has been eliminated by demeaning. Data for the January 1995 mean temperature recorded over a grid of 24×24 locations that covers Central America have been downloaded from the NASA Langley Research Center Atmospheric Science web-site.

To simulate from the joint posterior distribution of β_1, β_2, σ^2 and θ, we compare the Neal multivariate algorithm against its antithetic version applied to the symmetrized distribution. Each sampler is implemented in a Fortran code which is run on a 64-bit computer equipped with a CPU of 2.4GHz and 32Gb of RAM. One hundred thousand simulations are stored after a burn-in of one thousand. For the antithetic algorithm, the Box-Cox transformation is estimated using ten thousand extra simulations, and these simulations are also used to estimate the centre of the transformed distribution; all these operations are taken into account when measuring the run-time. For the two competing simulators, Table 7 reports the inefficiency factor for each parameter, the average number of evaluation of the joint posterior distribution per iteration, the relative inefficiency factor as well as the run-time. In this application the run-time deserves a particular monitoring because each evaluation of the joint posterior distribution involves the inversion of a $24^2 \times 24^2$ covariance matrix.

Table 7. Slice samplers efficiency in Example 6

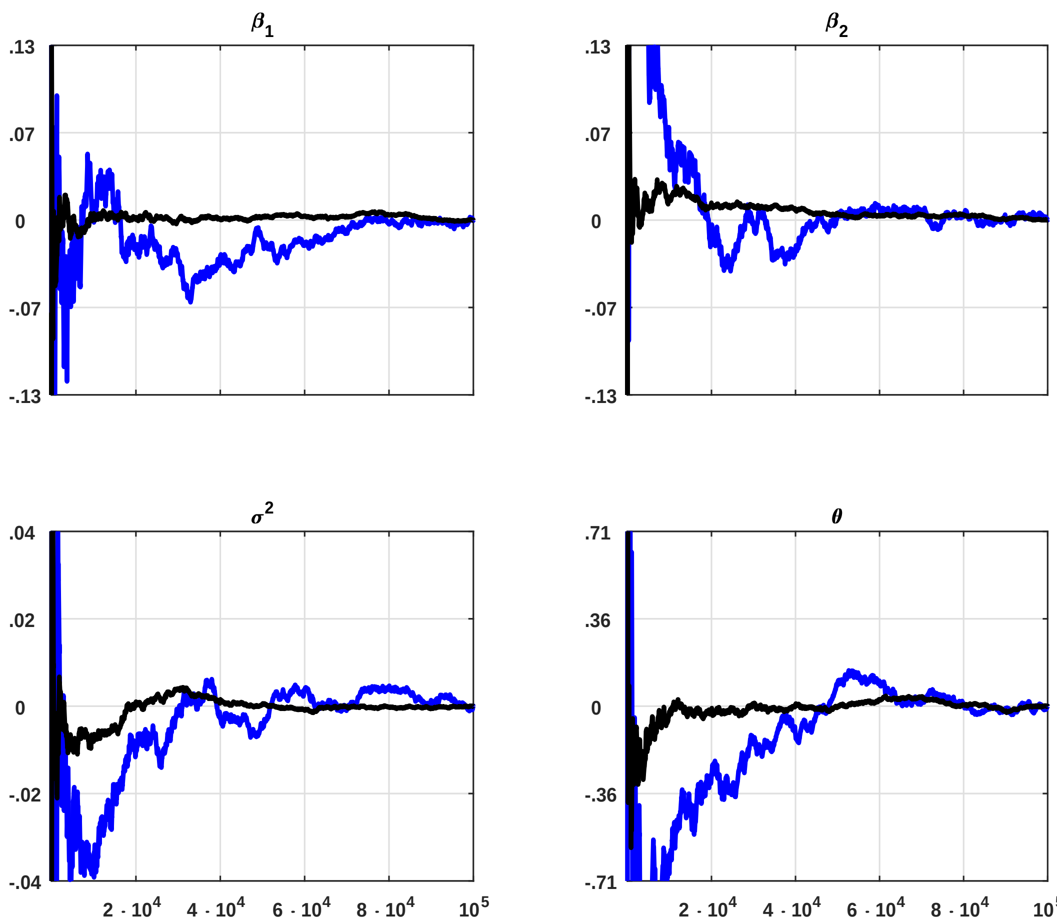
		β_1	β_2	σ^2	θ
Multivariate sampler with hyperrectangles					
IF		8.72	9.40	8.45	7.85
# eval	4.17				
run-time	28.52				
Antithetic multivariate sampler after symmetrizing					
IF		0.39	0.36	0.56	0.69
# eval	4.65				
run-time	37.30				
RIF		17.08	19.51	11.39	8.63

Notes: IF refers to the inefficiency factor defined by setting $h(x) = x$ in (2.2) and calculated using one hundred thousand of simulations after discarding the first one thousand; the autocorrelations are weighted by a Parzen window of length one hundred; #eval gives the average number of evaluations of the target density used to simulate one draw; run-time is the CPU time in seconds used per hundred simulated iterations; the relative inefficiency factor RIF is defined in (2.3) with Box-Cox transformation plus antithetic sampler taken as algorithm B.

The Neal sampler yields inefficiency factors about 8, and requires an average number of evaluations equal to 4.2. The Box-Cox transformation reduces the Mardia skewness from 2.7 on the original scale to 0.07. The antithetic sampler benefits from symmetrizing, yielding inefficiency factors below one, and this advantage is obtained at a moderate cost since the average number of evaluations increases only moderately from 4.2 to 4.7. For the four parameters, Figure 1 shows the cumulative posterior mean in deviation from the full sample average: the antithetic sampler appears to be quite effective in reducing the variance of the posterior mean. This result is obtained at the cost of increasing the run-time by one-third. Overall, this yields a RIF between nine and seventeen in favour of the antithetic algorithm. Given that one hundred thousand simulations are obtained in about eight hours, the Neal algorithm should run about seventy-two hours to yield the accuracy achieved with the antithetic algorithm.

FIGURE 1 Cumulative posterior means in Example 6,

in deviation from full sample averages



Notes: the x-axis shows the number of simulations; the blue line refers to multivariate slice sampling with hyperrectangles and the black one to the antithetic version applied to the symmetrized distribution.

4 Conclusion

We point out that the simple slice sampler generates chains with zero-correlations when the target distribution is centrally symmetric. This property is shared by the product slice sampler under the stronger condition that all factors of the target distribution are centrally symmetric around the same point. This explains the outcome of several comparisons between the simple and the product slice samplers which have appeared in the literature. The optimal behaviour of the simple slice sampler in the symmetric case also sheds some light on the algorithms that Neal (2003a) has proposed to circumvent the slice inversion problem. In the case of symmetric distributions which are univariate and unimodal, we could indeed see that stepping out preserves the zero-correlations property. Some correlations instead arise in the univariate multimodal and multivariate cases. In these situations we propose a slight amendment to stepping out and to Neal multivariate sampling with hyperrectangles which consists in constructing the slice interval around the mirror image of the current state with respect to centre of symmetry. The examples we report suggest that this antithetic strategy greatly enhances efficiency.

Since skewness of the target distribution impairs the efficiency of the slice sampler, symmetrizing the target distribution is worth considering. We focus on the Box-Cox transformation for its limited computational cost, with the objective to minimize the skewness. The antithetic algorithm is then implemented on the transformed distribution, the draws being then reset to the original scale by inverse transformation. The examples we report show that this simple and almost costless strategy yields important efficiency gains, in particular when the target distribution is strongly asymmetric. It thus provides a valid alternative to overrelaxation and reflective methods for improving efficiency. Of course the use of skewness to symmetrize the target distribution must be made with some caution since, if symmetry implies zero skewness, the opposite is not true. For such cases, alternative measures of symmetry should be considered. Should the Box-Cox transformation fail symmetrizing, alternative transformations can be found in Yeo and Johnson (2000) and in Yang (2006).

Appendix A

Given a positive integer d, the first moment of the distribution $\pi(x) = e^{-x^{1/d}} / d!$ verifies:

$$E_\pi(x) = \frac{1}{d!}\int_0^\infty x\, e^{-x^{1/d}} dx = \frac{1}{(d-1)!}\int_0^\infty z^{2d-1}\, e^{-z} dz = \frac{(2d-1)!}{(d-1)!}$$

where the second equality follows from the change of variable $z = x^{1/d}$ and the third one from the gamma function. The k-th moment verifies similarly:

$$E_\pi(x^k) = \frac{((k+1)d-1)!}{(d-1)!}$$

Given the first three moments the skewness $sk(x) \equiv E_\pi[(x - E_\pi(x))^3]/V_\pi(x)^{3/2}$ is obtained as:

$$sk(x) = \frac{(4d-1)!/(d-1)! - 3(3d-1)!(2d-1)!/(d-1)!^2 + 2((2d-1)!/(d-1)!)^3}{((3d-1)!/(d-1)! - ((2d-1)!/(d-1)!)^2)^{3/2}}$$

To calculate the first autocorrelation of the chain generated by the slice sampler first notice that the slice interval $S(u) = \{x : u \leq \pi(x)\}$ amounts to $x \in (0, [-\ln(u\ d!)]^d)$, with $u \in (0, 1/d!)$. The conditional moment $E(x|u)$ thus verifies $E(x|u) = [-\ln(u\ d!)]^d/2$. The marginal distribution of the auxiliary variable u is given by $\pi(u) = [-\ln(u\ d!)]^d$. The first covariance of the transition kernel is then obtained as:

$$\begin{aligned} V_\pi[E_\pi(x|u)] &= \frac{1}{4}\int_0^{\frac{1}{d!}} [-\ln(u\ d!)]^{3d} du - ((2d-1)!/(d-1)!)^2 \\ &= (3d)!/(4d!) - ((2d-1)!/(d-1)!)^2 \end{aligned}$$

where the integral is solved by substitution setting $v = -\ln(ud!)$ and using the gamma function. The first correlation of the transition kernel which verifies $Corr(x_i, x_{i+1}) = V_\pi[E_\pi(x|u)]/V_\pi(x)$ is such as:

$$\begin{aligned} Corr(x_i, x_{i+1}) &= \frac{(3d)!/(4d!) - ((2d-1)!/(d-1)!)^2}{(3d-1)!/(d-1)! - ((2d-1)!/(d-1)!)^2} \\ &= \frac{(3d)!/(4d!) - ((2d-1)!/(d-1)!)^2}{(3d)!/(3d!) - ((2d-1)!/(d-1)!)^2} \\ &= \frac{3}{4} - \frac{(2d-1)!^2}{4((3d-1)!(d-1)! - (2d-1)!^2)} \end{aligned}$$

References

Agarwal D.K. and Gelfand A.E. (2005), 'Slice sampling for simulation based fitting of spatial data models', *Statistics and Computing*, 15, 1, 61 – 69.

Andrews D.F., Gnanadesikan R., and Warner J.L. (1971), 'Transformations of multivariate data', *Biometrics*, 27, 825 – 840.

Banerjee S. (2005), 'On geodetic distance computations in spatial modeling', *Biometrics*, 6, 2, 617 – 625.

Berger J.O., De Oliveira V., and Sanso B. (2001), 'Objective Bayesian analysis of spatially correlated data', *Journal of the American Statistical Association*, 96, 456, 1361 – 1374.

Bishop C.M. (2006), *Pattern Recognition and Machine Learning*, Springer Verlag: Singapore.

Box G.E.P. and Cox D.R. (1964), 'An analysis of transformations', *Journal of the Royal Statistical Society B*, 26, 2, 211 – 252.

Damien P., Wakefield J.C., and Walker S. G. (1999), 'Gibbs sampling for nonconjugate and hierarchical models by using auxiliary variables', *Journal of the Royal Statistical Society B*, 61, 331 – 344.

De Oliveira V., Kedem B., and Short D.A. (1997), 'Bayesian prediction of transformed gaussian random fields', *Journal of the American Statistical Association*, 92, 1422 – 1433.

Dunson D.B. and C. Xing (2009), 'Nonparametric Bayes modeling of multivariate categorical data', *Journal of the American Statistical Association*, 104, 487, 1042 – 1051.

Edwards R.G. and A.D. Sokal (1988), 'Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm', *Physical Review D*, 38, 6, 2009 – 2012.

Gupta, R.D. and D. Kundu (2010), 'Generalized logistic distributions', *Journal of Applied Statistical Sciences*, 18, 51 – 66.

Hollander M. (1968), 'Certain uncorrelated nonparametric test statistics', *Journal of the American Statistical Association*, 63, 707 – 714.

Jasa T. and Xiang N. (2009), 'Efficient estimation of decay parameters in acoustically coupled-spaces using slice sampling', *Journal of the Acoustical Society of America*, 126, 3, 1269 – 1279.

Jones G.L. (2004), 'On the Markov chain central limit theorem', *Probability Surveys*, 1, 299 – 320.

Kline B. and Tamer E. (2016), "Bayesian inference in a class of partially identified models", *Quantitative Economics*, 7, 329 – 366.

Li J. (2011), "Volatility components, leverage effects, and the return–volatility relations", *Journal of Banking & Finance*, 35, 1530 – 1540.

Liechty M. W. and Lu J. (2010), "Multivariate normal slice sampling", *Journal of Computational and Graphical Statistics*, 19, 2, 281 – 294.

Liu J.S, Wong W.H., and Kong A. (1994), 'Covariance structure of the Gibbs sampler with applications to the comparison of estimators and data augmentation schemes', *Biometrika*, 81, 1, 27 – 40.

Mardia K.V. (1970), 'Measures of multivariate skewness and kurtosis with applications', *Biometrika*, 57, 3, 519 – 530.

Marron J. S. and Wand M.P. (1992), 'Exact integrated squared error', *The Annals of Statistics*, 20, 2, 712 – 736.

Mira A. and Roberts G.O. (2003), 'Discussion', *The Annals of Statistics*, 31, 3, 748 – 753.

Mira A. and L. Tierney (2002), 'Efficiency and convergence properties of slice samplers', *Scandinavian Journal of Statistics*, 29, 1, 1 – 12.

Neal R.M. (2003a), 'Slice sampling', *The Annals of Statistics*, 31, 3, 705 – 741.

Neal R.M. (2003b), 'Slice sampling: rejoinder', *The Annals of Statistics*, 31, 3, 758 – 767.

Nelsen R.B. (1993), 'Some concepts of bivariate symmetry', *Nonparametric Statistics*, 3, 95 – 101.

Robert C.P. and Casella G. (2004), *Monte Carlo statistical methods*, 2nd ed., Springer: New York.

Roberts G.O. and Rosenthal J. (2002), 'The polar slice sampler', *Stochastic Models*, 18, 2, 257 – 280.

Roberts G.O. and Rosenthal J. (1999), 'Convergence of slice sampler Markov chains', *Journal of the Royal Statistical Society B*, 31, 643 − 660.

Serfling R. J. (2006), Multivariate Symmetry and Asymmetry, in *Encyclopedia of Statistical Sciences*, ed. S. Kotz, N. Balakrishnan, C. B. Read, and B. Vidakovic, New York: Wiley, 5338 − 5345.

Tarasov L., Dyke A.S., Neal R.M., and Peltier W.R. (2012), 'A data-calibrated distribution of de-glacial chronologies for the North American ice complex from glaciological modeling', *Earth and Planetary Science Letters*, 315 − 316, pp. 30 − 40.

Taylor J. M. (1985), 'Power transformations to symmetry', *Biometrika*, 72, 1, 145 − 152.

Yang Z. (2006), 'A modified family of power transformations', *Economics Letters*, 92, 14 − 19.

Yeo I. and Johnson R.A. (2000), 'A new family of power transformations to improve normality or symmetry', *Biometrika*, 87, 4, pp. 954 − 959.

KJ-AE-18-011-EN-N

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/10835

ISBN 978-92-79-93405-6