ON GLOBAL COSTS AND NYQUIST'S THEOREM IN RANDOM VARIATE GENERATION*†

LUC DEVROYE

McGill University

We consider a large class of densities defined in terms of smoothness and tail conditions. Assume that we want to generate n iid random variables from a given density f in this class, and that the global cost of the generator is equal to the total number of evaluations of f. We demonstrate with the help of several examples how one can proceed to make the expected cost grow at a sublinear ($o(n)$) rate. Examples include the class of Lipschitz densities on $[0,1]$ with known Lipschitz constant, the class of bounded monotone densities on $[0,1]$, and the class of all densities with a characteristic function of bounded support and kth moment bounded by a given constant. In the last example, we proceed to show how Nyquist's theorem can be exploited to yield a generator with $O(n^{1/(k-1)})$ expected cost.

1. Global cost.

We consider a large class of densities defined in terms of smoothness and tail conditions. Assume that we want to generate n iid random variables from a given density f in this class, and that the global cost of the generator is equal to the total number of evaluations of f. This cost measure is appropriate when f is assumed to be present in a black box, and evaluations of f are time-consuming. In any case, modulo proportionality constants, the real time cost of all generators is bounded from below by the global cost as defined here.

We demonstrate with the help of several examples how one can proceed to make the expected cost grow at a sublinear ($o(n)$) rate. Examples include the class of Lipschitz densities on $[0,1]$ with known Lipschitz constant, and the class of all densities with a characteristic function of bounded support and kth moment bounded by a given constant. In the last example, we proceed to show how Nyquist's theorem can be exploited to yield a generator with $O(n^{1/(k-1)})$ expected cost.

It is important to note that the proposed methods can be used on densities for which we may already have good or expedient random variate generators. The novelty is only in the way the expected global cost can be made small for large n.

The examples chosen in this paper illustrate two guiding principles that can be used in the design of efficient generators. In the case of the Lipschitz class, information about the value of $f(x)$ can be inferred from the values of f at nearby points, because the density is known to be smooth. The Lipschitz class clarifies what happens when classes are defined in terms of smoothness properties. It is also possible to deduce information about the value of $f(x)$ from far-away points, as in classes of unimodal or monotone densities, or in the class of densities whose characteristic function ϕ vanishes outside a finite interval $[-a, a]$. Densities in this class are extremely smooth, and, in

224

fact, the value of $f(x)$ can be obtained from the value of f at all the points of a regular grid with grid size smaller than π/a (this is known as Nyquist's theorem). The fact that this grid can be chosen once and for all before the generator is started up makes the last class of densities special.

Most of the literature on random variate generation is concerned with explicitly specified densities f (see e.g. Knuth 1969, Schmeiser 1980, Ripley 1983 or Bratley, Fox and Schrage 1987 for surveys and definitions). Here we assume only that enough is known about f to determine some of its properties (bound on the moments, unimodality, smoothness properties, position of some quantiles, and so forth). Furthermore, a black box is available that can be consulted for the computation of $f(x)$ for every x.

If we employ the standard rejection method n times to generate our sample, then the expected global cost grows as this constant times n, because the expected number of evaluations of f per variate usually is a positive constant. To obtain sublinear growth in n, it is necessary to consider rejection constants that vary as $1 + o(1)$ as $n \to \infty$ and thus to update the dominating curve as n grows large. Since evaluations of f are usually the main contribution to the time in a rejection-based algorithm, the savings in computer time can be substantial for large n. In fact, since the time needed to generate n random variates always grows at least linearly with n, we note that the (presumably sublinear) expected global cost is asymptotically negligible compared to the overall time taken by the generator. In effect, this means that we have gone a long way towards making the algorithm's performance insensitive to the time required to evaluate f.

2. First example: Lipschitz densities.

In this section we consider densities that are Lipschitz with known constant C, i.e. $|f(x) - f(y)| \leq C|x - y|$ for all $x, y \in R$. For the sake of convenience, assume that f vanishes off $[0, 1]$. We begin by setting up a table of m entries for some integer m to be determined further on. We evaluate $f_i = f(i/m)$ for $i = 0, 1, \ldots, m$. Let

$$g_i = \max(f_i, f_{i+1}) + \frac{C}{2m}, \quad h_i = \min(f_i, f_{i+1}) - \frac{C}{2m}, \quad i = 0, 1, \ldots, m-1.$$

Then f is bounded from above and below by the histogram-shaped functions taking the values g_i and h_i on $[i/m, (i+1)/m)$ respectively. We can use the following rejection-urn method, in which elements of von Neumann's rejection method and Walker's alias method (Walker 1977; Kronmal and Peterson 1979) are mixed together:

Rejection-urn method for Lipschitz densities.

[SET – UP]
Construct an alias table with entries $h_i, g_i - h_i, i = 0, 1, \ldots, m-1$.
NOTE: it does not matter that the entries do not sum to one. Sampling from this table means that an entry is selected with probability proportional to its value. By Walker's method, this can be done in constant time.
[GENERATOR]
REPEAT
    Select an item (i.e. an index i and a flag indicating whether h_i or $g_i - h_i$ is picked) from the alias table by Walker's method.
    Generate a uniform $[0, 1]$ random variate U.
    $X \leftarrow \dfrac{i + U}{m}.$

CASE
      Item selected is h_i: RETURN X
      Item selected is $g_i - h_i$:
            Generate a uniform $[0, 1]$ random variate V.
            $T \leftarrow h_i + V(g_i - h_i)$.
            IF $T < f(X)$ THEN RETURN X
UNTIL False

The construction of the alias table takes time proportional to m. A generalization, called the alias-urn method (Peterson and Kronmal 1982) usually is more efficient (at the expense of space). The expected number of iterations in the rejection algorithm is $(1/m)\Sigma g_i \leqslant 1 + C/m$ and the expected number of evaluations of f is $m + 1$ (in the set-up step) plus $(n/m)\Sigma(g_i - h_i)$. This is smaller than $2nC/m$ but at least equal to nC/m. Recall that n is the number of variates needed. Thus, in the course of generating $X_1, \ldots, X_n$, the algorithm requires on the average not more than $m + 1 + 2nC/m$ evaluations of f. Disregarding truncation, this is minimal when m is $\sqrt{2nC}$, and the minimal value is $1 + \sqrt{8nC}$. Hence, the expected global time is at least 1 $+ \sqrt{4nC}$, and with $m = \lceil \sqrt{2nC} \rceil$, it is at most $2 + \sqrt{8nC}$. For Lipschitz densities with infinite tails, the rejection-urn method needs adjusting, and the best obtainable expected complexity in terms of evaluations of f can grow faster than $\sqrt{n}$.

The class of Lipschitz densities on $[0, 1]$ can be artificially enlarged via nonlinear transformations. One of the referees suggested the following example: the densities

$$f(x) = \frac{1}{\Gamma(1 + a)(1 - x)^2} e^{-(x/(1-x))^{1/a}}, \qquad 0 < x < 1,$$

for $0 < a < 1$ are Lipschitz on $[0, 1]$, so they can be dealt with by the technique described above. Interestingly, if X has density f, then $(X/(1 - X))^{1/a}$ is gamma with parameter a (i.e. its density is $x^{a-1}e^{-x}/\Gamma(a)$).

3. Second example: Monotone densities. Assume that the density f is bounded and nonincreasing on $[0, 1]$ and zero outside this interval. If we set up the same table as in the previous section, with $h_i = f((i + 1)/m)$ and $g_i = f(i/m)$ for $0 \leqslant i < m$, then it is easy to see that the expected number of evaluations of f, or the expected global cost, is $m + 1 + (n/m)\Sigma_{i=0}^{m-1}(g_i - h_i)$, which telescopes to $m + 1 + (n(f(0) - f(1)))/m$. The algorithm given in the previous section requires no modification.

Arguing as in the previous section, we see that for this algorithm, the expected global cost is at least $1 + \sqrt{4n(f(0) - f(1))}$. By taking m equal to $\lceil \sqrt{n(f(0) - f(1))} \rceil$, it is easy to verify that the expected global cost can be kept below $2 + \sqrt{4n(f(0) - f(1))}$.

If we use intervals with geometrically increasing sizes, as advocated in Devroye (1986, pp. 362–365), the expected global cost can be kept below $(1 + o(1))\sqrt{4nB}$ where $B = \log(1 + f(0) + f(0)\log(f(0)))$. This improvement should be used whenever $f(0)$ is large, i.e. the monotone density is very peaked.

4. Third example: Bounded spectrum characteristic functions.

4.1. *Bounded spectrum characteristic functions.* In the first two examples, we exploited a local smoothness property and an order restriction respectively. It is possible to obtain good global time from what could be called global smoothness properties. The global smoothness of a density is usually described in terms of the tail behavior of its characteristic function ϕ. For example, when $\int |t|^k |\phi(t)| \, dt < \infty$, f has a bounded continuous kth derivative tending to 0 as $|x| \to \infty$ (see e.g. Kawata 1972, pp. 438–439). Stronger still, if $|\phi(t)| = O(e^{-c|t|})$ as $|t| \to \infty$, for some $c > 0$, f is

analytic on the line (see e.g. Kawata 1972, pp. 439–440). That is, all the derivatives of f exist and are continuous, and f can be obtained at every point by computing the Taylor series centered at the origin (say). The strongest possible smoothness condition is obtained by requiring that $\phi(t) = 0$ outside some finite interval $[-a, a]$, for then, by Nyquist's theorem (see below), f can be represented as an infinite sum involving the values of the density at a fixed grid of width smaller than π/a. It is clear that Nyquist's theorem should be useful in reducing the global time, because the grid can be held fixed. On the other hand, one should be warned that the densities in this class necessarily have a substantial infinite tail. For example, by the theorems of Ingham and Levinson (Kawata 1972, pp. 288–298), it is impossible that $f(x) \leqslant Ce^{-D|x|}$ for some constants C and D. Thus, the tails in the infinite sum in Nyquist's theorem can be important, and the convergence can be slow.

The tails of f thus do matter. In fact, we will see that the expected global cost with the method to be proposed below is heavily influenced by the size of the tail, as it is $(O(n^{1/(k-1)}))$ for $k \geqslant 3$ whenever the kth absolute moment, μ_k, is finite. Note that the method below uses the information about the distribution very well, since merely using the fact that the densities are Lipschitz with constant $C = a^2/(2\pi)$ (which is easy to verify) could at best give an expected global cost of $O(\sqrt{n})$ by the method outlined in §2.

The densities $h_r \triangleq c_r(\sin(x)/x)^{2r}$ (where $r \geqslant 1$ is an integer and c_r is a constant) have bounded support characteristic functions and can easily be dealt with by the standard rejection method (see Devroye 1984). The support is contained in $[-r/2, r/2]$, so we can take $a = r/2$. Furthermore, for $r > (k + 1)/2$, the kth absolute moment is finite. The point of this paper is that even if we know very well how to generate such random variables, we can gain a lot from the present method if we are asked to generate a big batch of such random variables.

Sometimes, the distributions are truly difficult to work with. Consider, for example, densities defined in terms of their characteristic function, and in particular Polya characteristic functions (which are real and convex on the positive halfline) with bounded support. Evaluations of f require inversions of a characteristic function and are thus extremely costly.

Consider next the convolution density $f = h_r^* \overline{h}$ where $\overline{h}$ is an arbitrary density. Here too evaluations of f are costly since they involve numerical integration. The characteristic function of f is the product of the individual characteristic functions, and hence has support contained in $[-r/2, r/2]$ as well. And the kth absolute moment is bounded by 2^{k-1} times the sum of the individual kth absolute moments. Such densities are thus also covered by the method developed below.

Finally, there is a large class of densities related to Bessel functions (and their generalizations such as Struve functions) that have compact support characteristic functions. For example, if J_ν is the Bessel function of order ν with $\nu > 1/2$, then the Fourier transform of the function $(1 - t^2)^{\nu - 1/2}$ on $[-1, 1]$ is $J_\nu(x)(2/x)^\nu\Gamma(\nu + 1/2)\Gamma(1/2)$ (see Gradshteyn and Ryzhik 1980, formula 8.41.10). By the boundedness of the Bessel functions when $\nu \geqslant 0$, and the fact that $J_\nu(x)\sqrt{x}$ remains bounded as $x \to \infty$, we see that even powers of $J_\nu(x)/x^\nu$ are proportional to densities with bounded spectrum.

4.2. Nyquist's theorem.

THEOREM 1. *Nyquist's theorem. Let f be a density whose characteristic function vanishes outside $[-a, a]$. Then, for $b > a$,*

$$f(x) = \sum_{j=-\infty}^{\infty} a_j \triangleq \sum_{j=-\infty}^{\infty} f\left(\frac{j\pi}{b}\right)\frac{(-1)^j \sin(bx)}{bx - j\pi}.$$

(*For the sake of convenience, let us call this Nyquist's series even though this is not its official name in the literature.*)

PROOF OF THEOREM 1.   See Feller (1971, p. 631).   ■

Theorem 1 is a particular form of a property sometimes attributed to Shannon in the information theory literature, which states that a function with bounded support characteristic function can be reconstructed from its values at the points jl where j is integer and l is a small enough positive constant. It should be noted that when x is a multiple of π/b, the series seems to be nonsensical because we are dividing by zero. That this is not so follows from the fact that $(-1)^j \sin(bx) = \sin(bx - j\pi)$, so that the series can be represented as

$$f(x) = \sum_{j=-\infty}^{\infty} f\left(\frac{j\pi}{b}\right) S(bx - j\pi),$$

where $S(u) \triangleq \sin(u)/u$ is a function bounded by 1.

4.3.   *A generator based on the series method.*   A random variate with density f can be generated by the rejection method provided that we know a density g having the property that $f \leqslant cg$ for some constant $c \geqslant 1$:

Rejection method.
REPEAT
    Generate a pair (X, T) where X has density g, T is distributed as $Ucg(X)$ and U
    is a uniform $[0, 1]$ random variable.
UNTIL $T \leqslant f(X)$
RETURN X

This algorithm requires on the average c iterations and thus c evaluations of f. For n random variates, the expected cost is thus about cn. For a cost reduction, we can turn to Theorem 1. Indeed, notice that for the rejection method to work properly, we need an indicator I which is equal to one when $Ucg(X) \leqslant f(X)$ and to zero otherwise. Since we have a convergent series expansion for f (Theorem 1), the value of I can be determined without ever computing the entire series. This observation is at the basis of the series method, developed and analyzed in Devroye (1981), adapted here to our example:

The series method.
REPEAT
    Generate a pair (X, T) where X has density g, T is distributed as $Ucg(X)$ and U
    is a uniform $[0, 1]$ random variable.
    $T \leftarrow T - f(0)S(bX), j \leftarrow 1$
    REPEAT
        $T \leftarrow T - \left(f\left(\frac{j\pi}{b}\right) S(bX - j\pi) + f\left(\frac{-j\pi}{b}\right) S(bX + j\pi) \right)$
        $j \leftarrow j + 1$
        Compute W, an upper bound on the absolute value of the tail sums in Nyquist's
        series defined from j upwards and $-j$ downwards.
    UNTIL $|T| \geqslant W$
UNTIL $T < 0$
RETURN X

Let us call X generated at the top of one iteration good if for this value, $\lim_{j \leftarrow \infty} W = 0$. The algorithm halts with probability one and is correct if $P(X$ is good$) = 1$. Thus, two things are needed, a convergent series (without this, it is futile to look for a converging upper bound), and a converging upper bound. We do not normally compute W by computing the tail sums, for if we did, we would in fact be evaluating f exactly, and this is too costly at this stage. The first requirement therefore

is for a simple expression for W as a function of easy-to-evaluate quantities. In §§5, 6 and 7, we will give a few pointers.

There is but one other stumbling block, the determination of c and g in the inequality $f \leqslant cg$. We will not treat this issue lightly because, after all, f is not assumed to be given explicitly. Thus, preferably, c and g should depend upon general quantities such as the mean, variance, median, square integral, etcetera.

We also analyze the performance of the series method based upon Nyquist's series and verify under which conditions the number of evaluations of f is drastically reduced. To see how such savings are possible, we suggest using a dynamic array in which values $f(j\pi/b)$ are stored for integer j. Let J_n be the largest absolute value of the index j in the algorithm encountered in the process of generating $X_1, \ldots, X_n$. The dynamic array holds $2J_n + 1$ entries, and f is evaluated precisely $2J_n + 1$ times. The efficiency study boils down to a comparison between cn, the expected cost for the rejection method, and $E(2J_n + 1)$, the expected cost of the series method. In many cases, the latter expression grows at an $o(n)$ rate. This implies that the contribution from the evaluation of f to the total expected time is asymptotically negligible compared to the contribution from other sources such as the generation of (X, T), the management of the loops and so forth, and should therefore lead to relatively f-invariant timings.

4.4. *The dominating curve.* To tuck f under a dominating curve, we recall first that f has an infinite tail (see e.g. Kawata 1972, pp. 278–279). In fact, f is an analytic function on the real line. Both observations should guide us in the search for a dominating curve. We have to look at dominating curves with possibly fairly large tails. One possible upper bound can be obtained as follows: if the kth moment of a distribution function exists, for $k \geqslant 1$, then ϕ is k times differentiable, and $\phi^{(k)}$ is continuous and is given by

$$\phi^{(k)}(t) = \int (ix)^k e^{itx} f(x)\, dx.$$

Since ϕ has bounded support, $\phi^{(k)}$ is absolutely integrable, and $x^k f(x)$ can thus be recovered by inversion:

$$\left| x^k f(x) \right| = \frac{1}{2\pi} \left| \int e^{-itx} i^{-k} \phi^{(k)}(t)\, dt \right|$$

$$\leqslant \frac{1}{2\pi} \int \left| \phi^{(k)}(t) \right| dt$$

$$\leqslant \frac{a}{\pi} \sup_t \left| \phi^{(k)}(t) \right|$$

$$\leqslant \frac{a}{\pi} \int |x|^k f(x)\, dx.$$

We can summarize as follows.

THEOREM 2. *Let $k \geqslant 1$ be a fixed integer, and let f be a density with finite kth absolute moment μ_k and with characteristic function vanishing outside $[-a, a]$. Then*

$$f(x) \leqslant \frac{\int \left| \phi^{(j)}(t) \right| dt}{2\pi |x|^j}$$

$$\leqslant \frac{a\mu_j}{\pi |x|^j}, \qquad j = 0, 1, \ldots, k.$$

One single bound in Theorem 2 yields a useless dominating curve with infinite integral. One could for example combine the 0th and kth bounds with $k \geqslant 2$. Using the notation $c_j|x|^{-j}$ for the jth upper bound in Theorem 2, we thus have

$$f(x) \leqslant \min\left(c_0, c_k|x|^{-k}\right) = cg(x).$$

Here, c is a positive constant and g is a density. The dominating curve consists of a flat part near the origin, and two monotonically decreasing tails. The area under cg is

$$c = \frac{2k}{k-1}\left(c_0^{(k-1)/k}c_k^{1/k}\right).$$

For our rejection method, we need a random pair of random variables $(X, T) = (X, Ucg(X))$, where X has density g, and U is an independent uniform $[0, 1]$ random variate. This can best be obtained as follows.

Generator of (X, T).
Generate iid uniform $[-1, 1]$ and $[0, 1]$ random variates U, V.
$T \leftarrow c_0 V^{k/(k-1)}$
$X \leftarrow U\left(\dfrac{c_k}{T}\right)^{1/k}$
RETURN (X, T).

This method is based upon the fact that (X, T) is uniformly distributed under the curve of cg. Using the symmetry and monotonicity of g, we argue by swapping the coordinate axes. Thus, T can be generated by inversion of a uniform random variate, and X can then be obtained as $Ug^{-1}(T)$. See Devroye (1986) for more generators of this type. This generator can be used in the first line of the series algorithm of the previous section. We will base the remainder of the paper on the premise that f is bounded from above by

$$\min\left(c_0, c_k|x|^{-k}\right) = cg(x).$$

4.5. *Estimate of tail sum in Nyquist's series.* We can write Nyquist's series as

$$f(x) = \sum_{j=-\infty}^{\infty} a_j \triangleq \sum_{j=-\infty}^{\infty} f\left(\frac{j\pi}{b}\right)\frac{(-1)^j\sin(bx)}{bx - j\pi}.$$

Recall that $b > a$ is a constant picked by the user. Let J be a positive integer. Let us first bound $|\Sigma_{j \geqslant J}a_j|$. We have for $0 < bx < J\pi$,

$$\left|\sum_{j \geqslant J} a_j\right| \leqslant \frac{1}{J\pi - bx}\sum_{j=0}^{\infty}\left|f\left(\frac{(J+2j)\pi}{b}\right) - f\left(\frac{(J+2j+1)\pi}{b}\right)\right|$$

$$+ \sum_{j \geqslant J}f\left(\frac{(j+1)\pi}{b}\right)\left|\frac{1}{bx - j\pi} - \frac{1}{bx - (j+1)\pi}\right|$$

$$\leqslant \frac{1}{J\pi - bx}\int_{J\pi/b}^{\infty}|f'| + \sum_{j \geqslant J}f\left(\frac{(j+1)\pi}{b}\right)\frac{\pi}{(bx - j\pi)^2}.$$

Until now, we have only used the fact that f is absolutely continuous (and this follows from the fact that f is analytic). The bound on the tail sum obtained thus far is useless since it involves the indefinite integral of $|f'|$. We suggest two solutions to this

problem. The first solution is applicable to *unimodal f* only (with a mode at 0), while the second is universally applicable, provided that an upper bound for the kth absolute moment of f is available.

4.6. *Unimodal densities.* For unimodal f with a mode at zero, the upper bound for the tail sum becomes

$$\leqslant \frac{1}{J\pi - bx} f(J\pi/b) + f\left(\frac{(J+1)\pi}{b}\right) \sum_{j > J} \frac{\pi}{(bx - j\pi)^2}$$

$$\leqslant \frac{1}{J\pi - bx} f(J\pi/b) + f\left(\frac{(J+1)\pi}{b}\right) \int_{J-1}^{\infty} \frac{\pi}{(y\pi - bx)^2} \, dy$$

$$= \frac{f(J\pi/b)}{J\pi - bx} + \frac{f\left(\frac{(J+1)\pi}{b}\right)}{(J-1)\pi - bx}$$

$$\leqslant 2 \frac{f(J\pi/b)}{(J-1)\pi - bx}.$$

For the antisymmetric negative tail sum, the corresponding bound is

$$2 \frac{f(-J\pi/b)}{(J-1)\pi + bx}.$$

It is easy to see that both bounds are simultaneously valid whenever $(J-1)\pi > b|x|$. We can summarize the algorithm:

The series method for unimodal densities based upon Nyquist's series.
Employ the series method in which W is computed as follows:
   IF $(j-1)\pi \leqslant b|X|$
      THEN $W \leftarrow \infty$
      ELSE

$$W \leftarrow 2 \frac{f(-j\pi/b)}{(j-1)\pi + bX} + 2 \frac{f(j\pi/b)}{(j-1)\pi - bX}$$

This algorithm can be streamlined to some extent by re-using values of f previously computed. Typically, one sets up a dynamic array in which in position j is stored $f(j\pi/b)$ for all integer j up to a dynamically adjusted maximum. If in the course of the run a value of j is encountered which is smaller in absolute value than the maximum, f need not be reevaluated. Otherwise, f is evaluated, its value is stored in the array, and the maximal array index is adjusted by one. Notice that for every X, the bound W used in the algorithm decreases to zero with j. Thus, the algorithm is valid (i.e., halts with probability one) for *all unimodal densities with a mode at 0 and bounded support characteristic function with support on* $[-a, a]$ *for some a < b*. The only a priori knowledge required is that of a bounding curve cg. Note that for c_0 we can take $f(0)$ (by unimodality), and that for c_k we can take $k\mu_{k-1}$ (Devroye, 1986) or $a\mu_k/\pi$ (see above).

4.7. *Tail estimates for general distributions.* To be able to translate the upper bound for the tail sum in general, we need a good estimate for the tail integral of $|f'|$. In analogy with Theorem 2, we offer

THEOREM 3. *Let $k \geqslant 1$ be a fixed integer, and let f be a density with finite kth absolute moment μ_k and with characteristic function vanishing outside $[-a, a]$. Then*

$$|f'(x)| \leqslant \frac{(j+1)\int|t|\,|\phi^{(j)}(t)|\,dt}{2\pi|x|^j}$$

$$\leqslant \frac{(j+1)a^2\mu_j}{2\pi|x|^j}, \qquad j = 0, 1, \ldots, k.$$

PROOF OF THEOREM 3. We note first that

$$f'(x) = \frac{1}{2\pi}\int e^{-itx}(-it\phi(t))\,dt.$$

Thus, using inversion and the techniques of Theorem 2 again, we have

$$|f'(x)| \leqslant \frac{\int\left|(-it\phi(t))^{(j)}\right|dt}{2\pi|x|^j}$$

$$= \frac{\int\left|(-it)\phi^{(j)}(t) - ji\phi^{(j-1)}(t)\right|dt}{2\pi|x|^j}$$

$$\leqslant \frac{(j+1)\int|t|\,|\phi^{(j)}(t)|\,dt}{2\pi|x|^j}$$

since $\int|\phi^{(j-1)}(t)|\,dt \leqslant \int|t|\,|\phi^{(j)}(t)|\,dt$. This inequality follows from the fact that $\phi^{(j-1)}$ is absolutely continuous, and thus equal to the indefinite integral of $\phi^{(j)}$. A simple change of integral argument gives us the inequality. ∎

Let us use the following inequalities in this section:

$$f(x) \leqslant \min(c_0, c_k|x|^{-k}), \quad |f'(x)| \leqslant d_k|x|^{-k},$$

where the various constants are given in Theorems 2 and 3. Then for $0 < bx < J\pi$,

$$\left|\sum_{j\geqslant J}a_j\right| \leqslant \frac{1}{J\pi - bx}\int_{J\pi/b}^{\infty}|f'| + \sum_{j\geqslant J}f\left(\frac{(j+1)\pi}{b}\right)\frac{\pi}{(j\pi - bx)^2}$$

$$\leqslant \frac{d_k}{(k-1)(J\pi - bx)}\left(\frac{b}{J\pi}\right)^{k-1} + \sum_{j\geqslant J}c_k\left(\frac{(j+1)\pi}{b}\right)^{-k}\frac{\pi}{(j\pi - bx)^2}$$

$$\leqslant \frac{d_k}{(k-1)(J\pi - bx)}\left(\frac{b}{J\pi}\right)^{k-1} + c_k\int_J^{\infty}\left(\frac{u\pi}{b}\right)^{-k}du\frac{\pi}{(J\pi - bx)^2}$$

$$\leqslant \frac{d_k}{(k-1)(J\pi - bx)}\left(\frac{b}{J\pi}\right)^{k-1} + \left(\frac{b}{J\pi}\right)^{k-1}\frac{c_k\pi}{(k-1)(J\pi - bx)^2}$$

$$= \left(\frac{b}{J\pi}\right)^{k-1}\left(\frac{d_k}{(k-1)(J\pi - bx)} + \frac{c_k\pi}{(k-1)(J\pi - bx)^2}\right).$$

The last upper bound is convenient for practical use for it only depends upon known quantities. Combining a similar bound for the left tail with the present bound, we see that for all $|x| < J\pi/b$,

$$\left| \sum_{|j| \geqslant J} a_j \right| \leqslant \left(\frac{b}{J\pi} \right)^{k-1} \left(\frac{d_k}{k-1} \left(\frac{1}{J\pi - bx} + \frac{1}{J\pi + bx} \right) \right.$$

$$\left. + \frac{c_k \pi}{k-1} \left(\frac{1}{(J\pi - bx)^2} + \frac{1}{(J\pi + bx)^2} \right) \right).$$

$$\triangleq W(J, x).$$

The algorithm can be summarized as follows.

The series method based upon Nyquist's series.
Employ the series method in which W is computed as follows:
   IF $j\pi \leqslant b|X|$
      THEN $W \leftarrow \infty$
      ELSE $W \leftarrow W(j, X)$ (see above for definition)
The algorithm is valid if $\lim_{j \to \infty} W(j, x) = 0$ for all x, i.e. for all $k \geqslant 2$. Thus, the algorithm is applicable to all densities with finite second moment whose characteristic function has bounded support $[-a, a]$ for some $a < b$.

4.8. *Analysis of performance.* In this section, we briefly look at the expected time performance of the algorithms when $X_1, \ldots, X_n$ are generated. There are many contributors to the expected time. Roughly speaking, most of these can be accounted for in one of three ways:

A. The expected number of iterations of the outer loop in the rejection algorithm, cn. For an explicit computation of c, see §2.

B. The expected number of queries of f, i.e. the expected number of instances that some value of f is required. The number of outer loop iterations in the generation of X_l is called N_l (this is geometrically distributed with mean c). Let $J_{i,l}$, $i = 1, \ldots, N_l$ be the largest value of j in the Nyquist series used in the ith iteration for X_l. This is equivalent to saying that $2J_{i,l} + 1$ queries of f are needed. The expected value is easily computed by Wald's equation (see Chow and Teicher 1978):

$$E\left(\sum_{l=1}^{n} \sum_{i=1}^{N_l} (2J_{i,l} + 1) \right) = nE(N_1)(2E(J_{1,1}) + 1) = cn(2E(J_{1,1}) + 1).$$

C. The expected number of actual evaluations of f when evaluations are stored in a dynamic array for future use. In the notation of point B, this is $E(2 \max_{1 \leqslant l \leqslant n; \, 1 \leqslant i \leqslant N_l} J_{i,l} + 1)$. This is obviously less than the quantity in B for all $n \geqslant 1$. In view of Lemma 1 presented below, it does not exceed

$$\inf_{\alpha \geqslant 0} 1 + 2\alpha + 2cnE\left(J_{1,1} I_{[J_{1,1} > \alpha]} \right).$$

The lemma needed above is

LEMMA 1. *Let $Z_1, \ldots, Z_l, \ldots$, be iid random variables, and let N be a stopping time (see Chow and Teicher 1978 for definitions). Then, for any constant $\alpha \geqslant 0$,*

$$E\left(\max(Z_1, \ldots, Z_N) \right) \leqslant \alpha + E(N)E\left(Z_1 I_{[Z_1 > \alpha]} \right),$$

where I is the indicator function.

PROOF OF LEMMA 1.   Note that

$$E\big(\max(Z_1,\ldots,Z_N)\big) \leqslant \alpha + E\left(\sum_{l=1}^{N} Z_l I_{[Z_l > \alpha]}\right)$$

$$= \alpha + E(N)E\big(Z_1 I_{[Z_1 > \alpha]}\big)$$

by Wald's equation.   ■

Everything boils down to a comparison between

$$Q_n \triangleq cn\big(2E(J_{1,1}) + 1\big) \quad \text{and} \quad R_n \triangleq \inf_{\alpha \geqslant 0} 1 + 2\alpha + 2cnE\big(J_{1,1} I_{[J_{1,1} > \alpha]}\big).$$

For $\alpha = 0$, the expression in the infimum is almost equal to Q_n. The infimum usually is much smaller. We need only be concerned with the properties of $J_{1,1}$ from here on.

THEOREM 4.   *Assume that f is a density with characteristic function vanishing outside $[-a, a]$ for some $0 \leqslant a < b$ (where b is a constant used in the algorithm and in Nyquist's series). If the series method based upon Nyquist's series is valid (i.e., $J_{1,1} < \infty$ with probability one), then $R_n \leqslant Q_n$. If $E(J_{1,1}) < \infty$, then $R_n = o(n)$. Finally, if f has a finite kth moment μ_k ($k \geqslant 3$) and this value (or an upper bound for it) is used in inequalities of Theorems 2 and 3 that lie at the basis of the algorithm, then*

$$R_n = O\big(n^{1/(k-1)}\big).$$

PROOF OF THEOREM 4.   The first inequality follows by considering that the infimum in the definition of R_n is at most equal to the value of the expression for one particular value such as $\alpha = 0$. The second statement follows after first taking α so large that $E(J_{1,1} I_{[J_{1,1} > \alpha]}) < \epsilon$ for some arbitrary $\epsilon > 0$. Thus, $R_n \leqslant 1 + 2\alpha + 2\epsilon cn$, and we are done because ϵ is arbitrary.

To prove the last statement, a careful analysis of $J_{1,1}$ is required. For fixed $X = x$ we have

$$P\big(J_{1,1} > j \mid X = x\big) \leqslant \min\left(\frac{2W(j, x)}{cg(x)}, 1\right) \quad (j \geqslant \max(1, b|x|/\pi)).$$

Here $W(j, x)$ is the tail estimate used in the algorithm. Taking the expected value with respect to X gives

$$P\big(J_{1,1} > j\big) \leqslant \int_{|x| \leqslant \pi j/(2b)} \min\left(\frac{2W(j, x)}{c}, g(x)\right) dx$$

$$+ \int_{|x| > \pi j/(2b)} g(x)\, dx \quad (j \geqslant 1).$$

Since $g(x) \leqslant c_k c^{-1}|x|^{-k}$, the last integral is $O(j^{-(k-1)})$. Recall the definition of

$$W(j, x) \triangleq \left(\frac{b}{j\pi}\right)^{k-1}\left(\frac{d_k}{k-1}\left(\frac{1}{j\pi - bx} + \frac{1}{j\pi + bx}\right)\right.$$

$$\left. + \frac{c_k \pi}{k-1}\left(\frac{1}{(j\pi - bx)^2} + \frac{1}{(j\pi + bx)^2}\right)\right), \quad |x| < j\pi/b.$$

Thus, for $|x| \leqslant j\pi/(2b)$, we have the following uniform bound:

$$W(j, x) \leqslant \left(\frac{b}{j\pi}\right)^{k-1} \left(\frac{3d_k}{(k-1)j\pi} + \frac{5c_k\pi}{(k-1)(j\pi)^2}\right) = O(j^{-k}).$$

Thus, the first integral that we had to bound is

$$\int_{|x| \leqslant \pi j/(2b)} \min\left(\frac{2W(j, x)}{c}, g(x)\right) dx \leqslant \frac{\pi j}{b} \sup_{|x| \leqslant \pi j/(2b)} \frac{2W(j, x)}{c}$$

$$= O(j^{-(k-1)}).$$

Therefore, $P(J_{1,1} > j) \leqslant Kj^{-(k-1)}$ for some constant K. Now,

$$E(J_{1,1}I_{[J_{1,1} > \alpha]}) \leqslant \sum_{j > \alpha} P(J_{1,1} > j) \leqslant \sum_{j > \alpha} Kj^{-(k-1)} \leqslant \int_\alpha^\infty Kt^{-(k-1)} dt$$

$$= \frac{K}{(k-2)\alpha^{k-2}}.$$

Optimizing $1 + 2\alpha + 2Kcn/(k-2)\alpha^{k-2}$ with respect to α shows that it is advantageous to choose $\alpha = (Kcn)^{1/(k-1)}$. Thus,

$$R_n \leqslant 1 + 2\frac{k-1}{k-2}(Kcn)^{1/(k-1)}. \quad \blacksquare$$

We note first that the expected number of evaluations of f can grow at an arbitrarily slow polynomial rate (just take k large enough). One can now work out the details and compute explicit values for all the constants. By using a dynamic array, we do in fact replace on the average Q_n evaluations of f by Q_n array accesses plus at most R_n evaluations of f. Taking $\alpha = 0$ in the proof of Theorem 4 shows that

$$Q_n = cn\big(2E(J_{1,1}) + 1\big) \leqslant cn\left(3 + 2\sum_{j>0} Kj^{-(k-1)}\right) \leqslant cn\left(3 + 2K\frac{k-1}{k-2}\right).$$

All the performance measures, c, Q_n and R_n, depend upon the smoothness and the size of the tail of f. For some choices of the constants in g, the performance improves when a decreases (the density is smoother) and/or the kth absolute moment of the density decreases and/or k in the moment condition increases (the density has a smaller tail).

References

Bratley, P., Fox, B. L. and Schrage, L. E. (1987). *A Guide to Simulation.* Springer-Verlag, New York.

Chow, Y. S. and Teicher, H. (1978). *Probability Theory. Independence. Interchangeability. Martingales.* Springer-Verlag, New York.

Devroye, L. (1981). The Series Method in Random Variate Generation and Its Application to the Kolmogorov-Smirnov Distribution. *Amer. J. Math. Management Sci.* **1** 359–379.

———— Methods for Generating Random Variates with Polya Characteristic Functions. *Statist. Probab. Lett.* **2** 257–261.

———— (1986). *Non-Uniform Random Variate Generation.* Springer-Verlag, New York.

_____ (1986b). An Automatic Method for Generating Random Variables with a Given Characteristic Function. *SIAM J. Appl. Math.* **46** 698–719.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Vol. 2. John Wiley, New York.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series and Products.* Academic Press, New York.

Kawata, T. (1972). *Fourier Analysis in Probability Theory.* Academic Press, New York.

Knuth, D. E. (1969). *The Art of Computer Programming. Vol. 2. Seminumerical Algorithms.* Addison-Wesley, Reading, MA.

Kronmal, R. A. and Peterson, A. V. (1979). On the Alias Method for Generating Random Variables from a Discrete Distribution. *Amer. Statist.* **33** 214–218.

Peterson, A. V. and Kronmal, R. A. (1982). On Mixture Methods for the Computer Generation of Random Variables. *Amer. Statist.* **36** 184–191.

Ripley, B. D. (1983). Computer Generation of Random Variables: A Tutorial. *Internat. Statist. Rev.* **51** 301–319.

Schmeiser, B. W. (1980). Random Variate Generation: A Survey. in *Proc. 1980 Winter Simulation Conf.* Orlando , FL 79–104.

Walker, A. J. (1977). An Efficient Method for Generating Discrete Random Variables with General Distributions. *ACM Trans. Math. Software.* **3** 253–256.

SCHOOL OF COMPUTER SCIENCE, MCGILL UNIVERSITY, 805 SHERBROOKE STREET WEST, MONTREAL, QUEBEC, CANADA H3A 2K6