ON RANDOM VARIATE GENERATION WHEN ONLY MOMENTS OR FOURIER COEFFICIENTS ARE KNOWN

Luc Devroye
School of Computer Science
McGill University

ABSTRACT.

We consider algorithms for generating random variates having a density, when only its Fourier coefficients or moments are known. We also study the expected time per random variate.

1991 MATHEMATICS SUBJECT CLASSIFICATIONS: 65C10

KEYWORDS AND PHRASES. Random variate generation. Moment problem. Fourier series. Orthogonal polynomials. Expected time analysis.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

1. INTRODUCTION.

As a result of some theoretical analysis, one sometimes ends up with distributions that are indirectly specified through a sequence of Fourier coefficients, or a characteristic function, or a sequence of moments, while neither the density nor the distribution function is explicitly known. In this paper, we look at algorithms for generating random variates with a density f on the real line in just these situations. The generation problem when only the characteristic function can be exactly computed is dealt with in a series of papers by the author (1981,1984,1986). When a sequence of moments is known, i.e. for each n, the n-th moment is available at unit computational cost, a solution is developed in Devroye (1988) for integer-valued random variables. Unfortunately, that method can not be generalized to the continuous case. So, we consider densities that are specified either through the sequence of Fourier coefficients, or through the sequence of moments. It is known that under some conditions the density can be written as an infinite series involving these Fourier coefficients or moments, e.g. by using Fourier series or smoothed Fourier series in the former case and Hermite (Gram-Charlier, Edgeworth), Laguerre, Jacobi or Legendre series in the latter case. Unfortunately, the evaluation of these series takes an infinite amount of computational effort when we assume that each computation of a Fourier coefficient or moment takes one unit of time. Hence, we can't compute the density at all in the algorithms. Instead, we resort to algorithms that compute a random but finite number of coefficients or moments per generated random variate. The fact that we can do this and still insure theoretical exactness shows once again that random variate generation is "easier" than computing quantities related to a distribution. Most methods discussed here boil down to the acceptance-rejection method combined with the series method developed in Devroye (1980, 1986, pp. 151-171). There is one interesting exception related to the Fourier coefficient problem. It is known that any Fourier cosine series with nonnegative decreasing convex coefficients is the Fourier series of a density modulo a constant (Young, 1913; Zygmund, 1959, p. 183). This characterization yields a rich source of distributions that are specified by their Fourier coefficients. The local structure in the sequence of coefficients allows us to develop a universal inversion-based generator, i.e. one for which nothing else about the distribution is known. Other such shortcuts should be possible under various conditions on the local structure of the Fourier coefficients or moments.

Throughout, we assume that real numbers can be stored with infinite precision; that all standard arithmetic operations take one unit of time; and that we have a source capable of generating an infinite sequence of iid uniform $[0,1]$ random variates $U_1, U_2, \ldots$. Some of these conditions are controversial (e.g., which operations are "standard" ?), but they are consistent with those found in many recent articles and books on the subject (see e.g. Devroye (1986)).

2. THE GENERAL STRATEGY.

The problem dealt with here can in general be described as follows: the density f can be represented as

$$f(x) = \sum_{j=0}^{\infty} a_j \varphi_j(x),$$

where $\{\varphi_j, j \geq 0\}$ is a given fixed family of functions, independent of f, and the a_j's are coefficients in the expansion that could be Fourier coefficients, linear combinations of Fourier coefficients, or linear combinations of finite numbers of moments. Some of these representations are universally valid, while others apply only to a limited class of densities. We assume that the a_j's are given to us. Assume furthermore that we have enough information at hand to be able to construct an integrable function g and a sequence of (error) estimates R_n such that

$$f(x) \leq g(x) (\text{all } x)$$

and

$$|f(x) - \sum_{j=0}^{n} a_j \varphi_j(x)| \leq R_n(x) (\text{all } x, \text{all } n).$$

It is also assumed that for almost all x, $\lim_{n\to\infty} R_n(x) = 0$. Then the following algorithm halts with probability one and returns a random variate with density f:

`Incremental form of the algorithm.`

```
repeat generate X with density proportional to g.
       generate a uniform [0,1] random variate U.
       set T ← Ug(X), n ← 0, S ← 0.
       repeat S ← S + a_n φ_n(X), n ← n + 1
       until |S − T| > R_{n−1}(X)
until T < S
return X
```

```
Dyadic form of the algorithm.
repeat
  generate X with density proportional to g.
  generate a uniform [0,1] random variate U.
  set T ← Ug(X), n ← 1.
  repeat S ← ∑_{i=0}^{n} a_i φ_i(X), n ← 2n
  until |S − T| > R_{n/2}(X)
until T < S
return X
```

It is easy to verify that these algorithms halt with probability one. Also, when we exit from the inner loop, we have $T < S$ if and only if $T < f(X)$, so that the algorithm is indeed equivalent to the acceptance-rejection algorithm.

The choice of the dominating curve g is often dictated by the circumstances; of course, we always have the possibility of choosing $g(x) = a_0\varphi_0(x)+R_0(x)$, $g(x) = a_0\varphi_0(x)+a_1\varphi_1(x) + R_1(x)$, and so forth. This could be especially attractive when f is known to have support in a finite interval. In the case of Fourier coefficients, we will use a uniform dominating curve g on the interval $[-\pi, \pi]$, and in the case of the Legendre series for the moment problem, this strategy brings about a nonuniform dominating curve g on the interval $[-1, 1]$. With Fourier coefficients, it seems possible to derive g based upon information present in the sequence of coefficients. In the moment problem however, additional smoothness assumptions are usually required to construct g. This should not surprise us, since it is not clear how to quickly decide, just by looking at the moment sequence, whether the underlying distribution is atomic, singular continuous, absolutely continuous, or a blend of two or three types.

Example 1. (Fourier series.) A function f on an interval $[-\pi, \pi]$ has **Fourier coefficients**

$$a_n = 1/\pi \int_{-\pi}^{\pi} f(x)\cos(nx)dx,$$
$$b_n = 1/\pi \int_{-\pi}^{\pi} f(x)\sin(nx)dx.$$

The Fourier coefficients uniquely determine f. Note that $a_0 = 1/\pi$. We will write

$$f(x) \sim \frac{1}{2\pi} + \sum_{n=1}^{\infty} (a_n\cos(nx) + b_n\sin(nx))$$

to reflect the fact that f has the given Fourier coefficients, and that under certain conditions, the series on the right hand side converges and is equal to f. We will call the series on the right-hand-side the **Fourier series** for f. If it converges, its value is denoted by $\sigma(f)$ (which can but need not equal f). By assumption, the values of a_n and b_n can be obtained with infinite precision in one unit of time. □

Example 2. (Legendre series.) Assume that f vanishes off $[-1,1]$. If we take for $\{\varphi_j\}$ the **Legendre polynomials** defined by

$$\varphi_j(x) = \frac{1}{2^j j!}\frac{\partial^j}{\partial x^j}(x^2-1)^j = \sum_{k=0}^{[j/2]}(-1)^k 2^{-j}\binom{j}{k}\binom{2j-2k}{j}x^{j-2k},$$

(the middle expression is known as Rodrigues' formula), then we should use as coefficients in the expansion of f,

$$a_j = \frac{2j+1}{2}\int_{-1}^{1} f(x)\varphi_j(x)dx.$$

This is but a linear combination of the first j moments μ_i of the distribution:

$$a_j = \frac{2j+1}{2}\sum_{k=0}^{[j/2]}(-1)^k 2^{-j}\binom{j}{k}\binom{2j-2k}{j}\mu_{j-2k}.$$

Note that for densities with compact support, all the moments taken together uniquely determine the distribution. So, not surprisingly, the Legendre coefficients uniquely determine f, but the series does not necessarily converge for all f. Both φ_j and a_j can be computed at $O(j)$ time cost. Luckily, there is a simple recursive formula that allows us to incrementally compute each φ_j at $O(1)$ cost (see e.g. Sansone (1977, p. 178)):

$$(j+1)\varphi_{j+1}(x) - (2j+1)x\varphi_j(x) + j\varphi_{j-1}(x) = 0,$$

with $\varphi_0 = 1$ and $\varphi_1 = x$.

It should be noted that the Legendre series is but a special case of the Gegenbauer, Ferrer and Jacobi series (Erdelyi, Magnus, Oberhettinger and Tricomi (1953), Sansone (1977), Szego (1975)). □

Example 3. (Gram-Charlier series of type A) The Legendre series of example 2 is only applicable to functions with compact support. If f has support on the real line, one can use a special form of the Hermite series, the Gram-Charlier series of type A. The strategy followed in this note is only applicable to those densities for which all the moments are finite. In addition, we have to require that the moment sequence uniquely determines the distribution, and that this distribution has a density. Classical references on the moment problem include Shohat and Tamarkin (1943) and Widder (1941). For more references and a discussion pertinent to random variate generation, we refer to section

14.2 of Devroye (1986). Thus, the method proposed below comes with many limitations, which we will deal with further on. The **Hermite polynomials** H_n are defined by $H_0 \equiv 1$ and

$$H_n(x) = e^{x^2} \frac{\partial^n}{\partial x^n} e^{-x^2} = (-1)^n \sum_{k=0}^{[n/2]} \frac{(-1)^k n!}{k!(n-2k)!} (2x)^{n-2k} \ .$$

The coefficients that go with this in the expansion are

$$\begin{aligned} a_n &= \frac{1}{2^n n! \sqrt{\pi}} \int f(x) H_n(x) dx \\ &= \frac{1}{2^n n! \sqrt{\pi}} (-1)^n \sum_{k=0}^{[n/2]} \frac{(-1)^k n!}{k!(n-2k)!} 2^{n-2k} \mu_{n-2k}. \end{aligned}$$

The Gram-Charlier type A series uses $\varphi_n = e^{-x^2} H_n$. It should be noted that the series thus obtained, when truncated to four moments, has been a time-honored method for introducing skewness and kurtosis in families of distributions. The type A series, when rearranged, leads to Edgeworth's series (see e.g. Kendall and Stuart (1977, pp. 168-175)). In fact, the Gram-Charlier type A series is only one of many possible Hermite polynomial-based series (see e.g. Sansone, 1977, p. 368).

As in example 2, there are recursions for φ_n that can be used to save storage. In particular,

$$H_{n+1}(x) + 2xH_n(x) + 2nH_{n-1}(x) = 0, \varphi_{n+1}(x) + 2x\varphi_n(x) + 2n\varphi_{n-1}(x) = 0,$$

with $H_0 = 1$ and $H_1 = -2x$. □

Example 4. (Laguerre series.) The Laguerre series parallel in many respects the Hermite series, with the exception that it is geared towards functions with support on the positive halfline. The same remarks about the moment sequence mentioned in the previous example apply. The **Laguerre polynomials** $L_n^{(\alpha)}$ are defined by $L_0^{(\alpha)} = 1$, $L_{-1}^{(\alpha)} = 0$, and

$$L_n^{(\alpha)}(x) = \frac{x^{-\alpha} e^x}{n!} \frac{\partial^n}{\partial x^n} [x^{n+\alpha} e^{-x}] = (-1)^n \sum_{k=0}^{n} (-1)^k \frac{\Gamma(\alpha + n + 1)}{k!(n-k)!\Gamma(\alpha + n - k + 1)} x^n - k.$$

Here $\alpha > -1$ is a parameter of the series. The coefficients that go with this in the expansion are

$$a_n = (-1)^n \sum_{k=0}^{n} (-1)^k \frac{1}{\Gamma(\alpha + n - k + 1)} \binom{n}{k} \mu_n - k,$$

and the series uses $\varphi_n(x) = x^\alpha e^{-x} L_n^{(\alpha)}(x)$.

As in example 2, there are recursions for φ_n that can be used to save storage. In particular,

$$nL_n^{(\alpha)}(x) = (-x+2n+\alpha-1)L_{n-1}^{(\alpha)}(x) - (n+\alpha-1)L_{n-2}^{(\alpha)}(x) \quad (n \geq 2),$$
$$n\varphi_n(x) = (-x+2n+\alpha-1)\varphi_{n-1}(x) - (n+\alpha-1)\varphi_{n-2}(x) \quad (n \geq 2),$$

with $L_0^{(\alpha)} = 1$ and $L_1^{(\alpha)} = -x+\alpha+1$, and $\varphi_0 = x^\alpha e^{-x}$ and $\varphi_1 = x^\alpha e^{-x}(-x+\alpha+1)$. □

The expected number of outer loop iterations is $\int g(x)dx$, but this is hardly an appropriate mesure of the complexity of the algorithm. In most of the examples that follow, the function value $\varphi_n(X)$ can be computed from a bounded number of values $\varphi_j(X)$ $(j < n)$ in a constant amount of time, and a_n is also available from the given coefficients or moments and possibly previous a_j values in constant time. In Theorem 1, we obtain upper bounds for the the expected number of evaluations of functions φ_n before the algorithms halt. Note that this is half the expected number of evaluations of Fourier coefficients, provided that no coefficients are stored. These quantities should give us an appropriate idea of the dependence of the real time upon f. We should, however, issue a caveat: in some cases, the evaluation of φ_n costs an increasing number of time units as n grows large; a case in point is the Legendre series. Luckily, in most of these cases, we have at our disposal a simple recurrence relation that allows us to compute the next function value in a series in time $O(1)$. In fact, for all the methods dealt with in this paper, the thesis that φ_n and a_n can be computed in one universal time unit can be supported.

The bounds relate performance to the rate of decrease of the error functions R_n, and thus to the rate of decrease of the a_n's. It is well-known that in all the cases handled here, the size of the coefficients a_n is roughly determined by the "smoothness" of f. For example, in the Fourier series, analytic functions f give rise to exponentially decaying a_n's, while densities with $k-1$ absolutely continuous derivatives yield coefficients that are $O(n^{-k})$. Thus, smoother densities result in faster generators. Via Theorem 1, we can use classical results from Fourier analysis to obtain performance guarantees for the algorithms.

Theorem 1. *Let N be the number of evaluations of functions $\varphi_n, n > 0$ before the algorithm halts. Then, for the incremental form of the algorithm,*

$$\mathbf{E}N \leq \int g + 2\sum_{i=1}^{\infty} \int R_i.$$

For the dyadic form of the algorithm, assuming that R_n is $\downarrow$ for all x, we have

$$\mathbf{E}N \leq 2\int g + 4\sum_{i=1}^{\infty} \int R_i$$

when, within one run of the outer loop, evaluations of $\varphi_n(X)$ are saved. If evaluations of $\varphi_n(X)$ are not saved, then

$$\mathbf{E}N \leq 4\int g + 8\sum_{i=1}^{\infty} \int R_i.$$

In all cases, we have $\mathbf{E}N \geq \int g + \sum_{i=1}^{\infty} \int \min(R_i, g)$.

Proof. Note first that by Wald's equation (see Devroye (1986, p.50)),

$$\mathbf{E}N = \int g(x) E_x(N^*) dx,$$

where E_x is a conditional expectation given that $X = x$, and N^* are the number of evaluations of Fourier coefficients and φ_n functions respectively in the first outer loop of the algorithm. Also,

$$\frac{\min(R_i(x), g(x))}{g(x)} \leq P_x(N^* > i) \leq \frac{2R_i(x)}{g(x)} \ .$$

Thus, using $E_x(N^*) = \sum_{i=0}^{\infty} P_x(N^* > i)$, we have

$$\mathbf{E}N^* = \frac{\int g(x) E_x(N^*) dx}{\int g(x) dx} \leq \frac{\int g + 2\sum_{i=1}^{\infty} R_i}{\int g}.$$

This proves the first part of the Theorem and the lower bound. Consider now the dyadic form of the algorithm. Assume first that computations of φ_n and a_n can be stored within one run of an outer loop (i.e. for fixed $X = x$). Then, the monotonicity of the functions R_n implies that N^* is at least equal to and at most double the value obtained in the incremental algorithm. However, if values of $\varphi_n(X)$ are not stored, then we need a small additional argument. Indeed, assume that given X, we exit after having evaluated $a_1, \ldots, a_M$, where $M = 2^K$ is a power of 2. Then, the number of evaluations of functions φ_n is

$$\sum_{i=1}^{K} i 2^{K-i} \leq M \sum_{i=0}^{\infty} i 2^{-i} = 2M.$$

Thus, once again, we at most double the expected number of evaluations. □

3. FOURIER SERIES: FIRST AN UNSOPHISTICATED ALGORITHM.

The functions

$$S_n(x,f) = S_n(f) = \frac{1}{2\pi} + \sum_{i=1}^{n} (a_i \cos(ix) + b_i \sin(ix))$$

are the partial sums of the ordinary **trigonometric series**. Clearly,

$$|S_n(x,f) - f(x)| \le \sum_{k=n+1}^{\infty} |a_k \cos(kx) + b_k \sin(kx)| \le \sum_{k=n+1}^{\infty} \sqrt{a_k^2 + b_k^2}.$$

This rather trivial uniform bound allows us to use the algorithms of section 2 with

$R_n(x) = R_n = \sum_{k=n+1}^{\infty} \sqrt{a_k^2 + b_k^2}$

$g(x) = \frac{1}{2\pi} + R_0$ (so X is uniform on $[-\pi, \pi]$)

"$a_n \varphi_n(x)$"$= a_n \cos(nx) + b_n \sin(nx)$ $(n > 0)$; $= 1/(2\pi)$ $(n = 0)$

Even though no information is required besides accessibility to a program that computes each a_n or b_n, it is not clear that we always have access to the tail sums needed in the formula for $R_n(x)$. An important group of problems in which we do know $R_n(x)$ include those in which the a_n's and b_n's are explicitly known and the tail sums of the coefficients can be evaluated in constant time. And in any case, if the tail sums are awkward, it is always possible to replace R_n by upper bounds.

The algorithm halts with probability one if and only if the coefficients are absolutely summable:

$$\sum_{n=1}^{\infty} (|a_n| + |b_n|) < \infty.$$

Note that the absolute summability of the Fourier coefficients implies that the trigonometric series is absolutely and uniformly convergent to f for all x, and that the function f is necessarily continuous on $[-\pi, \pi]$ (see e.g. Zygmund (1959, pp. 232-245, or Bary, 1964, p. 331). Sufficient conditions for the absolute summability of the coefficients include (i) f is Lipschitz of order $\alpha > 1/2$; (ii) f is of bounded variation and Lipschitz of order $\alpha > 0$; (iii) f is absolutely continuous and $\int f' \log_+ |f'| < \infty$ (see Zygmund (1959)). It is easy to see that not many interesting continuous densities are excluded. For the incremental form of the series method, we have by Theorem 1,

$$\mathbf{E}N \le 1 + 4\pi \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} \sqrt{a_k^2 + b_k^2}$$

$$= 1 + 4\pi \sum_{k=1}^{\infty} k\sqrt{a_k^2 + b_k^2} \le 1 + 4\pi \sum_{k=1}^{\infty} k(|a_k| + |b_k|)),$$

where N is the number of (a_i, b_i) pairs that are evaluated before halting. To make this finite, it suffices to have coefficients that are $O(n^{-2-\epsilon})$ for some $\epsilon > 0$. It is also known that $\sum_{k=n}^{\infty}(|a_k| + |b_k|)) = O(1/n)$ implies that f has bounded variation (see p. 211 of Zygmund, 1959). Thus, it is almost true that the bounded variation of f is a necessary condition for $\mathbf{E}N < \infty$.

4. FOURIER COSINE SERIES WITH CONVEX COEFFICIENTS.

This section is a digression from the main stream of thoughts. We include it here to show that if additional structure is assumed in the sequence of Fourier coefficients, then some special algorithms may be available that do not require the use of the series method or any other acceptance-rejection method. We will make this point with the aid of the Fourier cosine series for symmetric functions f on $[-\pi, \pi]$,

$$(1/2)a_0 + \sum_{n=1}^{\infty} a_n \cos(nx), (|x| \le \pi),$$

where $a_n = 1/\pi \int_{-\pi}^{\pi} f(x)\cos(nx)dx$. If $a_n \downarrow 0$ and the sequence $a_0, a_1, \dots$ is convex, then the Fourier series converges, except possibly at $x = 0$, to a nonnegative integrable function f, and is the Fourier series of f (Young, 1913; see Zygmund, 1959, p. 183). The limit, for $x \ne 0$, is

$$f(x) = \sum_{n=0}^{\infty} \pi(n+1)\Delta^2 a_n K_n(x),$$

where $\Delta^2 a_n = a_{n+2} - 2a_{n+1} + a_n$ is the second order difference, and K_n is the **Fejer kernel** (or Fejer density)

$$K_n(x) = \begin{cases} \frac{1}{2\pi(n+1)} \left(\frac{\sin((n+1)x/2)}{\sin(x/2)}\right)^2 & |x| \le \pi \\ 0 & |x| > \pi \end{cases}$$

If the function f is a density, then from the fact that K_n is a density for each n, we deduce that f is a mixture of Fejer densities with mixture weights given by

$$p_n = \pi(n+1)\Delta^2 a_n, n \ge 0.$$

A random variate with density f can thus simply be generated by first generating a discrete random variate Z with probability vector $\{p_n, n \geq 0\}$, and then exiting with a random variate X having the Fejer density K_Z. A random variate with density K_n can be generated in expected time essentially independent of n (see below), and a random variate with probability vector p_n can be generated in a myriad of ways. The first one that should be considered is the inversion method:

```
Inversion method for convex Fourier cosine series.
generate a uniform [0,1] random variate U.
Z ← 0, S ← πΔ²a_0.
while U > S do
  Z ← Z + 1
  S ← S + π(Z + 1)Δ²a_Z
generate a random variate X with Fejer density K_Z.
return X.
```

The observation leading to the above algorithm can be considered as the discrete analog of a similar observation we made for generating continuous random variates with a Polya characteristic function (Devroye, 1984). The number of steps in the inversion method, i.e. comparisons $U > S$, is $Z + 1$. We have

THEOREM 2. *For the inversion algorithm shown above,*

$$\mathbf{E}(Z+1) = 1 + 2\pi \sum_{n=1}^{\infty} a_n,$$

and for $k > 0$,

$$\mathbf{P}\{Z+1 > k\} = \pi(k+1)(a_k - a_{k+1}) + \pi a_{k+1}.$$

PROOF.

$$\begin{aligned}
\mathbf{E}(Z+1) &= 1 + \pi \sum_{n=0}^{\infty} n(n+1)(a_{n+2} - 2a_{n+1} + a_n) \\
&= 1 - 2\pi \sum_{n=1}^{\infty} n(a_{n+1} - a_n) \\
&= 1 + 2\pi \sum_{n=1}^{\infty} a_n,
\end{aligned}$$

where we twice applied Abel's transformation for summation by parts. Furthermore, for $k > 0$, with $\Delta a_n = a_{n+1} - a_n$,

$$\begin{aligned}
\mathbf{P}\{Z+1 > k\} &= \sum_{n=k}^{\infty} \pi(n+1)\Delta^2 a_n \\
&= \sum_{n=k}^{\infty} \pi \sum_{j=0}^{n} (\Delta a_{n+1} - \Delta a_n) \\
&= \pi \sum_{j=0}^{\infty} \sum_{n=k}^{\infty} (\Delta a_{n+1} - \Delta a_n) \\
&= -\pi(k+1)\Delta a_k + \pi \sum_{j=k+1}^{\infty} (-\Delta a_j) \\
&= \pi(k+1)(a_k - a_{k+1}) + \pi a_{k+1}. \square
\end{aligned}$$

Theorem 2 shows that the average time behavior of the inversion method is directly related to the rate of decrease to zero of the Fourier coefficients. It is instructive to compare the performance with that of the "unsophisticated" version of the series method applied to Fourier series in the previous section. In the notation of that section, we had $\mathbf{E}N \leq \rho \stackrel{\text{def}}{=} 1 + 4\pi \sum_{k=1}^{\infty} k a_k$. From the proof of Theorem 1, we recall that often $\mathbf{E}N = \rho$, while in most cases, $\mathbf{E}N \geq c\rho$ for a positive constant c. Assuming that N and $Z+1$ measure roughly comparable units of computation, the improvement in expected performance is truly remarkable. Not only is $\mathbf{E}(Z+1) \leq (\rho+1)/2$, but it is possible to have $\mathbf{E}(Z+1) < \infty$ while $\rho = \infty$.

This leaves us with the problem of generating a random variate with density K_n. It is easy to show that

$$K_n(x) \leq \min\left(\frac{n+1}{4}, \frac{\pi}{2(n+1)x^2}\right).$$

A random variate with density proportional to the dominating curve can easily be obtained by the rejection method described on p.315 of Devroye (1986). Interestingly, the rejection constant is $\sqrt{2\pi}$, independent of n.

```
Rejection algorithm for the Fejer density.
repeat
  generate a uniform [0,1] random variate V and an independent
  uniform [−1,1] random variate U.
  set Y ← ((n+1)/4) V², X ← U √(π / (2(n+1)Y)).
  ((X,Y) is uniformly distributed under the dominating curve.)
until Y < K_n(X)
return X
```

5. LEGENDRE SERIES.

The relationship between the moments and the coefficients a_j used in the series is given in Example 2. The dyadic form of the algorithm is recommended. Theorem 1 is relevant since the n-th partial sum of the Legendre series can be obtained at cost n.

Not surprisingly, we need to restrict the densities somewhat in order to obtain useful error estimates $R_n(x)$. Assume for example that f is absolutely continuous on $[-1,1]$ and that f' is of bounded variation in $[-1,1]$. Then, by a theorem due to Jackson (1930), in a form explicitly computed by Sansone (1977), we can use

$$R_n(x) = \frac{16\sqrt{2}(\sup_{y\in[-1,1]}|f'(y)| + \text{“Variation” } (f'))}{(1-x^2)^{1/4}\pi n} \stackrel{\text{def}}{=} \frac{C}{(1-x^2)^{1/4}n}.$$

The real problem here is that the quantities appearing in the upper bound are usually not directly available. Since $a_0 = 1/2$ and $a_1 = (3/2)\mu_1$, we can take the following dominating curve:

$$g(x) = \max(0, 1/2 + 3/2\mu_1 x) + \frac{C}{(1-x^2)^{1/4}}, |x| < 1.$$

A random variate with density proportional to $(1-x^2)^{-1/4}$ can be obtained by Ulrich's polar method for symmetric beta random variates (Ulrich, 1984) as $\sin(2\pi V)\sqrt{1-U^4}$, where U, V are independent uniform $[0,1]$ random variables. It is not difficult to verify that the following mixture algorithm can be used to generate a random variate with density proportional to g:

```
Generator for density proportional to g.
[SET-UP]
compute the weights p ← 1, q ← √(8/π)Γ²(3/4)C (note:  q = 2.396280467...C).
If |μ1| > 1/3, set p ← 1 + (1 − 3|μ1|)²/(12|μ1|).
[GENERATOR.]
generate a uniform [0,1] random variate W.
if W ≤ p/(p+q)
  then if |μ1| ≤ 1/3
          then generate X uniform on [−1,1] and U uniform on [0,1].
               if 1/2 + 3/2μ1X − U/2 < 0, then X ← −X.
          else generate iid uniform [0,1] random variates U,V.
               set Y ← min(U,V).
               X ← −W/(3μ1) + (1 − W)sign (μ1)
  else generate iid uniform [0,1] random variates U,V.
       X ← sin(2πV)√(1 − U⁴)
return X
```

For other bounds, see Picone (1930) or Sansone (1977, pp. 203-205). It is worthwhile to mention however that if f has r derivatives, and $f^{(r)}$ is absolutely continuous on $[-1,1]$, we have

$$R_n(x) \le \frac{C\int_{-1}^{1}|f^{(r+1)}(y)|dy}{n^{r-1/2}}, n \ge r,$$

where r is a positive integer, and C is a universal constant depending upon r only (Jackson, 1930, p. 76).

6. GRAM-CHARLIER AND LAGUERRE SERIES.

Conditions under which the Gram-Charlier series of type A converges are given in Cramer (1926) and Stone (1928). See also Szego (1975) and Kendall and Stuart (1977, p. 173). Most of the known results are collected or simplified in Muckenhoupt (1970). For example, assume that

$$\int_n^\infty (f(x)+f(-x))e^{x^2/2}x^{-5/3} = o(1/n)$$

and that $\int f \log_+ f < \infty$. Then the Gram-Charlier type A series converges at almost all x to f (combine Szego (1975, p. 247) with section 4 of Muckenhoupt (1970)). Basically, the density should drop off faster than a normal density. Stronger conditions are needed to obtain useful error estimates.

The Laguerre series suggested in this paper converges under similar conditions. Again, assume that $\int f \log_+ f < \infty$. Then, if $\int_0^1 x^{-\alpha/2-1/4} f(x)dx < \infty$, and

$$\int_n^\infty x^{-\alpha/2-13/12} e^{x/2} f(x)dx = o(n^{-1/2}),$$

then the Laguerre series converges at almost all x (combine arguments from Szego (1975) and Muckenhoupt (1970)).

7. FOURIER SERIES WITH THETA FACTORS.

The advantages and disadvantages of the trigonometric series are captured in Lemma 1:

LEMMA 1.

A. *For all $f \in L_1[-\pi,\pi]$, $S_{m_k}(f) \to f$ almost everywhere for some subsequence m_k. Furthermore, $S_n(f) = o(\log(n))$ for almost all x, and $\int |S_n(f)| = o(\log(n))$.*

B. *If $f \in L_p[-\pi,\pi]$ for some $p > 1$, then $S_n(f) \to f$ for almost all x.*

C. *There exists an $f \in L_1[-\pi,\pi]$ with the property that $\limsup |S_n(f)|/\log\log(n) = \infty$ for all x.*

D. *For all sequences $m_k \uparrow \infty$, there exists an $f \in L_1[-\pi,\pi]$ such that $\limsup S_{m_k}(f) = \infty$ for almost all x.*

Part A is proved all standard texts, see e.g. Zygmund (1959, Section 7.3) or Edwards (1979, pp. 167, pp. 180). For $p = 2$, property B is known as Carleson's theorem (Carleson, 1966). The general statement for $p > 1$ was proved by Hunt in 1968, and is now known as the Carleson-Hunt theorem (see also Mozzochi (1971) or Jorsboe and Mejlbro (1982) for proofs). Part C strengthens Kolmogorov's counterexample (Kolmogorov, 1926; see Zygmund, 1959, Section 8.4), and is due to Korner (1981), who also proved part D based upon ideas of Kahane and Stein.

Part A tells us that there exists an almost everywhere convergent subsequence for every f: thus, there is hope of obtaining a universally applicable algorithm. Unfortunately, we need to have bounds on the error. In view of the negative results in parts C and D, this is an ambitious exercise. For certain subclasses of densities (such as the densities that are in $L_p[-\pi,\pi]$ for some $p > 1$), there is hope of obtaining just such a method. There exist smoothed versions of the trigonometric series that have better consistency properties and lead to better error estimates: these are based on a sequence of $m(n)$-th degree trigonometric polynomials of the following form:

$$S_n(f) = \frac{1}{2\pi} + \sum_{k=1}^{m}(n)\theta_n(k)\left(a_k\cos(kx) + b_k\sin(kx)\right)$$

where $m(n)$ increases with n and is such that $\theta_n(k) = 0$ for $|k| > m(n)$. Butzer and Nessel (1971, p. 47) call the $\theta_n(k)$'s row-finite **theta factors**. The theta factors control the degree of smoothing. A table is given below:

Factor	$m(n)$	$\theta_n(k)$
Dirichlet factor	n	1
Fejer factor	n	$1 - \frac{k}{n+1}$
de la Vallee Poussin factor	$2n-1$	$1, 1 \leq k \leq n$ $2 - \frac{k}{n}, n+1 \leq k \leq 2n-1$
Rogosinski factor	n	$\cos(\frac{k\pi}{2n+1})$
Fejer-Korovkin factor	n	$\frac{(n-k+3)\sin(\frac{k+1}{n+2}\pi)-(n-k+1)\sin(\frac{k-1}{n+2}\pi)}{2(n+2)\sin(\pi/(n+2))}$

These examples are all discussed in more detail in Butzer and Nessel (1971). For all these theta factors, except the Dirichlet factor, we know that $S_n(f) \to f$ almost everywhere and in the L_1 sense, for all f. For all continuous f (continuity is defined for the periodically continued f on the real line), in all cases, the Dirichlet factor excepted, the convergence is uniform. Thus, for continuous f, we can realistically expect to obtain uniform error bounds. We call these $R_n(f)$'s:

$$|S_n(f) - f| \leq R_n(f).$$

Note that the $R_n(f)$'s are uniform in x but not in f. Typically, they depend upon n and some smoothness property of f, such as a Lipschitz constant or a uniform bound on some derivative of f. The fact that the $R_n(f)$'s or upper bounds for them have to be known explicitly is perhaps the greatest drawback in the algorithm. At this point, we do not see how this is possible to avoid in one form or another. The dyadic form of the algorithm of section 2 could be used with the following substitutions:

$R_n(x) = R_n(f)$, the bounds given in this section

$g(x) = \frac{1}{2\pi} + R_0$ (so X is uniform on $[-\pi, \pi]$)

$S \leftarrow \frac{1}{2\pi} + \sum_{k=1}^{m(n)} \theta_n(k)(a_k \cos(kx) + b_k \sin(kx))$

It should be noted that in most cases, $\theta_n(k)$ depends upon both n and k. Therefore, every evaluation of $S_n(f)$ in the inner loop takes time at least proportional to $m(n)$ (which is usually equal to n). Also, the computations cannot be saved from one iteration to the next for most theta factors. Nevertheless, the results of Theorem 1 remain relevant.

7.1. Uniform error bounds.

It is helpful to recall some results from Fourier analysis about just how good Fourier series can be.

Let f be the periodic continuation of a density on $[-\pi, \pi]$, and let f be continuous with **modulus of continuity**

$$\omega(\delta, f) = \sup_{x, 0 \leq |h| \leq \delta} |f(x+h) - f(x)|.$$

The **generalized modulus of continuity** (Butzer and Nessel, 1971, p. 67) is

$$\omega^*(\delta, f) = \sup_{x, 0 \leq |h| \leq \delta} |f(x+h) + f(x-h) - 2f(x)| \ .$$

It is also helpful to define a few classes of densities on $[-\pi, \pi]$. When r is a natural number, we define

$$W_r = \{f : f \text{ is } r \text{ times differentiable, and the } r\text{-th derivative is continuous } \}$$

and

$$\mathrm{Lip}(C, \alpha) = \{f : \omega(\delta, f) \leq C\delta^\alpha\},$$
$$\mathrm{Lip}^*(C, \alpha) = \{f : \omega^*(\delta, f) \leq C\delta^\alpha\},$$

Finally, we define $W_r(C, \alpha)$ as the subset of W_r consisting of all densities which have $r-1$ absolutely continuous derivatives and for which $f^{(r)} \in \mathrm{Lip}(C, \alpha)$. Similarly, we define $W_r^*(C, \alpha)$ as the subset of W_r consisting of all densities which have $r-1$ absolutely continuous derivatives and for which $f^{(r)} \in \mathrm{Lip}^*(C, \alpha)$. Without loss of generality we may restrict the order α in these definitions to $0 < \alpha \leq 1$ for $\mathrm{Lip}(C, \alpha)$, and to $0 < \alpha \leq 2$ for $\mathrm{Lip}^*(C, \alpha)$. By controlling r and α, we now have a continuum of classes of ever smoother densities.

LEMMA 2. *Let T_n denote any trigonometric polynomial of order not higher than n (i.e., any linear function of 1, $\sin(ix)$ and $\cos(ix)$ for $1 \le i \le n$), and define*

$$E_n(f) = \inf_{T_n} \max_x |f(x) - T_n(x)|.$$

Then, the sequence $E_n(f)$ decreases monotonically to 0 (Weierstrass's theorem, see Bary (1964, p. 35)). Furthermore,

$$E_n(f) \le 12\omega(1/n, f) \text{ and } E_n(f) \le 18\omega^*(1/n, f)$$

(Jackson's first theorem, see Bary (1964, p. 533) and Butzer and Nessel (1971, p. 97)). Also, for all $f \in W_r$,

$$E_n(f) \le (36/n)^r \|f^{(r)}\|_\infty \text{and} E_n(f) \le 1297(36/n)^r \omega^*(1/n, f^{(r)})$$

(Jackson's second theorem, see Butzer and Nessel (1971, p. 97-98)). Also, for $f \in W_r$,

$$E_n(f) \le 4C/\pi \frac{\int |f^{(r)}|}{(n+1)^r}$$

where $C = \sum_{k \ge 0} (-1)^k (r-1)(2k+1)^{-(r+1)}$ is the best possible constant in the inequality (Favard, 1937).

Lemma 2 establishes the connection between the best possible supremum error with order n polynomials and the smoothness of f. Note that $E_n(f)$ measures the uniform error. Some errors $R_n(f)$ for theta factors given in the previous section can be directly related to $E_n(f)$:

LEMMA 3.

A. *For the Dirichlet factor, we have Lebesgue's inequality, $R_n(f) \le (3 + 4/\pi^2 \log(n)) E_n(f)$.*

B. *For the Rogosinski factor, $R_n(f) \le (2\pi + 1) E_n(f) + 2\omega^*(1/n, f)$.*

C. *For the de la Vallee Poussin factor, $R_n(f) \le 4E_n(f)$.*

PROOF. For part A, combine Butzer and Nessel (1971, p. 105) with an estimate of Devroye and Gyorfi (1985, p. 297). Part B is shown in Butzer and Nessel (1971, p. 106). For part C, see Butzer and Nessel (1971, p. 108). □

The Dirichlet factor, which leads to the ordinary trigonometric series, has a performance bound at most $\log(n)$ times $E_n(f)$. It is thus possible that the bound given in Lemma 3 does not tend to 0 even though $E_n(f)$ tends to 0. This, of course, is tied to the non-consistency of the trigonometric series estimate exposed in Lemma 1. The main advantage of the trigonometric series is that computations of a_n, b_n can be saved between iterations: $S_{n+1}(f)$ can be computed from $S_n(f)$, a_{n+1} and b_{n+1}. This property is not shared with the other series.

Let us next look at the Rogosinski and Fejer-Korovkin factors. For both, the bounds on $R_n(f)$ depend on $\omega^*(1/n, f)$. From Butzer and Nessel (1971, p. 76), it is perhaps worthy to recall that for $f \in W_2$ we have

$$\omega^*(1/n, f) \leq 1/n\omega(1/n, f') \leq \frac{\|f''\|_\infty}{n^2}.$$

Note that the same bound remains valid for all $f \in W_r$ regardless of how large r is. Thus, the bounds of Lemma 3 do not allow us to obtain a rate which is better than $1/n^2$. In fact, there is no hope of obtaining a better rate because the error is bounded from below by a constant times n^{-2} for all densities. We note in passing that the Fejer factor has a disappointing $1/n$ lower bound for all densities.

There exist factors without built-in limitations, such as the de la Vallee Poussin factor. The bound on $R_n(f)$ is at most a constant (in this case 4) times the best possible bound ($E_n(f)$). Note that $E_n(f)$ can decrease to 0 at any possible rate, so the smoother f, the smaller $E_n(f)$, and the smaller the error bound for the Fourier series.

7.2. The Fejer-Korovkin factor.

Combining some computations on p. 70 and p. 80 of Butzer and Nessel (1971), and using $1 - \cos(x) \leq x^2/2$ yields the bound

$$R_n(f) \leq (1/\sqrt{2} + \pi/2)^2 \omega^*(\frac{\pi}{\sqrt{2}(n+2)}, f)$$

when the Fejer-Korovkin factor is employed. On W_2, this can further be bounded by

$$(1/\sqrt{2} + \pi/2)^2 \left(\frac{\pi}{\sqrt{2}(n+2)}\right)^2 \|f''\|_\infty = \frac{D\|f''\|_\infty}{(n+2)^2},$$

where the constant D equals 35.194114087.... With the given choice for $R_n(f)$, we see from Theorem 1 that the expected time taken by the algorithm is bounded by qa constant times

$$1 + 4\pi D\|f''\|_\infty \sum_{n=0}^{\infty} \frac{1}{(n+2)^2} = 1 + 4\pi(\pi^{2/6} - 1)D\|f''\|_\infty.$$

It is important to observe that the expected time is finite when the absolute value of the second derivative of f is bounded. In fact, the expected time grows in proportion to $\|f''\|_\infty$. For the users, we recall that in the algorithm, one should use

$R_n(f)$	$m(n)$	$\theta_n(k)$
$\frac{D\|f''\|_\infty}{(n+2)^2}$	n	$\frac{(n-k+3)\sin(\frac{k+1}{n+2}\pi)-(n-k+1)\sin(\frac{k-1}{n+2}\pi)}{2(n+2)\sin(\pi/(n+2))}$

7.3. The de la Vallee-Poussin factor.

If one wants to have very small error bounds for very smooth f, it is perhaps best to work with the de la Vallee Poussin factor, or factors with similar optimality properties. For $f \in W_r$, we have useful bounds from Lemma 2 combined with Lemma 3, part C. By using estimates obtained in the previous section for the Fejer-Korovkin factor, and substituting the appropriate terms in the proof of Jackson's second theorem (Butzer and Nessel, 1971, p. 98), we have in fact,

$$R_n(f) \le 4E_n(f) \le 4\left(\frac{B}{n+2}\right)^r \|f^{(r)}\|_\infty,$$

where $B = \pi\sqrt{2}(1/\sqrt{2}+\pi/2)^2 = 20.373053357....$ In conjunction with Theorem 1, this shows that the performance is not greater than a constant times

$$1 + 4\pi\|f^{(r)}\|_\infty 4B^r \sum_{n=0}^{\infty}(n+2)^{-r} \le 1 + 16\pi\|f^{(r)}\|_\infty \frac{B^r}{r-1}$$

when $r > 1$. The bound deteriorates with increasing values of r, and when $r \downarrow 1$. Although we are not resolving the question of how large r should be (this obviously depends upon f), the value $r = 2$ certainly seems like a good compromise even if bounds on higher derivatives are available. For the users, we recall that in the algorithm, one should use

$R_n(f)$	$m(n)$	$\theta_n(k)$
$4(\frac{B}{n+2})^r\|f^{(r)}\|_\infty$	$2n-1$	$\begin{cases} 1, 1 \le k \le n \\ 2-k/n, n+1 \le k \le 2n-1 \end{cases}$

7.4. Lipschitz densities.

For the important class $\mathrm{Lip}(C,\alpha)$, all factors discussed so far are about equally good; that is, the error rate is either $n^{-\alpha}$ or $\log(n)n^{-\alpha}$. Just to illustrate the use of the recovery process for the ordinary trigonometric series, we proceed with the entire algorithm. Note that it is no longer necessary to double n in every iteration. We can use the inequality

$$R_n(f) \le (3 + 4/\pi^2 \log(n))\frac{12C}{n^\alpha}.$$

Also, by the Lipschitz condition, it is easy to bound the density from above; a trivial bound is $1/(2\pi) + C\pi^\alpha$. The expected time of the algorithm is ∞ for all choices of $\alpha \in (0,1]$, but the algorithm nevertheless halts with probability one.

```
Trigonometric series method for Lipschitz densities
repeat
  generate a uniform [−π, π] random variate X.
  generate a uniform [0, 1] random variate U.
  n ← 0
  V ← U(1/(2π) + Cπ^α)
  S ← 1/(2π)
  repeat n ← n + 1
         S ← S + a_n cos(nX) + b_n sin(nX)
  until |V − S| ≥ (3 + 4/π² log(n)) 12C/n^α
until V < S
return X
```

For the Fejer kernel, and $f \in \mathrm{Lip}^*(C, \alpha)$,

$$R_n(f) \leq \begin{cases} \frac{C\pi^2}{2m(\alpha-1)}(\alpha \in (1,2]) \\ \frac{C(1+\pi/2\log(\pi m))}{n}\ (\alpha = 1) \\ \frac{C(1+\frac{\pi}{2(1-\alpha)})}{n^\alpha}(\alpha \in (0,1)) \end{cases}$$

(Butzer and Nessel, 1971, p.81). It is well-known that the Fejer factor yields a series whose error cannot decrease faster than $1/n$, regardless of how smooth the density is. It should thus only be considered for classes W_1 or $\mathrm{Lip}(C, \alpha)$: noting that $\mathrm{Lip}(C, \alpha) \subseteq \mathrm{Lip}^*(2C, \alpha)$, we see that for the simple Lipschitz class with $\alpha = 1$, the best rate $1/n$ is obtained. Also observe that for $f \in W_1$, we have $f \in \mathrm{Lip}(\|f'\|_\infty, \alpha)$, so that once again we obtain the rate $1/n$.

8. IMPROVEMENTS.

The algorithms' expected time performance can be enhanced by starting from a nonuniform dominating density. Note that, in the general setting of section 2,

$$f(x) \leq \sum_{j=0}^{n} a_j \varphi_j(x) + R_n(x)$$

for all n. The upper bound is either a polynomial of degree n (as in the case of Legendre or Hermite-related series), or a trigonometric polynomial of degree n (as in the trigonometric series), or a trigonometric polynomial of degree $m(n)$ (as with the Fourier series when theta factors are used). If we can easily generate random variates with this nonuniform dominating density, then the following modification of the algorithm could be feasible:

```
Series method with nonuniform dominating density
[NOTE: n is a parameter of the algorithm.]
repeat
  generate a random variate X with density proportional to g(x) = sum_{j=0}^{n} a_j φ_j(x) + R_n(x).
  generate a uniform [0,1] random variate U.
  V ← U g(X)
  m ← n
  repeat m ← 2m
         S ← sum_{j=0}^{m} a_j φ_j(X)
  until |V − S| ≥ R_m(X)
until V < S
return X
```

The only modification we made here to the original algorithm is in the choice of the dominating curve. Sampling from it can be done very efficiently by table-based methods (see e.g. Marsaglia's rectangle-wedge-tail method (Marsaglia, Maclaren and Bray, 1964), the method of Ahrens and Kohrt (1981), the alias-rejection-mixture method (Kronmal and Peterson, 1980), or one of the methods of chapter VIII of Devroye (1986)). The investment in a table method will only pay benefits when many random variates from the same distribution are needed.

9. REFERENCES.

J. H. Ahrens and K. D. Kohrt, "Computer methods for efficient sampling from largely arbitrary statistical distributions," *Computing*, vol. 26, pp. 19–31, 1981.

N. K. Bary, *A Treatise on Trigonometric Series, Vol. I*, Macmillan, New York, N.Y., 1964.

N. K. Bary, *A Treatise on Trigonometric Series, Vol. II*, Macmillan, New York, N.Y., 1964.

P. L. Butzer and R. J. Nessel, *Fourier Analysis and Approximation*, Birkhauser Verlag, Basel, Switzerland, 1971.

L. Carleson, "On convergence and growth of partial sums of Fourier series," *Acta Mathematica*, vol. 116, pp. 135–157, 1966.

H. Cramer, "On some classes of series used in mathematical statistics," Skandinaviske Matematikercongres, Copenhagen, 1926.

L. Devroye, "The computer generation of random variables with a given characteristic function," *Computers and Mathematics with Applications*, vol. 7, pp. 547–552, 1981.

L. Devroye, "The series method in random variate generation and its application to the Kolmogorov-Smirnov distribution," *American Journal of Mathematical and Management Sciences*, vol. 1, pp. 359–379, 1981.

L. Devroye, "On the use of probability inequalities in random variate generation," *Journal of Statistical Computation and Simulation*, vol. 20, pp. 91–100, 1984.

L. Devroye, "Methods for generating random variates with Polya characteristic functions," *Statistics and Probability Letters*, vol. 2, pp. 257–261, 1984.

L. Devroye, *Nonuniform Random Variate Generation*, Springer Verlag, New York, N.Y., 1986.

L. Devroye, "An automatic method for generating random variables with a given characteristic function," *SIAM Journal of Applied Mathematics*, vol. 46, pp. 698–719, 1986.

L. Devroye, "Algorithms for generating discrete random variables with a given generating function or a given moment sequence," *SIAM Journal on Scientific and Statistical Computing*, 1988. To appear..

L. Devroye and L. Gyorfi, *Nonparametric Density Estimation: The Lsub*1 *View*, John Wiley, New York, 1985.

R. E. Edwards, *Fourier Series. A Modern Introduction, Vol. 1*, Springer Verlag, New York, N.Y., 1979.

A. Erdelyi, W. Magnus, F. Oberhettinger, and F. Tricomi, "Higher Transcendental Functions," Bateman Manuscript Project, II, McGraw-Hill, New York, 1953.

J. Favard, "Sur les meilleurs procedes d'approximation," *Bulletin de la Societe Mathematique de France*, vol. 61, pp. 209–224,243-256, 1937.

R. A. Hunt, "On the convergence of Fourier series," in: *Orthogonal Expansions and Their Continuous Analogues, Proceedings of a Conference held at Edwardsville, IL.*, pp. 235–255, Southern Illinois University Press, Carbondale, IL., 1968.

D. Jackson, "The Theory of Approximation," Colloquia Publications, vol. 11, American Mathematical Society, Providence, RI, 1930.

O. G. Jorsboe and L. Mejlbro, *The Carleson-Hunt Theorem of Fourier Series*, Springer Verlag, Berlin, 1982.

M. Kendall and A. Stuart, *The Advanced Theory of Statistics, Vol. 1*, Macmillan, New York, 1977.

A. N. Kolmogorov, "Une serie de Fourier-Lebesgue divergente partout," *Comptes Rendus de l'Academie des Sciences de Paris*, vol. 183, pp. 1327–1328, 1926.

T. W. Korner, "Everywhere divergent Fourier series," *Colloquium in Mathematics*, vol. 45, pp. 103–118, 1981.

R. A. Kronmal and A. V. Peterson, "Generating normal random variables using the alias-rejection-mixture method," in: *Proceedings of the 1979 ASA Annual Meeting, Computer Section*, Washington, D. C., 1980.

G. Marsaglia, M. D. Maclaren, and T. A. Bray, "A fast procedure for generating normal random variables," *Communications of the ACM*, vol. 7, pp. 4–10, 1964.

C. J. Mozzochi, *On The Pointwise Convergence of Fourier Series*, Springer Verlag, Berlin, 1971.

B. Muckenhoupt, "Equiconvergence and almost everywhere convergence of Hermite and Laguerre series," *SIAM Journal of Mathematical Analysis*, vol. 1, pp. 295–321, 1970.

M. Picone, *Appunti di Analise Superiore*, Naples, 1940.

G. Sansone, *Orthogonal Functions*, Krieger, Huntington, NY, 1977.

J. A. Shohat and J. D. Tamarkin, "The Problem of Moments," Mathematical Survey No. 1, American Mathematical Society, New York, 1943.

M. H. Stone, "Developments in Hermite polynomials," *Annals of Mathematics*, vol. 29, pp. 1–13, 1928.

G. Szego, "Orthogonal Polynomials (4th ed.)," Colloquia Publications vol. 23, American Mathematical Society, Providence, RI, 1975.

G. Ulrich, "Computer generation of distributions on the m-sphere," *Applied Statistics*, vol. 33, pp. 158–163, 1984.

D. V. Widder, *The Laplace Transform*, Princeton University Press, Princeton, N.J., 1941.

A. Zygmund, *Trigonometrical Series, Vols. 1, 2*, Cambridge University Press, New York, 1959.