EXPLAINING ADAPTIVE RADIAL-BASED DIRECTION SAMPLING

Econometric Institute Report EI 2003-37

Luc Bauwens, Charles S. Bos, Herman K. van Dijk, and Rutger D. van Oest

CORE and Department of Economics, Université catholique de Louvain
Tinbergen Institute and D. of Econometrics & O.R., Free University Amsterdam
Econometric and Tinbergen Institutes, Erasmus University Rotterdam
Econometric and Tinbergen Institutes, Erasmus University Rotterdam

Keywords: Markov Chain Monte Carlo, importance sampling, radial coordinates

Introduction

In recent decades Monte Carlo methods, such as Metropolis-Hastings [MH] (Metropolis et al. 1953, and Hastings 1970), Gibbs sampling (Geman and Geman 1984) and importance sampling, have been applied extensively and successfully within Bayesian analyses of statistical and econometric models. However, although Monte Carlo methods revolutionized the applicability of Bayesian inference, the literature indicates a substantial variation in their convergence behavior. For instance, the performance of the Gibbs sampler may be hampered by strong correlation in the target (or posterior) distribution. A multimodal target density may pose problems to all methods. A common difficulty encountered in all samplers is the choice of an importance or candidate density when little is known about the shape of the target density.

In order to sample effectively from ill behaved target densities, we have proposed the class of adaptive radial-based direction sampling (ARDS) methods, see Bauwens et al. (2003). 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*, that is, they can deal with a large variety of features of target distributions, in particular multimodality, strong correlation, extreme skewness and heavy tails. Thirdly, the algorithms can deal with multiple *linear inequality conditions* on the parameter space without any additional complications for the implementation.

The ARDS algorithms can be characterized as 'black-box' samplers (Chen et al. 2000). They extend earlier methods like the method of Box and Muller (1958), the adaptive direction sampling algorithms proposed by Gilks et al. (1994), the mixed integration method by Van Dijk et al. (1985), and the spherical integration method by Monahan and Genz (1997).

In this paper we briefly outline the computational steps of ARDS. For an in-depth discussion, the reader is referred to Bauwens et al. (2003).

Overview of ARDS

Most simulation algorithms for posterior distributions generate random drawings in the original parameter space. However, several researchers advocate to simulate in a transformed space, where the simulation is more efficient in some sense. For example, if there exists strong correlation between two random variables, an orthogonalizing transformation reduces serial dependence in a Gibbs sampling scheme. The ARDS algorithms rely on this general idea. They are based on a composite transformation to radial coordinates. For expository purpose, we treat this transformation explicitly in two steps. The first step is an m-dimensional location-scale transformation from x to y, which aims at standardizing the candidate density with respect to the location, scale and correlations of the target density. The second step is a radial transformation, which maps an m-dimensional point y to an $(m-1)$-dimensional direction η and a one-dimensional distance ρ. After applying this composite transformation to a candidate drawing x from the normal distribution (or any other elliptically-contoured distribution), the direction η is evaluated by either an MH step or an importance sampling step. Next, a new distance ρ is generated, conditional on the direction η, from the exact target density by means of the numerical inverse transformation method. Finally, the resulting

direction and distance are transformed to the original space by inverting the composite radial transformation. This procedure is repeated in many iterations.

The ARDS algorithms have a natural interpretation. Basically, in one iteration, first a direction η, which uniquely determines a line in the m-dimensional original parameter space, is generated from an elliptically-contoured *candidate* density. This line is either accepted or rejected in an MH step, or it is assigned an importance weight. Next, a distance ρ, which uniquely determines the position of one point on the line, is directly generated from the *target* distribution. As sampling on any given line mimics exactly the target density, ARDS should result in robust Monte Carlo estimates. We refer to the MH variant of ARDS as adaptive radial-based Metropolis-Hastings sampling (ARMHS), while we refer to the importance sampling variant as adaptive radial-based importance sampling (ARIS).

In ARDS, an adaptive procedure is considered to generate more efficient directions, with higher acceptance rates or a more uniform weight distribution. In this adaptive procedure, the location and covariance matrix of the candidate density are updated by replacing them by their Monte Carlo estimates. These Monte Carlo estimates are obtained from a previous run of the algorithm. So, ARDS is adaptive in the sense that it aims to improve the approximation of the candidate distribution to the target distribution by considering several sampling rounds.

The radial transformation

The radial transformation is at the heart of the ARDS algorithms. In a general form, the radial transformation from y to (ρ, η) may be characterized by its inverse transformation

$$\begin{aligned} y_1 &= \rho\, h_1(\eta_1, \ldots, \eta_{m-1}), \\ &\vdots \\ y_m &= \rho\, h_m(\eta_1, \ldots, \eta_{m-1}), \end{aligned} \tag{1}$$

where $h(\eta) = (h_1(\eta), \ldots, h_m(\eta))$ are differentiable functions, see for example Muirhead (1982, Section 1.5). The radial transformation maps the Cartesian coordinates y to a position $h(\eta)$ on the unit circle and a stretching factor ρ which further determines the position of y given $h(\eta)$. It turns out that ARDS can be applied to any transformation satisfying (1).

A well-known special case of the general transformation described above is the two-dimensional polar transformation from $y \in R^2$ to $(\rho, \eta) \in R^+ \times (0, 2\pi)$. Figure 1 illustrates the relationship between Cartesian coordinates and polar coordinates. A feature of the polar transformation is that $\rho \in R^+$, implying that the direction η only defines one half of a line, and not an entire line. However, in ARDS the distances ρ are sampled from the target density on *entire* lines, so that information from the *whole* target density is considered given a direction η. Two other drawbacks of using the polar transformation in ARDS are the following. Firstly, it is possible but not straightforward to generalize it to $m > 2$ dimensions. Secondly, the transformation is computationally not very efficient. We therefore propose a transformation which satisfies (1), is easy to generalize to more than two dimensions, is more efficient than the polar transformation, and which is such that the direction η does define an entire line. The proposed transformation maps y to $m-1$ Cartesian coordinates on the unit circle and a stretching factor ρ. This is illustrated in Figure 2 for the two-dimensional case. More details are given in Bauwens et al. (2003, Subsection 2.1).

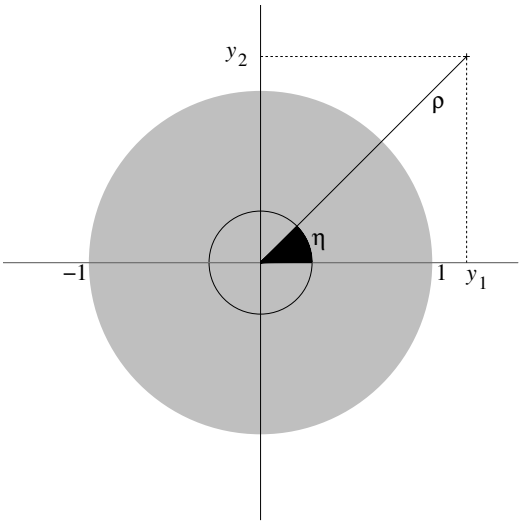

Figure 1: The relationship between Cartesian coordinates and polar coordinates in the two-dimensional case.

One iteration of ARDS

An iteration of ARDS consists of 5 steps.

Step 1 of the i-th iteration: A candidate drawing x_i^* is generated from a normal candidate distribution with mean μ and covariance matrix Σ.

Step 2: This drawing x_i^* is transformed to a candidate direction η_i^* and a candidate distance ρ_i^* by applying the composite radial transformation.

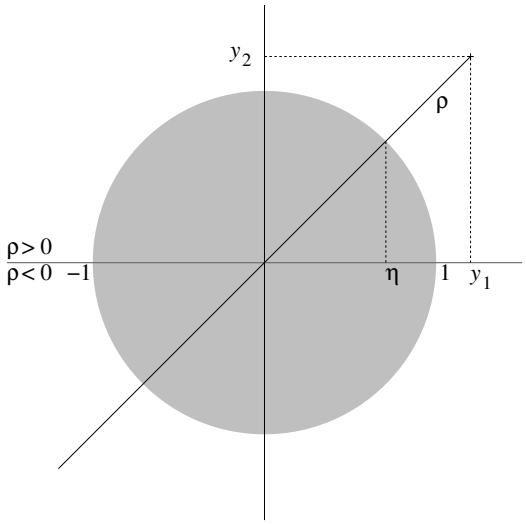

Figure 2: The relationship between Cartesian coordinates and radial coordinates in the two-dimensional case.

Step 3: An MH step or an importance sampling is applied to η_i^*, resulting in the direction η_i. In ARMHS, this direction is given by

$$\eta_i = \begin{cases} \eta_i^* & \text{with probability } \alpha(\eta_{i-1}, \eta_i^*) \\ \eta_{i-1} & \text{with probability } 1 - \alpha(\eta_{i-1}, \eta_i^*) \end{cases} \tag{2}$$

for some acceptance probability $\alpha(\eta_{i-1}, \eta_i^*)$, which is defined below. In ARIS, this direction $\eta_i = \eta_i^*$ is assigned importance weight $w(\eta_i)$, which is also defined below.

Step 4: A distance ρ_i is sampled from $p(\rho|\eta_i)$, which is the target density of ρ conditional on the generated direction η_i.

Step 5: The resulting direction η_i and distance ρ_i are transformed to the original parameter space by inverting the composite radial transformation. This yields one effective realization x_i from the target density.

It can be shown that the acceptance probability in the MH step of ARMHS is given by

$$\alpha(\eta_{i-1}, \eta_i^*) = \min\left\{ \frac{w(\eta_i^*)}{w(\eta_{i-1})}, 1 \right\}, \tag{3}$$

where the importance function $w(\eta)$, which is also considered in ARIS, is defined by

$$w(\eta) = \int_{-\infty}^{\infty} \kappa(\rho|\eta)\, d\rho, \tag{4}$$

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

$$p(\rho|\eta) \propto \kappa(\rho|\eta) = p(x(\rho, \eta|\, \mu, \Sigma))\, |\rho^{m-1}|. \tag{5}$$

Here $x(\rho, \eta|\, \mu, \Sigma)$ results from inverting the composite radial transformation for the arguments ρ and η.

In order to obtain the direction η_i, both in ARMHS and ARIS, the importance function $w(\eta)$ has to be evaluated. Its value can be computed using a deterministic integration rule, such as the adaptive Simpson's rule. As the density of ρ conditional on η is proportional to the integrand of $w(\eta)$, evaluations of this integrand, gathered during the deterministic integration phase, can be used to construct a grid for $p(\rho|\eta)$. So, using the numerical inverse transformation method, sampling the distance ρ conditional on the direction η is straightforward. We note that one can capitalize on the constructed grid for $p(\rho|\eta)$ by generating several drawings of ρ for each drawing of η.

The integral $w(\eta)$ has infinite integration bounds. However, in practice finite integration bounds have to be considered for its numerical evaluation. In order to obtain these finite bounds, we propose that minimum and maximum values are imposed for each parameter in the original parameter space. It is often possible to find sensible bounds, such that practically all density mass of ρ given η is included, by either theory and/or common sense. We note that our method is in particular suited to deal with multiple linear inequality restrictions, which sometimes have to be imposed on the parameters. These linear restrictions reduce the feasible region and hence the integration interval of $w(\eta)$, while they do not imply any additional complications for the implementation of ARDS. So, linear restrictions on the parameter space do not put a burden on the algorithm, but they might result in an efficiency gain.

An illustration

In order to illustrate the performance of ARDS, 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

$$\begin{aligned} \mu_1 &= 1.5\,(1,2,3,4,5,6,7,8)', \\ \mu_2 &= 1.5\,(5,6,7,8,1,2,3,4)', \\ \mu_3 &= 1.5\,(8,7,6,5,4,3,2,1)', \\ \Sigma_1 &= \Sigma_2 = \Sigma_3 = I_8, \\ p_1 &= p_2 = p_3 = 1/3. \end{aligned}$$

The marginal densities[1] for the eight components are displayed in Figure 3. It is seen

[1]These marginal densities are constructed using 250000 *directly* sampled drawings.

that the target distribution is ill behaved. We estimate the first and second moments for this distribution using ARMHS, 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, in the experiment we allow for at most 8 rounds. In any round, ARMHS and ARIS draw 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 density 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 target density.

The estimates of the location and scaling parameters are reported in Table 1. Furthermore, the table also contains the "true" parameter values, obtained from 250000 *directly* sampled drawings. It is seen that ARMHS and ARIS do a very good job, whereas MH and importance sampling fail. ARIS only needs 5 rounds to reach convergence (according to our definition), whereas the other three algorithms need the maximum number of rounds. However, after eight rounds also ARMHS 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 ARMHS 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 29.2% of the total weight.

Finally, in order to see whether the favorable results are merely a coincidence, we repeat the experiment above 10 times for different seeds of the random number generator. The results are robust: ARMHS 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.

Contact information

Corresponding author: Rutger van Oest, Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands (e-mail `vanoest@few.eur.nl`, tel. +31-10-4088946, FAX +31-10-4089162). The working paper is available as report EI 2003-22 at

http://www.few.eur.nl/few/research/pubs/ei/2003/reports.htm

Acknowledgement

The authors are indebted to Michel Lubrano for very helpful comments.

References

Bauwens L., C.S. Bos, H.K. van Dijk and R.D. van Oest (2003), "Adaptive Radial-Based Direction Sampling: Some Flexible and Robust Monte Carlo Integration Methods", forthcoming in *Journal of Econometrics*.

Box G. and M. Muller (1958), "A Note on the Generation of Random Normal Deviates", *Annals of Mathematical Statistics*, 29, 610−611.

Chen M.-H., Q.-M. Shao and J.G. Ibrahim (2000), *Monte Carlo Methods in Bayesian Computation*, Springer-Verlag, New York.

Geman S. and D. Geman (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721−741.

Gilks W., G. Roberts and E. George (1994), "Adaptive Direction Sampling", *The Statistician*, 43, 179−189.

Hastings W.K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", "Biometrika", 57, 97−109.

Metropolis N. A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller (1953), "Equation of State Calculations by Fast Computing Machines", *The Journal of Chemical Physics*, 21, 1087−1092.

Monahan J. and A. Genz (1997), "Spherical-Radial Integration Rules for Bayesian Computation", *Journal of the American Statistical Association*, 92, 664−674.

Muirhead R. (1982), *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Van Dijk H.K., T. Kloek and C.G.E. Boender (1985), "Posterior Moments Computed by Mixed Integration", *Journal of Econometrics*, 29, 3−18.

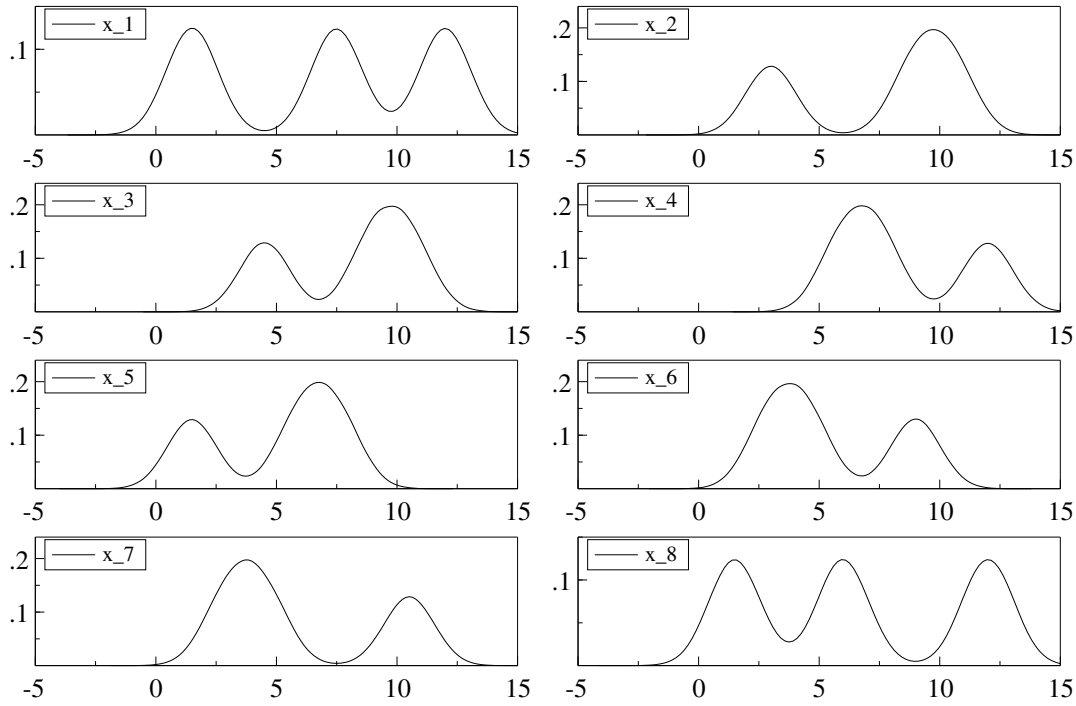

Figure 3: Marginal densities for the trimodal mixture distribution.

Table 1: Sampling results for the trimodal mixture distribution.

	bounds ARDS		initialization		ARMHS		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 ($\eta \times \rho$)					5000×5		5000×5		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			