View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elektronische Publikationen der Wirtschaftsuniversität Wien

Sampling from Linear Multivariate Densities

Wolfgang Hörmann, Josef Leydold

Research Report Series
Report 111, September 2009

Institute for Statistics and Mathematics
http://statmath.wu.ac.at/

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

Sampling from Linear Multivariate Densities

Wolfgang Hörmann and Josef Leydold

Abstract It is well known that the generation of random vectors with non-independent components is difficult. Nevertheless, we propose a new and very simple generation algorithm for multivariate linear densities over point-symmetric domains. Among other applications it can be used to design a simple decomposition-rejection algorithm for multivariate concave distributions.

1. Introduction

Exact sampling from (arbitrary) multivariate distributions over (subsets of) $\mathbb{R}^d$ is a challenging task. Only sampling from a distribution with independent components is simple. As we can generate each component independently, the sampling time for such random vectors scales linearly with dimension; see Hörmann et al. (2004) or Devroye (1986) for surveys of generation methods for univariate distributions. It is thus also very easy to sample from a distribution with constant density over a hyperrectangular domain (i.e., sampling uniformly from a box). For distributions with dependent components the situation is much more difficult. Only for the multinormal and the multi-t distributions well known generation algorithms are available that scale quadratically with the dimension. But even these distributions are not easily generated when their domain is restricted to a subset of $\mathbb{R}^d$.

There exist two general approaches for generating random vectors with a given probability density function (PDF) (see Hörmann et al. 2004): The rejection method and the conditional distribution method. The latter is only applicable in very simple

Wolfgang Hörmann
Department of Industrial Engineering, Boğaziçi University, 34342 Bebek-Istanbul, Turkey, e-mail: hormannw@boun.edu.tr

Josef Leydold
Department of Statistics and Mathematics, WU Wien, Augasse 2–6, A-1090 Vienna, Austria, e-mail: josef.leydold@wu-wien.ac.at

situations as marginal distributions are required. Rejection on the other hand works for arbitrary dimension but its applicability is limited by the fact that the rejection constant often grows exponentially with the dimension.

When designing new rejection algorithms for multivariate distributions we experimented with linear densities and were astonished to realize that they can be generated very easily. We were even more astonished by the fact that we did not find any hints to such methods in the literature. We therefore present our new method for generating random vectors with multivariate linear density over a bounded point-symmetric domain D and some of its applications.

This paper is organized as follows: Section 2 describes the new idea to sample from a multivariate linear density over point-symmetric domains. In Section 3 an improved rejection algorithm for multivariate concave densities over point-symmetric domains is introduced. Applications of the new algorithm are presented in Section 4.

2. Linear Densities over Point-symmetric Domains

A domain $D \subset \mathbb{R}^d$ is called *point-symmetric with center* $\mathbf{c}$ if $\mathbf{x} \in D$ implies that $\mathbf{x}^* = \mathbf{c} - (\mathbf{x} - \mathbf{c}) \in D$. The quantity $\mathbf{x}^*$ is the reflection of $\mathbf{x}$ in $\mathbf{c}$. For our purposes hyperrectangles $[x_{1l}, x_{1r}] \times \cdots \times [x_{dl}, x_{dr}]$ are the most important case of point-symmetric domains with center $\mathbf{c} = \frac{1}{2}\big((x_{1l}, \dots, x_{dl})' + (x_{1r}, \dots, x_{dr})'\big)$.

Let $\ell\colon D \subset \mathbb{R}^d \to [0,\infty)$, $\mathbf{x} \mapsto \ell(\mathbf{x}) = \mathbf{a}'(\mathbf{x} - \mathbf{c}) + f_c$ be a linear PDF with parameter $\mathbf{a}$, some constant f_c, and point-symmetric domain D. We denote the region below the graph of density ℓ by $\mathscr{F} = \{(\mathbf{x}, u) \in \mathbb{R}^{d+1} \colon \mathbf{x} \in D \text{ and } 0 \le u \le \ell(\mathbf{x})\}$. Notice that $\ell(\mathbf{c}) = f_c$. For such distributions we can easily show the following property.

Theorem 1. *Let $\mathbf{x} \in D$ and $u \in [0, 2f_c]$. Then $u < \ell(\mathbf{x})$ if and only if $u^* > \ell(\mathbf{x}^*)$, where $(\mathbf{x}^*, u^*) = 2(\mathbf{c}, f_c) - (\mathbf{x}, u)$, i.e., $(\mathbf{x}, u)$ reflected in $(\mathbf{c}, f_c)$.*

Proof. Notice that $(\ell(\mathbf{x}^*) - f_c) = -(\ell(\mathbf{x}) - f_c)$ and $(u^* - f_c) = -(u - f_c)$. Hence $u - f_c \le \ell(\mathbf{x}) - f_c$ if and only if $u^* - f_c \ge \ell(\mathbf{x}^*) - f_c$. Thus the statement follows.

Figure 1 sketches the situation for one dimension. As an immediate consequence we find that $\mathrm{Vol}_{d+1}(\mathscr{F}) = f_c \cdot \mathrm{Vol}_d(D)$, where $\mathrm{Vol}_d(\cdot)$ denotes the d-dimensional volume. Moreover, the reflection $(\mathbf{x}, u) \mapsto (\mathbf{x}^*, u^*)$ is a volume-preserving transformation that maps $\mathscr{F} \setminus (D \times [0, f_c])$ one-to-one onto $(D \times [0, f_c]) \setminus \mathscr{F}$. Algorithm 1 (`linearPDF-reflect`) compiles the relevant steps to sample from a linear density $\ell(\mathbf{x})$ using this property. Notice that Steps 2, 5, and 6 ("squeeze") reduce the average number of evaluations of the density and can speed up the algorithm in higher dimensions. They can also be entirely omitted.

Remark 1. Step 3 in Algorithm 1 is crucial and can be difficult for point-symmetric but irregular shaped domains[1]. Nevertheless, for a hyperrectangle $[x_{1l}, x_{1r}] \times \cdots \times$

[1] In fact, *every* non-uniform generation problem can be reduced to sampling uniformly from some domain in $\mathbb{R}^{d+1}$.

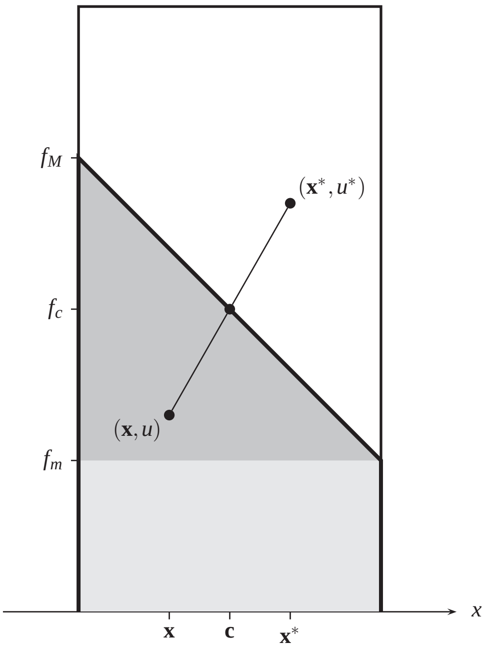

Fig. 1 Linear density on point-symmetric set D with center $\mathbf{c}$ (in one dimension).

Algorithm 1 `linearPDF-reflect`

```
Input: Linear density $\ell(\mathbf{x})$ on point-symmetric domain $D$ with center $\mathbf{c}$
    ($\ell(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in D$).
Output: Random vector $\mathbf{X}$ with density $\ell$.
    /* Setup */
    Compute $f_c \leftarrow \ell(\mathbf{c})$. /* squeeze */
    Compute $f_m \leftarrow \min_{\mathbf{x} \in D} \ell(\mathbf{x})$.
    /* Generator */
    Generate $\mathbf{X}$ uniformly in $D$.
    Generate $U$ uniformly in $[0, f_c]$.
    if $U \leq f_m$ then /* below squeeze */
        return $\mathbf{X}$.
    else if $U \leq \ell(\mathbf{X})$ then /* below density */
        return $\mathbf{X}$.
    else /* reflect point on center */
        return $\mathbf{X}^* = 2\mathbf{c} - \mathbf{X}$.
```

$[x_{dl}, x_{dr}]$ this is quite simple: $\mathbf{X} = \big(U_1 x_{1l} + (1-U_1)x_{1r}, \dots, U_d x_{dl} + (1-U_d)x_{dr}\big)$, where $U_1, \dots, U_d$ are i.i.d. $(0,1)$ uniform random numbers.

Let us look at the performance gain of this approach compared to Algorithm 2, which describes the simplest method, rejection from a constant hat. Both the setup and one iteration of the acceptance/rejection loop require the same number of density evaluations and uniform random numbers as Algorithm 1. However, the expected number of repetitions of this loop is $f_M/f_c = 2f_M/(f_M + f_m) \leq 2$, where

Algorithm 2 `rejection`

```
Input: Linear density $\ell(\mathbf{x})$ on domain $D$.
Output: Random vector $\mathbf{X}$ with density $\ell$.
    /* Setup */
 1: Compute $f_M \leftarrow \max_{\mathbf{x}\in D} \ell(\mathbf{x})$.
 2: Compute $f_m \leftarrow \min_{\mathbf{x}\in D} \ell(\mathbf{x})$.
    /* Generator */
 3: loop
 4:     Generate $\mathbf{X}$ uniformly in $D$.
 5:     Generate $U$ uniformly in $[0, f_M]$.
 6:     if $U \leq f_m$ then /* below squeeze */
 7:         return $\mathbf{X}$.
 8:     else if $U \leq \ell(\mathbf{X})$ then /* below density */
 9:         return $\mathbf{X}$.
```

f_M and f_m denote the maximum and minimum of the density ℓ on D, respectively. Hence, on average we save at most one iteration. In the best case the new approach saves 50% of the marginal generation time; in the worst case it is not slower than simple rejection with squeeze. Note that the new algorithm scales linearly with dimension.

2.1 An Extension

Algorithm 1 also works when $\mathbf{c} \notin D$ as long as $\ell(\mathbf{X})$ can be extended to a linear function ℓ' on some point-symmetric superset $D' \supset D$ with center $\mathbf{c}$. A simple example for this situation is a linear density on a ball restricted to its boundary (a sphere). Points $\mathbf{X}$ are still sampled uniformly in D in Step 4 (and not in the superset D').

3. Improved Rejection

Algorithm 2 (`rejection`) also works for distributions where the density is some linear function $\ell(\mathbf{x})$ restricted to its nonnegative part, i.e., $\max(0, \ell(\mathbf{x}))$. Algorithm 1 is not directly applicable for such densities but it can be easily adapted. We only have to add a rejection step to eliminate points with negative U-coordinates. Algorithm 3 (`linearPDF-general`) shows the details. Notice that there is no squeeze when $\min_{\mathbf{x}\in D} \ell(\mathbf{x}) \leq 0$.

This algorithm is based on the following modification of Theorem 1.

Theorem 2. *Let* $\mathbf{x} \in D$ *and* $u \in [f_m, f_M] = [\min_{\mathbf{x}\in D} \ell(\mathbf{x}), \max_{\mathbf{x}\in D} \ell(\mathbf{x})]$. *Then* $u < \ell(\mathbf{x})$ *if and only if* $u^* > \ell(\mathbf{x}^*)$, *where* $(\mathbf{x}^*, u^*) = 2(\mathbf{c}, f_c) - (\mathbf{x}, u)$.

It is easy to see that the rejection constant of Algorithm 3 is smaller than that of Algorithm 2 (`rejection`) if and only if $f_c = \ell(\mathbf{c}) > 0$. The performance gain

Algorithm 3 `linearPDF-general`

```
Input: Linear function $\ell(\mathbf{x})$ on point-symmetric domain $D$ with center $\mathbf{c}$.
Output: Random vector $\mathbf{X}$ with density $\max(0, \ell(\mathbf{x}))$.
/* Setup */
Compute $f_c \leftarrow \ell(\mathbf{c})$.
Compute $f_m \leftarrow \min_{\mathbf{x}\in D} \ell(\mathbf{x})$ and $f'_m \leftarrow \min(0, f_m)$.
/* Generator */
loop
    Generate $\mathbf{X}$ uniformly in $D$.
    Generate $U$ uniformly in $[f'_m, f_c]$.
    if $U > \ell(\mathbf{X})$ then /* above $\ell \to$ reflect point on center */
        $\mathbf{X} \leftarrow \mathbf{X}^* = 2\mathbf{c} - \mathbf{X}$,   $U \leftarrow U^* = 2\, f_c - U$.
    if $U \geq 0$ then /* density must be nonnegative */
        return $\mathbf{X}$.
```

is again limited by a factor of 2. We have to note here that the rejection constants of both algorithms can be arbitrarily large. Even when we restrict the domain of the density to its smallest bounding hyperrectangle, the rejection constant grows exponentially with dimension d if f is not linear on its entire domain.

3.1 Concave Densities

Let us look at a concave differentiable density f over a point-symmetric domain D with center $\mathbf{c}$. We can then construct a linear hat function ℓ over D by means of a tangent at some construction point $\mathbf{p}$. Thus we find

$$\ell(\mathbf{x}) = \nabla f(\mathbf{p})(\mathbf{x} - \mathbf{c}) + \big(\nabla f(\mathbf{p})(\mathbf{c} - \mathbf{p}) + f(\mathbf{p})\big) \,.$$

By the concavity of f, $\ell(\mathbf{x}) \geq 0$ and we can apply Algorithm 1 (`linearPDF-reflect`) for sampling from the majorizing density, Algorithm 4 (`concavePDF`). The acceptance probability of the rejection step is maximized if the area below the hat is minimized, i.e., when $\ell(\mathbf{c})$ is minimized. Such a construction point $\mathbf{p}$ can be easily found.

Theorem 3. *Let D be a point-symmetric domain with center* $\mathbf{c}$ *and let* f *be a density that is the restriction of some concave function to D. Then the rejection constant of a rejection algorithm based on the linear hat* $\ell(\mathbf{x})$ *is minimized if we choose center* $\mathbf{c}$ *as the construction point of* ℓ.

Proof. By Theorem 1 the volume of $\mathscr{F} = \{(\mathbf{x}, u) \in \mathbb{R}^{d+1} \colon \mathbf{x} \in D \text{ and } 0 \leq u \leq \ell(\mathbf{x})\}$ is given by $\mathrm{Vol}_{d+1}(\mathscr{F}) = \ell(\mathbf{c}) \cdot \mathrm{Vol}_d(D)$. Thus the rejection constant is minimized if $\ell(\mathbf{c})$ is minimized. By the concavity of f, $\ell(\mathbf{c}) \geq f(\mathbf{c})$ where equality holds for $\mathbf{p} = \mathbf{c}$. Thus the statement follows.

Algorithm 4 concavePDF

Input: Concave density $f(\mathbf{x})$ on point-symmetric domain D with center $\mathbf{c}$.
Output: Random vector $\mathbf{X}$ with density f.

```
/* Setup */
Compute $\mathbf{a} \leftarrow \nabla f(\mathbf{c})$ and $f_c \leftarrow f(\mathbf{c})$.
Compute $f_m \leftarrow \min_{\mathbf{x}\in D} f(\mathbf{x})$. /* use concavity of $f$ */
/* Generator */
loop
    Generate $\mathbf{X}$ uniformly in $D$.
    Generate $U$ uniformly in $[0, f_c]$.
    if $U \leq f_m$ then /* below squeeze */
        return $\mathbf{X}$.
    if $U > \mathbf{a}\cdot(\mathbf{X}-\mathbf{c}) + f_c$ then /* above $\ell \rightarrow$ reflect point on center */
        $\mathbf{X} \leftarrow \mathbf{X}^* = 2\mathbf{c} - \mathbf{X}, \quad U \leftarrow U^* = 2 f_c - U$.
    if $U \leq f(\mathbf{X})$ then /* accept */
        return $\mathbf{X}$.
```

3.2 General Densities

The concavity property of f above is only necessary to guarantee a simple set-up, which is even simpler and faster than the set-up of the Ahrens algorithm (see next paragraph) as no minimization procedure is required. The concept of Algorithm 4 also works for arbitrary densities. To use it in practice it is necessary to construct a linear upper bound to the density that is nonnegative for every point of the domain D. For example, secants can be used as hat functions for a convex density in one dimension.

4. Applications

4.1 Concave Densities and Ahrens Method

The simplest method for sampling from arbitrary multivariate distributions with given density are multigrid methods, which we call “multivariate Ahrens methods” as they generalize a method for univariate distributions proposed by Ahrens (1993). For this approach the domain of the distribution is partitioned into hyperrectangles. (If necessary the domain has to be extended to a union of proper hyperrectangles.) On each of these the maximum of the given density is estimated and a piecewise constant hat is computed. Thus the region below the density is covered by a union of bars and hence it is extremely simple to draw a random sample by rejection using Algorithm 2 for each of the hyperrectangles. This simple method has again and again attracted the interest of researchers who had to solve multivariate generation problems; see Jadach (2003) and Karawatzki (2006) for two recent examples. However, there are significant drawbacks associated with this approach. First, finding the maximum in each hyperrectangle requires either strong constraints on the given target distribution (e.g., uniorthomodal in Karawatzki 2006) or it is very time consuming when arbitrary densities are assumed (e.g., Jadach 2003).

The most prohibitive downside however is the slow convergence of the rejection constant. For a distribution in d dimensions with a bounded gradient, the rejection constant tends to 1 with rate $1+O(N^{-1/d})$ for increasing number N of hyperrectangles. Thus this leads to a large number N, and consequently large memory requirements and slow setup times even when computing the maximum is fast. Nevertheless, acceptance probabilities usually remain very small even for a moderate number of dimensions; see Karawatzki (2006). For distributions with concave density functions f we can formulate an alternative to the Ahrens method. It avoids the computation of the maximum in each hyperrectangle by using tangents in lieu of a constant hat and applies Algorithm 4 on each of the hyperrectangles. Then we find that the convergence rate is $1+O(N^{-2/d})$. The following theorem summarizes this observation.

Theorem 4 (Hörmann et al. 2004). *Let f be a bounded two times differentiable concave density f with bounded domain $D \subset \mathbb{R}^d$.*

(i) Construct a hat function h_1 using constant hats on N subintervals of equal size and shape. Then the volume between the hat and density tends to 0 with rate $O(N^{-1/d})$, i.e., we find for the rejection constant $1+O(N^{-1/d})$.

(ii) Likewise, when we construct a hat function h_2 using tangents in the center of each of N subintervals of equal size and shape, we find for the rejection constant $1+O(N^{-2/d})$.

Proof. (ii) As h_2 and f are both two times differentiable functions with the same first-order Taylor expansion at the center $\mathbf{c}$, we have $|h_2(\mathbf{x}) - f(\mathbf{x})| = O(r^2)$ around each $\mathbf{c}$, where $r = \|\mathbf{x}-\mathbf{c}\|$ is the distance from the center. Since we have N design points on a regular grid, the average radius is $r = O(N^{-1/d})$, which implies that the average distance $|h_2(\mathbf{x}) - f(\mathbf{x})| = O(N^{-2/d})$. As we have assumed a bounded domain D, we get $\int_D |h_2(\mathbf{x}) - f(\mathbf{x})|\, d\mathbf{x} = O(N^{-2/d})$.
(i) For constant hats we analogously find $|h_1(\mathbf{x}) - f(\mathbf{x})| = O(r) = O(N^{-1/d})$.

The performance gain of using Algorithm 4 compared to the Ahrens method (i.e., rejection from a piecewise constant hat) is twofold:

(1) There is no need to estimate the maximum of f in each of the (many) rectangles, except those in the boundary region of D.
(2) The rejection constant is reduced by some factor that is 1 in the worst case and $1/2$ in the best.

Of course, the latter is of practical relevance only if the rejection constant is not too large (at least below 100). In particular, for a rejection constant close to 1 the benefits become obvious. However, the faster (asymptotic) convergence of the new method is not of great help here. Unfortunately, to achieve rejection constants below 10 requires many hyperrectangles even in a moderate number (6–10) of dimensions; see Karawatzki (2006) for some computational experiences with the Ahrens method.

Remark 2. The improvement by a factor of at most 2 is rather disappointing. Nevertheless, if we try to accomplish the same improvement by refining the partition of the Ahrens method instead of using linear non-constant hats in lieu of constant ones, we need about 2^d times more rectangles.

4.2 Computational Experience

We implemented and tested our algorithms. For generating from linear densities we were astonished to see that the speed-up was even higher than the reduction of the expected number of repetitions reached by the reflection principle. Depending on the value of f_m, Algorithm 1 was up to three times faster than Algorithm 2. In one dimension Algorithm 1 is also faster than the inversion method so we can call it the fastest method to generate from linear densities.

For concave densities the speed-up is less spectacular. Depending mainly on the reduction of the expected number of repetitions we observed speed-ups in the range of 10 to 20 percent.

4.3 Importance Sampling

It has been shown that in the computation of expectations of functions of random variates by Monte Carlo methods it is often more efficient to replace the rejection algorithm by importance sampling (IS) with the hat function as the importance density; see e.g., Hörmann and Leydold (2005). This is in particular the case when the evaluation of the density $f(x)$ is expensive compared to the integrand, as by using IS we keep all information about our density f. This is even more inevitable when the rejection constant is high (which becomes very likely when the dimension increases).

Some methods like VEGAS (Lepage 1978) approximate the integrand in question by a piecewise constant function. Again using tangents decreases the approximation error. The rejection constant can be seen as a rough measure for the deviation from the target distribution. Notice that in the case of IS the hat function need not be a majorizing function as for the rejection algorithm. Thus we can safely drop the concavity assumption.

5. Conclusion

We have introduced a simple and fast algorithm for efficient sampling from linear multivariate densities with point-symmetric domains. It was demonstrated that the new idea can be utilized to considerably simplify and speed up both the set-up and the sampling of table methods to generate from multivariate concave distributions. The new idea may also be applied to importance sampling and for generating non-uniform distributions on the sphere.

References

Ahrens, J. H. 1993. Sampling from general distributions by suboptimal division of domains. *Grazer Math. Berichte* 319.

Devroye, L. 1986. *Non-uniform random variate generation*. New York: Springer-Verlag.

Hörmann, W., and J. Leydold. 2005. Monte Carlo integration using importance sampling and Gibbs sampling. In *Proceedings of the International Conference on Computational Science and Engineering*, ed. H. Dag and Y. Deng, 92–97.

Hörmann, W., J. Leydold, and G. Derflinger. 2004. *Automatic nonuniform random variate generation*. Berlin Heidelberg: Springer-Verlag.

Jadach, S. 2003. Foam: A general-purpose cellular Monte Carlo event generator. *Computer Physics Communications* 152:55–100.

Karawatzki, R. 2006. The multivariate Ahrens sampling method. Technical Report 30, Department of Statistics and Mathematics, WU Wien. `http://epub.wu-wien.ac.at/`.

Lepage, G. P. 1978. A new algorithm for adaptive multidimensional integration. *J. Comput. Phys.* 27 (2): 192–203.