ePub^WU Institutional Repository

Roman Karawatzki

The Multivariate Ahrens Sampling Method

Paper

http://epub.wu.ac.at/

The Multivariate Ahrens Sampling Method

Roman Karawatzki
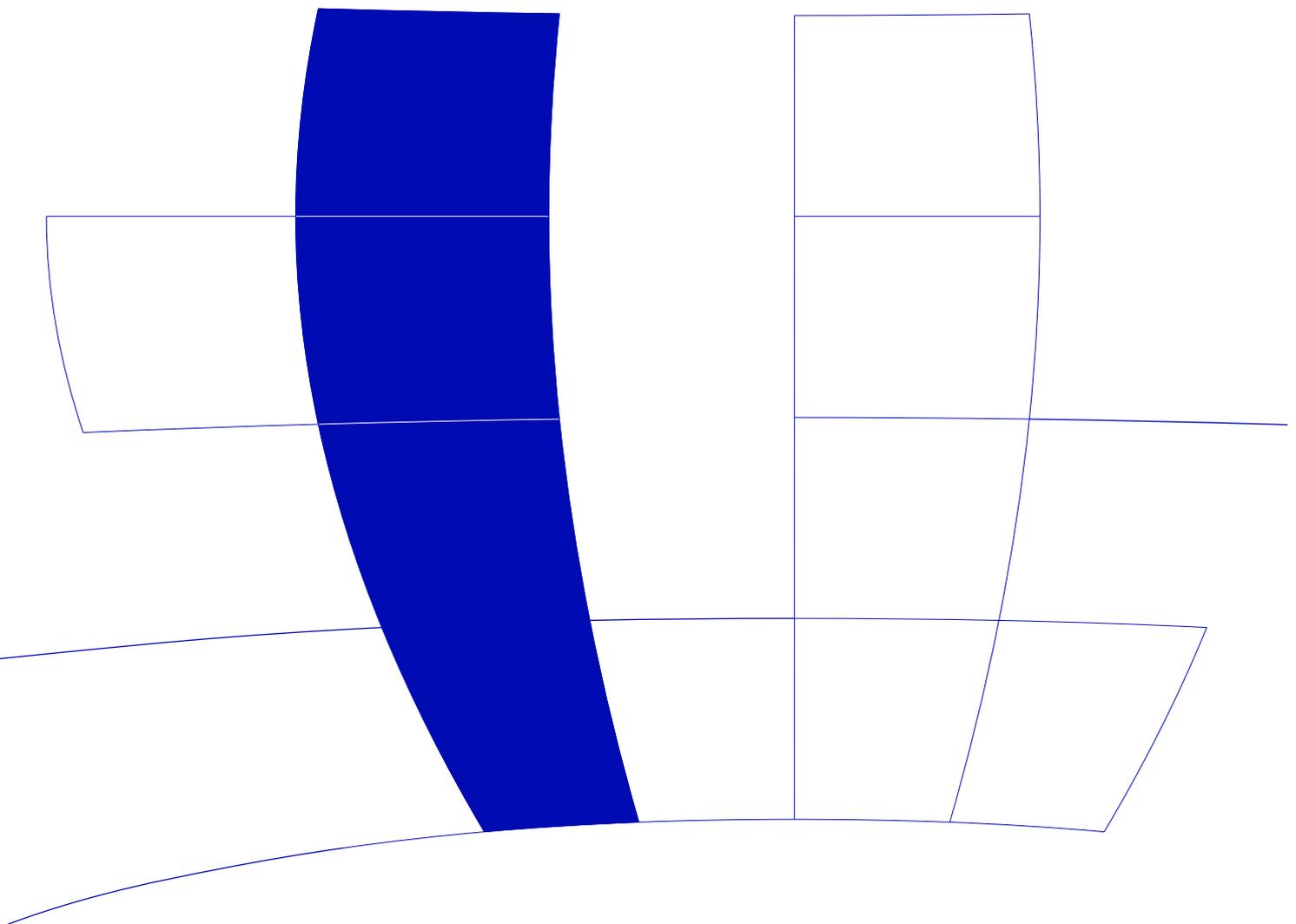
Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 30
March 2006

http://statmath.wu-wien.ac.at/

The Multivariate Ahrens Sampling Method

Roman Karawatzki

Department of Statistics and Mathematics,
University of Economics and Business Administration,
Augasse 2-6, A-1090 Vienna, Austria

Abstract

The "Ahrens method" is a very simple method for sampling from univariate distributions. It is based on rejection from piecewise constant hat functions. It can be applied analogously to the multivariate case where hat functions are used that are constant on rectangular domains. In this paper we investigate the case of distributions with so called orthounimodal densities. Technical implementation details as well as their practical limitations are discussed. The application to more general distributions is considered.

1   Introduction

There exist many methods to sample from a univariate distribution when only the probability density function (pdf) f is given (see [1, 2] for a survey). Among them the "Ahrens method" is one of the simplest, see [2, 3]. It is a rejection/acceptance method that uses piecewise constant hats and squeezes. On some interval $[b_l, b_r]$ the hat function is simply defined by the maximum of f and the squeeze by its minimum. For a monotone density these extrema are located on the boundary points b_l and b_r. The performance can be increased by subdividing the domain $[b_l, b_r]$ into N intervals $[a_0, a_1]$, $[a_1, a_2]$, ..., $[a_{N-1}, a_N]$ where the hat and squeeze is constructed on each interval which results in staircase-shaped hat and squeeze, see Fig. 1.

Thus for the setup we have to compute maximum h_j and minimum s_j for each subinterval $[a_{j-1}, a_j]$. The sampling algorithm is then quite simple:

1. Choose an interval $[a_{j-1}, a_j]$ at random with probability vector proportional to the areas below the hat, $(a_j - a_{j-1}) \cdot h_j$.
2. Generate a point X uniformly distributed in the selected interval $[a_{j-1}, a_j]$.
3. Generate a point $U \sim U(0, h_j)$.
4. If $U \leq s_j$ (below squeeze) accept and return X
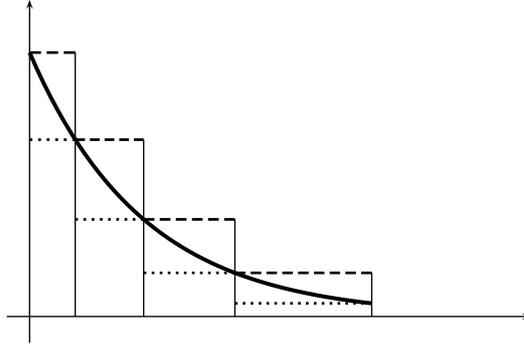
10 March 2006

Fig. 1. Ahrens method for the truncated exponential distribution. Probability density (solid line), staircase-shaped hat (dashed line) and squeeze (dotted line)

5. If $U \leq f(X)$ (below density) accept and return X.
6. Otherwise reject and try again.

Step 4 is optional but can speed up the algorithm when the evaluation of the density is expensive. Indeed, when the rejection constant is close to one the marginal generation time of this algorithm does hardly depend on the density f and is extremely fast. Step 1 can be done in constant time, i.e., independently of the number of subintervals by means of indexed search [4] or the alias method [5].

The Ahrens method can also be applied to an arbitrary distribution with given density provided that the domain can be split into intervals where f behaves monotonically (called *slopes* in [3]). Such slopes can be easily computed if the extrema of f are known. However, when the domain of the density is unbounded it must be truncated in such a way that the tail regions are computationally not relevant, i.e., the probability of falling into these tail regions must be much too small to happen in any long running simulation.

The design points $a_0, a_1, \ldots, a_N$ can be chosen such that the rejection constant (i.e., the ratio α between the area below the hat and the area below the density) is as small as desired. However, in practice it is often more convenient to look at the ratio ϱ between the area below the hat and the area below the squeeze since densities are often not normalized (e.g. for posterior densities in Bayesian inference) and the area below the density function is not available. There exist many approaches for the task of finding good construction points. Such methods are either deterministic and minimize the rejection constant [3, 6, 7] or the memory consumption [8], or use adaptive rejection sampling [9]. A very simple and powerful method is derandomized adaptive rejection sampling where iteratively those intervals are subdivided where the area between hat and squeeze is too large, see [2, §5.1.4].

The advantages of the Ahrens method are its extreme simplicity and its good

performance when the rejection constant is small. On the other hand its drawbacks are that it requires many subintervals when a small rejection constant is required (then $\alpha = 1 + \mathcal{O}(1/N)$). Moreover, without detailed knowledge of the positions of the maxima and minima of the density function, relatively time-intensive optimization problems must be solved.

The generation of multivariate random vectors from general distributions is much more difficult. Only a few methods exist, see [2, §11]. However, efficient sampling methods to obtain random points from large classes of multivariate probability density functions are especially important, as the applications thereof in Monte Carlo integration and other stochastic simulation problems are increasing.

It is evident that the Ahrens method can be generalized to multivariate distributions. However, optimization problems are much more difficult then and we need the right concept of a monotone function. Moreover, it is not clear how a domain can be subdivided such that the performance of the sampling algorithm is not too bad. For our task we will look at so called *orthomonotone* and *orthounimodal* densities. In Sect. 2 we describe these concepts and compile a multivariate Ahrens algorithm for this class of distributions. Computational experiences are reported in Sect. 3. At last we discuss the difficulties of applying the Ahrens method to more general distributions in Sect. 4.

2 Orthomonotone and orthounimodal densities

The Ahrens algorithm could be easily generalized to multivariate distributions in $\mathbb{R}^d$ with a rectangular domain $D = [b_{l1}, b_{r1}] \times \cdots \times [b_{ld}, b_{rd}]$: Split D into rectangular subdomains $R_j = [l_{j1}, u_{j1}] \times \cdots \times [l_{jd}, u_{jd}]$ and compute maximum h_j and minimum s_j in each of the hyperrectangles. Sampling is done analogously to the univariate case where a particular hyperrectangle R_j is chosen at random with probability vector proportional to the volumes of the bars $R_j \times [0, h_j]$. However, this approach has the drawback that we have to solve an optimization problem for each subdomain. Simply computing all maxima and minima of the density on the entire domain D and derive the location of the extrema in each subdomain as in the univariate case does not work in general. Thus we have to look at appropriate classes of multivariate distributions.

The notion of orthomonotone densities is a proper multivariate generalization of univariate monotone density functions. A density $f(\mathbf{x}) = f(x_1, \ldots, x_d)$ on a hyperrectangle $[b_{l1}, b_{r1}] \times \cdots \times [b_{ld}, b_{rd}]$ (or, more generally, on $\mathbb{R}^d$) is called *orthounimodal* if within each orthant (quadrant) defined by the mode $\mathbf{m} = (m_1, \ldots, m_d)$, the density is a monotone function of each of its arguments

individually [10]. The mode **m** partitions $\mathbb{R}^d$ into 2^d quadrants. On each quadrant, the density f is *orthomonotone*, i.e., $f(x_1, \ldots, x_d)$ is non-increasing in each of the d variables $\pm x_i$, with the 2^d sign combinations according to the position of the quadrant. The class of orthounimodal distributions is of particular interest as many densities of unimodal distributions can be transformed such that it is orthounimodal by some linear transformation.

Orthomonotone densities have the very important property that the maximum and minimum of the density in each hyperrectangle R_j are found at the vertices which have smallest and largest distance to mode **m**, respectively. This property enables us to construct hat and squeeze with relative ease.

In [2, §11.4.3] a sampling method for a bounded and orthomonotone quasi-density on the unit-cube $D = [0, 1]^d$ is presented, wherein the original cube is split into $N = N_1^d$ smaller cubes of equal size. The authors also propose an improved version using derandomized adaptive rejection sampling in which each cube is splitted into 2^d smaller cubes whenever the difference between the hat and squeeze is bigger than a certain threshold.

The main idea of the present strategy in the multivariate Ahrens method is to recursively split each hyperrectangle R_j into only **two** smaller cells.

We start with one single hyperrectangle R_1 covering the whole domain of an orthomonotone density – or alternatively – with 2^d hyperrectangles R_j decomposing the original domain into its orthants for an orthounimodal density.

Each subdomain is then splitted whenever the area between hat and squeeze is too large. This is realized by the following scheme: In each iteration compute the average volume B over all bars $R_j \times [s_j, h_j]$. If for such a bar the volume is above some threshold value τB, it is split along some direction by halving the length of a particular size (see also Fig. 2). The direction can be chosen at random or deterministically using some simple rule.

The threshold constant τ is introduced to overcome an initial deadlock and numerical issues in the case of a symmetric orthounimodal density, where the 2^d initial subdomains R_j all have the same bar-volume $R_j \times [s_j, h_j]$ being equal to the average bar volume B. This constant should be chosen to be $\tau < 1$. We have fixed its value to be $\tau = 0.9$ in all our runs.

Algorithm 1 compiles the general idea of our implementation of the multivariate Ahrens algorithm using standard abstract programming notation. The hyperrectangles R_j are stored by their left lower and right upper vertices, $(l_{j1}, \ldots, l_{jd})$ and $(u_{j1}, \ldots, u_{jd})$, maximum (hat) h_j and minimum (squeeze) s_j.

The algorithm can also be applied to densities with an arbitrary mode **m** in combination with an initial translation of **m** into the coordinate origin. For a

distribution with bounded but non-hyperrectangular domain the density has to extended to an hyperrectangle such that $f(\mathbf{x}) = 0$ for every $\mathbf{x}$ not in the original domain.

Algorithm 1 Multivariate Ahrens Method for Orthounimodal Densities

---

Input: Orthounimodal density f in hyperrectangle $D = [b_{l1}, b_{r1}] \times \cdots \times [b_{ld}, b_{rd}]$ with mode $\mathbf{m} = (0, \dots, 0)$; maximal number of subdomains N_D; minimal ratio ϱ; threshold constant $\tau < 1$.

Output: Random vector $\mathbf{X}$ with distribution f.

   [Setup]

 1: Initialize set of subdomains $\mathcal{R} = \{R_j\}$ by all 2^d orthants.

 2: Set number of vertices $N \leftarrow 2^d$.

 3: For each R_j compute volume $A_j = \text{Vol}_d(R_j)$, maximum h_j and minimum s_j (at vertex with smallest norm and largest norm, resp.), and volume of bars $V_j = A_j \cdot h_j$.

 4: **repeat**

 5:    Calculate the average volume between the hat and squeeze
      $B = \frac{1}{N} \sum_{j=1}^{N} A_j \cdot (h_j - s_j)$.

 6:    **for all** $R_j \in \mathcal{R}$ with $A_j \cdot (h_j - s_j) \geq \tau B$ **do**

 7:       Split hyperrectangle R_j along its longest side into two smaller subdomains, R'_j and R''_j.

 8:       Compute volumes A_j, V_j, maximum, and minimum of R'_j and R''_j.

 9:       Replace R_j by R'_j and R''_j in $\mathcal{R}$ and increment counter N.

10: **until** $N \geq N_D$ or $\sum_{j=1}^{N} A_j h_j / \sum_{j=1}^{N} A_j s_j \leq \varrho$

   [Sampling]

11: **loop**

12:    Generate J with probability vector proportional to $(V_1, \dots, V_N)$ (by indexed search).

13:    Generate a point $\mathbf{X}$ uniformly in hyperrectangle R_J.

14:    Generate $U \sim U(0, h_J)$.

15:    **if** $U \leq s_j$ **then** [below squeeze]

16:       Return $\mathbf{X}$.

17:    **if** $U \leq f(\mathbf{X})$ **then** [below pdf]

18:       Return $\mathbf{X}$.

The performance of this algorithm (the expected number of repetitions to get one random point) depends on the rejection constant α which converges to one with $\alpha = 1 + \mathcal{O}(N^{-1/d})$. (This follows from an argument similar to [2, Thm. 11.7].)

3 Computational experiences

Our numerical experiments with orthomonotone distributions covered the multinormal distribution

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{|\mathbf{\Sigma}|}}\, e^{-\frac{1}{2}\mathbf{x}^t\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{x}}$$

and the multicauchy distribution

$$f(\mathbf{x}) = \frac{\Gamma(\frac{d+1}{2})}{\pi^{\frac{d+1}{2}}\sqrt{|\mathbf{\Sigma}|}}\frac{1}{(1+\mathbf{x}^t\cdot\mathbf{\Sigma}^{-1}\cdot\mathbf{x})^{\frac{d+1}{2}}}$$

3.1 Orthomonotone distributions

We ran our experiences on multinormal and multicauchy distributions with positive diagonal matrices $\mathbf{\Sigma}$ and with their domains restricted to hyperrectangles of the form $[0,b]^d$.

In Fig. 3 and Fig. 4 the convergence of the hat and squeeze volumes is displayed for the multinormal distribution. We find a rapid decrease of the relative hat volumes (rejection constant) also in the case where the extent of the domain D is large.

From our runs and the following figures, where the setup and sample times have been depicted, we may draw some conclusions as to the efficiency of the multivariate Ahrens method. Firstly we note, that the setup times are increasing nearly linear with the number of subdomains N_D (Fig. 5).

The decrease of the rejection constant would theoretically lead to an equivalent decrease of the sample times, but as we have found in our practical implementations using guide-tables of length $2\,N_D$, the sampling times are not always decreased as expected due to memory caching effects, where for small dimensions we notice an increase in the times to pick the right subdomain using longer guide-tables (Fig. 6).

3.2 Orthounimodal distributions

There is only a marginal increase of complexity when applying the multivariate Ahrens method to orthounimodal distributions when compared with the case of orthomonotone distributions. In each of the 2^d quadrants (orthants) the
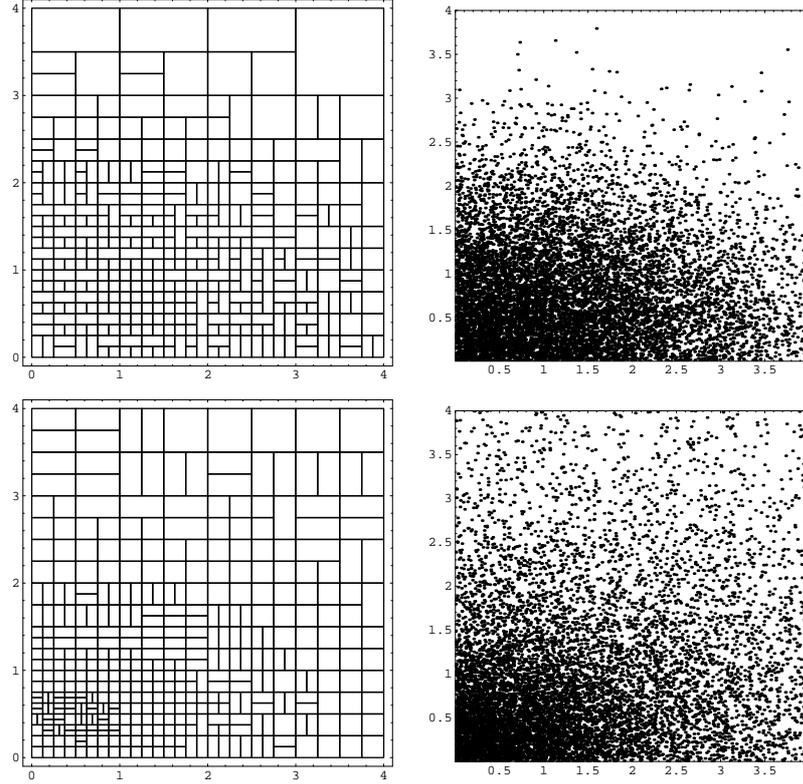
6

Fig. 2. Subdomain splittings (left) and random samples (right) for the multinormal distribution (top) and the multicauchy distribution (bottom). We may clearly see how the domain splittings adapt to the different shapes of the distributions. Note that a high density of the subdomains is by construction expected to reside in areas where the absolute value of the gradient of the pdf is large. In the present example ($d = 2$), the only non-vanishing elements of $\mathbf{\Sigma}$ were the two diagonal elements having the values 3 and 1 respectively.
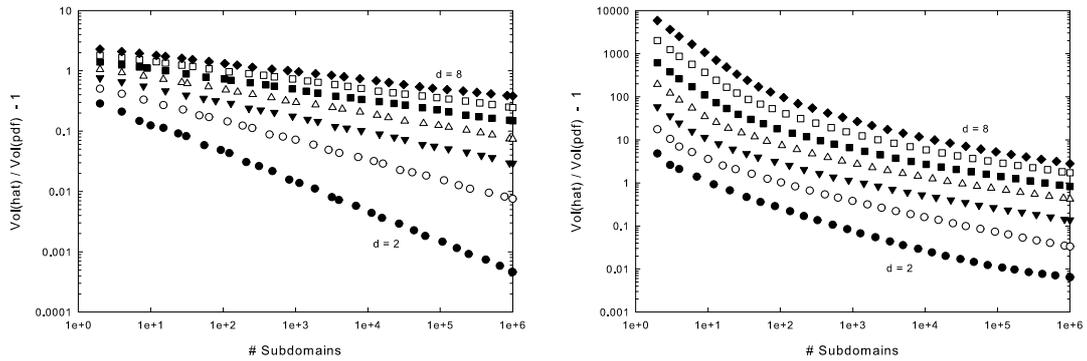
Fig. 3. Convergence of hat volume for the orthomonotone multinormal distribution with unit covariance-matrix and dimension $2 \leq d \leq 8$. The domain D covers one quadrant, $D = [0, 1]^d$ (left) and $D = [0, 4]^d$ (right).
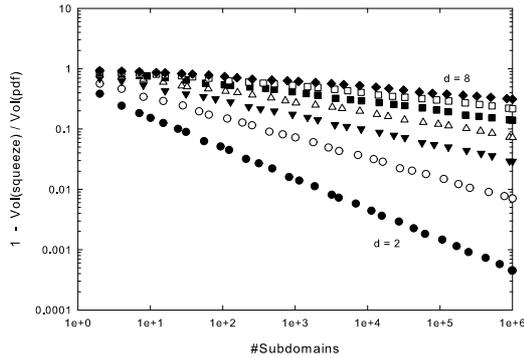
7

Fig. 4. Convergence of squeeze volume for the orthomonotone multinormal distribution with unit covariance-matrix and dimension $2 \leq d \leq 8$. The domain D covers one quadrant $D = [0, 1]^d$.
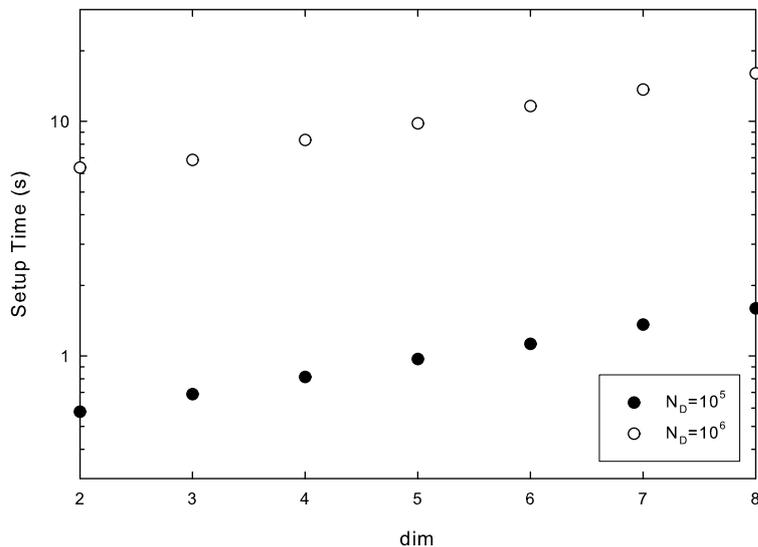

Fig. 5. Absolute setup times for the orthomonotone multinormal distribution with unit covariance matrix and domain $D = [0, 1]^d$ as a function of the dimension. (measured on a Mobile AMD Turion 64 ML-32 running at 800 MHz)

orthounimodal density function is orthomonotone and we may thus reduce this case to the previous one.

We present a typical result for the hat and squeeze convergence for a multinormal distribution with diagonal variance-covariance matrix and with $d \in \{2, \ldots, 8\}$ restricted to same domains $[-a, a]^d$. We would clearly need 2^d times more subdomains to achieve the same rejection constant as in the orthomonotone case. Nevertheless it is interesting to compare these two cases as the domains are being splitted in a different order.
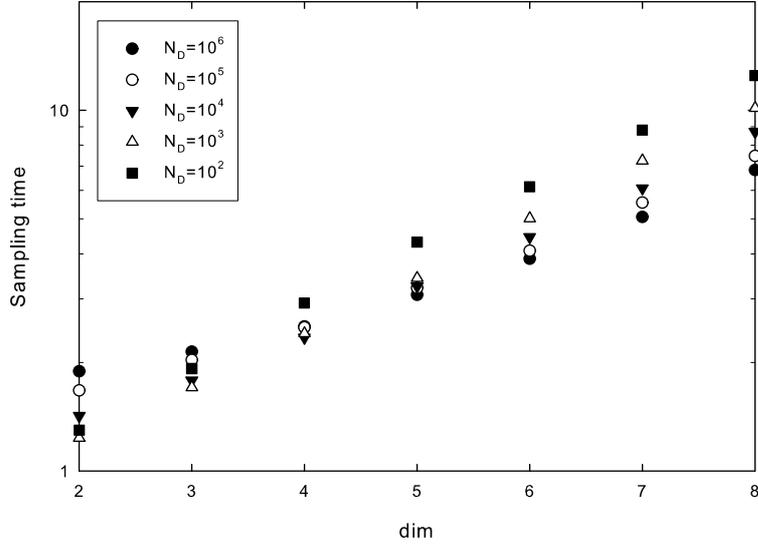
8

Fig. 6. Absolute sampling times (μs/sample) for the orthomonotone multinormal distribution with unit covariance matrix and domain $D = [0,1]^d$ as a function of the dimension. (measured on a Mobile AMD Turion 64 ML-32 running at 800 MHz) The number of samples was kept fixed at $N = 10^6$, whereas the number of subdomains N_D were varied between 10^2 and 10^6. For small dimensions, the sampling times are seen to *increase* with the number of subdomains due to architecture dependent memory caching effects.
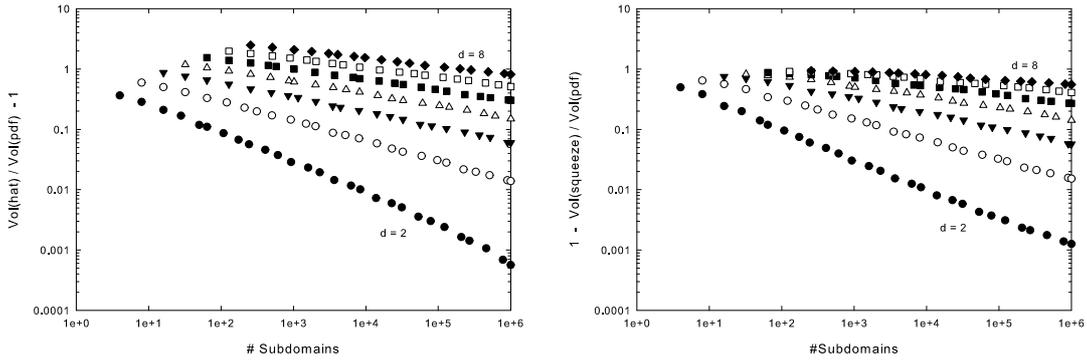

Fig. 7. Convergence of hat volume (left) and squeeze volume (right) for the orthounimodal multinormal distribution with unit covariance-matrix and dimension $2 \leq d \leq 8$. The domain D covers all quadrants $D = [-1,1]^d$.

9

4 General distributions

In this section we study the performance of the Ahrens method in cases, where the positions of the minima or maxima of the probability density function are not necessarily located at the domain corners.

We can model such cases by e.g. using a special form of the covariance matrix of an AR1-process

$$\Sigma_d(\rho) = [\rho^{|i-j|}], \quad 1 \le i, j \le d$$

with $\rho \neq 0$. We were thus interested to see to which extent our multivariate Ahrens implementation might be able to cope with a "slightly" non-orthounimodal distribution, i.e., for relatively small values of ρ.
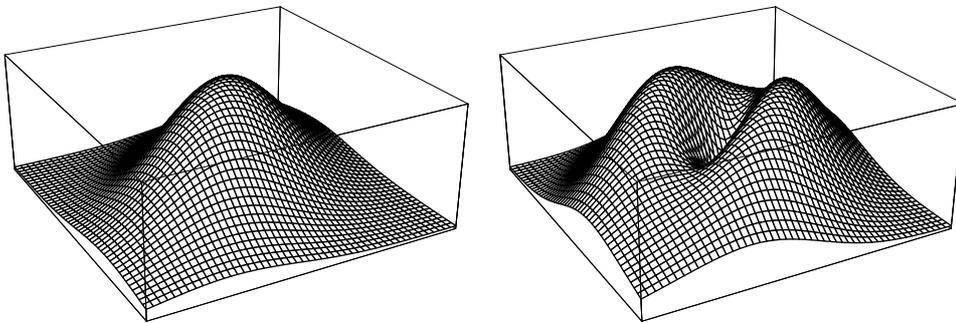

Fig. 8. Probability distribution function (left) and the norm of its gradient (right) for a multivariate normal distribution with covariance-matrix $\Sigma_2(0.5)$
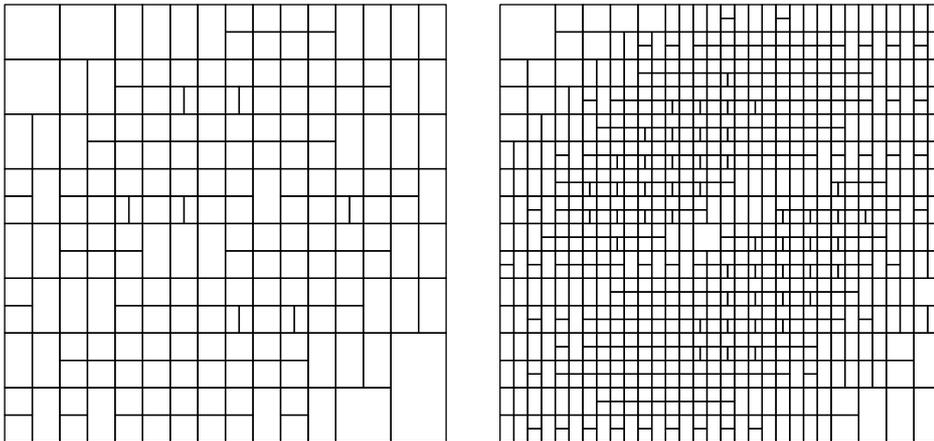

Fig. 9. Domain splittings for a multivariate normal distribution with covariance-matrix $\Sigma_2(0.5)$. The numbers of subdomains are 200 and 800 respectively. As the subdomains are constructed to minimize the volume–differences between the hat and squeeze, we can recognize an increased density of subdomains in areas where the absolute value of the gradient of the pdf is large (c.f. Fig. 8)

10

In Fig. 10 the convergence of the hat volumes is displayed for a non-orthounimodal distribution. It might at first sight look as if the algorithm is behaving properly in that case too, but we cannot be sure if the constructed hat is a true hat at all. The reason for this is, that in the case of non-orthounimodal distributions, our variant of the multivariate Ahrens method (Algorithm 1) is not providing a correct hat (and squeeze) as the maxima (and minima) of the pdf are not necessarily residing at the vertices of the constructed subdomains. A time-intensive Monte-Carlo min/max search for each subdomain might be performed – this is the approach currently implemented in the ROOT/FOAM package [11]. However, Fig. 11 displays very clearly what might possibly go wrong when the minima and maxima are way off the vertices – the total hat volume decrease as more and more subdomains are added, but at certain points the total hat volume may suddenly start to increase as a splitting is performed of a "wrong" subdomain leading to two "less wrong" subdomains.
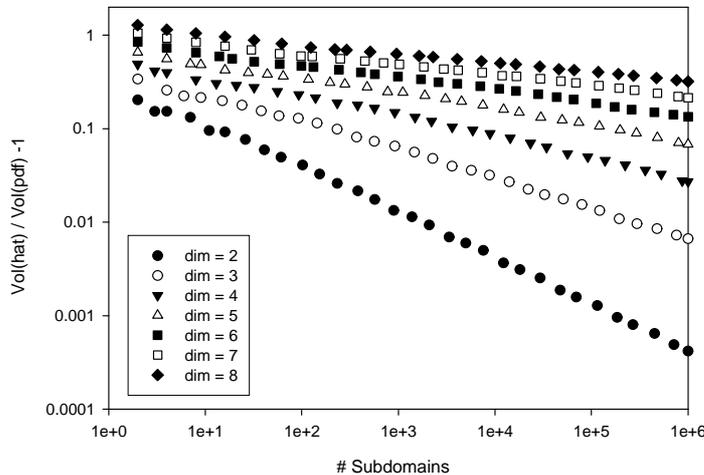

Fig. 10. Convergence of hat volume for a multivariate normal distribution with covariance-matrix $\Sigma_2(0.5)$. The domain D covers one quadrant $D = [0, 1]^d$.

5 Conclusion

Our computational experiences and results with the multivariate Ahrens method as described in Algorithm 1, indicate that it is useful for strictly orthounimodal distributions and small to medium size dimensions ($2 \leq d \leq 8$). However much care must be taken to assure that compiler and cache–size dependent effects are not overshadowing the theoretical gain and efficiency of the method. Furthermore, the error committed by applying the multivariate Ahrens method to non-orthounimodal distributions may possibly become large without any noticeable increases of the total hat-volume during the splitting process.
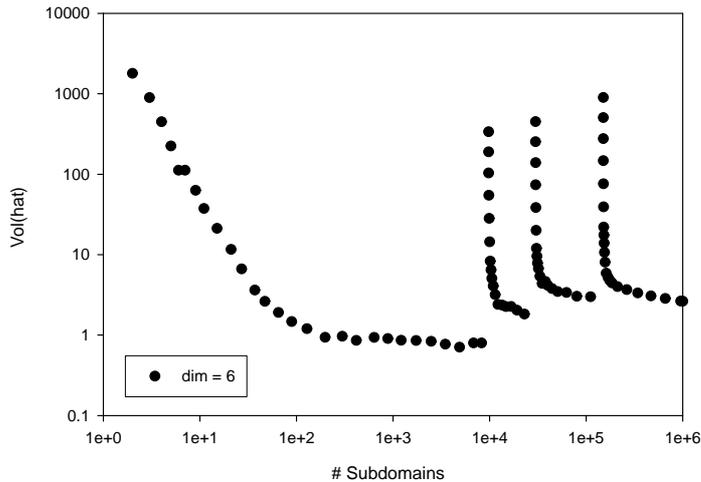
11

Fig. 11. Convergence of hat volume for a multinormal distribution with covariance-matrix $\Sigma_2(0.5)$. The domain D covers all quadrants $D = [-4, 4]^d$ and the mode is shifted a distance $m = 0.01$ off the center. In this case, the multivariate Ahrens algorithm is not providing a correct hat according to the fact, that the maxima of the pdf are not residing at the vertices of the constructed subdomains.

A quick and necessary (but not sufficient) indicator for a wrongly assumed hat – in the case that the maximum of the distribution function is not found at one of the vertices – should be tested at each domain-splitting, by ensuring that the sum of the hat volumes for the split-domains is not larger than the original hat volume before the splitting.

In the general case, there is no unique and simple solution to the problem of finding the global maximum and minimum of an arbitrary multivariate density distribution on a subdomain – from the knowledge of its values at its 2^d corners only. One might be tempted to construct n-point interpolating multivariate polynomials by adding additional support points or nodes in between the existing 2^d corners. Alternatively, a Monte-Carlo search for the maxima and minima in each domain may be performed. However, both approaches are very time consuming, i.e. leading to very long setup times of the domains. We may suggest to use other methods to sample from general multivariate densities, like for instance the very fast HITRO–method [12].

Acknowledgments

References

[1] L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, New-York, 1986.

[2] W. Hörmann, J. Leydold, G. Derflinger, Automatic Nonuniform Random Variate Generation, Springer-Verlag, Berlin Heidelberg, 2004.

[3] J. H. Ahrens, Sampling from general distributions by suboptimal division of domains, Grazer Math. Berichte 319 (1993) 20 pp.

[4] H. C. Chen, Y. Asau, On generating random variates from an empirical distribution, AIIE Trans. 6 (1974) 163–166.

[5] A. J. Walker, An efficient method for generating discrete random variables with general distributions, ACM Trans. Math. Software 3 (1977) 253–256.

[6] W. Hörmann, A note on the performance of the "Ahrens algorithm", Computing 69 (2002) 83–89.

[7] W. Hörmann, G. Derflinger, Asymptotically optimal design points for rejection algorithms, Communications in Statistics: Simulation and Computation 34 (4) (2005) to appear.

[8] J. H. Ahrens, A one-table method for sampling from continuous and discrete distributions, Computing 54 (2) (1995) 127–146.

[9] A. Zaman, Generation of random numbers from an arbitrary unimodal density by cutting corners, unpublished manuscript (1996).

[10] L. Devroye, Random variate generation for multivariate unimodal densities, ACM Trans. Model. Comput. Simul. 7 (4) (1997) 447–477.

[11] S. Jadach, Foam: A general-purpose cellular Monte Carlo event generator, http://arxiv.org/abs/physics/0203033 (2002).

[12] R. Karawatzki, J. Leydold, K. Pötzelberger, Automatic Markov chain Monte Carlo procedures for sampling from multivariate distributions, Tech. Rep. 27, Department of Statistics and Mathematics, WU Wien, http://statmath.wu-wien.ac.at/papers/2005-08.kar-ley-poe.pdf (2005).