J. Statist. Comput. Simul., 1984, Vol. 20, pp. 91–100
0094-9655/84/2002-0091 $18.50/0

On the Use of Probability Inequalities in Random Variate Generation

LUC DEVROYE
McGill University, Montreal, Canada

(*Received January 5, 1984*)

We show by example how various well-known probability inequalities can aid in the design of general nonuniform random variate generators, i.e. generators that can be used for large classes of densities. We obtain universal algorithms for the following classes: (1) all unimodal densities with mode at 0 and absolute rth moment not exceeding a known constant; (2) all densities satisfying a Lipschitz condition with known constant, having an absolute rth moment or a moment generating function in a neighborhood of 0 not exceeding a known constant. The algorithms assume that the density f can be computed at every point, but the distribution function is not needed.

KEY WORDS: Probability inequalities; random variate generation; simulation; acceptance/rejection method; unimodality; Chebyshev's inequality; algorithms.

1. INTRODUCTION

The general principles of non-uniform random variate generation (inversion, acceptance/rejection, acceptance/complement, aliasing, guide tables, etc.) have been successfully applied in the design of efficient algorithms for selected families of distributions, such as the normal, gamma, beta, Poisson, binomial and Student's t families. Statisticians, in their continuous quest for better models have come up with ever more sophisticated families of distributions to fit "real-life" situations. These can be separated into two overlapping groups:

i) The genesis of the distribution is simple to describe, e.g. it is a discrete or continuous mixture of simpler distributions (and is then called a "generalized" distribution), or the random variable can be written as a simple function of other random variables.

ii) The distribution itself is simple to describe, e.g. the density has a simple analytical form. In most cases, the parameters, moments and/or quantiles are linked by simple equations.

Case (i) is sometimes introduced artificially to facilitate random variate generation (see for example Schmeiser (1977) for a survey). In case (ii), the simplicity forced upon the model creates difficulties from the random variate generation point of view. In fact, the number of distributions of type (ii) has proliferated to the point that it is unreasonable to expect that statistical packages and computer libraries carry efficient procedures for all of them. Thus, there is a need for acceptably efficient algorithms that are general enough to be applied to an entire group of distributions.

The inversion method is of course truly universally applicable, but its application is limited by the fact that often a density f is given and that the distribution function F can only be obtained by numerical integration. In this note, we will assume throughout that the density f can be computed for a certain distribution. We will show, mainly by example, how one can go about developing general algorithms for groups of distributions. General groups of interest to the average user include

A. All monotone densities on $[0, \infty)$.
B. All unimodal densities on R with mode at m.
C. All monotone densities on $[0, \infty)$ with known rth moment μ_r.
D. All log-concave densities on R with mode at m.
E. All densities with a decreasing hazard rate (DHR).
F. All densities with an increasing hazard rate (IHR).
G. All densities with one ore more known moments, satisfying a known Lipschitz condition.

Some of these groups can be handled by employing special properties: for group E, we can make use of a variation of the thinning algorithm of Lewis and Shedler (1979) (see Devroye (1983)). For groups A and B, a combination of the inversion and rejection methods yields fast algorithms whenever the distribution function F is easy to compute (Devroye, 1984). For group D, a general bound for all log-concave densities in terms of m and $f(m)$ has led to the development of fast universal rejection algorithms (Devroye, 1984). In this note, we will illustrate for groups A–C and G how some well-known facts from probability theory (such as Chebyshev's inequality) can help in the design of generally applicable algorithms. We assume that the reader is familiar with the inversion and acceptance/rejection methods in random variate generation (see for example some standard texts in simulation, e.g. Bratley, Fox and Schrage (1983, Ch. 5), Fishman (1978, Ch. 9), Kennedy and Gentle (1980, Ch. 6), Law and Kelton (1982, Ch. 7) or Rubinstein (1981)).

2. MONOTONE DENSITIES

Consider the following group of densities:

$$G_r = \{f : f \text{ is monotone on } [0, \infty), \text{ bounded (by } b = f(0));$$
$$\int x^r f(x)\,dx \leq \mu_r < \infty\}, \qquad r > 0.$$

We assume that μ_r is explicitly known.

PROPERTY 2.1 *For every $f \in G_r$, we have*

$$f(x) \leq \min(b, c/x^{r+1}) = g(x),$$

where $b = f(0)$ and $c = (r+1)\mu_r$. The area $\int_0^\infty g(x)\,dx$ equals

$$\left(1 + \frac{1}{r}\right)(r+1)^{1/(r+1)}(b^r \mu_r)^{1/(r+1)}.$$

Proof By assumption, $f(x) \leq b$. Also, $\mu_r \geq \int_0^x y^r f(y)\,dy \geq f(x)x^{r+1}/(r+1)$. Let x^* be the solution of $b = c/x^{r+1}$. Then the area under g is

$$bx^* + c/(rx^{*r}) = b(c/b)^{1/(r+1)} + c(b/c)^{r/(r+1)}/r$$
$$= \left(1 + \frac{1}{r}\right) b^{r/(r+1)} c^{1/(r+1)}.$$

This property is all that is needed to write a general purpose algorithm for G_r. We could for example proceed as follows:

(*Preprocessing.*) Let $b \leftarrow f(0), x^* \leftarrow ((r+1)\mu_r/b)^{1/(r+1)}$.
(*Generation.*) *Repeat* Generate U, V independent uniform $[0,1]$ random variates.
If $U \leqq (r/(r+1))$ *then* $X \leftarrow U(r/(r+1))x^*$
(X is uniform on $[0, x^*]$)
else $X \leftarrow x^*/(U(r+1)-r)^{1/r}$
($(r+1)U-r$ is again uniform on $[0,1]$, and thus X has density proportional to $1/x^{r+1}$ on $[x^*, \infty)$).
Until $Vg(X) \leqq f(X)$.
(Note that $g(X)=b$, $X \leqq x^*$;
$g(X)=(r+1)\mu_r/X^{r+1}$, $X > x^*$.)
Exit with X.

In the *Repeat* loop, X can also be obtained as $U_1 x^*/U_2^{1/r}$, where U_1, U_2 are independent uniform $[0,1]$ random variates.

The expected number of executions of the *Repeat* loop is $\int_0^\infty g(x)\,dx$, which is given in Property 2.1. We will write this number as

$$A_r B_r, \text{ where } A_r = 1 + \frac{1}{r},\ B_r = ((r+1)b^r \mu_r)^{1/(r+1)}.$$

The factor B_r is scale invariant and small for most densities that are of interest to the general user. In particular, we have:

PROPERTY 2.2

A. $\inf_{f \in G_r} B_r = 1$.

B. $\sup_{f \in G_r,\ f \text{ concave on its support.}} B_r = 2/(r+2)^{1/(r+1)} \leqq 2$.

C. $\sup_{f \in G_r,\ f \text{ log-concave.}} B_r = (\Gamma(r+2))^{1/(r+1)} \sim \frac{r+1}{e}$ *as* $r \to \infty$.

Proof Statement A follows from the fact that if we force b to be 1, then μ_r is minimized by taking the uniform density f on $[0,1]$. Under the same restriction on b, we maximize μ_r among all densities f that are concave on their support by taking $f(x) = 1-(x/2)$,

$0 \leqq x \leqq 2$. Thus, for this f,

$$\mu_r = \int_0^2 x^r \left(1 - \frac{x}{2}\right) dx = 2^{r+1}\left(\frac{1}{r+1} - \frac{1}{r+2}\right) = 2^{r+1}\frac{1}{(r+1)(r+2)}.$$

Statement B follows directly.

Finally, when f is log-concave on its support, i.e. $\log f$ is concave on $[0,\infty) \cap \{x : f(x) > 0\}$, and f is monotone down, then bX is stochastically smaller than an exponential random variate Y (this follows from a trivial geometric argument based upon the log-concavity of f). In particular, $E(b^r X^r) = b^r \mu_r \leqq E(Y^r) = \Gamma(r+1)$. Statement C now follows without further work.

A short discussion of Property 2.2 is in order here. First, we note that $\int g \geqq 1 + (1/r)$ for all $f \in G_r$. Thus, the algorithm cannot possibly be efficient for r near 0, at least not uniformly over all f in G_r. For quite a few densities, $A_r B_r$ is sufficiently close to 1 for the algorithm to be useful. When f is concave on its support, then the upper bound for $A_r B_r$ obtained by using Property 2.2B varies from $4/\sqrt{3} = 2.3094\ldots (r=1)$ to $3/4^{1/3} = 1.88988\ldots (r=2)$ to $8/(3.5^{1/4}) = 1.7833\ldots (r=3)$ to $5/(2.6^{1/5}) = 1.7470\ldots (r=4)$ to $12/(5.7^{1/6}) = 1.7352\ldots (r=5)$ to $7/(3.8^{1/7}) = 1.73366\ldots (r=6)$ to $16/(7.9^{1/8}) = 1.7367\ldots (r=7)$ and monotonically back up to 2 as $r \uparrow \infty$. Thus, the upper bound is minimal for $r=6$: our algorithm is guaranteed to perform at its best when the sixth moment is known. Unfortunately, very few densities are concave on their support. In contrast, the class of log-concave densities that are monotone on $[0,\infty)$ includes the halfnormal, gamma, Weibull, generalized inverse gaussian, exponential power, logistic and hyperbolic secant densities (see Devroye, 1984). From the proof of Property 2.2C we see that the worst density in this class is the exponential density, and that it is possible that $A_r B_r$ is large when r is large. The upper bound for $A_r B_r$ now varies from $2.2^{1/2} = 2.82\ldots (r=1)$ to $(3/2)\cdot 6^{1/3} = 2.7256\ldots (r=2)$ to $(4/3)\cdot 24^{1/4} = 2.9511\ldots (r=3)$ and monotonically up to ∞ as $r \uparrow \infty$. Here the optimal value is $r=2$. We note that the average time of our algorithm is uniformly bounded over all log-concave densities for any value of r. If μ_r is not known, we can replace it by the larger quantity $\Gamma(r+1)/b^r$ in the algorithm while keeping the upper bound for $A_r B_r$. For $r=2$, the algorithm comes very close in performance to the algorithm of Devroye (1984) that was specially designed for log-concave densities *only*.

From this discussion, we gather that the value $r=2$ is optimal or nearly optimal for most monotone densities of general interest. The *If* statement in the algorithm can of course be improved for such special cases. Thus, in

$$\textit{If } U \leqq \tfrac{2}{3} \textit{ then } X \leftarrow \tfrac{2}{3}\, Ux^* \textit{ else } X \leftarrow x^*/\sqrt{3U-2}$$

we can replace the *else* part by *else* $X \leftarrow x^*/\max(3U-2, W)$ where W is another uniform $[0,1]$ random variate, thereby avoiding the costly square root. Such streamlining is normally not possible for non-integer values of r.

We could also use functionals of f instead of moments to get bounds, as in the obvious inequality $f(x) \leqq (\delta f^{\alpha})^{1/\alpha}/x^{1/\alpha}$, valid for all monotone f on $[0,\infty)$, and all $\alpha > 0$.

3. DENSITIES SATISFYING A LIPSCHITZ CONDITION

We say that f is Lipschitz (C) (and write $f \in \mathrm{Lip}(C)$) when

$$\sup_{x,y} |f(x) - f(y)| \leqq C|x-y|.$$

The constant C, or an upper bound for it, is known, if we know for example an upper bound for $\sup_x |f'(x)|$.

PROPERTY 3.1 *For every $f \in \mathrm{Lip}(C)$, we have*

$$f(x) \leqq \sqrt{2C \min(F(x), 1-F(x))} = g(x),$$

where F is the distribution function corresponding to f.

Proof By a geometrical argument, we see that the triangle with base $[x, x+(y/C)]$ and $f(x) = y$, $f(x+(y/C)) = 0$ is the function f (not necessarily density function) yielding the largest value of y at x while not violating the Lipschitz condition. But the area of the triangle must be at most $1-F(x)$. Thus, since the area is $y^2/(2C)$, we have $y \leqq \sqrt{2C(1-F(x))}$. Property 3.1 now follows by symmetry.

Remark If we merely satisfy the Lipschitz condition on $(0,\infty)$ (and thus, may have a discontinuity at 0), it is still true that $f(x) \leqq \sqrt{2C(1-F(x))}$, $x > 0$.

Example 3.1 (*Convex densities on* $[0,\infty)$) Because $C \leqq f'(0)$, the bound becomes

$$f(x) \leqq \sqrt{2f'(0)(1-F(x))}.$$

Because convex densities on $[0, \infty)$ are necessarily monotone, we can employ inequalities for $1-F(x)$ to obtain a useful upper bound for random variate generation. We have for example,

$$1-F(x) \leqq \left(\frac{r}{r+1} \cdot \frac{\mu_r}{x}\right)^r, \qquad \text{all } x, r > 0$$

(*Narumi's inequality*, see Savage (1961) for a discussion and Andreev (1981) for a proof), valid for all monotone f on $[0, \infty)$. The difference with Chebyshev's inequality is the factor $(r/(r+1))$. Thus, combining inequalities gives the bound

$$f(x) \leqq g(x) = \min\left(f(0), \left(\frac{r}{r+1} \cdot \frac{\mu_r}{x}\right)^{r/2} \sqrt{2f'(0)}\right).$$

This form is similar to the form of g for monotone densities (Property 2.1). Thus, if we know $f(0)$, $f'(0)$ and μ_r, we can produce a simple rejection algorithm for this class of densities. It should be noted that for g to be integrable, we must have $r > 2$.

Example 3.2 (*Densities with known moment generating function*) The moment generating function for a random variable X is defined by $M(t) = E(e^{tX})$, $t \in R$ (see Patel, Kapadia and Owen (1976) for several examples). By *Jensen's inequality* $1-F(x) \leqq M(t)e^{-tx}$, $t > 0$, we obtain

$$f(x) \leqq g(x) = \begin{cases} \sqrt{2Ce^{-t|x|}M(t)}, & x \geqq 0, \\ \sqrt{2Ce^{-t|x|}M(-t)}, & x < 0. \end{cases}$$

This inequality is valid for all $t > 0$ for which $M(t)$ and $M(-t)$ are finite. To minimize the upper bound, we could attempt to find the optimal t for each x. This would require the knowledge of $M(t)$ in an analytically manageable form of course, an assumption which we are not willing to make.

Let us consider an important special case: f is symmetric about 0.

Then the bound becomes

$$f(x) \leqq g(x) = \sqrt{32CM(t)/t^2} \cdot \left(\frac{t}{4} \exp\left(-\frac{t}{2}|x| \right) \right), \qquad t \geqq 0.$$

The function in parentheses is a Laplace density function with parameter $t/2$. The algorithm suggested in this section thus becomes very simple:

Repeat Generate E, U, independent exponential and uniform $[0,1]$ random variates. Set $E \leftarrow 2E/t$.

Until $\sqrt{2CM(t)}\, U \exp\left(-\frac{t}{2} E \right) \leqq f(E)$.

Exit with $X \leftarrow SE$ where S is a random sign.

The expected number of iterations is the area under g, i.e. $\sqrt{32CM(t)/t^2}$. Thus, the best value for t is the value that minimizes $M(t)/t^2$. We would like to point out that the smaller $M(t)$ and the smaller C, the faster the algorithm. In other words, speed is linked to small tails (small $M(t)$) and smoothness (small C).

Example 3.3 (*The generalized gaussian family*) The generalized gaussian family of distributions contains all distributions for which for some $s \geqq 0, M(t) \leqq \exp(s^2t^2/2)$, all t. (Chow, 1966). The mean of all these distributions exists and is zero. One important property of this family is that both $1-F(x)$ and $F(-x)$ do not exceed $\exp(-x^2/(2s^2))$, $x>0$ (Chow, 1966). Thus, we have

$$f(x) \leqq g(x) = \sqrt{2C}\, s\sqrt{4\pi} \cdot \left(\frac{1}{s\sqrt{4\pi}} \exp(-x^2/(4s^2)) \right).$$

The function in parentheses is a normal $(0, s\sqrt{2})$ density. Our algorithm now reads:

Repeat Generate N, E, independent normal and exponential random variates. Set $X \leftarrow Ns\sqrt{2}$.

Until $-N^2/2 - E \leqq \log(f(X)/\sqrt{2C})$.

Exit with X.

Example 3.4 (*Densities with known moments*) Examples 3.1–3.3 assume quite a bit of knowledge about the distribution besides C. In the majority of the cases, if we know one quantity, it is a moment, say $\mu_r = E(|X|^r)$. Assume that we do not know that f is unimodal, for if we do, we can refer to the method of Section 2. *Chebyshev's inequality* $1-F(x) \leqq \mu_r/|x|^r$, $r>0$, leads to the fundamental inequality

$$f(x) \leqq g(x) = \sqrt{2C}\min(1, \sqrt{\mu_r}/|x|^{r/2}),$$

which is only useful to us for $r>2$ (otherwise g is not integrable). For random variate generation, we need only a small modification of the algorithm of Section 2 because the form of g is similar to that discussed there. The integral of g over R is easily computed:

$$\int g = \sqrt{8C}\frac{r}{r-2}\mu_r^{1/r}.$$

Note that $r/(r-2)$ decreases monotonically to 1 and that $\mu_r^{1/r}$ is nondecreasing as $r \uparrow \infty$ (the last fact follows directly from Jensen's inequality). Thus, to obtain the best value for r, we need a careful compromise.

Example 3.5 (*Log-concave densities*) If we only know that f is log-concave with a mode at 0 and support on $[0, \infty)$, then we can use the inequality of Example 3.4. Indeed, we do know that $f(0)^r \mu_r \leqq \Gamma(r+1)$. In fact, we know much more: $1-F(x) \leqq \exp(-xf(0))$: thus,

$$f(x) \leqq g(x) = \frac{\sqrt{8C}}{f(0)} \cdot \left(\frac{f(0)}{2}\exp\left(-\frac{f(0)}{2}x\right)\right), \qquad x>0.$$

The algorithm can be written as follows:

Repeat Generate E_1, E_2, independent exponential random variates.
Set $X \leftarrow E_1 \cdot (2/f(0))$.
Until $-E_2 - E_1 \leqq \log(f(X)/\sqrt{2C})$.
Exit with X.

The expected number of loops is $\sqrt{8C}/f(0)$.

References

Andreev, N. I. (1981). A method of obtaining Chebyshev-type inequalities. *Automation and Remote Control* **42,** 594–603.

Bratley, P., Fox, B. L. and Schrage, L. E. (1983). *A Guide to Simulation.* Springer-Verlag, New York.

Chow, Y. S. (1966). Some convergence theorems for independent random variables. *Annals of Mathematical Statistics* **37,** 1482–1493.

Devroye, L. (1984). Random variate generation for unimodal and monotone densities. *Computing* **32,** 43–68.

Devroye, L. (1983). The analysis of some algorithms for generating random variates with a given hazard rate. Technical Report, School of Computer Science, McGill University, Montreal.

Devroye, L. (1984). A simple algorithm for generating random variates with a log-concave density. *Computing* **32,** to appear.

Fishman, G. S. (1978). *Principles of Discrete Event Simulation.* John Wiley, New York.

Kennedy, W. J. and Gentle, J. E. (1980). *Statistical Computing.* Marcel Dekker, New York.

Law, A. M. and Kelton, D. E. (1982). *Simulation Modeling and Analysis.* McGraw-Hill, New York.

Lewis, P. A. W. and Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* **26,** 403–413.

Patel, J. K., Kapadia, C. H. and Owen, D. B. (1976). *Handbook of Statistical Distributions.* Marcel Dekker, New York.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method.* John Wiley, New York.

Savage, I. R. (1961). Probability inequalities of the Thebycheff type. *J. Res. Nat. Bur. Stand.* **65,** 211–222.

Schmeiser, B. (1977). Methods for modeling and generating probabilistic components in digital computer simulation when the standard distributions are not adequate: a survey. *Proceedings of the 1977 Winter Simulation Conference*, Orlando, Florida.