Random Variate Generation for Multivariate Unimodal Densities

LUC DEVROYE
McGill University

A probability density on a finite-dimensional Euclidean space is orthounimodal with a given mode if within each orthant (quadrant) defined by the mode, the density is a monotone function of each of its arguments individually. Up to a linear transformation, most of the commonly used random vectors possess orthounimodal densities. To generate a random vector from a given orthounimodal density, several general-purpose algorithms are presented; and an experimental performance evaluation illustrates the potential efficiency increases that can be achieved by these algorithms versus naive rejection.

Categories and Subject Descriptors: G.3 [**Mathematics of Computing**]: Probability and Statistics—*random number generation*; I.6.1 [**Simulation and Modeling**]: Simulation Theory

General Terms: Algorithms, Experimentation, Measurement, Verification

Additional Key Words and Phrases: Multivariate densities, nonparametric classes, random variate generation, unimodality

1. INTRODUCTION

A multivariate density f on $\mathbb{R}^d$ is orthounimodal with mode at $m = (m_1, \ldots, m_d)$ if for each i, $f(x_1, \ldots, x_d)$ is a decreasing function of x_i as $x_i \uparrow \infty$ for $x_i \geq m_i$, and as $x_i \downarrow -\infty$ for $x_i \leq m_i$, when all other components are held fixed. For such multivariate densities, we propose general random variate generators that are hopefully of universal utility. Orthounimodal densities belong to the class of orthounimodal distributions (which include distributions without densities as well).

With multivariate distributions, one is often faced with enormous problems for random variate generation. Von Neumann's rejection method [von Neumann 1963; Devroye 1986b] requires a case-by-case study. The conditional method (generate one random variate; generate the next one conditional on the first one, and so forth) requires often difficult-to-compute

This work was supported by NSERC grant A3456 and by FCAR grant 90-ER-0291.
Author's address: School of Computer Science, McGill University, Montreal, Canada H3A 2K6; email: luc@cs.mcgill.ca.

marginal densities. Recently, Gibbs samplers have been proposed. These algorithms output an infinite sequence whose limit is a random variable with density f under certain regularity conditions. For example, when $d = 2$, one starts with $X_1 = x$, where x is arbitrary. X_2 is obtained by randomly sampling from the conditional density of X_2 given $X_1 = x$. Given $X_2 = y$, a new value for X_1 is obtained by randomly sampling from the conditional density of X_1 given $X_2 = y$. This process is continued for a while, after which the last pair (X_1, X_2) is returned. Its popularity is due to the fact that conditional densities are readily derivable from f, and no inconvenient marginal densities need to be computed. However, the returned pair is only approximately correct. See Bélisle et al. [1993], Gilks and Wild [1992] or Roberts and Polson [1994] for various aspects of Gibbs samplers.

We feel, therefore, that there is a real need for exact algorithms for large nonparametric classes of multivariate distributions. On the real line, first steps in this direction were taken by the author with respect to unimodal densities [Devroye 1984b], log-concave densities [Devroye 1984c], Lipschitz densities [Devroye 1984d], and densities with given characteristic functions or moment sequences [Devroye 1984a; 1986a; 1989; 1991]. In all these examples, efficient generators were derived that did not require any knowledge about the underlying densities beyond certain general parameters (the location of a mode, a Lipschitz constant, and so forth). The present article is in the same mold, and is only a first timid step into the vast world of multivariate distributions.

The main text on multivariate random variate generation is by Johnson [1987]. The emphasis in that text and in the literature is on modeling—finding a family of distributions that is flexible enough to fit the situation at hand, yet not too complex so as to make simulation unmanageable. This does not cover cases where one is asked by outsiders to generate random variates from a given distribution.

The mode m, which must be known (for now), partitions $\mathbb{R}^d$ into 2^d quadrants. On each quadrant, the density f is orthomonotone, that is, $f(x_1, x_2, \ldots, x_d)$ is nonincreasing in $s_1x_1, s_2x_2, \ldots, s_dx_d$, where the s_is are $+1$ or -1 according to the position of the quadrant. Such densities are best handled by taking care of each quadrant separately (see Section 8). The main part of this article deals with the fundamental problem for orthomonotone densities defined on the first quadrant (which is defined by the inequalities $x_i \geq 0$, $1 \leq i \leq d$). We assume that the mode is at the origin and that f is zero off the first quadrant. Examples include the following (C is a normalization constant that may vary from equation to equation):

(A)

$$f(x_1, \ldots, x_d) = \prod_{i=1}^{d} f_i(x_i), \quad x_i \geq 0,\ 1 \leq i \leq d,$$

where the f_is are nonincreasing and supported on $[0, \infty)$.

(B)

$$f(x_1, \ldots, x_d) = Ce^{-\prod_{i=1}^{d} g_i(x_i)}, \quad x_i \geq 0,\ 1 \leq i \leq d,$$

where the g_is are increasing functions of their arguments. This class includes densities such as the Gumbel bivariate exponential family [Gumbel 1960] $f(x_1, x_2) = C \exp(-ax_1 - bx_2 - cx_1x_2)$, $a, b, c > 0$.

(C) If $Z = (Z_1, \ldots, Z_d)$ is any random vector supported on $(0, \infty)^d$ (so that $P\{Z_i \leq 0\} = 0$ for any i), and $(U_1, \ldots, U_d)$ are i.i.d. uniform [0, 1] random variables independent of the Z_i's, then $X = (Z_1U_1, \ldots, Z_dU_d)$ is called a block monotone random variable [Shepp 1962]. For fixed Z with nonzero components, X is uniformly distributed on the hyperrectangle with vertices at Z and the origin, and is thus orthomonotone. If Z is random, we obtain a mixture of orthomonotone densities, which remains orthomonotone. Thus, block monotonicity implies orthomonotonicity. The converse, however, is false.

2. OTHER NOTIONS OF UNIMODALITY

Various notions of unimodality may be entertained. In this section, we briefly discuss these, and discuss their relevance and relationship with orthounimodality. For a general survey, we refer to Dharmadhikari and Joagdev [1982].

Intuitively, various notions of unimodality of a density f may be defined by the shapes or properties of the sets $A_c = \{f(x) \geq c\}$. For example, if for some c, A_c consists of two disjoint closed sets, f must have two humps. As we are not interested in topological properties of these sets, we will restrict the shape of A_c to some suitably large subclass of connected sets.

A density is star-unimodal at the origin if for any x, $f(tx)$ is nonincreasing as t varies from 0 to ∞. Even this class is too large to be of practical interest.

Orthounimodal densities are star unimodal, and both definitions above generalize the standard notion of univariate unimodality. Most bivariate densities, for example, that we are aware of, are either orthounimodal, or orthounimodal after a linear transformation. This includes the bivariate normal distribution, the non-standard normal conditionals distribution [Bhattacharyya 1942; Arnold et al. 1992, p. 31], the bivariate exponential distribution [Arnold and Strauss 1988], the bivariate Pareto distribution [Arnold et al. 1992, p. 58], the bivariate Cauchy distribution [Arnold et al. 1992, p. 68], the multivariate t distribution (or multivariate Pearson distribution), the multivariate Pearson II distribution, and many others too numerous to mention.

There are narrower notions than orthounimodality. For example, a density f is convex unimodal at the origin if all sets $T_c = \{x : f(x) \geq c\}$ are

Table I. A List of Useful Multivariate Families of Densities

Family	*Form of Density*
Liouville distributions	$\psi(\Sigma\, x_i)$
MP (or: min-product) distributions	$c \min(1, \Pi\, \psi(x_i))$
generalized MP distributions	$c \min(1, \Pi\, \psi_i(x_i))$
multivariate max distributions	$\psi(\max x_i)$
multivariate min distributions	$\psi(\min x_i)$
platymorphous distribution	$c \min(1, be^{-\Sigma x_i})$
schizomorphous distribution	$c \min(1, b/(\Pi\, x_i(\Sigma\, x_i)^a))$
bathymorphous distribution	$c \min(1, b/(\Pi\, x_i\, (\max x_i)^a))$

either empty or are convex sets that contain the origin. Some, but not all, of the densities mentioned above are convex unimodal.

Log-concave densities, monotone unimodal densities, axially unimodal densities, linear unimodal densities and α-unimodal densities are all discussed in Dharmadhikari and Joagdev [1988]. These are not considered here. Of these, we believe that the log-concave densities are by far the most important. They deserve a special study altogether.

So, orthounimodal densities are widely applicable and form a class with a rich structure for which beautiful performance bounds can be obtained. Finally, it is a robust class: for example, all lower-dimensional marginals of orthounimodal densities are orthounimodal. The same is not true for convex unimodal densities [Dharmadhikari and Joagdev 1988, p. 63].

3. OVERVIEW OF THE RESULTS

The discussion is greatly helped if we begin with the definition of several families of mostly new multivariate distributions. For these distributions, simple efficient random variate generation algorithms are proposed. For easy reference, the list of these basic families is given in Table I. In all cases, the domain of the densities is the positive quadrant $x_i \geq 0$, $1 \leq i \leq d$. The symbols ψ, ψ_i denote general nonnegative functions on $[0, \infty)$. The symbols a, b, c denote positive constants.

The next section deals with these families. In the remainder of the article, we propose general rejection algorithms for orthomonotone densities f. These are all based upon inequalities of the form $f \leq cg$, where g belongs to one of the basic families of Table I, and c is a constant. All algorithms thus take the form

```
repeat
    generate U uniform [0, 1]
    generate X with density g on [0, ∞)^d
  until Ucg(X) ≤ f(X)
  return X
```

The expected number of iterations is precisely c [Devroye 1986b]. To obtain useful rejection inequalities, it suffices to assume that a little extra information is known about the distribution. Table II summarizes the

Table II. Under Certain General Conditions of f, Rejection with Dominating Densities from the Basic Families of Multivariate Densities become Feasible

Condition on f	*Form of g in $f \le cg$*
Bounded; support on $[0, 1]^d$	platymorphous (after transformation)
Generator available for $f_i = f(0, \ldots, x, 0, \ldots, 0)$	product density
symmetric in x_is	multivariate max
bounded; $\mu_i = \mathrm{E}\{X_i^a\}$ known and finite for some $a > 0$	bathymorphous
bounded; $\mathrm{E}\{(\Sigma X_i)^a\}$ known and finite for some $a > 0$	schizomorphous
bounded; moment generating function known	MP family

algorithms developed in this article based upon inequalities that are derived further on.

Finally, we develop table methods and adaptive algorithms that are useful when many random vectors from the same f are needed. A limited simulation concludes the article.

4. BASIC FAMILIES

In this section, the following simple facts will be used repeatedly.

LEMMA 1. *For $A, B, a, b > 0$, we have*

$$\int_0^\infty \min(Au^{a-1}, B/u^{b+1})du = \frac{a+b}{ab}(A^b B^a)^{1/(a+b)}.$$

If $p, q > 0$ and if U, V are i.i.d. uniform $[0, 1]$, then the density of $Y = U^p/V^1$ is

$$\frac{1}{p+q}\min(y^{(1/p)-1}, 1/y^{(1/q)+1}).$$

The density of $W \stackrel{\text{def}}{=} (B/A)^{pq/(p+q)}Y$ is

$$\frac{1}{(p+q)(B^q A^p)^{1/(p+q)}}\min(Aw^{(1/p)-1}, B/w^{(1/q)+1}).$$

4.1 Liouville Distributions

The class of Liouville distributions is studied at length by Sivazlian [1981a; 1981b]. A particular subfamily of densities is of the form

$$f(x_1, \ldots, x_d) = \psi(x_1 + \cdots + x_d),\ x_i \ge 0,$$

where ψ is a general function. Whenever densities are written in this form, the following fundamental property holds (check section XI.4.2 of Devroye [1986b] or Sivazlian [1981a]:

If X has density $\xi(x) = x^{d-1}\psi(x)/(d-1)!$ on $(0, \infty)$, and $(S_1, \ldots, S_d)$ is a collection of uniform of uniform spacings (that is, the spacings of [0, 1] defined by $d - 1$ i.i.d. uniform [0, 1] random variables), then $(XS_1, \ldots, XS_d)$ has density $\psi(x_1 + \ldots + x_d)$ on the positive quadrant. The density ξ is called the generator for f.

Generating uniform spacings is trivial—see Devroye [1986 Ch. 6]. Thus, one need only worry about the one-dimensional random variate generation problem for g. We also point to the generalized Liouville distributions of Sivazlian [1981b], a particular instance of which has densities of the form

$$\psi\left(\sum_{i=1}^{d}(x_i/a_i)^b\right).$$

This class includes the class of radially symmetric distributions with a density (take $b = 2$). In the context of orthomonotone densities, this more general family will not be needed.

4.2 The Min-Product Family

The MP (min-product) family consists of all densities on the positive quadrant that can be written as

$$f(x) = c\ \min\left(1, \prod_{i=1}^{d} \psi(x_i)\right),$$

where $x = (x_1, \ldots, x_d)$, ψ is an unspecified (not necessarily integrable) function, and $c > 0$ is a normalization constant. The generalized MP family consists of all densities on the positive quadrant that can be written as

$$f(x) = c\ \min\left(1, \prod_{i=1}^{d} \psi_i(x_i)\right),$$

where the ψ_is are positive functions, and $c > 0$ is a normalization constant.

4.3 Multivariate Max Distributions

If a multivariate density is of the form $\psi(\max(x_i))$ for some function ψ, where $x_i \geq 0$, we say that it is a multivariate max density. It is trivial to verify that for such densities, we may generate a random variate by the following method.

If X has density $\xi(x) = dx^{d-1}\psi(x)$ on $(0, \infty)$, Z is uniformly distributed on $\{1, \ldots, d\}$, and $(U_1, \ldots, U_d)$ are d i.i.d. uniform [0, 1] random variables, and if we replace U_Z by 1, then $(XU_1, \ldots, XU_d)$ has density proportional to $\psi(\max_{i=1}^{d} x_i)$ on the positive quadrant. The density ξ is called a generator.

To check the validity, let $(X_1, \ldots, X_d)$ have density $\psi(\max(x_1, \ldots, x_d))$. Set $X^* = \max(X_1, \ldots, X_d)$. Then, if A_i is the set on which $x_i = \max(x_i, \ldots, x_d)$,

$$\mathrm{P}\{X^* \le u\} = \int_{[0,u]^d} \psi(\max(x_1, \ldots, x_d)) \prod_{j=1}^{d} dx_j$$

$$= \sum_{i=1}^{d} \int_{A_i \cap [0,u]^d} \psi(\max(x_1, \ldots, x_d)) \prod_{j=1}^{d} dx_j$$

$$= \sum_{i=1}^{d} \int_0^u x_i^{d-1} \psi(x_i) dx_i .$$

Thus, X^* has density $du^{d-1}\psi(u)$. Given $X^* = X_i$, the X_j's $j \neq i$, are independent and uniformly distributed on $[0, X^*]^{d-1}$.

Example 1. Consider $f(x_1, \ldots, x_d) = e^{-\max(x_1, \ldots, x_d)}/\Gamma(d+1)$, $x_i \ge 0$, a multivariate distribution of exponential flavor. Then the density of $X^* = \max(X_1, \ldots, X_d)$ is

$$\frac{dx^{d-1}e^{-x}}{\Gamma(d+1)} = \frac{x^{d-1}e^{-x}}{\Gamma(d)}, \quad x \ge 0,$$

which is the gamma (d) density. In the remainder of this article, G_d denotes a gamma (d) random variable. Thus,

$$(X_1, \ldots, X_d) \stackrel{\mathcal{L}}{=} (U_1 G_d, U_2 G_d, \ldots, U_d G_d),$$

where the U_i's are i.i.d. uniform [0, 1], with the exception of $U_Z = 1$, where Z is a uniformly picked integer from $\{1, \ldots, d\}$. Gamma variate generation is discussed in Devroye [1986b]. Recent high quality gamma generators were obtained and implemented by Cheng and Feast [1979; 1980], Ahrens et al. [1983], and Stadlober and Kremer [1992]. Equivalently, by considering order statistics, it is easy to see that random variates with the given density f may be obtained as follows, avoiding gamma generators altogether:

generate $E_1, E_2, \ldots, E_d$, i.i.d. exponential random variates
set $X_1 \leftarrow E_1$, $X_2 \leftarrow X_1 + E_2, \ldots, X_d \leftarrow X_{d-1} + E_d$
return a random permutation of $(X_1, \ldots, X_d)$

Example 2. Consider $f(x_1, \ldots, x_d) = c \min(A, B/\max^a(x_1, \ldots, x_d))$ where c, A, B and a are positive constants. The density of X^* is $dc \min(Ax^{d-1}, Bx^{d-1-a})$. This is a valid density on $(0, \infty)$ and if $a > d$. By

Lemma 4.1, we note that

$$X^* = \max(X_1, \ldots, X_d) \stackrel{\mathcal{L}}{=} \left(\frac{B}{A}\right)^{1/a} \frac{U^{1/d}}{V^{1/(a-d)}}.$$

With X^* in hand, we have:

generate U, V i.i.d. uniform [0, 1]
  $X \leftarrow (B/A)^{1/a} U^{1/d} / V^{1/(a-d)}$
  generate $(U_1, \ldots, U_d)$ uniformly in $[0, 1]^d$
  generate a random integer $Z \in \{1, \ldots, d\}$
  set $U_Z \leftarrow 1$
  return $(XU_1, \ldots, XU_d)$

4.4 Multivariate Min Distributions

If a multivariate density on $[0, 1]^d$ is of the form $\psi(\min(x_i)$ for some function ψ, where $x_i \geq 0$, we say that it is a multivariate min density. For such densities, we may generate a random vector by the following method:

If Y has density $\xi(y)$ proportional to $(1 - y)^{d-1}\psi(y)$ on [0, 1], Z is uniformly distributed on $\{1, \ldots, d\}$ and $(U_1, \ldots, U_d)$ are d i.i.d. uniform [0, 1] random variables, and if we replace U_Z by 0, then $(Y, Y, \ldots, Y) + ((1 - Y)U_1, \ldots, (1 - Y)U_d)$ has density proportional to $\psi(\min_{i=1}^{d} x_i)$ on the positive quadrant. Note that given Y, Z, $d - 1$ components of this random vector are uniformly distributed on the cube $[Y, 1]^{d-1}$.

The proof of the validity is exactly as for multivariate max distributions.

4.5 The Platymorphous Distribution

The platymorphous distribution has density

$$h_{b,d}(x) = c \, \min(1, be^{-\sum_{i=1}^{d} x_i}),$$

where $x = (x_1, \ldots, x_d)$, and $x_i \geq 0$ for $i \geq 1$. It is a two-parameter family with parameters $b \geq 1$ and d a positive integer. Furthermore, c is a normalization constant defined by

$$c = \frac{1}{I_1 + I_2}, \quad I_1 = \frac{\log^d b}{d!}, \quad I_2 = \sum_{i=0}^{d-1} \frac{\log^i b}{i!}.$$

The notation $\log^d b$ is used to denote $(\log b)^d$. The name platymorphous comes from "platy"—the density is flat near the origin—, and "morphous"—shape. An important special case occurs when $b = 1$: the density reduces to

$$h_{1,d}(x) = \prod_{i=1}^{d} e^{-x_i}, \quad x_i \geq 0,$$

the density of d independent exponential random variables. The platymorphous distribution is a Liouville distribution with generator given by

$$g_{b,d}(u) = \frac{cu^{d-1}}{(d-1)!}\min(1, be^{-u}) = \begin{cases} \dfrac{cu^{d-1}}{(d-1)!} & \text{if } 0 < u < \log b \\ \dfrac{bcu^{d-1}e^{-u}}{(d-1)!} & \text{if } \log b \le u. \end{cases}$$

The generator consists of two pieces, a polynomial part and a tail of a gamma distribution. We let $G_d(t)$ denote a gamma (d) random variable restricted to $[t, \infty)$, or, a tail-gamma variate. It is easy to check that

$$\int_0^{\log b} g_{b,d}(u)du = \frac{c \log^d b}{d!} = \frac{I_1}{I_1 + I_2},$$

and that

$$\int_{\log b}^{\infty} g_{b,d}(u)du = bc\mathrm{P}\{G_d \ge \log b\} = bc \sum_{i=0}^{d-1} \frac{\log^i b e^{-\log b}}{i!} = \frac{I_2}{I_1 + I_2}.$$

The following composition algorithm may thus be used to generate a random variate with density $g_{b,d}$:

generate a uniform [0, 1] random variate U
if $U < I_1/(I_1 + I_2)$
  then generate V uniformly on [0, 1]
    return $X \leftarrow V^{1/d} \log b$
  else return $X \leftarrow G_d(\log b)$

Tail-gamma variate generation. $G_d(t)$ can be generated in $O(1)$ expected time via the algorithm given in Devroye [1986, Sect. IX.3.7] when $t > d$. For $d \ge 1$, the expected time is bounded from above by

$$\frac{1}{1 - ((d-1)/t)^2}$$

(this follows from Devroye [1986, p. 422]. It is uniformly bounded if $t \ge 2d$, say. For $t < d$, the algorithm

repeat generate G_d
until $G_d \ge t$
return $X \leftarrow G_d$

takes expected number of iterations $1/\mathrm{P}\{G_d \ge t\} \le 1/\mathrm{P}\{G_d \ge d\}$. But $\mathrm{P}\{G_d \ge d\} \to 1/2$ by the convergence of $(G_d - d)/\sqrt{d}$ to the normal distribution as $d \to \infty$. This shows that the expected number of iterations stays uniformly bounded. Finally, for $d \le t < 2d$, one may employ rejection based upon a dominating curve that consists of a constant piece near the origin and an exponential tail starting at $d + \sqrt{d}$. This too takes uniformly bounded expected time.

The above algorithm suffers from a major drawback, the computation of I_1 and I_2, two constants which involve summations and factorials. Merely computing them takes time proportional to d. We will show that the following efficient modification provides exact variates as well, and is less cumbersome:

(algorithm *gdb*)
generate a uniform [0, 1] random variate U
generate a tail-gamma variate $G_{d+1}(\log b)$
set $G_d \leftarrow U^{1/d} G_{d+1}(\log b)$
if $G_d \geq \log b$
  then return $X \leftarrow G_d$
  else return $X \leftarrow V^{1/d} \log b$ where V is uniform [0, 1]

Algorithm gdb requires at most two uniform and one tail-gamma variate per returned X, and should be acceptable in most cases. Needless to say that further improvements are possible. Its correctness is shown in Appendix A. If $(S_1, \ldots, S_d)$ are uniform spacings on [0, 1], then $(XS_1, \ldots, XS_d)$ is platymorphous. The expected time is $O(d)$ if the uniform spacings are generated by the algorithms described by Devroye [1986, Section V.3.1], if X has density $g_{b,d}$ and is generated by the method given above, and if tail-gamma variates are obtained as described above.

4.6 The Schizomorphous Density

The schizomorphous density has three parameters, a positive integer d, a positive shape parameter a and a scale parameter b:

$$\varphi_{b,a,d}(x) = c \min\left(1, \frac{b}{\prod_{i=1}^{d} x_i (x_1 + \cdots + x_d)^a}\right).$$

The constant c normalizes the integral to one. The prefix schizo refers to the dependence of the density upon both $\Sigma\, x_i$ and $\Pi\, x_i$. In previous sections, we have seen how functions of the sum or the product alone may be handled, but not functions of both. A generator for the schizomorphous density is given below. The proof of the validity is given in Appendix C.

generate i.i.d. uniform [0, 1] random variates U, V
generate independent gamma $(a/(d + a))$ random variates $Y_1, \ldots, Y_d$
$Y \leftarrow \Sigma_{i=1}^{d} Y_i$
$(T_1, \ldots, T_d) \leftarrow ((Y_1/Y), \ldots, (Y_d/Y))$
$T \leftarrow \Pi_{i=1}^{d} T_i$
$S \leftarrow (b/T)^{1/(a+d)} \times (U^{1/d}/V^{1/a})$
return $X = (X_1, \ldots, X_d) \leftarrow (ST_1, \ldots, ST_d)$

4.7 The Bathymorphous Density

The bathymorphous density is

$$\zeta_{b,a,d}(x) = c \min\left(1, \frac{b}{\prod_{i=1}^{d} x_i (\max_{i=1}^{d} x_i)^a}\right),$$

where $a > 0$ is a shape parameter, $b > 0$ is a scale parameter, and d is a positive integer. Once again, c is a normalization constant. The prefix bathy (deep) refers to the shape of the density in dimension 2, where it resembles a flat beach in a cove, and a quickly dropping sea level. Let $X = (X_1, \ldots, X_d)$ have this density. Set $X_i = T_i M$, $M = \max_j X_j$ for all i, where $0 \le T_i \le 1$ are random variables. With probability one, all T_is are pairwise different. The joint density of $(T_2, \ldots, T_d, M)$ given $M = X_1 = m$ is

$$dcm^{d-1} \min\left(1, \frac{b}{m^{a+d} \times \prod_{j=2}^{d} t_j}\right).$$

If we take the integral with respect to m, then we obtain the marginal density of $(T_2, \ldots, T_d)$. Recalling Lemma 4.1, we note that the conditional density of $(T_2, \ldots, T_d)$ given $X_1 = M$ is given by a constant times

$$\left(\prod_{j=2}^{d} t_j\right)^{-d/d+a}, \quad 0 \le t_j \le 1, \quad 2 \le j \le d.$$

This is proportional to the product of $d - 1$ densities on [0, 1] of the form $(a/(a + d))t_j^{-d/(a+d)}$. A random variate with the latter density may be generated as $T_j \leftarrow U_j^{(a+d)/a}$ where U_j is uniform [0, 1]. Therefore, given $X_1 = M$,

$$(T_2, \ldots, T_d) \stackrel{\mathcal{L}}{=} (U_2^{(a+d)/a}, \ldots, U_d^{(a+d)/a})$$

where the U_is are independent uniform [0, 1] random variables.

To generate M (still conditional on $X_1 = M$), we use the conditional method. Let $T = \prod_{i=2}^{d} T_i$. Then M has density proportional to

$$m^{d-1} \min\left(1, \frac{b}{Tm^{a+d}}\right).$$

By Lemma 1, it is easily seen that

$$M \stackrel{\mathcal{L}}{=} \left(\frac{b}{T}\right)^{1/(a+d)} \times \frac{U^{1/d}}{V^{1/a}}$$

where U, V are i.i.d. uniform [0, 1]. Finally, by symmetry, we note that all of the above may be repeated for the conditions $M = X_i$. In fact, then, $M =$

X_N where N is uniformly distributed on $\{1, \ldots, d\}$. We summarize the algorithm for the bathymorphous density:

> generate N uniformly on $\{1, \ldots, d\}$
> generate i.i.d. uniform [0, 1] random variates $U, V, U_1, \ldots, U_d$
> $U_N \leftarrow 1$
> $(T_1, \ldots, T_d) \leftarrow (U_1^{(a+d)/a}, \ldots, U_d^{(a+d)/a})$
> $T \leftarrow \prod_{i=1}^d T_i$
> $M \leftarrow (b/T)^{1/(a+d)} \times (U^{1/d}/V^{1/a})$
> return $X = (X_1, \ldots, X_d) \leftarrow (MT_1, \ldots, MT_d)$

5. GENERAL REJECTION ALGORITHMS

In this section we develop five rejection algorithms for different general subclasses of orthomonotone densities, as listed in Table II. In each case, the algorithm follows easily from an appropriate inequality.

5.1 Bounded Densities On The Unit Cube

Let f be orthomonotone on $[0, \infty)^d$. Then, by orthomonotonicity,

$$1 = \int f(y)dy \geq f(x) \prod_{i=1}^d x_i.$$

Therefore,

$$f(x) \leq \min\left(f(0), \frac{1}{\prod_{i=1}^d x_i}\right).$$

The bounding function is not integrable. However, if f is also bounded and vanishes off $[0, 1]^d$, then we may apply a transformation $x_i = e^{-y_i}$, and restrict each y_i to $[0, \infty)$. Under this transformation, $f(x)$ is transformed into $g(y)$, with $y = (y_1, \ldots, y_d)$ and

$$g(y) = f(x)e^{-\sum_{i=1}^d y_i}, y_i \geq 0.$$

In particular,

$$g(y) \leq \min(1, f(0)e^{-\sum_{i=1}^d y_i}) = \frac{h_{f(0),d}(y)}{c},$$

where $h_{f(0),d}$ is the platymorphous density with parameters $f(0)$ and d, and

$$c = \frac{1}{I_1 + I_2} = \frac{1}{\sum_{i=0}^d [\log^i f(0)]/i!}.$$

Table III. Expected Number of Iterations for Algorithm Universal-1 as a Function of $f(0)$ and d

f(0)	1	2	4	8	16	32	64	128	256	512	1024
d=1	1.00	1.69	2.38	3.07	3.77	4.46	5.15	5.85	6.54	7.23	7.93
d=2	1.00	1.93	3.34	5.24	7.61	10.47	13.81	17.62	21.91	26.69	31.95
d=3	1.00	1.98	3.79	6.74	11.16	17.40	25.79	36.66	50.33	67.15	87.45
d=4	1.00	1.99	3.94	7.51	13.63	23.42	38.26	59.75	89.73	130.26	183.64
d=5	1.00	1.99	3.98	7.84	14.99	27.58	48.62	82.16	133.42	208.99	316.97
d=6	1.00	1.99	3.99	7.95	15.62	29.99	55.81	100.28	173.80	290.85	471.01
d=7	1.00	2.00	3.99	7.98	15.87	31.18	60.08	112.84	205.79	363.81	623.53
d=8	1.00	2.00	3.99	7.99	15.96	31.70	62.30	120.46	227.96	420.69	755.69
d=9	1.00	2.00	3.99	7.99	15.99	31.90	63.33	124.57	241.62	460.13	857.47
d=10	1.00	2.00	4.00	7.99	15.99	31.96	63.75	126.56	249.20	484.73	928.02

Algorithm universal-1 is a generalization of the one-dimensional method of Devroye [1984b].

This inequality may be used straightforwardly with the rejection method.

(algorithm universal-1)
repeat
  generate U uniform [0, 1]
  generate Y with platymorphous density $h_{f(0),d}$
  set $Z = (e^{-Y_1}, \ldots, e^{-Y_d})$
until $U \min(1, f(0)e^{-\Sigma_{i=1}^d Y_i}) \leq f(Z)e^{-\Sigma_{i=1}^d Y_i}$
return Z

The expected number of iterations before halting is

$$\frac{1}{c} = \sum_{i=0}^{d} \frac{\log^i f(0)}{i!}.$$

Clearly, this number never exceeds $f(0)$ as it sums over only the first $d + 1$ terms of the Taylor series expansion of $f(0) = e^{\log f(0)}$. The number $f(0)$ measures the expected number of iterations for the rejection method based on the trivial inequality $f(x) \leq f(0)$. If $d \ll \log f(0)$, the expected number of iterations is much smaller than $f(0)$. For fixed d, it grows as $[\log^d f(0)]/d!$ as $f(0) \to \infty$. The Table III shows the usefulness for large values of the ratio $\log(f(0))/d$:

Scaling. If f is orthomonotone on $\times_{i=1}^d [0, s_i]$, then we first set $w_i = x_i/s_i$, so that $(W_1, \ldots, W_d)$ has density $f''(w) = f(s_1 w_1, \ldots, s_d w_d) \prod_{i=1}^d s_i$ on $[0, 1]^d$. We may apply the above algorithm to obtain $(W_1, \ldots, W_d)$ and then set $X_i = s_i W_i$ for all i. Note that $f''(0) = sf(0)$, where $s = \prod_{i=1}^d s_i$. The expected number of iterations before halting is

$$\sum_{i=0}^{d} \frac{\log^i (sf(0))}{i!},$$

Note that $sf(0)$ is a scale-invariant factor, and thus that the complexity is determined solely by the shape of the density.

5.2 Bounding by Product Densities

In multivariate Gibbs sampling, one assumes that the explicit form of f is known before the Gibbs sampler is designed. This situation differs from the others in this section, where f can only be computed at fixed points but its analytic form is not available or manageable. The starting point is the obvious inequality

$$f(x_1, \ldots, x_d) \le \min(f(x_1, 0, \ldots, 0), \ldots, f(0, \ldots, 0, x_d)).$$

Setting $f_i(u) = f(0, \ldots, 0, u, 0, \ldots, 0)$ with u in the ith position for f, and using $x = (x_1, \ldots, x_d)$, we have

$$f(x) \le \min_i f_i(x_i) \le \prod_i f_i^{1/d}(x_i).$$

Thus, the following algorithm is valid for generating random variates with density f.

(algorithm universal-2)
repeat
  for $1 \le i \le n$, generate X_i with density of the form $Qf_i^{1/d}$ (for some constant Q)
  generate U uniformly on [0, 1]
until $U \prod_i f_i^{1/d}(X_i) < f(X)$
return X

The expected number of iterations until halting is $\prod_i \int f_i^{1/d}$. As an example, consider the density $f(x) = \exp-\sum_{i=1}^d x_i)$, the density of the product of d independent exponential random variables. Here $f_i(u) = \exp(-u)$ and $\int_0^\infty f_i^{1/d}(u)du = d$. Thus, the expected number of iterations before halting is d^d. If, however, f is uniform on $[0, a]^d$, then $f_i = 1/a^d$ on $[0, a]$, and the expected number of iterations is 1. Thus, the efficiency of the algorithm depends heavily on the density.

5.3 Symmetric Densities

Consider the important subclass of densities that are symmetric in all x_is. In that case, we have

$$f(x) \le g\left(\max_{1\le i\le d} x_i\right),$$

where $g(u) = f(u, 0, 0, \ldots, 0)$. The following algorithm is applicable since $g(\max_{1 \le i \le d} x_i)$ is proportional to a multivariate max density.

```
(algorithm symmetric)
repeat
  generate X = (X_1, ..., X_d) with multivariate max
    density of the form Qg(max_{1≤i≤d} x_i) (for some constant Q)
  generate U uniformly on [0, 1]
until Ug(max_{1≤i≤d} X_i) < f(X)
return X
```

The expected number of iterations is $1/Q = \int_0^\infty dt^{d-1}g(t)dt$. The algorithm above is applicable if and only if this quantity is finite.

Example 1. Take the symmetric density $f(x) = C/(1 + x_1^a + \cdots + x_d^a)$ on the positive quadrant, and note that $g(u) = C/(1 + u^a)$. For the multivariate max generator, it suffices that one can generate random variates with the univariate density proportional to $t^{d-1}/(1 + t^a)$ on $[0, \infty)$. This is only possible if $a > d$.

Example 2. Consider the symmetric density $f(x) = C \exp(-a \prod_{i=1}^d (1 + x_i))$ on the positive quadrant. Observe that $g(u) = Ce^{-a(1+u)}$, so that the univariate density for the multivariate max generator becomes proportional to $t^{d-1}e^{-at}$, which is the density of G_d/a.

Example 3. For $a > d$, consider the symmetric densities $f(x) = C/\prod_{i=1}^d (1 + x_i)^a$ and $f(x) = C/(1 + \sum_{i=1}^d x_i)^a$ on the positive quadrant. Observe that in both cases, $g(u) = C/(1 + u)^a$, and the univariate density for the multivariate max generator becomes $\xi(x) = Qx^{d-1}/(1 + x)^a$ for some constant Q. Note that ξ is proportional to the beta II density. A beta II random variate may be generated as a ratio G_d/G_{a-d} of two independent gamma random variates [Devroye 1986, p. 427].

Example 4. Consider the symmetric density $f(x) = C/\prod_{i=1}^d (1 + x_i^d)^2$ on the positive quadrant. Observe that the multivariate max generator is $\xi(x) = CQx^{d-1}/(1 + x^d)^2$, which is proportional to the density of $(U/(1 - U))^{1/d}$ [Devroye 1986, p. 437] when U is uniform [0, 1].

5.4 Bounded Densities With Some Known Moments

Assume X is orthomonotone on $[0, \infty)^d$ but possibly of infinite support. No rejection method is possible unless we are given some information about the rate of decrease of the tail of f. In particular, assume that one of the multivariate moments is known, such as

$$\mu = \mathrm{E}\{X_1^{a_1} \cdots X_d^{a_d}\}$$

for some $a_1, \ldots, a_d \geq 0$. By orthomonotonicity,

$$
\begin{aligned}
f(x) \prod_{i=1}^{d} \frac{x_i^{a_i+1}}{a_i+1} &= f(x) \int_{0 \leq y_i \leq x_i} \prod_{i=1}^{d} y_i^{a_i} \, dy_1 \ldots dy_d \\
&\leq \int_{0 \leq y_i \leq x_i} f(y) \prod_{i=1}^{d} y_i^{a_i} \, dy_1 \ldots dy_d \\
&\leq \mathrm{E}\{X_1^{a_1} \cdots X_d^{a_d}\} \\
&\quad \text{(where } (X_1, \ldots, X_d) \text{ has density } f\text{)} \\
&= \mu.
\end{aligned}
$$

Therefore,

$$
f(x) \leq \min\left(\frac{\mu \prod_{i=1}^{d} (a_i + 1)}{\prod_{i=1}^{d} x_i^{a_i+1}}, f(0) \right).
$$

Several examples may be considered:

(A) If $a_i \equiv 0$, then $\mu = 1$, and we obtain an inequality like that involving the platymorphous distribution. It is only useful if the X_is have compact support.

(B) If $a_1 = a$, $a_i = 0$ for $i > 1$, then

$$
f(x) \leq \min\left(\frac{(a+1)\mu}{x_1^a \prod_{i=1}^{d} x_i}, f(0) \right).
$$

If $\mu_i = \mathrm{E}X_i^a$ for all i, and the inequality above is applied for all i individually, we obtain the bound

$$
f(x) \leq \min\left(f(0), \min_{1 \leq i \leq d} \frac{(a+1)\mu_i}{x_i^a \prod_{j=1}^{d} x_j} \right).
$$

Case B proves useful and will be discussed in the next section.

(C) If $a_i \equiv a$, for $i \geq 1$, then

$$
f(x) \leq \min\left(\frac{(a+1)^d \mu}{\prod_{i=1}^{d} x_i^{a+1}}, f(0) \right).
$$

Unfortunately, the upper bound is not integrable in x for any value of a, so this inequality is useless for us.

Let us consider case B in more detail, and set $\mu_i = \mathrm{E}X_i^a$ for all i. Define the normalized random variables $Y_i = X_i/\mu_i^{1/a}$ so that $\mathrm{E}Y_i^a = 1$ for all i. If f is the (orthomonotone) density of $X = (X_1, \ldots, X_d)$, then $\mu f(y_1\mu_1^{1/a}, \ldots, y_d\mu_d^{1/a})$ is the (orthomonotone) density of $(Y_1, \ldots, Y_d)$, where $\mu = \Pi_i \mu_i^{1/a}$. Call this density g. If $y = (y_1, \ldots, y_d)$, then

$$g(y) \le \min\left(\mu f(0), \frac{(a+1)}{\max_i y_i^a \times \Pi_{j=1}^d y_j}\right).$$

The upper bound is proportional to the bathymorphous density $\zeta_{b,a,d}$ with $b = (a+1)/(\mu f(0))$.

We summarize the rejection algorithm based upon the above inequality:

(algorithm universal-3)
$b \leftarrow (a+1)/(\mu f(0))$
repeat
  generate U uniform [0, 1]
  generate Y with bathymorphous density $\zeta_{b,a,d}$
until $U \min(\mu f(0), [(a+1)/(\max_i Y_i^a \times \Pi_{j=1}^d Y_j)]) \le g(Y)$
  (where $g(Y) = \mu f(Y_1\mu_1^{1/a}, \ldots, Y_d\mu_d^{1/a})$)
return $X = (X_1, \ldots, X_d) \leftarrow (Y_1\mu_1^{1/a}, \ldots, Y_d\mu_d^{1/a})$

5.5 Bounded Densities With Known Moment Of A Sum

By orthomonotonicity

$$f(x)(x_1 + \cdots + x_d)^a 2^{-a} \prod_{i=1}^d x_i$$

$$\le f(x)\mathrm{E}\left\{(U_1x_1 + \cdots + U_dx_d)^a \prod_{i=1}^d x_i\right\}$$

(by Jensen's inequality if $a \ge 1$); (where $(U_1, \ldots, U_d)$ are i.i.d. uniform [0, 1])

$$= f(x) \int_{0 \le t_i \le 1} (t_1x_1 + \cdots + t_dx_d)^a \prod_{i=1}^d x_i dt_1 \ldots dt_d$$

$$= f(x) \int_{0 \le y_i \le x_i} (y_1 + \cdots + y_d)^a dy_1 \ldots dy_d$$

$$\le \int_{0 \le y_i \le x_i} f(y)(y_1 + \cdots + y_d)^a dy_1 \ldots dy_d$$

$$\le \mathrm{E}\{(X_1 + \cdots + X_d)^a\}$$

(where $(X_1, \ldots, X_d)$ has density f)

$$= \mu.$$

From this, we have

$$f(x) \le \min\left(f(0), \frac{\mu 2^a}{\prod_{i=1}^d x_i(x_1 + \cdots + x_d)^a} \right).$$

The right-hand-side is proportional to the schizomorphous density $\varphi_{b,a,d}$ with $b = 2^a \mu / f(0)$. The rejection algorithm in its entirety may be summarized as follows.

(algorithm universal-4)
$b \leftarrow 2^a \mu / f(0)$
repeat
  generate an i.i.d. uniform [0, 1] random variate W
  generate $X = (X_1, \ldots, X_d)$ with the schizomorphous density $\varphi_{b,a,d}$
until $W \min [f(0), (\mu 2^a/(\Pi X_i(\Sigma X_i)^a)] \le f(X)$
return X

5.6 Bounded Densities With Finite Moment Generating Function

Assume X is orthomonotone on $[0, \infty)^d$ and that the moment generating function ϕ is known and finite at one or more points $(t_1, \ldots, t_d)$ with nonzero components, where

$$\phi(t_1, \ldots, t_d) \stackrel{\text{def}}{=} \mathrm{E} e^{t_1 X_1 + \cdots t_d X_d}.$$

By orthomonotonicity,

$$f(x) \prod_{i=1}^d \frac{e^{t_i x_i} - 1}{t_i} = f(x) \int_{0 \le y_i \le x_i} e^{t_1 y_1 + \cdots t_d y_d} dy_1 \ldots dy_d$$

$$\le \int_{0 \le y_i \le x_i} f(y) e^{t_1 y_1 + \cdots t_d y_d} dy_1 \ldots dy_d$$

$$\le \mathrm{E} e^{t_1 Y_1 + \cdots t_d Y_d}$$

(where $(Y_1, \ldots, Y_d)$ has density f)

$$= \phi(t_1, \ldots, t_d).$$

Therefore,

$$f(x) \le \min\left(f(0),\ \phi(t_1, \ldots, t_d) \prod_{i=1}^{d} \frac{t_i}{e^{t_i x_i} - 1} \right)$$

$$\le \begin{cases} \min\left(f(0),\ \phi(t_1, \ldots, t_d) \prod_{i=1}^{d} \dfrac{1}{x_i(1 + t_1 x_i/2)} \right) & \\ \min\left(f(0),\ \phi(t, \ldots, t) \dfrac{t \max_i x_i}{(e^{t \max_i x_i} - 1) \prod_{i=1}^{d} x_i} \right) & (\text{if } t_i \equiv t > 0 \text{ for all } i). \end{cases}$$

The first two bounding functions are proportional to generalized MP densities which we call MP I and MP II, respectively. The third bounding density (for $t_i \equiv t > 0$) depends upon $\max_i x_i$ and $\prod_{i=1}^{d} x_i$ only, a property shared with the bathymorphous density. Random variate generation for f may thus be done by an algorithm that resembles (but is different from) universal-3.

Proof of the last two inequalities. Set $t_i = t$ for all i. It suffices to bound the product

$$\prod_{i=1}^{d} \frac{t}{e^{t x_i} - 1},$$

where all variables are positive. Assume without loss of generality that $x_1 = \max(x_1, \ldots, x_n)$. By Taylor's series expansion,

$$e^{t x_i} - 1 \ge t x_i(1 + t x_i/2) \ge t x_i.$$

Apply the first inequality d times to get

$$\prod_{i=1}^{d} \frac{t}{e^{t x_i} - 1} \le \prod_{i=1}^{d} \frac{t}{t x_i(1 + t x_i/2)} = \prod_{i=1}^{d} \frac{1}{x_i(1 + t x_i/2)}.$$

Apply the second inequality for $2 \le i \le d$, and obtain

$$\prod_{i=1}^{d} \frac{t}{e^{t x_i} - 1} \le \frac{t}{e^{t x_1} - 1} \prod_{i=2}^{d} \frac{t}{t x_i}$$

$$= \frac{t x_1}{e^{t x_1} - 1} \prod_{i=1}^{d} \frac{1}{x_i}$$

$$= \frac{t \max(x_1, \ldots, x_d)}{e^{t \max(x_1, \ldots, x_d)} - 1} \prod_{i=1}^{d} \frac{1}{x_i}. \qquad \square$$

5.7 Densities With Compact Support

In this section, we assume that X is orthomonotone on $[0, 1]^d$ and that additionally, f is analytically known so that $g(u) = f(u, u, \ldots, u)$ (the diagonal function) may be used in an auxiliary random variate generator. Clearly, by orthomonotonicity,

$$f(x_1, \ldots, x_d) \le g\left(\min_{1\le i\le d} x_i\right).$$

If this inequality is used in the rejection method, we need to obtain random vectors with density proportional to the multivariate min density

$$g\left(\min_{1\le i\le d} x_i\right)$$

on the unit cube. Using properties of multivariate min densities, we obtain the following algorithm:

```
(algorithm universal-5)
repeat
  generate U uniform [0, 1]
  generate Y with density proportional
    to (1 − u)^(d−1) f(u, u, . . . , u) on [0, 1]
    generate i.i.d. uniform [0, 1] random variates U_1, . . . , U_d
    set U_Z = 0 where Z is uniform on {1, . . . , d}
    for i = 1 to d, set X_i = Y + (1 − Y)U_i
until Uf(Y, Y, . . . , Y) ≤ f(X_1, . . . , X_d)
return X = (X_1, . . . , X_d)
```

6. MANY RANDOM VARIATES FROM THE SAME DISTRIBUTION

If many random variates from the same distribution are required, we may increase the efficiency in two ways:

6.1 Adaptive Table Methods

Develop increasingly better bounds for rejection as new data are generated. If $X_1, \ldots, X_n$ have been generated, and if we set $X_0 = 0$, then we may bound $f(x)$ as follows:

$$f(x) \le g_n(x) \stackrel{\text{def}}{=} \min_{i\le n: X_i \le x} f(X_i).$$

Here $\le$ denotes inequality for all d components. Assume for simplicity that f is orthomonotone with support on $[0, 1]^d$. The dominating curve is piecewise constant and thus could be dealt with by an appropriate discrete random variate generation method. Sampling is done from g_n, and the X_is used in improving g_n include all points generated thus far, both those rejected and accepted. The expected area under the dominating curve decreases to one as n tends to ∞. This fact is shown in Appendix B.

6.2 Static Table Methods Via Precomputation

Do some preprocessing and precompute f on a carefully selected grid. Assume once again that f is supported on [0, 1]. Compute f at all points

$$(i_1/n, \ldots, i_d/n),$$

where $0 \le i_j \le n$ for all j, and $n > 1$ is integer-valued. The grid consists of $(n+1)^d$ points. If grid cells are characterized by their lower left vertex, we see that exactly n^d cells are present. A table is set up (see Devroye [1986b, Ch. VIII]) from which a cell with lower left vertex x_i is selected with probability proportional to $f(x_i)$ in expected time $O(1)$. A point X is generated with a uniform distribution in the cell. Afterwards, the rejection method is applied in a routine manner. If $B = f(0)$, the expected number of iterations is

$$\begin{aligned}
\sum_i f(x_i)/n^d &= \sum_i (f(x_i) - f(y_i))/n^d + \sum_i f(y_i)/n^d \\
&\quad \text{(where } y_i \text{ is the upper right vertex of cell } i\text{)} \\
&\le \sum_{i:\ \text{some component of } x_i \text{ is } 0} f(x_i)/n^d + 1 \\
&\quad \text{(by telescoping)} \\
&\le B(n+1)^{d-1}/n^d + 1 \\
&= \frac{B}{n+1}\left(1 + \frac{1}{n}\right)^d + 1.
\end{aligned}$$

By adjusting n as a function of B, the efficiency can thus be controlled without any problems. The efficiency gain with respect to naive uniform rejection is about n (at the expense of n^d storage). In the next section, we verify that this method compares favorably with the other methods. Further speed-ups are possible by taking geometrically increasing grid cells (as described by Devroye [1986b, Ch. VIII]) and by introducing squeeze steps (see same reference).

7. EXPERIMENTAL RESULTS

Four simple experiments on three platforms show the usefulness of the above bounds. Tests A and B were carried out in an interpreted language (PostScript) as these are increasingly important and often neglected in computer studies. A Sun 4 workstation without floating point acceleration was used for test A and a Pentium 120 processor for test B. Tests C and D were programmed in C, compiled with gcc, and run on a machine with a Pentium 120 processor.

7.1 Test A: Table Methods Versus Rejection

The results are given in a general time scale relative to the expected time to generate one gamma random variate, which in turn takes 0.7 milliseconds on average. Uniform variates are obtained by the PostScript linear congruential generator ($x_{n+1} = 16807 x_n \mathrm{mod}(2^{31} - 1)$), even though we realize its limitations in real-world simulations. Exponential random variates are obtained by inversion, and gamma variates by the ratio-of-uniforms method [Devroye 1986b, pp. 197, 203].

Tests A and B are designed using the observation that if $Z = (Z_1, Z_2, \ldots, Z_d)$ is any random variable on $[0, \infty)^d$, then $(U_1 Z_1, \ldots, U_d Z_d)$ is block monotone (and thus orthomonotone) when $U_1, \ldots, U_d$ are independent uniform [0, 1] random variables.

We four test densities on $[0, 1]^3$:

(1) f_1 is block monotone with mixing random variable

$$Z = \begin{cases} [1, 1, 1] & \text{with probability } 1/2 \\ [0.1, 1, 1] & \text{with probability } 1/4. \\ [0.01, 0.1, 1] & \text{with probability } 1/4 \end{cases}$$

(2) f_2 is block monotone with mixing random variable

$$Z = \begin{cases} [1, 1, 1] & \text{with probability } 1/2 \\ [0.01, 0.1, 1] & \text{with probability } 1/2 \end{cases}.$$

(3) f_3 is orthomonotone but not block monotone, and is of the form $CI_A(x)$ for a constant C and a set

$$A = R[0.01, 1, 1] \cup R[1, 0.01, 1] \cup R[1, 1, 0.01]$$

where $R[a, b, c]$ denotes the rectangle with the origin and (a, b, c) as opposite

(4) Similarly, $f_4 = CI_A(x)$, where now

$$A = R[0.01, 0.01, 1] \cup R[0.01, 1, 0.01] \cup R[1, 0.01, 0.01].$$

The values of $f(0)$ matter a lot for most of the algorithms—they are 253, 500.5, 33.6689 and 3333.33, respectively. These numbers are roughly proportional to the difficulty of each density for the naive rejection method.

Timings are reported for the following three methods (after subtracting the time for empty loops):

(A) The naive rejection method, with uniform bound $f(x) \le f(0)$ on $[0, 1]^3$. The expected time grows as $f(0)$.

(B) Algorithm universal-1. The expected time grows as $\Sigma_{i=0}^{d} [\log^i f(0)]/i!$. This number is roughly equal to $f(0)$ unless d is much less than log $f(0)$. This is only the case for densities 1, 2 and 4.

Table IV. Mean Times per Random Variate and Set-up Times per Distribution Divided by the Mean Time per Gamma Random Variate

function	1	2	3	4
naive rejection	254.54	421.99	38.22	3778.91
guide table	5.84	9.13	14.41	156.80
guide table set-up	1586.00	1471.00	1475.00	1838.00
universal-1	201.60	269.78	134.02	603.63

(C) The table method described earlier based upon a $9 \times 9 \times 9$ grid of points. Selection of a grid cell was done via the guide table method of Chen and Asau [1974] (see Devroye [1986, pp. 96–98]). The expected time still grows proportionally with $f(0)$, but the constant is much smaller. This method requires considerable set-up times and extra storage. However, as soon as at least 10 variates with the same density are needed, the amortized set-up time becomes manageable.

It is noteworthy that one platymorphic random variate ($h_{f(0),d}$) takes between 1.5 and 4.5 time units for any value of $f(0)$ when $d = 3$.

For density f_4, the Gibbs sampler stays in one of the three arms of the distribution for on average 300 iterations. Sufficient mixing can only occur if switching between arms is repeated sufficiently often. Suppose we do this only 20 times. Then about 6000 iterations are needed to even get close to the true limit distribution. This requires about 3000 of our relative time units. All methods except naive uniform rejection are more efficient.

7.2 Test B: Comparison Between Universal Algorithms

We compared all the universal algorithms with each other. Again, we consider block monotone densities in $\mathbb{R}^5$ of the form

$$f(x) = p\,\frac{I_A(x)}{r^5} + (1-p)\sum_{i=1}^{5} \frac{I_{A_i}(x)}{r^4},$$

where $A = R[r, r, r, r, r]$, $A_1 = R[1, r, r, r, r]$, $A_2 = R[r, 1, r, r, r]$, $A_3 = R[r, r, 1, r, r]$, $A_4 = R[r, r, r, 1, r]$, $A_5 = R[r, r, r, r, 1]$, and $p \in [0, 1]$ and $r \in (0, 1)$ are parameters. A huge peak at the origin results when r is taken very small. In this manner, even though f has support in $[0, 1]^5$, we may simulate large tails.

What one finds, generally speaking, is that the timings decrease with the amount of knowledge employed. The algorithms universal-3 and universal-4 assume certain moments, but work quite generally on densities of unbounded support. Surely, these methods are the slowest, but for difficult densities, some of them become very competitive. Note in particular that different densities require different moments (parameter a in algorithm

Table V. Mean Times per Random Variate in Milliseconds per Random Vector

function (p,r)=	(0,0.5)	(0.999,0.6)	(0.7,0.1)	(0.9,0.2)
rejection	4.20e+00	5.20e+00	3.56e+04	6.68e+02
universal-1	2.21e+01	4.50e+01	1.63e+03	2.60e+02
universal-2	3.33e+01	7.40e+00	2.99e+03	1.92e+02
universal-3 (a=8)	3.95e+01	1.74e+01	1.76e+03	6.88e+01
universal-3 (a=4)	4.32e+01	4.70e+01	4.81e+02	1.34e+01
universal-3 (a=2)	2.41e+02	4.92e+02	8.10e+02	6.66e+02
universal-3 (a=1)	8.21e+03	4.44e+03	1.17e+04	5.84e+03
universal-4 (a=1)	2.59e+04	3.08e+04	3.45e+04	1.75e+04
universal-5	1.64e+01	1.51e+01	9.86e+03	1.89e+03

Table VI. Mean Times per Random Variate in Seconds. "Exact Method" Refers to the Method Based Upon the Genesis of the Distribution

dimension	d=3	d=5
exact method	1.70e-06	2.30e-06
symmetric	7.00e-05	2.13e-03
universal-3 (a=1)	6.20e-04	1.15e-01
universal-3 (a=2)	3.20e-04	1.80e-02
universal-3 (a=4)	3.70e-04	1.02e-02
universal-3 (a=8)	1.47e-03	2.54e-02
universal-4 (a=1)	1.60e-03	5.48e-01

Table VII. Mean Times per Random Variate in Seconds for the Algorithm Symmetric

dimension	time
d=3	7.00e-05
d=5	2.13e-03
d=7	7.42e-02
d=9	6.87e+00

universal-3) for optimal performance; in our case, $a = 4$ and $a = 8$ were often the best choices. The algorithms universal-2 and universal-5 assume analytic knowledge of f at some point, and have performances that are very much tied to f, lacking the robustness shown by universal-3, for example. The algorithm universal-1 requires knowledge of the compact support (which is much more than just knowledge of moments), and performs quite well for densities that are well spread out over the support.

7.3 Test C: Densities With Infinite Tails

In the third test, we see how the methods stack up for orthomonotone densities with infinite tails. The methods that apply are symmetric, universal-3 and universal-4. Uniform random variates are obtained by L'Ecuyer's combined generator [1988]. Gamma variates are obtained by Cheng's method GB [Cheng 1977]. The test density is that of $(E_1/Z_1, \ldots, E_d/Z_d)$, where $E_1, \ldots, E_d$ are i.i.d. exponential, and $(Z_1, \ldots, Z_d)$ is a vectors of ones with one "4" in a random position. The density is a mixture of densities, each of which is a product of univariate exponential densities. We will refer to it as an exponential mixture. For testing purposes and easy extensions to higher dimensions, this family is very convenient.

7.4 Test D: Expected Time Versus Dimension

In the last test, we merely point out that the expected time of our rejection algorithms typically grow very rapidly with the dimension. The example of test C is taken for dimensions 3, 5, 7 and 9. The algorithm was symmetric, as it performed well for this distribution at small dimensions. The timings show that past dimension 10, one should give serious consideration to other methods, foremost of which adaptive methods that store $(x, f(x))$ pairs for which the values are known. As this introduces interesting data structure problems, it will be dealt with elsewhere.

8. ORTHOUNIMODALITY FROM ORTHOMONOTONICITY

If $Q_1, \ldots, Q_{2^d}$ are the closed quadrants of $\mathbb{R}^d$ with center at the origin, then an orthounimodal density f with mode at the origin is almost everywhere equal to

$$\sum_{i=1}^{2^d} f(x) I_{Q_i}(x)$$

where I is the indicator function of a set. Note that on quadrant boundaries, we do not obtain f, but this does not matter as the boundaries have zero Lebesgue measure. Assume that bounding on all quadrants can be done as explained in the previous sections. Assume thus that we have $f(x) \le g(x)$ for all x, where g is obtained by piecing together 2^d functions, one for each quadrant. As g is known, we easily compute $p_i = \int_{Q_i} g$ for each $1 \le i \le 2^d$ (note again that the definition of g on quadrant boundaries is unimportant). One then applies the rejection method by noting that a random vector with density proportional to g is obtained as X_Z, where $P\{Z = i\} = p_i/\Sigma_j p_j$, and X_i is a random vector on the ith quadrant Q_i with density proportional to gI_{Q_i}.

As an example, consider the orthounimodal bivariate density $1/(x_1^4 + x_2^4 + 1)$. By symmetry, if we have $f(x) \le g(x)$ on the quadrant $Q_1 = \{x_1 \ge 0, x_2 \ge 0\}$, then $f(x) \le g(|x|)$ everywhere, where $|x| \stackrel{\text{def}}{=} (|x_1|, |x_2|)$. In this case, all p_is are equal. It suffices to generate X with orthomonotone density fI_{Q_1}, and to randomly and uniformly flip the signs of the components of X.

SUMMARY AND FUTURE WORK

We introduced various general algorithms for generating random variates with orthounimodal densities. All are based upon probability inequalities that require minimal information on the part of the user. For dimensions up to 10, the algorithms are shown to be efficient. For higher dimensions, other methods must be developed. We believe that the most promising among the exact methods are the adaptive methods that make clever use of all known $(x, f(x))$ pairs, perhaps using quadtrees or other multivariate tree structures to partition the space up into pieces.

The present article shows how orthounimodality may be dealt with. Convex unimodality and log concavity require separate study. Further research is also needed in the exploitation of symmetries that occur frequently in multivariate densities. For example, how would one generate random variates when the densities are of the form

$$\psi\left(\sum_{i \neq j} x_i x_j\right)?$$

APPENDIX A. Proof of Algorithm gdb

Lt G^*_{d+1} denote a gamma random variate with parameter $d + 1$. Set $G^*_d = U^{1/d} G^*_{d+1}$, where U is uniform [0, 1]. Observe that G^*_d is gamma (d) by a well-known property of the gamma distribution (see Devroye [1986, Sect. IV.6.4.]. The random variate G_{d+1} in the algorithm may be thought of as G^*_{d+1} given $G^*_{d+1} \geq \log b$. With this conditional distribution in mind, we have for $u > 0$.

$$\begin{aligned}
\mathrm{P}\{X \leq u\} &= \mathrm{P}\{G_d < \log b,\ V^{1/d} \log b \leq u\} + \mathrm{P}\{G_d \geq \log b,\ G_d \leq u\} \\
&= \frac{\mathrm{P}\{G^*_d < \log b,\ V^{1/d} \log b \leq u,\ G^*_{d+1} \geq \log b\}}{\mathrm{P}\{G^*_{d+1} \geq \log b\}} \\
&\quad + \frac{\mathrm{P}\{G^*_d \geq \log b,\ G^*_d \leq u,\ G^*_{d+1} \geq \log b\}}{\mathrm{P}\{G^*_{d+1} \geq \log b\}} \\
&= \mathrm{P}\{V^{1/d} \log b \leq u\} \mathrm{P}\{G^*_d < \log b \mid G^*_{d+1} \geq \log b\} \\
&\quad + \frac{\mathrm{P}\{u \geq G^*_d \geq \log b\}}{\mathrm{P}\{G^*_{d+1} \geq \log b\}}.
\end{aligned}$$

If $u < \log b$, only the first term is nonzero, while if $u \geq \log b$, the first term reduces to $p = \mathrm{P}\{G^*_d < \log b \mid G^*_{d+1} \geq \log b\}$, which is independent of u. The density of X is obtained by taking a derivative with respect to u. First, for

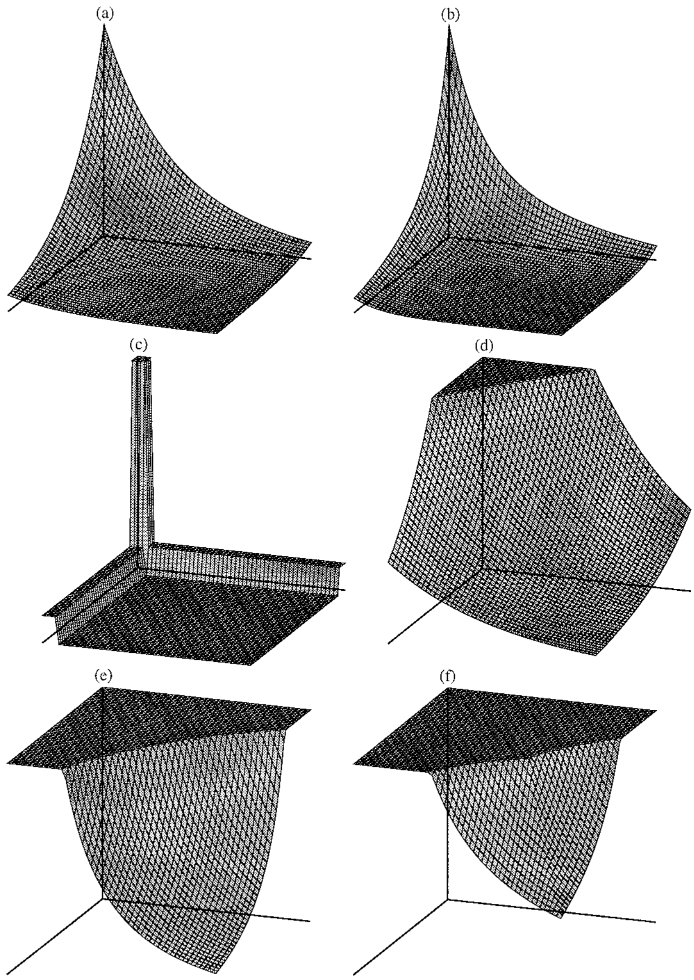

Fig. 1. Six orthomonotone functions are shown: (a) and (b) show two exponential mixture densities; (c) shows a density on the unit rectangle of the form tested in test B; (d) is a platymorphous density; (e) is a schizomorphous density; and (f) is a bathymorphous density. To some extent, the shapes explain our nomenclature.

$u < \log b$, the density is $pz(u)$, where $z(u) = \mathrm{du}^{\mathrm{d}-1}/\log^d b$ is a valid density on $[0, \log b)$. But

$$\mathrm{P}\{G_d^* < \log b,\ G_{d+1}^* \geq \log b\} = \mathrm{P}\{G_d^* < \log b\} - \mathrm{P}\{G_{d+1}^* < \log b\}$$

$$= \sum_{i=0}^{d} \frac{\log^i b e^{-\log b}}{i!} - \sum_{i=0}^{d-1} \frac{\log^i b e^{-\log b}}{i!}$$

$$= \frac{\log^d b}{bd!} = \frac{I_1}{b}.$$

As $P\{G^*_{d+1} \geq \log b\} = (I_1 + I_2)/b$, we see that $p = I_1/(I_1 + I_2)$.

Next, for $u \geq \log b$, the density is the gamma (d) density divided by $(I_1 + I_2)/b$. Equivalently, it is $I_2/(I_1 + I_2)$ times the tail-gamma (d) density on $[\log b, \infty)$. Thus, the density of X coincides with $g_{b,d}$. $\square$

APPENDIX B. Convergence Result for the Adaptive Sampler

We show that for each $x \in (0, 1)^d$,

$$\limsup_{n\to\infty} \mathrm{E}g_n(x) = f(x).$$

As $g_n(x)$ is monotone in n, this implies

$$\limsup_{n\to\infty} \int \mathrm{E}g_n(x)dx = \int f(x)dx = 1.$$

Let $\epsilon > 0$ be small and at the very least smaller than the smallest component of x. Let A_n be the event that X_n falls in the rectangle with $x - \epsilon$ and x as its opposite vertices where ϵ is the d-vector all of whose components are ϵ. Then

$$g_n(x) \leq g_{n-1}(x)(1 - I_{A_n}) + I_{A_n} f(x - \epsilon).$$

But if $\mathcal{F}_{n-1}$ denotes the history up to time $n - 1$, then $P\{A_n | \mathcal{F}_{n-1}\} \geq g_{n-1}(x)\epsilon^d / \int g_{n-1}$. Note that $\int g_{n-1} \leq \int g_0 = B$ if we start with the uniform density on the unit square. Thus, since $g_0(x) \geq f(x - \epsilon)$, the sequence $g_n(x)$ can only decrease monotonically.

$$\begin{aligned} P\left\{\liminf_{n\to\infty} g_n(x) > f(x - \epsilon) + 1/k\right\} &\leq P\{\cap_{n=1}^{\infty} A_n^c\} \\ &\leq \prod_{n=1}^{\infty}\left(1 - \frac{(f(x-\epsilon) + 1/k)\epsilon^d}{B}\right) \\ &= 0. \end{aligned}$$

Therefore, by the arbitrary nature of k and δ, we conclude that $\limsup_{n\to\infty} g_n(x) \leq f(x - \epsilon)$ almost surely. Thus, by the arbitrary nature of ϵ, and monotone convergence,

$$\limsup_{n\to\infty} \mathrm{E}g_n(x) \leq \inf_{y<x} f(y).$$

Since $\inf_{y<x} f(y) = f(x)$ at almost all x,

$$\limsup_{n\to\infty} \mathrm{E} \int g_n(x)dx \le \int f(x)dx = 1.$$

Thus, the acceptance rate increases to one monotonically with n.

APPENDIX C. Validity of the Generator for Schizomorphous Density

If we define $S = \sum_{i=1}^{d} X_i$, $X_i = T_i S$ for all i, with $0 \le T_i \le 1$, $\sum_{i=1}^{d} T_i = 1$, then $(S, T_1, \ldots, T_{d-1})$ has the following density on $[0, \infty) \times \mathcal{T}$, where $\mathcal{T}$ is the simplex of $\mathbb{R}^{d-1}$ defined by $0 \le t_i \le 1$, $\Sigma_i t_i \le 1$:

$$f'(s, t_1, \ldots, t_{d-1}) \stackrel{\text{def}}{=} \varphi_{b,a,d}(t_1 s, \ldots, t_{d-1}s, (1 - t_1 - \cdots - t_{d-1})s) \times s^{d-1}$$

$$= cs^{d-1} \min\left(1, \frac{b}{\prod_{i=1}^{d-1} t_i(1 - \sum_{i=1}^{d-1} t_i)s^{d+a}}\right).$$

Random variate generation can now proceed by the conditional method [Johnson 1987]. Taking the integral with respect to s in the upper bound yields a function of $\prod_{i=1}^{d-1} t_i(1 - \sum_{i=1}^{d-1} t_i)$ only (see Lemma 4.1):

$$\frac{d+a}{ad} (\varphi_{b,a,d}(0))^{a/(d+a)} \left(\frac{b}{\prod_{i=1}^{d-1} t_i(1 - \sum_{i=1}^{d-1} t_i)}\right)^{d/(d+a)}, \ t_i \ge 0, \ \sum_i t_i \le 1.$$

This function is proportional to the Dirichlet density with parameters $a/(d + a)$ for all variables [Devroye 1986b, Sect. XI.4.1]. We generate $(T_1, \ldots, T_{d-1})$ as

$$\frac{G_1}{G}, \ldots, \frac{G_{d-1}}{G},$$

where the G_is are independent gamma $(a/(d + a))$ random variables, and $G = \sum_{i=1}^{d} G_i$. Furthermore, it is known that G is independent of $(T_1, \ldots, T_{d-1})$ if we do this. For later reference, we define $T_d = 1 - \sum_{i<d} T_i = G_d/S$. Fixing $T = \prod_{i=1}^{d-1} T_i(1 - \sum_{i=1}^{d-1} T_i)$, we then generate S with density proportional to

$$s^{d-1} \min\left(1, \frac{b}{Ts^{d+a}}\right).$$

By Lemma 4.1, it is easily seen that

$$S \stackrel{\mathcal{L}}{=} \left(\frac{b}{T}\right)^{1/(a+d)} \times \frac{U^{1/d}}{V^{1/a}}$$

where U, V are i.i.d. uniform [0, 1].

ACKNOWLEDGMENTS

The author would like to thank the area editor and both referees for their invaluable and detailed feedback.

REFERENCES

AHRENS, J. H., KOHRT, K. D., AND DIETER, U. 1983. Algorithm 599. Sampling from gamma and Poisson distributions. *ACM Trans. Math. Softw. 9,* 255–257.

ARNOLD, B. C., CASTILLO, E., AND SARABIA, J.-M. 1992. Conditionally specified distributions. Lecture Notes in Statistics, *73,* Springer-Verlag, Berlin.

ARNOLD, B. C. AND STRAUSS, D. 1988. Bivariate distributions with exponential conditionals. *J. Am. Stat. Assoc. 83,* 522–527.

BARBU, GH. 1987. A new fast method for computer generation of gamma and beta random variables by transformations of uniform variables. *Statistics 18,* 453–464.

BÉLISLE, C. J. P., ROMEIJN, H. E., AND SMITH, R. L. 1993. Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res. 18,* 255–266.

BHATTACHARYYA, B. C. 1942. The use of McKay's Bessel function curves for graduating frequency distributions. *Sankhya Series A, 6,* 175–182.

CHEN, H. C. AND ASAU, Y. 1974. On generating random variates from an empirical distribution. *AIIE Transactions 6,* 163–166.

CHENG, R. C. H. 1977. The generation of gamma variables with non-integral shape parameter. *Appl. Stat. 26,* 71–75.

CHENG, R. C. H. AND FEAST, G. M. 1977. Some simple gamma variate generators. *Appl. Stat. 28,* 290–295.

CHENG, R. C. H. AND FEAST, G. M. 1980. Gamma variate generators with increased shape parameter range. *Commun. ACM 23,* 389–393.

DEVROYE, L. 1984a. Methods for generating random variates with Polya characteristic functions. *Stat. Probab. Lett. 2,* 257–261.

DEVROYE, L. 1984b. Random variate generation for unimodal and monotone densities. *Computing 32,* 43–68.

DEVROYE, L. 1984c. A simple algorithm for generating random variates with a log-concave density. *Computing 33,* 247–257.

DEVROYE, L. 1984d. On the use of probability inequalities in random variate generation. *J. Stat. Comput. Simul. 20,* 91–100.

DEVROYE, L. 1986a. An automatic method for generating random variables with a given characteristic function. *SIAM J. Appl. Math. 46,* 698–719.

DEVROYE, L. 1986b. *Non-Uniform Random Variate Generation.* Springer-Verlag, New York.

DEVROYE, L. 1989. On random variate generation when only moments or Fourier coefficients are known. *Math. Comput. Simul. 31,* 71–89.

DEVROYE, L. 1991. Algorithms for generating discrete random variables with a given generating function or a given moment sequence. *SIAM J. Sci. Stat. Comput. 12,* 107–126.

DHARMADHIKARI, S. AND JOAGDEV, L. 1988. *Unimodality, Convexity and Applications*. Academic Press, New York.

L'ECUYER, P. 1988. Efficient and portable combined random number generators. *Commun. ACM. 31,* 742–749, 774.

GILKS, W. R. AND WILD, P. 1992. Adaptive rejection for Gibbs sampling. *Appl. Stat. 41,* 337–348.

GUMBEL, E. J. 1960. Bivariate exponential distributions. *J. Am. Stat. Assoc. 55,* 698–707.

JOHNSON, M. E. 1987. *Multivariate Statistical Simulation.* Wiley, New York.

ROBERTS, G. O. AND POLSON, N. G. 1994. On the geometric convergence of the Gibbs sampler. *J. Royal Stat. Soc. B56*, 377–384.

SHEPP, L. 1962. Symmetric random walk. *Trans. Am. Math. Soc. 104,* 144–153.

SIVAZLIAN, B. D. 1981a. On a multivariate extension of the gamma and beta distributions. *SIAM J. Appl. Math. 41,* 205–209.

SIVAZLIAN, B. D. 1981b. A class of multivariate distributions. *Australian J. Stat. 23,* 251–255.

STADLOBER, E. AND KREMER, R. 1992. Sampling from discrete and continuous distributions with C-RAND. In *Simulation and Optimization*, G. Pflug and U. Dieter, Eds. *374,* 154–162. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin.

VON NEUMANN, J. 1963. Various techniques used in connection with random digits. *Collected Works 5,* Pergamon Press, 768–770. (Also in Monte Carlo Method, National Bureau of Standards Series, Vol. 12, 36–38, 1951).

ZECHNER, H. AND STADLOBER, E. 1992. Generating beta variates via patchwork rejection. *Computing, 43,* 1–19.

Received April 1995; revised February 1997; accepted March 1997