

ADAPTIVE RADIAL-BASED DIRECTION SAMPLING: Some flexible and robust Monte Carlo integration methods

Econometric Institute Report EI 2003-22

Luc Bauwens¹, Charles S. Bos², Herman K. van Dijk³, and Rutger D. van Oest³

July 10, 2003

Abstract

Adaptive radial-based direction sampling (ARDS) algorithms are specified for Bayesian analysis of models with nonelliptical, possibly, multimodal target distributions. A key step is a radial-based transformation to directions and distances. After the transformations a Metropolis-Hastings method or, alternatively, an importance sampling method is applied to evaluate generated directions. Next, distances are generated from the exact target distribution by means of the numerical inverse transformation method. An adaptive procedure is applied to update the initial location and covariance matrix in order to sample directions in an efficient way.

Tested on a set of canonical mixture models that feature multimodality, strong correlation, and skewness, the ARDS algorithms compare favourably with the standard Metropolis-Hastings and importance samplers in terms of flexibility and robustness. The empirical examples include a regression model with scale contamination and a mixture model for economic growth of the USA.

Keywords: Markov chain Monte Carlo, importance sampling, radial coordinates.

JEL classification: C11, C15, C63

¹CORE and Department of Economics, Université catholique de Louvain.

²Econometrics & O.R. Department, Free University of Amsterdam.

³Econometric and Tinbergen Institutes, Erasmus University Rotterdam.

Correspondence to H.K. van Dijk, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. Email: hkvdiijk@ect.few.eur.nl

We thank Michel Lubrano, Rodney Strachan, several referees, and participants of seminars at Cambridge University, CORE, Tinbergen Institute Rotterdam, ESEM, and STOCOM meetings for helpful comments on an earlier version of this paper (Bauwens, Bos, Van Dijk, and Van Oest (2002)). This led to a substantial revision. Of course, responsibility for errors remains with the authors.

Support from HCM grant ERBCHRXCT 940514 of the European Commission is gratefully acknowledged. This paper presents research results of the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. The scientific responsibility is assumed by the authors.

1 Introduction

In recent decades Markov Chain Monte Carlo (MCMC) methods, in particular Metropolis-Hastings (MH) and Gibbs sampling (GS) and, to a lesser extent, independent sampling methods like importance sampling (IS), have been applied extensively and successfully within Bayesian analyses of statistical and econometric models.

Although Monte Carlo (MC) methods revolutionized the applicability of Bayesian inference, there is, in practice, a substantial variation in their convergence behaviour. The special features of the sampling method, the complex structure of the model, or the nature of the data may be the culprit of such behaviour. Hobert and Casella (1996) show for instance that the Gibbs sampler does not converge for the case of a hierarchical linear mixed model when the prior is uniform. Other examples of complex models are the ones with reduced rank structures. Kleibergen and Van Dijk (1994,1998) demonstrate near reducibility of MCMC methods when there exists near nonidentifiability and nonstationarity in econometric models with flat priors. Justel and Peña (1996) emphasize the convergence problems of the Gibbs sampler when there are outliers in the data. The performance of the Gibbs sampler is also seriously hampered by strong correlation in the target distribution. Convergence problems of importance sampling using a simple normal or Student t candidate density have been documented by Van Dijk and Kloek (1984) and Geweke (1989). A multimodal target density may pose problems to all methods. If the MH candidate density is unimodal, with low probability of drawing candidate values in one of the modes, this mode may be missed completely, even when the sample size is large. More generally stated, the acceptance probability may be very low, as many candidate values lying between the modes have to be rejected. With the Gibbs sampler, reducibility of the chain may occur in this case. Using a unimodal normal or Student t candidate function the method of importance sampling ends up with many drawings having only negligible weights. A common difficulty encountered in all samplers is the choice of a candidate or importance density when little is known about the shape of the target density. In such a case, updating the candidate density sequentially is a partial solution.¹

In this paper we introduce the class of adaptive radial-based direction sampling (ARDS) methods to sample from a target (posterior) distribution which is possibly multi-modal, skew, and exhibits strong correlation, in summary it is nonelliptical. The ARDS algorithms feature a composite transformation. The key step is one where the m -dimensional parameter space is transformed into radial coordinates which consist of a distance measure and a $(m - 1)$ -dimensional vector of directions. A MH or an IS algorithm is applied to sample a direction. Next, distances are sampled conditionally on the directions, from the (transformed) exact target density, by the inverse transformation method. A location-scale transformation is used as part of the transformation to radial coordinates and this transformation is sequentially updated, using the posterior first and second order moments obtained in successive rounds of the algorithm. The adaptive procedure is intended to improve the acceptance rate within the MH step and to give a more uniform distribution of the weights in the IS step.

The advantages of the ARDS algorithms are threefold. Firstly, the algorithms are *parsimonious* in their use of information on the shape of the target density. Only location and scale need to be specified as initial values. Secondly, the algorithms are *flexible* and *robust*: they can handle a large variety of features of target distributions, in particular multimodality,

¹This corresponds to the experimental results obtained by local adaptive importance sampling when the posterior is ill behaved, see e.g. Van Dijk and Kloek (1980), Oh and Berger (1992), and Givens and Raftery (1996).

strong correlation, extreme skewness, and heavy tails. We claim that the ARDS algorithms avoid the often time-consuming and difficult task, especially for non-experts, of choosing and tuning a sampling algorithm for a specific application, such as computation of posterior moments in Bayesian inference. They can be applied without the need to study in depth the shape of the posterior density in order to design a sophisticated approximation to it. It can be argued that for any specific model and data combination, a more efficient algorithm than ARDS may be designed, but our viewpoint is that the extra effort (in research time) required to achieve this may not be rewarding. Furthermore, we point out that one may apply ARDS algorithms as a first step to explore the shape of the posterior. Given knowledge of this shape one may design a specific algorithm more tailored to the problem studied. Thirdly, the algorithms can handle multiple *linear inequality conditions* on the parameter space without any additional complications for the implementation. In practice researchers often make use of bounds on parameters.

The ARDS algorithms extend earlier methods like the method of Box and Muller (1958), the adaptive direction sampling (ADS) algorithms proposed by Gilks, Roberts, and George (1994), the mixed integration method by Van Dijk, Kloek, and Boender (1985), and the spherical integration method by Monahan and Genz (1997). Details are given in Subsection 2.4.

The outline of the paper is as follows. In Section 2 the algorithms are introduced. In Section 3 mixture models are used for experimenting with ARDS and for comparing its performance with that of the standard Metropolis-Hastings and importance sampling algorithms. The models feature multimodality, high correlation and skewness. As empirical examples we use a regression model with scale contamination in order to investigate a study from Justel and Peña (1996) concerning the oxidation of ammonia to nitric acid in a plant. Next, we analyze economic growth of the USA using a mixture model as in Frühwirth-Schnatter (2001). Conclusions are presented in Section 4.

2 Adaptive radial-based direction sampling

Most simulation algorithms for posterior distributions generate random drawings in the original parameter space. Several researchers advocate to simulate in a transformed space, where the simulation is more efficient in some sense, see e.g. Gilks and Roberts (1996). For example, if there exists a strong correlation between two random variables, an orthogonalising transformation reduces serial dependence in a Gibbs sampling scheme. Another example arises in the context of importance sampling: an efficient importance function is easier to construct when an adequate transformation yields a distribution that is much closer to a symmetric one than the original one. The adaptive radial-based sampling algorithms rely on this general idea. They are based on a composite transformation to radial coordinates. Heuristically, the original parameter space is transformed into a $(m - 1)$ -dimensional space of directions and a unidimensional complementary space of distances. When the target or candidate density is a member of the elliptical family one can use the following result. The transformed density of the $(m - 1)$ -dimensional directions is uniform on the unit sphere and this density is independent of the density of the unidimensional distance which has a known analytical form (e.g. a member of the gamma family when the candidate is normal); see e.g. Muirhead (1982) (section 1.5). This result is the basis for the Box-Muller method of generating normal random variables; see Box and Muller (1958). In our approach, one key step is the creation

of directions, or lines, where the target density has non-trivial mass. A second key step is generating distances, or points on the lines, from the exact target density by means of the numerical inverse transformation method.

In this section, we concentrate on two members of the radial-based algorithms, one is of the Metropolis-Hastings type, and the second one is of the importance sampling type. These variants will be referred to as adaptive radial-based Metropolis-Hastings sampling (ARMS) and adaptive radial-based importance sampling (ARIS), respectively.

2.1 The radial transformation

Since the radial transformation is at the heart of the algorithms, we start by describing the transformation from Cartesian coordinates to radial coordinates. For ease of exposition, we distinguish between the two-dimensional case and the general m -dimensional case. In our notation, y denotes the Cartesian coordinates of a point, and (ρ, η) denotes the corresponding radial coordinates. Here $\eta = (\eta_1, \dots, \eta_{m-1})$ indicates the direction of the point relative to the origin, and ρ is related to the Euclidean distance.

2.1.1 The two-dimensional case

The two-dimensional transformation from $(y_1, y_2) \in \mathbb{R}^2$ to $(\rho, \eta) \in \mathbb{R} \times (-1, 1)$ is given by

$$\rho = \operatorname{sgn}(y_2) \sqrt{y_1^2 + y_2^2}, \quad (1)$$

$$\eta = \frac{y_1}{\rho}, \quad (2)$$

with inverse transformation

$$y_1 = \rho \eta, \quad (3)$$

$$y_2 = \rho \sqrt{1 - \eta^2}. \quad (4)$$

Note that $(\eta, \sqrt{1 - \eta^2})$ is located on the unit circle. Figure 1 illustrates the one-to-one relationship between the Cartesian coordinate system and the radial coordinate system. The Jacobian of the transformation is

$$J_{y_1, y_2}(\rho, \eta) = \det \begin{pmatrix} \frac{\partial y_1(\rho, \eta)}{\partial \eta} & \frac{\partial y_1(\rho, \eta)}{\partial \rho} \\ \frac{\partial y_2(\rho, \eta)}{\partial \eta} & \frac{\partial y_2(\rho, \eta)}{\partial \rho} \end{pmatrix} = \det \begin{pmatrix} \rho & \eta \\ -\frac{\rho \eta}{\sqrt{1 - \eta^2}} & \sqrt{1 - \eta^2} \end{pmatrix} = \frac{\rho}{\sqrt{1 - \eta^2}}.$$

2.1.2 The m -dimensional case

The m -dimensional transformation from $(y_1, \dots, y_m) \in \mathbb{R}^m$ to $(\rho, \eta) = (\rho, \eta_1, \dots, \eta_{m-1}) \in \mathbb{R} \times \{\eta \in \mathbb{R}^{m-1} : \eta' \eta < 1\}$ is given by

$$\rho = \operatorname{sgn}(y_m) \sqrt{y' y}, \quad (5)$$

$$\eta_j = \frac{y_j}{\rho}, \quad j = 1, \dots, m - 1, \quad (6)$$

with inverse transformation

$$y_j = \rho \eta_j, \quad j = 1, \dots, m - 1, \quad (7)$$

$$y_m = \rho \sqrt{1 - \eta' \eta}. \quad (8)$$

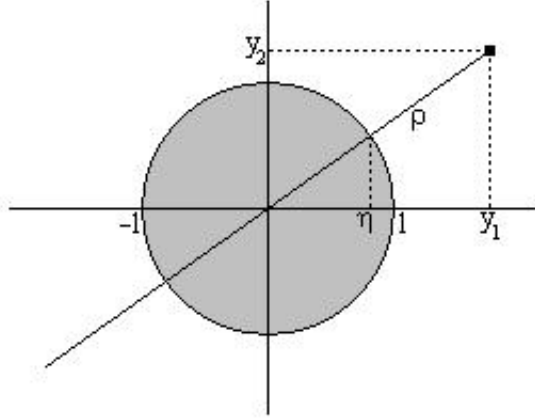


Figure 1: The relationship between Cartesian coordinates and radial coordinates in the two-dimensional case.

By defining $\tilde{y} = (y_1, \dots, y_{m-1})$, the Jacobian of the transformation is

$$\begin{aligned} J_y(\rho, \eta) &= \det \begin{pmatrix} \frac{\partial \tilde{y}(\rho, \eta)}{\partial \eta'} & \frac{\partial \tilde{y}(\rho, \eta)}{\partial \rho} \\ \frac{\partial y_m(\rho, \eta)}{\partial \eta'} & \frac{\partial y_m(\rho, \eta)}{\partial \rho} \end{pmatrix} = \det \begin{pmatrix} \rho I_{m-1} & \eta \\ -\frac{\rho \eta'}{\sqrt{1-\eta'\eta}} & \sqrt{1-\eta'\eta} \end{pmatrix} \\ &= \frac{\rho^{m-1}}{\sqrt{1-\eta'\eta}} = J_y(\rho) J_y(\eta), \end{aligned} \quad (9)$$

where

$$J_y(\rho) = \rho^{m-1}, \quad (10)$$

$$J_y(\eta) = (1 - \eta'\eta)^{-1/2}. \quad (11)$$

Thus, the generalization of (1)-(4) to the m -dimensional case is straightforward.

2.1.3 A generalization

In an earlier version (Bauwens, Bos, Van Dijk, and Van Oest (2002)), we made use of the polar transformation. We prefer the radial transformation for ease of exposition and for computational efficiency. However, both transformations can just be regarded as special cases of a more general transformation such that the inverse transformation is of the form $y = \rho h(\eta)$, where h is some m -dimensional differentiable function. In any case the Cartesian coordinates y are transformed to a distance ρ and a position $h(\eta)$ on the unit circle. The Jacobian of this general transformation is of the form $J_y(\rho, \eta) = J_y(\rho) J_y(\eta)$, where $J_y(\rho)$ is still defined by $J_y(\rho) = \rho^{m-1}$. As the algorithms turn out to depend only on the invariant Jacobian factor $J_y(\rho)$, and not on $J_y(\eta)$, our results for the radial transformation can be extended to any transformation of the form described above.

2.2 Adaptive radial-based Metropolis-Hastings sampling (ARMS)

We start by defining the radial-based Metropolis-Hastings algorithm (RMS), which is based on a candidate generating density that is taken to be multivariate normal with parameters

μ and Σ . Next, we illustrate the algorithm for a bimodal target density. Finally, we define the adaptive RMS algorithm (ARMS), where μ and Σ are updated using the sample of draws from a previous round of the RMS algorithm.

2.2.1 Definition of radial-based Metropolis-Hastings sampling (RMS)

RMS is based on an independence chain MH algorithm. It uses draws from a $N(\mu, \Sigma)$ candidate where hopefully μ and Σ provide good approximations to the unknown mean and covariance matrix of the target distribution, see below. We note that normality of the candidate density is only relevant to the extent that drawings should be generated from a member of the class of elliptical distributions; see also Remark 1 below. In contrast with the MH algorithm, the drawings are not used for construction of a Markov chain in the original parameter space. Instead, a composite transformation is made. For expository purpose we treat this transformation explicitly in two steps.

The first step is a location-scale transformation of a realization x to a realization y . This transformation aims at standardizing the candidate density with respect to the location, scale, and correlations of the target (posterior) density, denoted by $p(x)$. It is given by²

$$y = y(x | \mu, \Sigma) = \Sigma^{-1/2}(x - \mu), \quad (12)$$

with inverse transformation

$$x = x(y | \mu, \Sigma) = \mu + \Sigma^{1/2}y, \quad (13)$$

and Jacobian

$$J_x(y) = \det(\Sigma^{1/2}). \quad (14)$$

The second key step is the radial transformation, which is defined by (5) and (6), with inverse transformation given by (7) and (8), and Jacobian (9).

Combining the two transformations, one obtains the composite transformation

$$\begin{pmatrix} \rho \\ \eta \end{pmatrix} = \begin{pmatrix} \rho(x | \mu, \Sigma) \\ \eta(x | \mu, \Sigma) \end{pmatrix} = \begin{pmatrix} \rho(y(x | \mu, \Sigma)) \\ \eta(y(x | \mu, \Sigma)) \end{pmatrix} \quad (15)$$

with inverse transformation

$$x = x(\rho, \eta | \mu, \Sigma) = x(y(\rho, \eta) | \mu, \Sigma) \quad (16)$$

and Jacobian

$$J_x(\rho, \eta) = J_y(\rho, \eta)J_x(y) = J_y(\rho)J_y(\eta)\det(\Sigma^{1/2}). \quad (17)$$

Applying the two transformations to a candidate realization x_i^* from $N(\mu, \Sigma)$ yields a distance ρ_i^* and a vector of directions η_i^* (hereafter referred to as a ‘direction’).³ Ignoring

² $\Sigma^{1/2}$ denotes the Cholesky decomposition of Σ , and $\Sigma^{-1/2}$ denotes the inverse matrix of $\Sigma^{1/2}$.

³Henceforth, the index i (in x_i^* , η_i^* ...) does not indicate the i -th element of the corresponding vector, but indicates the number of the draw in a sequence of successive draws.

the distance, the candidate direction is either accepted or rejected in a MH step, that is, the direction becomes

$$\eta_i = \begin{cases} \eta_i^* & \text{with probability } \alpha(\eta_{i-1}, \eta_i^*) \\ \eta_{i-1} & \text{with probability } 1 - \alpha(\eta_{i-1}, \eta_i^*) \end{cases} \quad (18)$$

for some acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, which is given in Proposition 1 below. An iteration of RMS is completed by drawing from the target distribution on the line defined by the direction η_i . This can be done as follows. First, one draws a distance ρ_i from the transformed target distribution for given direction η_i using the numerical inverse transformation method, see Proposition 1. Next, ρ_i and η_i are transformed to the original space by inverting the radial transformation and the location-scale transformation. In Table 1, we summarize the steps of one iteration of RMS.

Table 1: One iteration of RMS

-
1. Generate x_i^* from $N(\mu, \Sigma)$
 2. Transform x_i^* to $y_i^* = \Sigma^{-1/2}(x_i^* - \mu)$
 3. Transform y_i^* to ρ_i^* and η_i^* , using (5) and (6)
 4. Apply MH step to η_i , see (18)
 5. Generate ρ_i from $p(\rho|\eta_i)$ by inverting numerically its cdf
 6. Transform ρ_i and η_i to y_i , using (7) and (8)
 7. Transform y_i to $x_i = \mu + \Sigma^{1/2}y_i$

Note that steps 1 and 2 amount to generating y_i^* from $N(0, I_m)$. We want to make explicit the dependence on μ and Σ .

Step 4 of a RMS iteration requires the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, and step 5 requires the distribution of the distance ρ conditional on the direction η_i . They are given in the next proposition.

Proposition 1 *For all elliptically contoured candidate distributions with mean μ and covariance matrix Σ , the acceptance probability of step 4 of the RMS algorithm, summarized in Table 1, depends only on the generated direction and not on the functional form of the candidate density. The value of $\alpha(\eta_{i-1}, \eta_i^*)$ is given by*

$$\alpha(\eta_{i-1}, \eta_i^*) = \min \left\{ \frac{I(\eta_i^*)}{I(\eta_{i-1})}, 1 \right\}, \quad (19)$$

where

$$I(\eta) = \int_{-\infty}^{\infty} \kappa(\rho|\eta) d\rho, \quad (20)$$

and where $\kappa(\rho|\eta)$ is a kernel of the conditional density $p(\rho|\eta)$ of step 5, defined by

$$p(\rho|\eta) \propto \kappa(\rho|\eta) = p(x(\rho, \eta | \mu, \Sigma)) |J_y(\rho)|. \quad (21)$$

Proof: See the Appendix. \diamond

Remark 1: A noteworthy property is that the acceptance probability does not depend on the functional form of the candidate density under the condition that this candidate density is of the form $f((x - \mu)' \Sigma^{-1} (x - \mu))$, i.e. an elliptically-contoured density. However, the acceptance probability depends on the generated direction η and thus on the location and scaling matrix of the candidate density.

Remark 2: In order to obtain the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, the integral $I(\eta)$ defined by (20) can be computed by a deterministic integration rule. Since the density of ρ conditional on η is proportional to the integrand of $I(\eta)$, evaluations of the integrand, gathered during the deterministic integration phase, can be used in order to construct a grid for $p(\rho|\eta)$. Using the numerical inverse transformation method, sampling the distance ρ conditional on the direction η , that is, step 5 of a RMS iteration, is straightforward.

Remark 3: The integral $I(\eta)$ has infinite integration bounds. However, in practice we use finite integration bounds for its numerical evaluation. Of course it is important to take these bounds such that practically all density mass of ρ given η is included, but for the sake of efficiency it is also desirable that the integration interval is as small as possible. In order to obtain bounds for the distance ρ we impose minimum and maximum values for each element of x in the original space. It is often possible to find sensible bounds by either theory and/or common sense, and in the applications below we will explicitly report the considered bounds. More generally, by imposing bounds on the values of x we impose linear inequality conditions on the original parameter space. These restrictions are of the form $c'_j x \leq b_j$, where j indicates the number of the restriction. The conditions imposed on the original space translate to bounds ρ_{\min} and ρ_{\max} for ρ through the relationships $\rho_{\min} = \max_j \{\rho_j : \rho_j < 0\}$ and $\rho_{\max} = \min_j \{\rho_j : \rho_j > 0\}$, where $\rho_j = \frac{b_j - c'_j \mu}{c'_j (\tilde{x} - \mu)}$ with $\tilde{x} = x(\rho = 1, \eta | \mu, \Sigma)$. As additional linear restrictions might reduce the integration interval for $I(\eta)$, making the evaluation of $I(\eta)$ more efficient, they do not put a burden on the algorithm, but they might result in an efficiency gain.

2.2.2 Convergence

RMS is a combination of a Metropolis-Hastings sampler for the directions η and the inverse transformation method for generating the distance ρ . The MH step on the directions introduces dependence on past drawings. Hence, the transition kernel of RMS is just the transition kernel of the MH step. The following theorem provides a sufficient condition for convergence of RMS.

Proposition 2 *If for every $\eta \in \{\eta \in \mathbb{R}^{m-1} : \eta' \eta < 1\}$ in the support of the target density $p(\eta)$ the candidate density $q(\eta)$ is non-null and continuous, then the sampled RMS chain converges in distribution to the target distribution.*

In the Appendix it is shown that for an elliptically contoured candidate, the density $q(\eta)$ is given by $q(\eta) \propto J_y(\eta) = (1 - \eta' \eta)^{-1/2}$. As $q(\eta)$ is clearly positive and continuous on $\{\eta \in \mathbb{R}^{m-1} : \eta' \eta < 1\}$, the sufficient condition in Proposition 2 is satisfied.

In practice, convergence should always be checked using diagnostics. One can monitor the sequential values of posterior moments, the acceptance rates, and the (tail) distribution and variance of the importance weights.

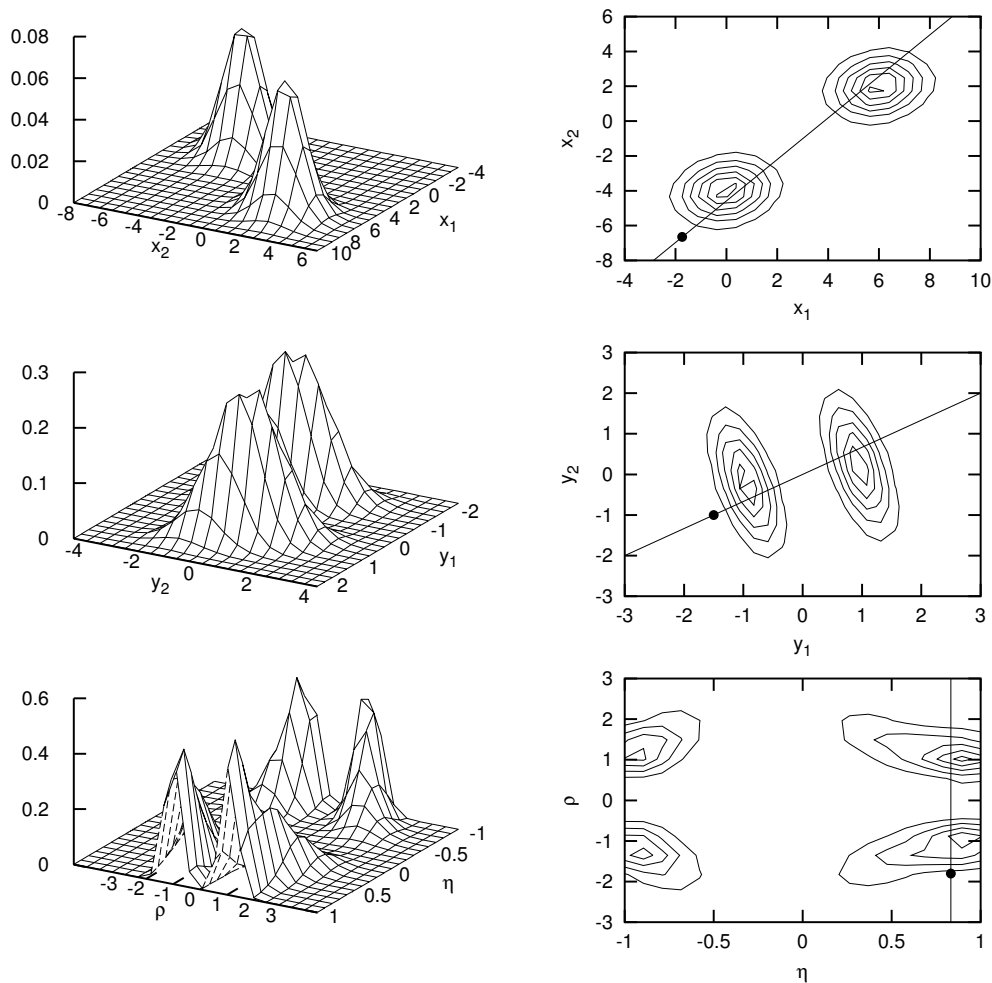


Figure 2: Adaptive radial-based direction sampling: target density in original space (above), target density after location-scale transformation (middle) and target density after radial transformation (below).

Remark 4: In practice we reduce the computational effort by generating several drawings of ρ for each drawing of η , i.e. we capitalize on the construction of the grid for $p(\rho|\eta)$ (see Remark 2). We emphasize that the computed integrals (MC estimators) still converge to the theoretical integrals. The main point is that although the generated drawings of y and x are dependent, the computed integrals are consistent estimates of the theoretical values of the integrals that one is interested in, see Geweke (1999) (p 44) and the references cited there.

2.2.3 Illustration

Figure 2 illustrates RMS for a bivariate bimodal target distribution. The upper two graphs display the target density in the original space. A point, representing a realization from the normal candidate distribution $N(\mu, \Sigma)$, is visible in the contour plot. If μ and Σ coincide with the mean and the covariance matrix of the target distribution, then the location-scale

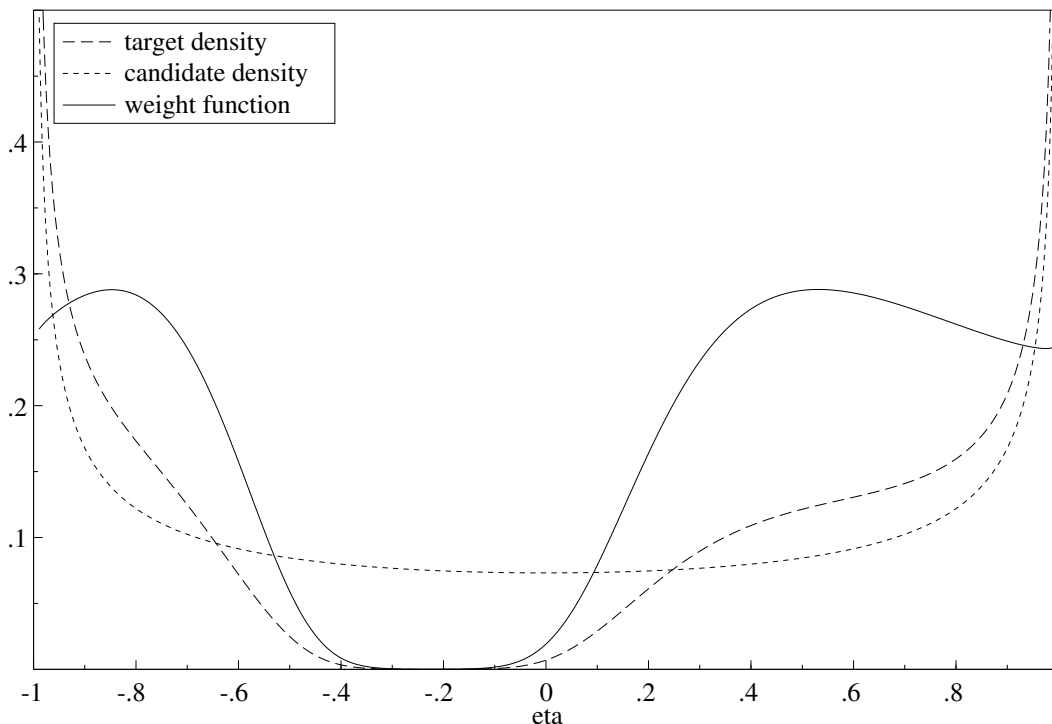


Figure 3: The marginal target density $p(\eta)$, the marginal candidate density $q(\eta)$, and the weight function $p(\eta)/q(\eta)$ (up to scaling constants). The original target density is the same as in the previous figure.

transformation would lead to the target density that is depicted in the middle graphs. The gain of the location-scale transformation is clear: the density mass is better located around the origin in the sense that a line through the origin, defined by some direction η , ‘hits the density mass’ more easily. Since RMS precisely considers such lines, the location-scale transformation may lead to a substantial improvement for appropriate μ and Σ . The target density after applying the radial transformation is depicted in the bottom two graphs.

Seven steps are distinguished in an iteration of RMS, see Table 1. The visualization of these steps in Figure 2 is as follows. In step 1, the point in the upper contour plot is drawn from $N(\mu, \Sigma)$. This point is transformed in step 2 to the point in the middle contour plot. Step 3 results in the point in the bottom contour plot. In step 4, the direction η , that is the horizontal position of the point in the bottom contour plot, is either accepted or rejected. If we assume for expository purpose that η is accepted, step 5 consists of drawing one (or several) distance(s) ρ on the vertical line through the point. Step 6 can be represented by the transformation of points generated on the line in the bottom contour plot to points generated on the line in the middle contour plot. Similarly, step 7 results in points generated on the line in the upper contour plot. In sum, RMS amounts to drawing from the exact target density on lines which are obtained through a MH acceptance-rejection mechanism. This acceptance-rejection mechanism takes into account the shape of the target density.

Figure 3 shows the marginal target density $p(\eta)$ that would result from the bottom two graphs in Figure 2. It looks quite ill behaved. However, similar to the relationship between

standard independence chain MH and standard importance sampling, it is not directly the target density that matters for the MH step in RMS. What matters is the ratio of the target density and the candidate density of η , that is $p(\eta)/q(\eta) \propto I(\eta)$ where $q(\eta)$ denotes the marginal candidate density and $I(\eta)$ has been defined in Proposition 1, see the Appendix for details. It is seen from Figure 3 that, although the target density $p(\eta)$ is ill behaved, the weight function $p(\eta)/q(\eta)$ is quite well behaved in the sense that variation in its values is only moderate. We further note from the bottom two graphs in Figure 2 that the transformed target density is bimodal (ill behaved) with respect to the distance ρ . However, this is not a problem, as a grid for $p(\rho|\eta)$ can always be constructed easily, see also Remark 2.

2.2.4 Adaptive radial-based Metropolis-Hastings sampling (ARMS)

For implementation of RMS, the mean μ and the covariance matrix Σ of the normal candidate distribution have to be specified. Good enough initial approximations are usually the posterior mode and minus the inverse Hessian of the log posterior evaluated at the mode. Heuristically, convergence of RMS should improve if μ and Σ are close to, rather than far from, the target mean and covariance matrix, respectively. Actually, if the target density is (approximately) normal, the acceptance probability should be (approximately) equal to 1 for appropriate μ and Σ . ARMS considers a sequential adaptive approach. Given a generated sample x_1, x_2, \dots, x_n from a previous run of the algorithm, μ and Σ are replaced by the Monte Carlo estimates of the posterior mean and variance, which are given by

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (22)$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})(x_i - \hat{\mu})', \quad (23)$$

respectively. Using these estimates, one can proceed with a new sampling round. This process can be repeated any number of times. We note that information, coming from a ‘wrong’ sample, may have a misleading effect and may worsen convergence. Thus, convergence should be monitored by usual tools, see Van Dijk and Kloek (1980) and Oh and Berger (1992). However, since only the direction η , and not the distance ρ , depends on the candidate distribution, the risk of collecting a ‘wrong’ sample is reduced. ARMS should be quite robust, as the distance ρ conditional on the direction η immediately comes from the target distribution, that is, sampling on a given line mimics exactly the target density.

In order to monitor convergence over sampling rounds, we find the Mahalanobis distance particularly useful. It is defined as $\text{MAH}_j = (\hat{\mu}^{(j)} - \hat{\mu}^{(j-1)})' [\hat{\Sigma}^{(j)}]^{-1} (\hat{\mu}^{(j)} - \hat{\mu}^{(j-1)})$, where j indicates the sampling round. The Mahalanobis distance measures the extent to which the estimated posterior mean changes between successive sampling rounds, while accounting for parameter uncertainty and the underlying correlation structure. In the applications below we will use the Mahalanobis distance extensively.

2.3 Adaptive radial-based importance sampling (ARIS)

Radial-based importance sampling (RIS) replaces the MH step of RMS for the direction η by an importance sampling step. So, step 4 of a RMS iteration (see Table 1) changes. In RIS, every sampled direction η_i is kept, a distance ρ_i is sampled conditional on it, and the

resulting radial coordinates are transformed to a draw x_i in the original space, which is weighted according to the appropriate importance weight $w(\eta_i)$. This importance weight is given by

$$w(\eta_i) = \frac{p(\eta_i)}{q(\eta_i)} \propto I(\eta_i), \quad (24)$$

where $I(\eta)$ is defined by (20), see the Appendix.

An interpretation of RIS is that one samples from the target distribution on lines with directions being derived from the elliptically contoured candidate distribution. Each line receives a weight, indicating the importance of the underlying direction. The weight of a line is carried over to any realization on that line. Alternatively, one may interpret RIS as just a special case of standard importance sampling. A realization x in the original space is a function of a realization (ρ, η) in the transformed space, see (16), implying that the importance weight of (ρ, η) is also the importance weight of x . Taken together, step 1 to step 5 of a RIS iteration can be regarded as yielding a realization (ρ_i, η_i) from a candidate distribution with density

$$q_{imp}(\rho, \eta) = q(\eta)p(\rho|\eta), \quad (25)$$

and providing the importance weight

$$w(\rho, \eta) = \frac{p(\rho, \eta)}{q_{imp}(\rho, \eta)} = \frac{p(\eta)p(\rho|\eta)}{q(\eta)p(\rho|\eta)} = \frac{p(\eta)}{q(\eta)}, \quad (26)$$

which coincides with (24). We note that the importance function q_{imp} takes into account the *shape* of the target density through $p(\rho|\eta)$. This is not the case in standard importance sampling with importance function $q(\eta)q(\rho|\eta)$ in the transformed space, which explains that the importance function of RIS should be more efficient than the importance function in standard IS with a normal or Student t candidate density. As RIS can be interpreted as a special case of importance sampling, convergence properties of RIS follow directly from those for the latter method. The distribution of the weights $w(\eta)$, in particular, a (bounded) variance of $w(\eta)$, are important diagnostics. For details, see Geweke (1989).

Similar to ARMS, the parameters μ and Σ of the location-scale transformation can be updated by replacing them by their Monte Carlo estimates. These estimates are given by

$$\hat{\mu}_w = \frac{\sum_{i=1}^n w(\eta_i)x_i}{\sum_{i=1}^n w(\eta_i)}, \quad (27)$$

$$\hat{\Sigma}_w = \frac{\sum_{i=1}^n w(\eta_i)(x_i - \hat{\mu}_w)(x_i - \hat{\mu}_w)'}{\sum_{i=1}^n w(\eta_i)}, \quad (28)$$

where x_1, x_2, \dots, x_n is the collected sample, and $w(\eta_1), w(\eta_2), \dots, w(\eta_n)$ are the corresponding importance weights.

2.4 On related methods and a classification

The polar transformation is the basis of the well known method of Box and Muller (1958) for generating normal random variables. Consider the angle θ formed by the y_1 -axis and a line passing through the origin. In polar coordinates, a point is described by this angle θ and its Euclidean distance from the origin. In the Box-Muller method, an angle (therefore a direction

η) is generated from a uniform distribution on the interval $(0, 2\pi)$, and a squared distance is generated from an exponential distribution. See e.g. Rubinstein (1981) (p. 86-87). In radial coordinates, the point is described by the radial coordinate η and the signed distance ρ , see (1) and (2). The ARDS method extends the Box-Muller algorithm by generating directions η using an MH or IS step, where the candidate density is compared with a target density. Given a generated candidate direction η , distances ρ are generated from a very accurate numerical approximation to the target distribution of the distances. Note that this exact distribution is model specific. If the normal candidate density in the original space is a good approximation to the target density in that space then the probability of acceptance in the MH step is close to one and the weight in the IS step is relatively constant. Non-normality can be evaluated using the weights computed in the one-dimensional integration step; see Hop and Van Dijk (1992) and Monahan and Genz (1997).

It is also of interest to compare ARDS with the class of adaptive direction sampling (ADS) algorithms, see Gilks, Roberts, and George (1994). Two well known members of ADS are the hit and run algorithm of Schmaier and Chen (1991) and the snooker algorithm of Gilks, Roberts, and George (1994). In ADS, directions are sampled in the original parameter space. Only information on the shape of the target density is used. In ARDS, we generalize ADS in two ways. Firstly, use is made of a MH or IS step where candidate and target are compared to generate a direction. Secondly, one generates a distance ρ from a numerically very accurate approximation to the target distribution. This step is not always spelled out in ADS. Since ARMS and ARIS are members of the MH and IS class of Monte Carlo methods, convergence properties of these methods are well established. This is not so transparent for the ADS methods.

We emphasize again that within ARDS one can make use of any candidate that belongs to the family of elliptical distributions. The advantage of the normal density is its simplicity and parsimony of parameters: location and scale determine the distribution. A good estimate of the location and scale is important for efficient generation of directions, that is, directions that generate lines which cover the region where the target has substantial probability mass, see e.g. the line in Figure 2. We note that Monahan and Genz (1997) use the terminology radial based integration in this context.

The ARDS class comprises several algorithms. One may distinguish between rejection sampling, importance sampling and Metropolis-Hastings sampling as Monte Carlo integration methods. So far, we have experimented with ARMS and ARIS. However, one may also define a radial-based rejection sampling algorithm (ARRES): the sampled direction is accepted if $w(\eta) > cu$ (and rejected otherwise), where u is uniformly drawn in $(0, 1)$, and c is a constant such that the importance function envelopes the target function. For such an algorithm, formulas (22) and (23) apply to the accepted transformed draws. Consider next the case where generating random drawings of ρ is replaced by only evaluating the unidimensional integral. We name this case deterministic integration with respect to ρ . One can combine this deterministic integration with respect to ρ with rejection sampling, importance sampling or Metropolis-Hastings sampling with respect to η and evaluate posterior moments and densities. For the case of importance sampling this has been done in the so-called mixed integration method of Van Dijk, Kloek, and Boender (1985), compare also Monahan and Genz (1997), who use the term spherical radial integration. Thus these methods are special cases of the ARDS class where the step of generating random drawings of ρ is reduced to evaluating only a unidimensional integral. The limitation of deterministic integration with respect to ρ is that one has to compute a different unidimensional integral for each moment of the target

distribution, see Hop and Van Dijk (1992). When one needs only to compute first and second-order posterior moments, one may use the mixed integration methods efficiently. We have summarized the six algorithms in Table 2.

We emphasize that the ARDS sampling methods can be applied to target densities where the region of integration is bounded by several inequality conditions. Linear inequality conditions play an important role in several econometric and statistical applications.

Table 2: Members of the ARDS class of algorithms

$\eta \setminus \rho$	Stochastic generation of ρ	Deterministic integration of ρ
Rejection sampling	ARRES	MIXRES
Importance sampling	ARIS	MIXIS
M-H sampling	ARMS	MIXMS

3 Applications

In this section, a set of models are used to illustrate the versatility of radial-based algorithms. We do not claim that the algorithms are necessarily the most efficient ones for the analyzed models and data, since careful analysis of the posterior density may help to design a more efficient algorithm (actually, the radial-based algorithms can also be useful in this perspective). We compare ARMS and ARIS to the (independence chain) Metropolis-Hastings algorithm and importance sampling. In all examples, MH and importance sampling use the Student t distribution with 5 degrees of freedom as the candidate. We start with two artificial examples involving mixtures of normal distributions in the parameter space. The experiments with these mixtures feature multimodality, skewness and high correlation. Next, we analyze two empirical examples that are used in the literature for illustrative purposes. The first example involves a regression model with scale contamination in order to investigate a study from Justel and Peña (1996) concerning the oxidation of nitric acid in a plant. Secondly, we consider a mixture model as in Frühwirth-Schnatter (2001) for the analysis of economic growth in the USA. We note that in these examples the mixture process refers to the data space. However, it is noteworthy that such mixture processes may give rise to bimodality, extreme correlation and skewness in the parameter space. Of course, when the sample of data is very large and the model is regular such ill behaviour of the posterior density will be less likely. The issue is that a researcher does not know this before the posterior analysis of a model.

We emphasize that all numerical results reported below depend on the design of the experiments and on the initial values. Different initial values (seed of the random number generator and starting values of the location and scale) may give different numerical results for the given number of drawings. However, the central message from the results reported below turns out to be quite robust.

3.1 The bimodal density from the illustration

As a first experiment we consider the bivariate bimodal density which was used in the illustration of RMS (see Figures 2 and 3). The distribution is given by

$$0.5 N\left(\begin{pmatrix} 0 \\ -4 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) + 0.5 N\left(\begin{pmatrix} 6 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

We compute the first and second moments using RMS, RIS, MH and importance sampling with three different Student t candidate densities. The candidates differ in their location and scaling parameters. They are given by

$$(a) \quad t_5\left(\begin{pmatrix} 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}\right),$$

$$(b) \quad t_5\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 25 & 0 \\ 0 & 25 \end{pmatrix}\right),$$

$$(c) \quad t_5\left(\begin{pmatrix} 0 \\ -4 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

We note that for RMS and RIS a normal candidate would give exactly the same results, as the only requirement is that the candidate density is elliptically contoured, see Remark 1. The first candidate has the location and scaling parameters of the true target density. Only the correlation parameter is incorrect. The second candidate density has incorrect location but also a relatively large scale so that MH and importance sampling consider candidate drawings over a sufficiently large region. The third candidate density coincides with the first mode of the target density, so that the scale is too small. For each method, the moment estimates are based on 1000 drawings. In RMS and MH, these drawings are preceded by a burn-in period of 100 drawings. In RMS and RIS, we set the minimum and maximum bounds at respectively -10 and 10 for both components.

The estimated first and second moments and the true values for these moments are reported in Table 3. Furthermore, Figure 4 displays the underlying RMS and MH samples. The main findings are as follows. For the first two candidates (a) and (b), all four methods come up with good estimates for the scaling parameters and the correlation parameter, although the estimates from RMS and RIS are slightly more accurate. However, the location parameters estimated using RMS and RIS are much more accurate than those obtained from MH and importance sampling. The reason for this is illustrated in Figure 4. The RMS plots concerning the candidate densities (a) and (b) display much more points than the corresponding MH plots, indicating that the acceptance rates for the RMS samples are much higher than the acceptance rates for the MH samples. This leads to a higher accuracy. Finally, it is seen for the third candidate (c) that MH and importance sampling fail to deliver a representative sample. They only find one mode and miss the other mode completely. This is in contrast to RMS and RIS, which still succeed to come up with good moment estimates. In sum, this example demonstrates that, *given the same number of drawings*, ARDS algorithms should

Table 3: Sampling results for the bimodal density considered in the illustration of RMS.

		RMS	RIS	MH	IS	true
(a)	mean(x_1)	2.83	2.96	2.70	3.35	3.00
	mean(x_2)	-1.16	-1.05	-1.37	-0.71	-1.00
	stdev(x_1)	3.15	3.16	2.98	3.09	3.16
	stdev(x_2)	3.15	3.17	3.21	3.28	3.16
	corr(x_1, x_2)	0.91	0.91	0.87	0.91	0.90
(b)	mean(x_1)	2.94	2.97	2.41	2.21	3.00
	mean(x_2)	-1.01	-1.01	-1.59	-1.59	-1.00
	stdev(x_1)	3.12	3.15	3.11	3.06	3.16
	stdev(x_2)	3.17	3.21	3.21	3.16	3.16
	corr(x_1, x_2)	0.89	0.91	0.93	0.91	0.90
(c)	mean(x_1)	2.90	3.19	0.01	0.02	3.00
	mean(x_2)	-1.14	-0.82	-4.02	-4.00	-1.00
	stdev(x_1)	3.11	3.11	1.04	1.02	3.16
	stdev(x_2)	3.15	3.18	1.04	1.01	3.16
	corr(x_1, x_2)	0.90	0.91	-0.03	0.00	0.90

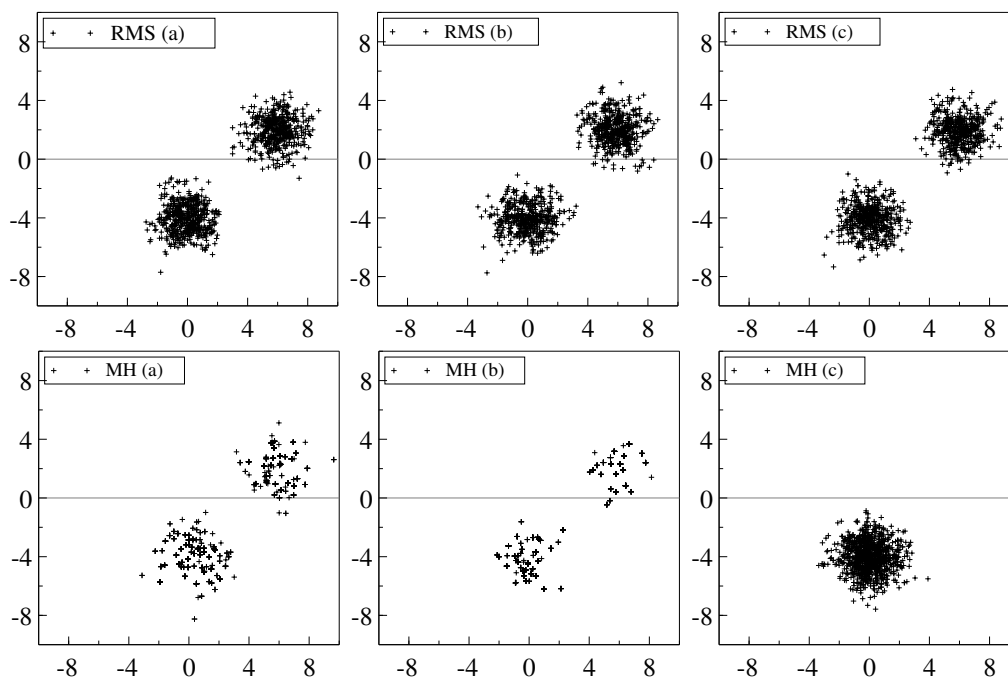


Figure 4: Samples obtained for the bimodal density considered in the illustration of RMS.

at least result in more accurate estimates than MH and importance sampling. Further, the example illustrates that ARDS algorithms should be more robust in the sense that the risk of overlooking density regions with nontrivial density mass (i.e. missing complete modes) is reduced. In this example, we conditioned on the number of drawings. However, as obtaining drawings from RMS or RIS is more time consuming than obtaining drawings from the other two methods, we will make the computing times comparable in the next example.

3.2 A trimodal mixture distribution

In this second example, we consider an 8-dimensional trimodal mixture distribution featuring skewness, some high correlations (varying between -0.95 and 0.90) and multimodality. It is given by

$$p_1 N(\mu_1, \Sigma_1) + p_2 N(\mu_2, \Sigma_2) + p_3 N(\mu_3, \Sigma_3),$$

where

$$\mu_1 = 1.5 (1, 2, 3, 4, 5, 6, 7, 8)',$$

$$\mu_2 = 1.5 (5, 6, 7, 8, 1, 2, 3, 4)',$$

$$\mu_3 = 1.5 (8, 7, 6, 5, 4, 3, 2, 1)',$$

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = I_8,$$

$$p_1 = p_2 = p_3 = 1/3.$$

The marginal densities⁴ for the eight components are displayed in Figure 5. It is seen that the target density is very ill behaved. We estimate the mean and the covariance matrix for this distribution using ARMS, ARIS, MH and importance sampling. This is done in several sampling rounds. In our adaptive approach, additional sampling rounds are considered as long as the Mahalanobis distance is larger than 0.02. However, we allow for at most 8 rounds. In any round, ARMS and ARIS sample 5000 directions and 5 distances per direction, resulting in a sample of size 25000. In order to make the computing times comparable, the MH and importance sampling algorithms are allowed to collect a larger sample of size 250000. The mean for the initial candidate is set at 6 for all eight components. Furthermore, the scale is taken sufficiently large so that MH and importance sampling can initially cover the whole density mass.

The estimates for the location and scale parameters are reported in Table 4. Furthermore, the table also contains the “true” parameter values, obtained from 250000 *directly* sampled drawings. It is seen that ARMS and ARIS do a very good job, whereas MH and importance sampling fail. ARIS only needs 5 iterations to reach convergence (according to our definition), whereas the other three algorithms need the maximum number of rounds. However, after eight rounds also ARMS has converged whereas the other two algorithms clearly have not, see the reported Mahalanobis distances which concern the final sampling round. We note that the average computing times per sampling round for the four algorithms are comparable.

The acceptance rates for ARMS and MH (reported for the final round) show a large difference in values. Furthermore, it is seen that the moment estimates, obtained from importance sampling, are almost completely determined (99.7%) by only 5% of the drawings. In contrast, in ARIS the 5% most influential drawings only have about 30% of the total weight.

⁴These marginal densities were constructed using 250000 *directly* sampled drawings.

Table 4: Sampling results for the trimodal mixture distribution.

	bounds ARDS		initialization		ARMS		ARIS		MH		IS		true	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
x_1	-20.00	20.00	6.00	10.00	7.03	4.41	7.07	4.40	1.83	0.04	7.44	0.11	7.00	4.42
x_2	-20.00	20.00	6.00	10.00	7.54	3.38	7.52	3.43	2.57	0.04	8.10	0.15	7.50	3.40
x_3	-20.00	20.00	6.00	10.00	8.03	2.73	7.98	2.70	4.09	0.35	10.21	0.08	8.00	2.74
x_4	-20.00	20.00	6.00	10.00	8.54	2.70	8.47	2.73	6.01	0.39	11.21	0.40	8.50	2.74
x_5	-20.00	20.00	6.00	10.00	4.99	2.71	5.06	2.70	6.13	0.45	2.96	0.91	5.00	2.74
x_6	-20.00	20.00	6.00	10.00	5.44	2.77	5.49	2.72	8.91	0.62	2.14	0.61	5.50	2.74
x_7	-20.00	20.00	6.00	10.00	6.07	3.38	5.96	3.41	10.21	0.32	4.24	0.48	6.00	3.40
x_8	-20.00	20.00	6.00	10.00	6.48	4.39	6.44	4.46	12.01	1.18	5.14	0.85	6.50	4.42
drawings per iteration ($\eta \times \rho$)					5000 \times 5		5000 \times 5		250000		250000			
number of iterations					8		5		8		8			
average time per iteration (in s)					112.9		82.0		110.4		93.9			
Mahalanobis distance					0.01		0.01		1.57		10.68			
acceptance rate (in %)					27.8				0.1					
5% most influential weights (in %)							29.2				99.7			

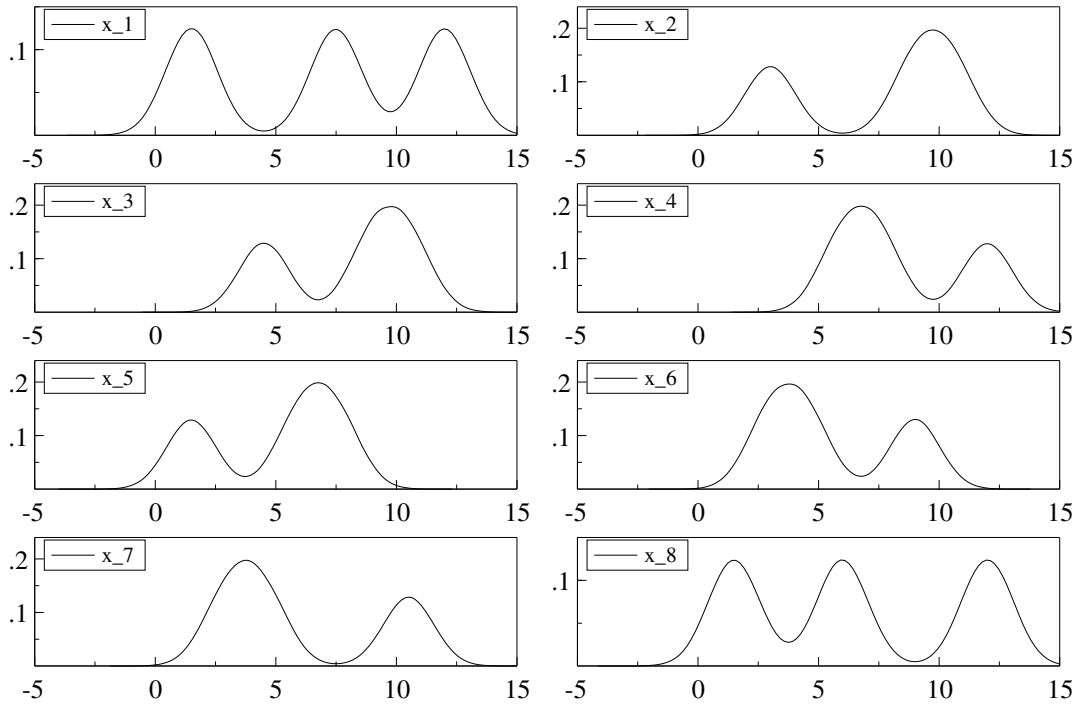


Figure 5: Marginal densities for the trimodal mixture distribution.

In order to see whether the favourable results are merely a coincidence, we repeat the experiment above 10 times for different seeds of the random number generator. The results are robust: ARMS succeeds 9 times and ARIS even succeeds for all 10 repetitions, whereas MH and importance sampling only have success rates of 30%, making the outcome of the latter two methods very unreliable.

3.3 Scale contamination

In Justel and Peña (1996), the authors investigate a data set from Brownlee (1965, pp. 491–500) concerning the oxidation of ammonia to nitric acid in a plant. The data set incorporates 21 daily observations on four variables. The stack loss rate y , that is the proportion of ingoing ammonia to the plant that escapes unabsorbed, is related to the amount of air flow x_1 (representing the rate of operation of the plant), the temperature of the cooling water x_2 , and the concentration of the circulating acid x_3 . In several investigations it was found that several observations might be classified as outliers, and therefore care should be taken in the analysis to allow for this. In a regression setting, it is sufficient to allow for scale contamination, as in the model

$$\begin{aligned}
 y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i \\
 \varepsilon_i &\sim N(0, \sigma_i^2) \\
 \sigma_i &= \begin{cases} \sigma & \text{with probability } 1 - p \\ \kappa\sigma & \text{with probability } p \end{cases}
 \end{aligned}$$

For identification of the two variance regimes, we impose that $\kappa > 1$. The prior for the parameter vector $(\beta_1, \beta_2, \beta_3, \sigma, \kappa, p)'$ is chosen proportional to $[(1 - p)\sigma + p\kappa\sigma]^{-1}$ within the parameter bounds reported in Table 5. The prior density is zero outside these bounds.

The six model parameters are sampled using ARMS, ARIS, MH and importance sampling. The sampling setup is the same as in the previous example. So, we consider an adaptive approach in which additional sampling rounds are considered as long as the Mahalanobis distance is larger than 0.02 with a maximum of 8 rounds. In any round, ARMS and ARIS sample 5000 directions and 5 distances per direction, whereas the MH and importance sampling algorithms collect a sample of size 250000.

The parameter estimates are reported in Table 5, together with the corresponding large sample values (computed from 250000 ARMS drawings). It is seen that all four methods trace back the response parameters β_1 , β_2 and β_3 accurately. However, this is not the case for the remaining three parameters. In particular, MH and importance sampling underestimate the left mode of the bimodal marginal density for σ , resulting in an overestimation of the mean of σ . This is also illustrated in Figure 6 in which the large sample marginal densities (250000 ARMS drawings) and the estimated densities from 25000 ARMS drawings and from 250000 MH drawings are shown. Furthermore, it is seen that MH and importance sampling underestimate the right mode of the bimodal marginal density for the mixture probability p , resulting in an underestimation of the mean of p . We note from Table 5 that the average times per sampling round are comparable for the four considered algorithms, and that they have all converged within 4 or 5 rounds.

The difference between the acceptance rates of ARMS and MH is striking. The acceptance rate of the former is about five times as high as the acceptance rate of the latter. Further,

Table 5: Sampling results for the scale contamination model.

	bounds ARDS		initialization		ARMS		ARIS		MH		IS		large sample	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
β_1	-10.00	10.00	0.00	2.00	0.81	0.19	0.81	0.19	0.81	0.20	0.81	0.20	0.81	0.19
β_2	-10.00	10.00	0.00	2.00	1.01	0.52	1.00	0.55	1.01	0.58	1.02	0.57	1.01	0.54
β_3	-10.00	10.00	0.00	2.00	-0.61	0.09	-0.61	0.10	-0.61	0.10	-0.61	0.10	-0.61	0.10
σ	0.00	10.00	5.00	2.00	2.84	1.32	2.77	1.38	3.06	1.38	3.09	1.39	2.82	1.36
κ	1.00	10.00	5.00	2.00	3.40	2.35	3.48	2.39	3.60	2.48	3.70	2.58	3.48	2.40
p	0.00	1.00	0.50	0.20	0.47	0.34	0.51	0.34	0.43	0.34	0.43	0.34	0.49	0.34
drawings per iteration ($\eta \times \rho$)					5000 \times 5		5000 \times 5		250000		250000			
number of iterations					5		4		4		4			
average time per iteration (in s)					39.8		37.6		43.1		47.0			
Mahalanobis distance					0.02		0.00		0.01		0.01			
acceptance rate (in %)					58.2				11.4					
5% most influential weights (in %)							19.6				62.8			

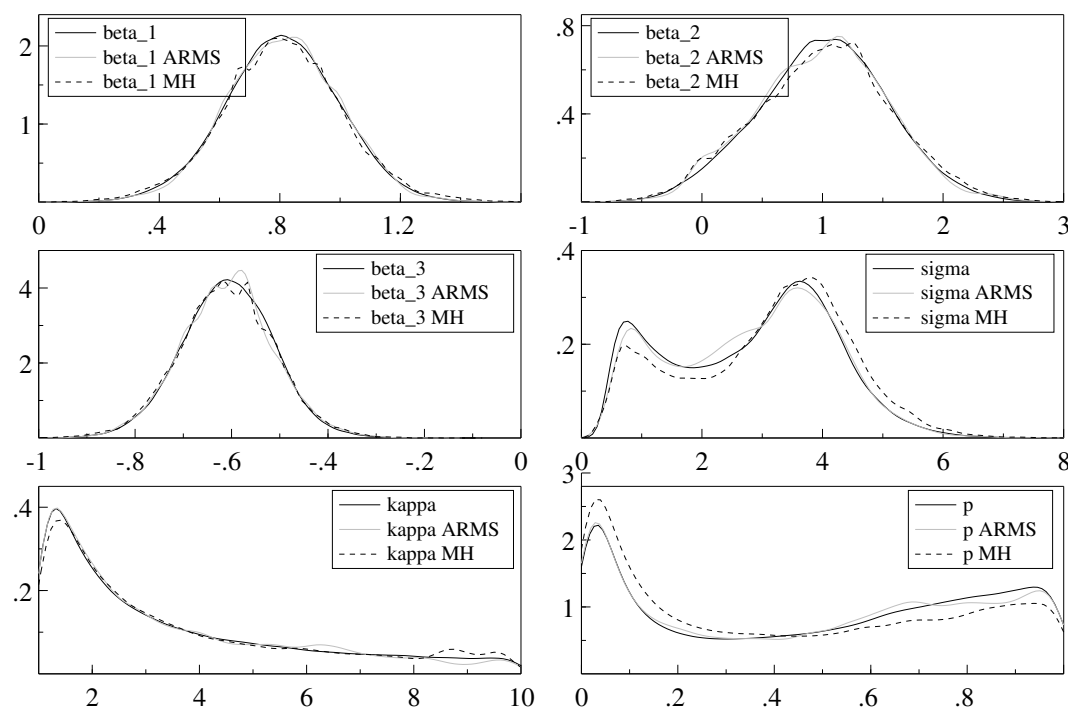


Figure 6: Marginal densities for the scale contamination model.

the distribution of the importance weights for ARIS is much more uniform than the weight distribution for importance sampling. In the last round of the importance sampling procedure, the 5% most influential drawings have 63% of the total weight, whereas this is only about 20% in ARIS. Again, this demonstrates the accuracy of the ARDS algorithms.

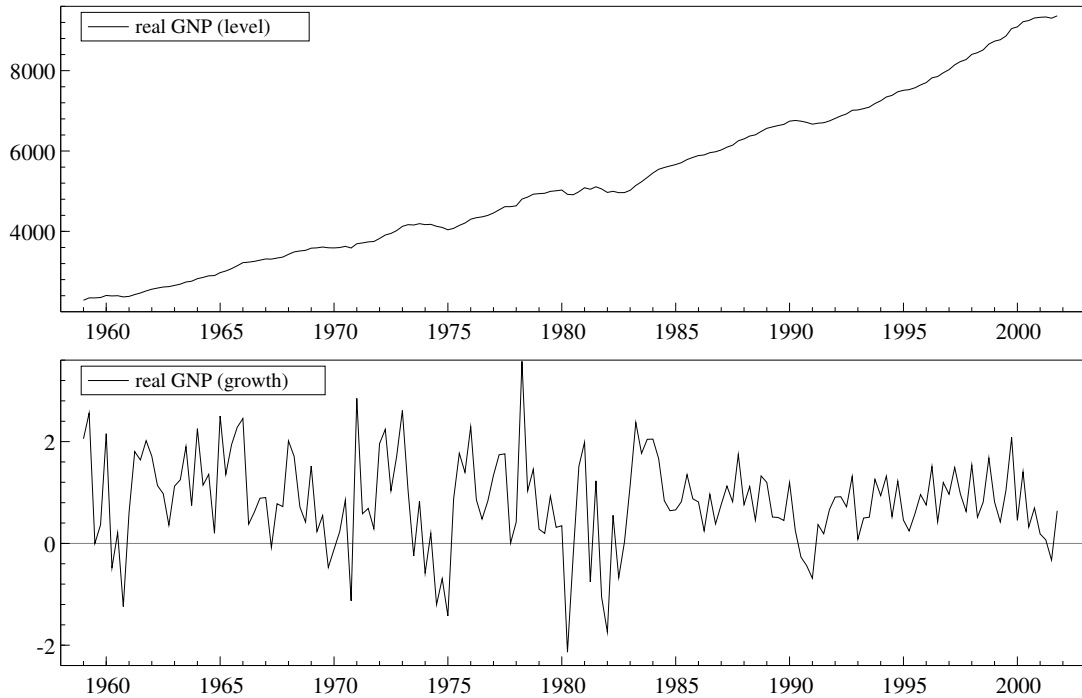


Figure 7: Real GNP of the United States in billions of dollars (above), and its quarterly growth rate in % (below).

3.4 A mixture model for the U.S. GNP growth rate

In models for the growth rate of the gross national product, great advances have been made by allowing for separate regimes in periods of recession and expansion. However, these models give rise to difficulties with respect to convergence of sampling methods due to multiple modes. As an illustration of our algorithms we consider a mixture model with two AR(1) regimes for real GNP growth. This model is similar to the model considered in Frühwirth-Schnatter (2001) where another recent sampling method is discussed.⁵ The model reads:

$$y_t = \begin{cases} \beta_{11} + \beta_{12}y_{t-1} + \varepsilon_t & \text{with probability } p \\ \beta_{21} + \beta_{22}y_{t-1} + \varepsilon_t & \text{with probability } 1 - p \end{cases}$$

$$\varepsilon_t \sim N(0, \sigma^2) \tag{29}$$

where y_t denotes the quarterly growth rate. We investigate data concerning U.S. real GNP (source: Economagic). The data consists of observations from the first quarter of 1959 to the last quarter of 2001. Figure 7 displays both the real GNP level and the quarterly growth rate

⁵Actually, Frühwirth-Schnatter (2001) also allows the variances to differ between regimes.

Table 6: Sampling results for the two-regime mixture model for U.S. real GNP.

	bounds ARDS		initialization		ARMS		ARIS		MH		IS		large sample	
	min.	max.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
β_{11}	-4.00	4.00	0.00	0.50	0.11	0.64	0.10	0.59	-0.14	0.88	0.01	0.72	0.07	0.70
β_{12}	-1.00	1.00	0.00	0.50	0.45	0.24	0.45	0.25	0.42	0.28	0.40	0.28	0.41	0.27
β_{21}	-4.00	4.00	1.00	0.50	1.32	0.74	1.27	0.78	1.22	0.83	1.28	0.85	1.30	0.79
β_{22}	-1.00	1.00	0.00	0.50	-0.07	0.39	-0.02	0.38	0.05	0.39	0.01	0.40	-0.04	0.41
σ	0.00	2.00	1.00	0.50	0.82	0.05	0.82	0.06	0.82	0.06	0.82	0.06	0.82	0.06
p	0.00	1.00	0.50	0.20	0.59	0.38	0.53	0.38	0.48	0.39	0.52	0.39	0.55	0.38
drawings per iteration ($\eta \times \rho$)					5000 \times 5		5000 \times 5		250000		250000			
number of iterations					8		5		8		8			
average time per iteration (in <i>s</i>)					81.4		82.9		89.6		85.1			
Mahalanobis distance					0.04		0.02		0.20		0.15			
acceptance rate (in %)					17.6				1.2					
5% most influential weights (in %)							57.9				99.7			

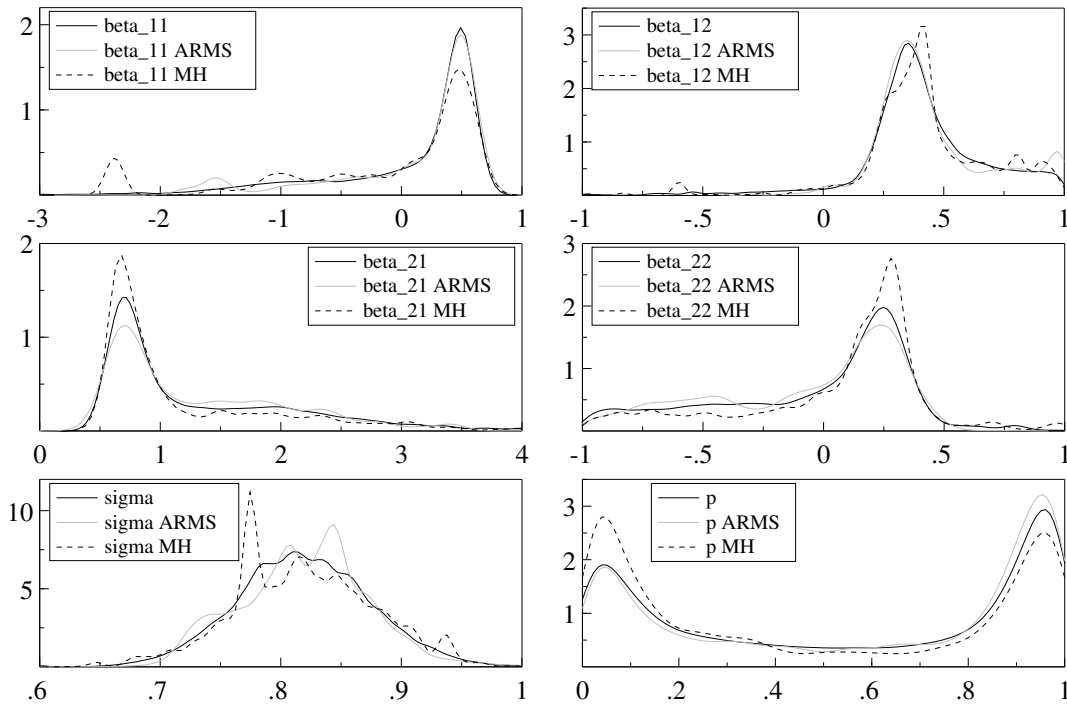


Figure 8: Marginal densities for the two-regime mixture model for U.S. real GNP.

(defined as 100 times the first difference of the logarithm). The priors for β_{11} , β_{12} , β_{21} , β_{22} and p are chosen uniform, and the prior for σ is taken proportional to $1/\sigma$. For identification, it is imposed that $\beta_{11} < \beta_{21}$. Again parameter bounds apply, see Table 6. The sampling setup is identical to the setup in the previously considered scale contamination example.

The parameter estimates are reported in Table 6, together with the corresponding large sample values (computed from 250000 ARMS drawings). Further, Figure 8 shows the large sample marginal densities, and the densities estimated from 25000 ARMS drawings and from 250000 MH drawings. It is seen that the large sample densities (obtained from 250000 ARMS drawings) are smooth, while this is not always the case for the “small sample” densities. However, we note that the “small sample” MH densities have been constructed from the same number of drawings as the large sample ARMS densities. In general, it is seen that the ARMS, ARIS and importance sampling estimates are quite close to each other, but the MH estimates are sometimes quite different. In particular, this is illustrated in the density of the mixture probability p : MH overestimates the left mode and underestimates the right mode. The average times per sampling round are comparable for the four considered algorithms, but only ARIS converged within 8 rounds, although the Mahalanobis distance for ARMS is also quite small. The acceptance rates (ARMS and MH) and the total weights of the 5% most influential drawings (ARIS and importance sampling) again provide support for the ARDS algorithms.

4 Conclusions

We have extended the Metropolis-Hastings and importance sampling methods by applying a radial transformation to the parameter space of the posterior (or target) density. Sampling does not take place in the m -dimensional parameter space directly, but in an $(m - 1)$ -dimensional subspace of directions. The final dimension, which corresponds to a distance measure, is sampled exactly from the target density (conditional on the directions), using the inverse transformation method. In this way the shape of the posterior density is taken into account perfectly along the sampled directions. For a given number of draws, this approach requires more functional evaluations of the posterior density than a traditional MH or IS algorithm. The usual type of tradeoff occurs: with a more sophisticated algorithm, one can hope to get ‘correct’ results with less draws than with a less sophisticated algorithm. It may also happen that a simple method cannot deliver reliable results. It would however be surprising when ARDS cannot deliver good results while the simpler, less computer intensive methods, can. This is confirmed by the (empirical) illustrations in Section 3. Moreover, a possible use of the ARDS algorithms is as a preliminary step to explore the posterior distribution and prepare a more sophisticated method.

We emphasize that there is no claim that ARDS algorithms are superior in theory to other kinds of algorithms (such a claim would make no sense). We believe that for any model/data combination, a sufficient research effort will usually allow to find a specific algorithm that performs better than ARDS or other algorithms. However, this is not necessarily guaranteed, and the specific algorithm may not be better even for a different data set (with the same model).

An interesting extension of this paper would be to embed ARMS in a Gibbs algorithm, where a subset of the parameters can be directly simulated from their conditional distribution, while the remaining parameters cannot. In this framework, special care should be given to

start sampling with sufficiently good initial guesses of the location and scale of the conditional distribution to be simulated. Examples where such an algorithm may be of great potential efficiency are in the Bayesian analysis of a linear simultaneous equation model, where the so-called simultaneity parameters induce a very nonelliptical shape of the posterior; a cointegration model, and a state space model as in Koop and Van Dijk (2000).

Appendix: Proof of Proposition 1

First, given (16) and (17), the target density $p(x)$ in terms of ρ and η is given by

$$p(\rho, \eta) = p(x(\rho, \eta) | \mu, \Sigma) |J_x(\rho, \eta)| = p(x(\rho, \eta) | \mu, \Sigma) |J_y(\rho)| J_y(\eta) \det(\Sigma^{1/2}), \quad (30)$$

implying that

$$\begin{aligned} p(\eta) &= \int_{-\infty}^{\infty} p(\rho, \eta) d\rho = \int_{-\infty}^{\infty} p(x(\rho, \eta) | \mu, \Sigma) |J_y(\rho)| J_y(\eta) \det(\Sigma^{1/2}) d\rho \\ &\propto J_y(\eta) \int_{-\infty}^{\infty} p(x(\rho, \eta) | \mu, \Sigma) |J_y(\rho)| d\rho \end{aligned} \quad (31)$$

(the last expression being a kernel of the marginal target density of η). Second, the normal candidate density, denoted by $q(x)$, becomes the following function in terms of ρ and η :

$$\begin{aligned} q(\rho, \eta) &= q(x(y(\rho, \eta) | \mu, \Sigma) | J_x(\rho, \eta)| \\ &\propto \exp\left(-\frac{1}{2}\rho^2\right) |J_y(\rho)| J_y(\eta), \end{aligned} \quad (32)$$

so that⁶

$$q(\eta) \propto J_y(\eta). \quad (33)$$

It follows from (31) and (33) that the acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, defined for an independence chain as

$$\alpha(\eta_{i-1}, \eta_i^*) = \min\left\{\frac{\frac{p(\eta_i^*)}{q(\eta_i^*)}}{\frac{p(\eta_{i-1})}{q(\eta_{i-1})}}, 1\right\}, \quad (34)$$

simplifies to the expression in (19). Further, it follows from (30) that the density of ρ conditional on η , that is $p(\rho|\eta) \propto p(\rho, \eta)$, is given by (21).

References

- BAUWENS, L., C. S. BOS, H. K. VAN DIJK, AND R. D. VAN OEST (2002): “Adaptive Polar Sampling: a class of flexible and robust Monte Carlo integration methods,” Report EI 2002-27, Econometric Institute, Erasmus University Rotterdam.
- BOX, G., AND M. MULLER (1958): “A note on the generation of random normal deviates,” *Annals of Mathematical Statistics*, 29, 610–611.

⁶If the candidate is elliptically contoured, ρ and η stay independent and the marginal distribution of η is the same as in (33), but the distribution of ρ is different from what appears in (32).

- BROWNLEE, K. A. (1965): *Statistical Theory and Methodology in Science and Engineering*. Wiley, New York, 2 edn.
- DOORNIK, J. A. (1999): *Object-Oriented Matrix Programming using Ox*. Timberlake Consultants Ltd, London, 3rd edn., See <http://www.nuff.ox.ac.uk/Users/Doornik>.
- FRÜHWIRTH-SCHNATTER, S. (2001): “Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching Models,” *Journal of the American Statistical Association*, 96(453), 194–209.
- GEWEKE, J. (1989): “Bayesian inference in econometric models using Monte Carlo integration,” *Econometrica*, 57, 1317–1339.
- (1999): “Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication,” *Econometric Reviews*, 18(1), 1–73.
- GILKS, W., AND G. ROBERTS (1996): “Strategies for Improving MCMC,” in *Markov Chain Monte Carlo in Practice*, ed. by R. S. Gilks, W.R., and D. Spiegelhalter. Chapman and Hall/CRC.
- GILKS, W., G. ROBERTS, AND E. GEORGE (1994): “Adaptive direction sampling,” *The Statistician*, 43, 179–189.
- GIVENS, G. H., AND A. E. RAFTERY (1996): “Local Adaptive Importance Sampling for Multivariate Densities With Strong Nonlinear Relationships,” *Journal of the American Statistical Association*, 91(433), 132–141.
- HOBERT, J. P., AND G. CASELLA (1996): “The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models,” *Journal of the American Statistical Association*, 91(436), 1461–1473.
- HOP, J., AND H. K. VAN DIJK (1992): “SISAM and MIXIN: Two algorithms for the computation of posterior moments and densities using Monte Carlo integration,” *Computer Science in Economics and Management*, 5, 183–220.
- JUSTEL, A., AND D. PEÑA (1996): “Gibbs Sampling Will Fail in Outlier Problems with Strong Masking,” *Journal of Computational & Graphical Statistics*, 5(2), 176–189.
- KLEIBERGEN, F. R., AND H. K. VAN DIJK (1994): “On the Shape of the Likelihood/Posterior in Cointegration Models,” *Econometric Theory*, 10(3-4), 514–551.
- (1998): “Bayesian Simultaneous Equations Analysis using Reduced Rank Structures,” *Econometric Theory*, 14(6), 701–743.
- KOOP, G., AND H. K. VAN DIJK (2000): “Testing for Integration using Evolving Trend and Seasonal Models: A Bayesian Approach,” *Journal of Econometrics*, 97, 261–291.
- MONAHAN, J., AND A. GENZ (1997): “Spherical-radial integration rules for Bayesian computation,” *Journal of the American Statistical Association*, 92, 664–6674.
- MUIRHEAD, R. (1982): *Aspects of Multivariate Statistical Theory*. Wiley, New York.

- OH, M. S., AND J. O. BERGER (1992): “Adaptive Importance Sampling in Monte Carlo Integration,” *Journal of Statistical Computation and Simulation*, 41, 143–168.
- RUBINSTEIN, R. (1981): *Simulation and the Monte Carlo Method*. Wiley, New York.
- SCHMAISER, B., AND M.-H. CHEN (1991): “General Hit-and-run Monte Carlo Sampling for Evaluating Multidimensional Integrals,” Discussion paper, School of Industrial Engineering, Purdue University.
- VAN DIJK, H. K., AND T. KLOEK (1980): “Further experience in Bayesian analysis using Monte Carlo integration,” *Journal of Econometrics*, 14, 307–328.
- VAN DIJK, H. K., AND T. KLOEK (1983): “Monte Carlo analysis of skew posterior distributions: An econometric example,” *The Statistician*, 32, 216–223.
- (1984): “Experiments with some alternatives for simple importance sampling in Monte Carlo integration,” in *Bayesian Statistics 2*, ed. by J. M. Bernardo, M. Degroot, D. Lindley, and A. F. M. Smith, Amsterdam. North Holland.
- VAN DIJK, H. K., T. KLOEK, AND C. G. E. BOENDER (1985): “Posterior Moments Computed by Mixed Integration,” *Journal of Econometrics*, 29, 3–18.