Chapter Fourteen
PROBABILISTIC SHORTCUTS AND ADDITIONAL TOPICS

A probabilistic shortcut in random variate generation is a method for reducing the expected time in a simulation by recognizing a certain structure in the problem. This principle can be illustrated in hundreds of ways. Indeed, there is not a single example that could be called "typical". It should be stressed that the efficiency is derived from the problem itself, and is probabilistic in nature. This distinguishes these shortcuts from certain techniques that are based upon clever data structures or fast algorithms for certain sub-tasks. We will draw our examples from three sources: the simulation of maxima and sums of iid random variables, and the simulation of regenerative processes.

Other topics briefly touched upon include the problem of the generation of random variates under incomplete information (e.g. one just wants to generate random variates with a unimodal density having certain given moments) and the generation of random variates when the distribution is indirectly specified (e.g. the characteristic function is given). Finally, we will briefly deal with the problem of the design of efficient algorithms for large simulations.

1. THE MAXIMUM OF IID RANDOM VARIABLES.

1.1. Overview of methods.

In this section, we will look at methods for generating $X=\max(X_1, \ldots, X_n)$, where the X_i's are iid random variables with common density f (the corresponding distribution function will be called F). We will mainly be interested in the expected time as a function of n. For example, the naive method takes time proportional to n, and should be avoided whenever possible. Because X has distribution function F^n, it is easy to see that the following algorithm is valid:

Inversion method

Generate a uniform [0,1] random variate U.

RETURN $X \leftarrow F^{-1}(U^{\frac{1}{n}})$.

The problem with this approach is that for large n, $U^{1/n}$ is close to 1, so that in regular wordsize arithmetic, there could be an accuracy problem (see e.g. Devroye, 1980). This problem can be alleviated if we use $G=1-F$ instead of F and proceed as follows:

Inversion method with more accuracy

Generate an exponential random variate E and a gamma (n) random variate G_n.

RETURN $X \leftarrow G^{-1}(\frac{E}{E+G_n})$.

Unless the distribution function is explicitly invertible, both inversion-based algorithms are virtually useless. In the remaining sections, we present two probabilistic shortcuts, one based upon the quick elimination principle, and one on the use of records. The expected times of these methods usually increase as $\log(n)$. This is not as good as the constant time inversion method, but a lot better than the naive method. The advantages over the inversion method are measured in terms of accuracy and flexibility (fewer things are needed in order to be able to apply the shortcuts).

1.2. The quick elimination principle.

In the quick elimination principle, we generate the maximum of a sequence of iid random variables after having eliminated all but a few of the X_i's without ever generating them. We need a threshold point t and the tail probability $p=1-F(t)$. These are picked before application of the algorithm. Typically, p is of the order of $(\log(n))/n$. The number of X_i's that exceed t is binomial (n,p). Thus, the following algorithm is guaranteed to work:

The quick elimination algorithm (Devroye, 1980)

Generate a binomial (n, p) random variate Z.

IF $Z=0$

THEN

RETURN $X \leftarrow \max(X_1, \ldots, X_n)$ where the X_i's are iid random variates with density $f/(1-p)$ on $(-\infty, t]$.

ELSE

RETURN $X \leftarrow \max(X_1, \ldots, X_Z)$ where the X_i's are iid random variates with density f/p on $[t, \infty)$.

To analyze the expected time complexity, observe that the binomial (n, p) random variate can be generated in expected time proportional to np as $np \rightarrow \infty$ by the waiting time method. Obviously, we could use $O(1)$ expected time algorithms too, but there is no need for this here. Assume furthermore that every X_i in the algorithm is generated in one unit of expected time, uniformly over all values of p. It is easy to see that the expected time of the algorithm is $T+o(np)$ where we define $T=aP(Z=0)n+b(1-P(Z=0))np+cnp$ for some constants $a, b, c>0$.

Lemma 1.1.

$$\inf_{0<p<1} T \sim (b+c)\log(n) \quad (n \rightarrow \infty).$$

If we set

$$p = \frac{\log(n)+\delta_n}{n},$$

then $T \sim (b+c)\log(n)$ provided that the sequence of real numbers δ_n is chosen so that

$$\lim_{n \rightarrow \infty} \delta_n + \log(\log(n)) = \infty, \quad \delta_n = o(\log(n)).$$

Proof of Lemma 1.1.

Note that

$$T = na(1-p)^n + bnp(1-(1-p)^n) + cnp$$
$$\leq (b+c)np + ane^{-np} \ .$$

The upper bound is convex in p with one minimum. Setting the derivative with respect to p equal to zero and solving for p gives the solution

$$p = \frac{1}{n}\log(\frac{an}{b+c}) \ .$$

Resubstitution in the upper bound for T shows that

$$T \leq (b+c)\log(\frac{ane}{b+c}) \ .$$

When $p=(\log(n)+\delta_n)/n$, then the upper bound for T is

$$ae^{-\delta_n} + (b+c)(\log(n)+\delta_n) \ .$$

This $\sim (b+c)\log(n)$ if $\delta_n = o(\log(n))$ and $e^{-\delta_n} = o(\log(n))$. The latter condition is satisfied when $\delta_n + \log(\log(n)) \to \infty$.

Finally, it suffices to work on a lower bound for T. We have for every $\epsilon > 0$ and all n large enough, since the optimal p tends to zero:

$$T \geq (na - bnp)e^{-\frac{np}{1-p}} + (b+c)np$$
$$\geq na(1-\epsilon)e^{-\frac{np}{1-\epsilon}} + (b+c)np.$$

We have already minimized such an expression with respect to p above. It suffices to formally replace n by $n/(1-\epsilon)$, a by $a(1-\epsilon)^2$, and $(b+c)$ by $(b+c)(1-\epsilon)$. Thus,

$$\inf_{0<p<1} T \geq (1-\epsilon)(b+c)\log(\frac{ane}{b+c})$$

for all n large enough. This concludes the proof of Lemma 1.1. ■

A good choice for δ_n in Lemma 1.1 is $\delta_n = \log(\frac{a}{b+c})$. When $Z=0$ in the algorithm, iid random variates from the density $f/(1-p)$ restricted to $(-\infty, t]$ can be generated by generating random variates from f until n values less than or equal to t are observed. This would force us to replace the term $aP(Z=0)n$ in the definition of T by $aP(Z=0)n/(1-p)$. However, all the statements of Lemma 1.1 remain valid.

The main problem is that of the computation of a pair (p,t). For if we start with a value for p, such as the value suggested by Lemma 1.1, then the value for t is given by $F^{-1}(1-p)$ (or $G^{-1}(p)$ where $G=1-F$, if numerical accuracy is of concern). This is unfortunately possible only when the inverse of the distribution function is known. But if the inverse of the distribution were known, we would have been able to generate the maximum quite efficiently by the inversion method. There is a subtle difference though: for here, we need one inversion, even if we would need to generate a million iid random variables all distributed as the maximum X. With the inversion method, a million inversions would be required. If on the other hand we were to start with a value for t, then p would have to be set equal to $\int_t^\infty f = G(t) = 1-F(t)$. This requires knowledge of the distribution function but not of its inverse. The value of t we start with should be such that p satisfies the conditions of Lemma 1.1. Typically, t is picked on theoretical grounds as is now illustrated for the normal density.

Example 1.1.

For the normal density it is known that $G(x) \sim f(x)/x$ as $x \to \infty$. A first approximate solution of $f(t)/t = p$ is $t = \sqrt{2\log(1/p)}$, but even if we substitute the value $p = (\log(n))/n$ in this formula, the value of $G(t)$ would be such that the expected time taken by the algorithm far exceeds $\log(n)$. A second approximation is

$$t = \sqrt{2\log(\frac{1}{p}) - \frac{\log(4\pi)+\log(\log(\frac{1}{p}))}{2\sqrt{2\log(\frac{1}{p})}}} ,$$

with $p = (\log(n))/n$. It can be verified that with this choice, $T = O(\log(n))$. ■

For other densities, one can use similar arguments. For the gamma (a) density for example, we have $G(x) \sim f(x)$ as $x \to \infty$, and $f(x) \le G(x) \le f(x)/(1-(a/x))$ for $a>1, x>a-1$. This helps in the construction of a useful value for t.

The computation of $G(t)$ is relatively straightforward for most distributions. For the normal density, see the series of papers published after the book of Kendall and Stuart (1977) (Cooper (1968), Hill (1969), Hitchin (1973)), the paper by Adams (1969), and an improved version of Adams's method, called algorithm AS66 (Hill (1973)). For the gamma density, algorithm AS32 (Bhattacharjee (1970)) is recommended: it is based upon a continued fraction expansion given in Abramowitz and Stegun (1965).

1.3. The record time method.

In some process simulations one needs a sequence $(Z_{n_1}, \ldots, Z_{n_k})$ of maxima that correspond to one realization of the experiment, where $n_1 < n_2 < \cdots < n_k$. In other words, for all i, we have $Z_i = \max(X_1, \ldots, X_i)$ where the X_i's are iid random variables with common density f. The inversion method requires k inversions, and can be implemented as follows:

Inversion method

$n_0 \leftarrow 0, Z \leftarrow -\infty$

FOR $i := 1$ TO k DO

Generate Z, the maximum of $n_i - n_{i-1}$ iid random variables with common density f.

$Z_{n_i} \leftarrow \max(Z_{n_{i-1}}, Z)$

The record time method introduced in this section requires on the average about $\log(n_k)$ exponential random variates and evaluations of the distribution function. In addition, we need to report the k values Z_{n_i}. When $\log(n_k)$ is small compared to k, the record time method can be competitive. It exploits the fact that in a sequence of n iid random variables with common density f, there are about $\log(n)$ records, where we call the n-th observation a record if it is the largest observation seen thus far. If the n-th observation is a record, then the index n itself is called a record time. It is noteworthy that given the value V_i of the i-th record, and given the record time T_i of the i-th record, $T_{i+1} - T_i$ and V_{i+1} are independent: $T_{i+1} - T_i$ is geometrically distributed with parameter $G(V_i)$:

$$P(T_{i+1} - T_i = j \mid T_i, V_i) = G(V_i)(1 - G(V_i))^{j-1} \quad (j \geq 1) .$$

Also, V_{i+1} has conditional density $f / G(V_i)$ restricted to $[V_i, \infty)$. An infinite sequence of records and record times $\{(V_i, T_i), i \geq 1\}$ can be generated as follows:

The record time method (Devroye, 1980)

$T_1 \leftarrow 1, i \leftarrow 1$

Generate a random variate V_1 with density f .

$p \leftarrow G(V_1)$

WHILE True DO

  $i \leftarrow i+1$

  Generate an exponential random variate E.

  $T_i \leftarrow T_{i-1} + \lceil -E / \log(1-p) \rceil$

  Generate V_i from the tail density $\frac{f(x)}{1-p} I_{[x \geq V_{i-1}]}$.

  $p \leftarrow G(V_i)$

It is a straightforward exercise to report the Z_{n_i} values given the sequence of records and record times. We should exit from the loop when $T_i \geq n_k$. The expected number of loops before halting is thus equal to the expected number of records in a sequence of length n_k , i.e. it is

$$\sum_{i=1}^{n_k} \frac{1}{i} = \log(n_k) + \gamma + o(1)$$

where γ=0.5772... is Euler's constant. We note that the most time consuming operation in every iteration is the evaluation of G . If the inverse of G is available, the lines

Generate V_i from the tail density $\frac{f(x)}{1-p} I_{[x \geq V_{i-1}]}$.

$p \leftarrow G(V_i)$

can be replaced by

Generate a uniform [0,1] random variate U.
$p \leftarrow pU$
$V_i \leftarrow G^{-1}(p)$

A final remark is in order here. If we assume that G can be computed in one unit of time for all distributions, then the (random) time taken by the algorithm is an invariant, because the distribution of record times is distribution-free.

1.4. Exercises.

1. **Tail of the normal density.** Let f be the normal density, let $t>0$ and define $p=G(t)$ where $G=1-F$ and F is the normal distribution function. Prove the following statements:
   A. **Gordon's inequality.** (Gordon (1941), Mitrinovic (1970)).
   $$\frac{t}{t^2+1} f(t) \leq p \leq \frac{1}{t} f(t).$$
   B. As $t \rightarrow \infty$, $G(t) \sim f(t)/t$.
   C. If $t=\sqrt{2\log(n/\log(n))}$, then for the quick elimination algorithm, $T = \Omega(n^{1-\epsilon})$ for every $\epsilon>0$ as $n \rightarrow \infty$.
   D. If $t=s-\frac{1}{2s}(\log(4\pi)+\log(\log(\frac{n}{\log}(n))))$, where s is as in point C, then for the quick elimination algorithm, $T = O(\log(n))$. Does $T \sim (b+c)\log(n)$ if b,c are the constants in the definition of T (see Lemma 1.1) ?
2. Let $T_1, T_2, \ldots$ be the record times in a sequence of iid uniform [0,1] random variables. Prove that $E(T_2)=\infty$. Show furthermore that $\log(T_n) \sim n$ in probability as $n \rightarrow \infty$.

2. RANDOM VARIATES WITH GIVEN MOMENTS

2.1. The moment problem.

The classical moment problem can be formulated as follows. Let $\{\mu_i ; 1 \leq i\}$ be a collection of moments. Determine whether there is at least one distribution which gives rise to these moments; if so, construct such a distribution and determine whether it is unique. Solid detailed treatments of this problem can be found in Shohat and Tamarkin (1943) and Widder (1941). The main result is the following.

Theorem 2.1.

If there exists a distribution with moments μ_i, $1 \leq i$, then

$$\begin{vmatrix} 1 & \mu_1 & \cdots & \mu_s \\ \mu_1 & \mu_2 & & \mu_{s+1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_s & & \cdots & \mu_{2s} \end{vmatrix} \geq 0$$

for all integers s with $s \geq 1$. The inequalities hold strictly if the distribution is nonatomic. Conversely, if the matrix inequality holds strictly for all integers s with $s \geq 1$, then there exists a nonatomic distribution matching the given moments.

Proof of Theorem 2.1.

We will only outline why the matrix inequality is necessary. Considering the fact that

$$E\left((c_0 + c_1 X + \cdots + c_s X^s)^2\right) \geq 0$$

for all values of $c_0, \ldots, c_s$, we have by a standard result from linear algebra (Mirsky (1955, p. 400)) that

$$\begin{vmatrix} 1 & \mu_1 & \cdots & \mu_s \\ \mu_1 & \mu_2 & & \mu_{s+1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_s & & \cdots & \mu_{2s} \end{vmatrix} \geq 0 . \blacksquare$$

Theorem 2.2.

If there exists a distribution on $[0,\infty)$ with moments μ_i , $1 \le i$, then

$$\begin{vmatrix} 1 & \mu_1 & \cdots & \mu_s \\ \mu_1 & \mu_2 & & \mu_{s+1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_s & \cdot & \cdots & \mu_{2s} \end{vmatrix} \ge 0 ,$$

$$\begin{vmatrix} \mu_1 & \mu_2 & \cdots & \mu_{s+1} \\ \mu_2 & \mu_3 & & \mu_{s+2} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_{s+1} & \cdot & \cdots & \mu_{2s+1} \end{vmatrix} \ge 0 ,$$

for all integers $s \ge 0$. The inequalities hold strictly if the distribution is nonatomic. Conversely, if the matrix inequality holds strictly for all integers $s \ge 0$, then there exists a nonatomic distribution matching the given moments.

The determinants in Theorems 2.1, 2.2 are called Hankel determinants. What happens when one or more of them are zero is more complicated (see e.g. Widder (1941)). The problem of the uniqueness of a distribution is covered by Theorem 2.3.

Theorem 2.3.

Let $\mu_1,\mu_2,\ldots$ be the moment sequence of at least one distribution. Then this distribution is unique if Carleman's condition holds, i.e.

$$\sum_{i=0}^{\infty} \mid \mu_{2i} \mid^{-\frac{1}{2i}} = \infty .$$

If we have a distribution on the positive halfline, then a sufficient condition for uniqueness is

$$\sum_{i=0}^{\infty} (\mu_i)^{-\frac{1}{2i}} = \infty .$$

When the distribution has a density f , then a necessary and sufficient condition for uniqueness is

$$\int_{-\infty}^{\infty} \frac{\log(f(x))}{1+x^2} \, dx = -\infty$$

(Krein's condition).

For example, normal distributions or distributions on compact sets satisfy Carleman's condition and are thus uniquely determined by their moment sequence. In exercises 2.2 and 2.3, examples are developed of distributions having identical infinite moment sequences, but widely varying densities. In exercise 2.2, a unimodal discrete distribution is given which has the same moments as the log-normal distribution.

The problem that we refer to as the moment problem is that of the generation of a random variate with a given collection of moments $\mu_1,\mu_2, \ldots , \mu_n$, where n can be ∞. Note that if we expand the characteristic function ϕ of a random variable in its Taylor series about 0, then

$$\phi(t) = \phi(0)+\frac{t}{1!}\phi^{(1)}(0)+ \cdots +\frac{t^{k-1}}{(k-1)!}\phi^{(k-1)}(0)+R_k$$

where the remainder term satisfies

$$\mid R_k \mid \leq \mu_k \frac{\mid t \mid^k}{k!} .$$

This uses the fact that if $\mid \mu_k \mid < \infty$, the k-th derivative of ϕ exists, and is a continuous function given by $E((iX)^k e^{itX})$. In particular, the k-th derivative is in absolute value not greater than $E(\mid X \mid^k)$. See for example Feller (1971, pp. 512-514). The remainder term R_k tends to 0 in a neighborhood of the origin when

$$\limsup \frac{\mid \mu_k \mid^{1/k}}{k} < \infty .$$

Thus, the Taylor series converges in those cases. It follows that ϕ is analytic in a neighborhood of the origin, and hence completely determined by its power series about the origin. The condition given above is thus sufficient for the moment sequence to uniquely determine the distribution. One can verify that the condition is weaker, but not much weaker, than Carleman's condition. The point of all this is that if we are given an infinite moment sequence which uniquely determines the distribution, we are in fact given the characteristic function in a special form. The problem of the generation of a random variate with a given characteristic function will be dealt with in section 3. Here we will mainly be concerned with the finite moment case. This is by far the most important case in practice, because researchers usually worry about matching the first few moments, and because the majority of distributions have only a finite number of finite moments. Unfortunately, there are typically an infinite number of distributions sharing the same first n moments. These include discrete distributions and distributions with densities. If some additional constraints are satisfied by the moments, it may be possible to pick a distribution from relatively small classes of distributions. These include:

A. The class of all unimodal densities, i.e. uniform scale mixtures.

B. The class of normal scale mixtures.

C. Pearson's system of densities.

D. Johnson's system of densities.

E. The class of all histograms.

F. The class of all distributions of random variables of the form $a+bN+cN^2+dN^3$ where N is normally distributed.

The list is incomplete, but representative of the attempts made in practice by some statisticians. For example, in cases C,D and F, we can match the first four moments with those of exactly one member in the class except in case F, where some combinations of the first four moments have no match in the class. The fact that a match always occurs in the Pearson system has contributed a lot to the early popularity of the system. For a description and details of the Pearson system, see exercise IX.7.4. Johnson's system (exercise IX.7.12) is better for quantile matching than moment matching. We also refer the reader to the Burr family (section IX.7.4) and other families given in section IX.7.5. These families of distributions are usually designed for matching up to four moments. This of course is their main limitation. What is needed is a general algorithm that can be used for arbitrary $n>4$. In this respect, it may first be worthwhile to verify whether there exists a uniform or normal scale mixture having the given set of moments. If this is the case, then one could proceed with the construction of one such distribution. If this attempt fails, it may be necessary to construct a matching histogram or discrete distribution (note that discrete distributions are limits of histograms). Good references about the moment problem include Widder (1941), Shohat and Tamarkin (1943), Godwin (1964), von Mises (1964), Hill (1969) and Springer (1979).

2.2. Discrete distributions.

Assume that we want to match the first $2n-1$ moments with those of a discrete distribution having n atoms located at $x_1, \ldots, x_n$, with respective weights $p_1, \ldots, p_n$. We know that we should have

$$\sum_{i=1}^{n} p_i (x_i)^j = \mu_j \quad (0 \leq j \leq 2n-1) .$$

This is a system of $2n$ equalities with $2n$ unknowns. It has precisely one solution if at least one distribution exists with the given moments (von Mises, 1964). In particular, if the locations x_i are known, then the p_i's can be determined from the first n linear equations. The locations can first be obtained as the n roots of the equation

$$x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0 = 0 ,$$

where the c_i's are the solutions of

$$\begin{vmatrix} \mu_0 & \cdot & \mu_{n-1} \\ \mu_1 & \cdot & \mu_n \\ \cdot & & \cdot \\ \mu_{n-1} & \cdot & \mu_{2n-2} \end{vmatrix} \begin{vmatrix} c_0 \\ c_1 \\ \cdot \\ c_{n-1} \end{vmatrix} = - \begin{vmatrix} \mu_n \\ \mu_{n+1} \\ \cdot \\ \mu_{2n-1} \end{vmatrix} .$$

To do this could take some valuable time, but at least we have a minimal solution, in the sense that the distribution is as concentrated as possible in as few atoms as possible. One could argue that this yields some savings in space, but n is rarely large enough to make this the deciding factor. On the other hand, it is impossible to start with $2n$ locations of atoms and solve the $2n$ equations for the weights p_i, because there is no guarantee that all p_i's are nonnegative.

If an even number of moments is given, say $2n$, then we have $2n+1$ equations. If we consider $n+1$ atom locations with $n+1$ weights, then there is an excess of one variable. We can thus choose one item, such as the location of one atom. Call this location a. Shohat and Tamarkin (1943) (see also Royden, 1953) have shown that if there exists at least one distribution with the given moments, then there exists at least one distribution with at most $n+1$ atoms, one of them located at a, sharing the same moments. The locations $x_0, \ldots, x_n$ of the atoms are the zeros of

$$\begin{vmatrix} 1 & 1 & \mu_0 & \cdot & \mu_{n-1} \\ x & a & \mu_1 & \cdot & \mu_n \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ x^{n+1} & a^{n+1} & \mu_{n+1} & \cdot & \mu_{2n} \end{vmatrix} = 0 .$$

The weights $p_0, p_1, \ldots, p_n$ are linear combinations of the moments:

$$p_i = \sum_{j=0}^{n} c_{ji} \mu_j .$$

The coefficients c_{ji} in turn are defined by the identity

$$\sum_{j=0}^{n} c_{ji} x^j \equiv \prod_{j \neq i} \frac{x-x_j}{x_i-x_j} \quad (0 \leq i \leq n) .$$

When the distribution puts all its mass on the nonnegative real line, a slight modification is necessary (Royden, 1953). Closely related to discrete distributions are the histograms: these can be considered as special cases of distributions with densities

$$f(x) = \sum_{i=1}^{n} \frac{p_i}{h_i} K\left(\frac{x-x_i}{h_i}\right) ,$$

where K is a fixed form density (such as the uniform [-1,1] density in the case of a histogram), x_i is the center of the i-th component, p_i is the weight of the i-th component, and h_i is the "width" of the i-th component. Densities of this form are well-known in the nonparametric density estimation literature: they are the kernel estimates. Archer (1980) proposes to solve the moment equations numerically for the unknown parameters in the histogram. We should point out that the density f shown above is the density of $x_Z + h_Z Y$ where Y has density K, and Z has probability vector $p_1, \ldots, p_n$ on $\{1, \ldots, n\}$. This greatly facilitates the computations and the visualization process.

2.3. Unimodal densities and scale mixtures.

A random variable X has a unimodal distribution if and only if there exists a random variable Y such that X is distributed as YU where U is a uniform [0,1] random variable independent of Y (Khinchine's theorem). If U is not uniform and Y is arbitrary then the distribution of X is called a scale mixture for U. Of particular importance are the normal scale mixtures, which correspond to the case when U is normally distributed. For us it helps to be able to verify whether for a given collection of n moments, there exists a unimodal distribution or a scale mixture which matches these moments. Usually, we have a particular scale mixture in mind. Assume for example that U has moments $\nu_1, \nu_2, \ldots$. Then, because $E(X^i) = E(Y^i)E(U^i)$, we see that Y has i-th moment μ_i / ν_i. Thus, the existence problem is solved if we can find at least one distribution having moments μ_i / ν_i.

Applying Theorem 2.1, then we observe that a sufficient condition for the moment sequence μ_i to correspond to a U scale mixture is that the determinants

$$\begin{vmatrix} 1 & \mu_1/\nu_1 & \cdots & \mu_s/\nu_s \\ \mu_1/\nu_1 & \mu_2/\nu_2 & & \mu_{s+1}/\nu_{s+1} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \mu_s/\nu_s & \cdot & \cdots & \mu_{2s}/\nu_{2s} \end{vmatrix} \geq 0$$

are all positive for $2s < n$, n odd. This was first observed by Johnson and Rogers (1951). For uniform mixtures, i.e. unimodal distributions, we should replace ν_i by $1/(i+1)$ in the determinants. Having established the existence of a scale mixture with the given moments, it is then up to us to determine at least one Y with moment sequence μ_i / ν_i. This can be done by the methods of the previous section.

By insisting that a particular scale mixture be matched, we are narrowing down the possibilities. By this is meant that fewer moment sequences lead to solutions. The advantage is that if a solution exists, it is typically "nicer" than in the discrete case. For example, if Y is discrete with no atom at 0, and U is uniform, then X has a unimodal staircase-shaped density with mode at the origin and breakpoints at the atoms of Y. If U is normal, then X is a superposition of a few normal densities centered at 0 with different variances. Let us illustrate briefly how restrictive some scale mixtures are. We will take as example the case of four moments, with normalized mean and variance, $\mu_1=0, \mu_2=1$. Then, the conditions of Theorem 2.1 imply that we must always have

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \mu_3 \\ 1 & \mu_3 & \mu_4 \end{vmatrix} \geq 0 .$$

Thus, $\mu_4 \geq (\mu_3)^2+1$. It turns out that for all μ_3, μ_4 satisfying the inequality, we can find at least one distribution with these moments. Incidentally, equality occurs for the Bernoulli distribution. When the inequality is strict, a density exists. Consider next the case of a unimodal distribution with zero mean and unit variance. The existence of at least one distribution with the given moments is guaranteed if

$$\begin{vmatrix} 1 & 0 & 3 \\ 0 & 3 & 4\mu_3 \\ 3 & 4\mu_3 & 5\mu_4 \end{vmatrix} \geq 0 ,$$

in other words, $\mu_4 \geq \frac{9}{5} + \frac{16}{15}(\mu_3)^2$. It is easy to check that in the (μ_3, μ_4) plane, a smaller area gets selected by this condition. It is precisely the (μ_3, μ_4) plane which can help us in the fast construction of moment matching distributions. This is done in the next section.

2.4. Convex combinations.

If Y and Z are random variables with moment sequences μ_i and ν_i respectively, then the random variable X which equals Y with probability p and Z with probability $1-p$ has moment sequence $p\,\mu_i+(1-p)\nu_i$, in other words, it is the convex combination of the original moment sequences. Assume that we want to match four normalized moments. Recall that the allowable area in the (μ_3,μ_4) plane is the area above the parabola $\mu_4 \geq (\mu_3)^2+1$. Every point (μ_3,μ_4) in this area lies on a horizontal line at height μ_4 which intersects the parabola at the points $(-\sqrt{\mu_4-1},\mu_4)$, $(\sqrt{\mu_4-1},\mu_4)$. In other words, we can match the moments by a simple convex combination of two distributions with third and fourth moments $(-\sqrt{\mu_4-1},\mu_4)$ and $(\sqrt{\mu_4-1},\mu_4)$ respectively.

The weight in the convex combination is determined quite easily since we must have, attaching weight p to the distribution with positive third moment,

$$(p-(1-p))\sqrt{\mu_4-1} = \mu_3 \ .$$

Thus, it suffices to take

$$p = \frac{1+\frac{\mu_3}{\sqrt{\mu_4-1}}}{2} \ .$$

It is also easy to verify that for a Bernoulli (q) random variable, we have normalized fourth moment

$$\frac{3q^2-3q+1}{q(1-q)}$$

and normalized third moment

$$\frac{1-2q}{\sqrt{q(1-q)}} \ .$$

Notice that this distribution always falls on the limiting parabola. Furthermore, by letting q vary from 0 to 1, all points on the parabola are obtained. Given the fourth moment μ_4, we can determine q via the equation

$$q = \frac{1}{2}\left(1 \pm \sqrt{\frac{\mu_4-1}{\mu_4+3}}\right) ,$$

where the plus sign is chosen if $\mu_3 \geq 0$, and the minus sign is chosen otherwise. Let us call the solution with the plus sign q. The minus sign solution is $1-q$. If B is a Bernoulli (q) random variable, then $(B-q)/\sqrt{q(1-q)}$ and $-(B-q)/\sqrt{q(1-q)}$ are the two random variables corresponding to the two intersection points on the parabola. Thus, the following algorithm can be used to generate a general random variate with four moments $\mu_1, \ldots, \mu_4$:

Generator matching first four moments

Normalize the moments: $\sigma \leftarrow \sqrt{\mu_2-(\mu_1)^2}$,

$$(\mu_3,\mu_4) \leftarrow \left(\frac{\mu_3-3\mu_2\mu_1+2(\mu_1)^3}{\sigma^3}, \frac{\mu_4-4\mu_3\mu_1+6\mu_2(\mu_1)^2-3(\mu_1)^4}{\sigma^4} \right)$$

$$q \leftarrow \frac{1}{2}\left(1+\sqrt{\frac{\mu_4-1}{\mu_4+3}}\right)$$

$$p \leftarrow \frac{1+\frac{\mu_3}{\sqrt{\mu_4-1}}}{2}$$

Generate a uniform [0,1] random variate U.

IF $U \le p$

THEN

$X \leftarrow I_{[U \le pq]}$ (X is Bernoulli (q))

RETURN $X \leftarrow \mu_1 + \sigma \frac{X-q}{\sqrt{q(1-q)}}$

ELSE

$X \leftarrow I_{[U \le p+(1-p)q]}$ (X is Bernoulli (q))

RETURN $X \leftarrow \mu_1 - \sigma \frac{X-q}{\sqrt{q(1-q)}}$

The algorithm shown above can be shortened by a variety of tricks. As it stands, one uniform random variate is needed per returned random variate. The point of this example is that it is very simple to generate random variates that match four moments if one is not picky. Indeed, few users will be pleased with the convex combination of two Bernoulli distributions used in the example. But interestingly, the example can also be used in the construction of the distribution of Y in scale mixtures of the form YU discussed in the previous section. In that respect, the algorithm becomes more useful, because the returned distributions are "nicer". The algorithm for unimodal distributions with mode at 0 is given below.

Simple unimodal distribution generator matching four moments

Readjustment of moments: $\mu_1 \leftarrow 2\mu_1$, $\mu_2 \leftarrow 3\mu_2$, $\mu_3 \leftarrow 4\mu_3$, $\mu_4 \leftarrow 5\mu_5$.

Generate a random variate Y having the readjusted moments (e.g. by the algorithm given above).

Generate a uniform [0,1] random variate U.

RETURN $X \leftarrow YU$.

The algorithms for other scale mixtures are similar.

One final remark about moment matching is in order here. Even with a unimodality constraint, there are many distributions with widely varying densities but identical moments up to the n-th moment. One should therefore always ask the question whether it is a good thing at all to blindly go ahead and generate random variates with a certain collection of moments. Let us make this point with two examples.

Example 2.1.(Godwin, 1964)

The following two densities have identical infinite moment sequences:

$$f(x) = \frac{1}{4} e^{-|x|^{\frac{1}{2}}} \quad (x \in R) .$$

$$g(x) = \frac{1}{4} e^{-|x|^{\frac{1}{2}}} (1 + \cos(\sqrt{|x|}) \quad (x \in R)$$

(Kendall and Stuart (1977), see exercise 2.3). Thus, noting that

$$\int_A f = 0.4656... ; \int_A g = 0.7328... ,$$

where $A = [-\pi^2/4, \pi^2/4]$, we observe that

$$\int | f - g | \geq 0.5344... .$$

Considering that the L_1 distance between two densities is at most 2, the distance 0.5344... is phenomenally large. ■

Example 2.2.

The previous example involves a unimodal and an oscillating density. But even if we enforce unimodality on our counterexamples, not much changes. See for example Leipnik's example described in exercise 2.2. Another way of illustrating this is as follows: for any symmetric unimodal density f with moments μ_2, μ_4, it is true that

$$\sup_g \int |f-g| \geq \omega^2(1-\omega)$$

where the supremum is taken over all symmetric unimodal g with the same second and fourth moments, and $\omega=\sqrt{(3\mu_2)^2/(5\mu_4)}$. It should be noted that $0\leq\omega\leq 1$ in all cases (this follows from the nonnegativity of the Hankel determinants applied to unimodal distributions). When f is normal, $\omega=\sqrt{3/5}$ and the lower bound is $\frac{3}{5}(1-\sqrt{\frac{3}{5}})$, which is still quite large. For some combinations of moments, the lower bound can be as large as $\frac{4}{27}$. There are two differences with Example 2.1: we are only matching the first four moments, not all moments, and the counterexample applies to any symmetric unimodal f, not just one density picked beforehand for convenience. Example 2.2 thus reinforces the belief that the moments contain surprisingly little information about the distribution. To prove the inequality of this example, we will argue as follows: let f,g,h be three densities in the given class of densities. Clearly,

$$\max(\int |f-h|, \int |f-g|) \geq \frac{1}{2}(\int |f-h| + \int |f-g|)$$
$$\geq \frac{1}{2}\int |h-g| .$$

Thus it suffices to prove twice the lower bound for $\int |h-g|$ for two particular densities h,g. Consider densities of random variables YU where U is uniformly distributed on [0,1] and Y is independent of U and has a symmetric discrete distribution with atoms at $\pm b, \pm c$, where $0<b<c<\infty$. The atom at c has weight $p/2$, and the atom at b has weight $(1-p)/2$. For h and g we will consider different choices of b,c,p. First, any choice must be consistent with the moment restrictions:

$$(1-p)b^2+pc^2 = 3\mu_2 ,$$
$$(1-p)b^4+pc^4 = 5\mu_4 .$$

Solving for p gives

$$1-p = \frac{5\mu_4-3\mu_2c^2}{b^4-b^2c^2} .$$

Forcing $p\in[0,1]$ gives us the constraints $0\leq 3\mu_2c^2-5\mu_4\leq b^2(c^2-b^2)$. It is to our advantage to take the extreme values for c. In particular, for g we will take $c=\sqrt{(5\mu_4)/(3\mu_2)}$, $b=0$, $p=\omega^2$. It should be noted that this not yield a density g since there will be an atom at the origin. Thus, we use an approximating argument with a sequence g_n approaching g in the sense that the atom at 0 is approached by an atom at $\epsilon_n \rightarrow 0$. Next, for h, we take the limit of the sequence h_n where as $n \rightarrow \infty$, $b \rightarrow \sqrt{3\mu_2}$, $p \rightarrow 0$, and $c \rightarrow \infty$. This is the case in which the rightmost atom escapes to infinity but has increasingly negligible weight p. Since $p \rightarrow 0$, the contribution of the rightmost atom to the L_1 distance is also $o(1)$. Thus, h can be considered as having one atom at $\sqrt{3\mu_2}$ of weight 1/2. We obtain by simple geometrical considerations,

$$\lim_{n \rightarrow \infty} \int | g_n - h_n | = 4(\sqrt{(5\mu_4)/(3\mu_2)} - \sqrt{3\mu_2})(\frac{1}{2}\omega^2 \frac{1}{\sqrt{(5\mu_4)/(3\mu_2)}})$$

$$= 2\omega^2(1-\omega) .$$

Since the sequences h_n, g_n are entirely in our class, we see that the lower bound for $\sup_g \int | f - g |$ is at least $\omega^2(1-\omega)$. ■

2.5. Exercises.

1. Show that for the normal density, the $2i$-th moment is

$$\mu_{2i} = (2i-1)(2i-3) \cdots (3)(1) \quad (i \geq 2) .$$

Show furthermore that Carleman's condition holds.

2. **The lognormal density.** In this exercise, we consider the lognormal density

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log(x))^2}{2\sigma^2}} \quad (x > 0) .$$

Show first that this density fails both Carleman's condition and Krein's condition. Hint: show first that the r-th moment is $\mu_r = e^{\sigma^2 r^2/2}$. Thus, there exist other distributions with the same moments. We will construct a family of such distributions, referred to hereafter as Heyde's family (Heyde (1963), Feller (1971, p. 227)): let $-1 \leq a \leq 1$ be a parameter, and define the density

$$f_a(x) = f(x)(1+a\sin(2\pi\log(x))) \quad (x > 0) .$$

To show that f_a is a density, and that all the moments are equal to the moments of $f_0 = f$, it suffices to show that

$$\int_0^\infty x^k f(x)\sin(2\pi\log(x))\, dx = 0$$

for all integer $k \geq 0$. Show this. Show also the following result due to Leipnik (1981): there exists a family of discrete unimodal random variables X having the same moments as a lognormal random variable. It suffices to let X take the value $ae^{\sigma i}$ with probability $ca^{-i} e^{-\sigma^2 i^2/2}$ for $i=0,\pm1,\pm2,\ldots$, where $a>0$ is a parameter, and c is a normalization constant.

3. **The Kendall-Stuart density.** Kendall and Stuart (1977) introduced the density

$$f(x)=\frac{1}{4}e^{-|x|^{\frac{1}{2}}} \quad (x \in R).$$

Following Kendall and Stuart, show that for all real a with $|a| \leq 1$,

$$f_a(x)=\frac{1}{4}e^{-|x|^{\frac{1}{2}}}(1+a\cos(\sqrt{|x|}) \quad (x \in R)$$

are densities with moments equal to those of f.

4. Yet another family of densities sharing the same moment sequence is given by

$$f_a(x)=e^{-x^{\frac{1}{4}}}\frac{(1-a\sin(x^{\frac{1}{4}}))}{24} \quad (x>0),$$

where $a \in [0,1)$ is a parameter. Show that f_0 violates Krein's condition and that all moments are equal to those of f_0. This example is due to Stieltjes (see e.g. Widder (1941, pp. 125-126)).

5. Let $p \in (0,\frac{1}{2})$ be a parameter, and let $c=(p\cos(p\pi))^{1/p}/\Gamma(1/p)$ be a constant. Show that the following two densities on $(0,\infty)$ have the same moments:

$$f(x)=c\ e^{-x^p\cos(p\pi)},$$

$$g(x)=f(x)(1+\sin(x^p\sin(p\pi)))$$

(Lukacs (1970, p. 20)).

6. **Fleishman's family of distributions.** Consider all random variables of the form $a+bN+cN^2+dN^3$ where N is a normal random variable, and a,b,c,d are constants. Many distributions are known to be approximately normal, and can probably be modeled by distributions of random variables of the form given above. This family of distributions, studied by Fleishman (1978), has the advantage that random variate generation is easy once the constants are determined. To compute the constants, the first four moments can be matched with fixed values μ_1,μ_2,μ_3,μ_4. For the sake of simplicity, let us normalize as follows: $\mu_1=0,\mu_2=1$. Show that b,d can be found by solving

$$1=b^2+6bd+15d^2+2c^2,$$

$$\mu_4-3=24(bd+c^2(1+b^2+28bd)+d^2(12+48bd+141c^2+255d^2)),$$

where

$$c = \frac{\mu_3}{2(b^2+24bd+105d^2+2)} .$$

Furthermore, $a=-c$. Show that not all combinations of normalized moments of distributions (i.e. all pairs (μ_3,μ_4) with $\mu_4 \geq (\mu_3)^2+1$) lead to a solution. Determine the region in the (μ_3,μ_4) plane of allowable pairs. Finally, prove that there exist combinations of constants for which the density is not unimodal, and determine the form of the distribution in these cases.

7. Assume that we wish to match the first six moments of a symmetric distribution (all odd moments are zero). We normalize by forcing μ_2 to be 1. Show first that the allowable region in the (μ_4,μ_6) plane is defined by the inequalities $\mu_4 \geq 1$, $\mu_6 \geq (\mu_4)^2$. Find simple families of distributions which cover the borders of this region. Rewrite each point in the plane as the convex combination of two of these simple distributions, and give the corresponding generator, i.e. the generator for the distribution that corresponds to this point.
8. Let the a-th and b-th absolute moments of a unimodal symmetric distribution with a density be given. Find a useful lower bound for

$$\inf_f \sup_g \int |f-g| ,$$

where the infimum and supremum is over all symmetric unimodal densities having the given absolute moments. The lower bound should coincide with that of Example 2.2 in the case $a=2,b=4$.

3. CHARACTERISTIC FUNCTIONS.

3.1. Problem statement.

In many applications, a distribution is best described by its characteristic function ϕ. Sometimes, it is outright difficult to invert the characteristic function to obtain a value for the density or distribution function. One might ask whether in those cases, it is still possible to generate a random variate X with the given distribution. An example of such a distribution is the stable distribution. In particular, the symmetric stable distribution with parameter $\alpha \in (0,2]$ has the simple characteristic function $e^{-|t|^\alpha}$. Yet, except for $\alpha \in \{\frac{1}{2},1,2\}$, no convenient analytic expression is known for the corresponding density f; the density is best computed with the help of a convergent series or a divergent asymptotic expansion (section IX.6.3). For random variate generation in this simple case, we refer to section IX.6. For $\alpha \in (0,1]$ the characteristic function can be written as a mixture of triangular characteristic functions. This property is shared by all real (thus, symmetric) convex characteristic functions, also called Polya characteristic functions. The mixture property can be used to obtain generators (Devroye, 1984; see also section IV.6.7). In a black box method one only assumes that ϕ belongs to a certain class of characteristic functions, and that $\phi(t)$ can be computed in finite time for every t. Thus, making use of the mixture property of Polya characteristic functions cannot lead to a black box method because ϕ has to be given explicitly in analytic form.

Under certain regularity conditions, upper bounds for the density can be obtained in terms of quantities (functionals, suprema, and so forth) defined in terms of the characteristic function (Devroye, 1981). These upper bounds can in turn be used in a rejection algorithm. This simple approach is developed in section 3.2. Unfortunately, one now needs to compute f in every iteration of the rejection algorithm. This requires once again an inversion of ϕ, and may not be feasible. One should note however that this can be avoided if we are able to use the series method based upon a convergent series for f. This series could be based upon the inversion formula.

A genuine black box method for a large subclass of Polya characteristic functions was developed in Devroye (1985). Another black box method based upon the series method will be studied in section 3.3.

3.2. The rejection method for characteristic functions.

General rejection algorithms can be based upon the following inequality:

Theorem 3.1.

Assume that a given distribution has two finite moments, and that the characteristic function ϕ has two absolutely integrable. Then the distribution has a density f bounded as follows:

$$f(x) \leq \begin{cases} \frac{1}{2\pi}\int |\phi| \\ \frac{1}{2\pi x^2}\int |\phi''| \end{cases} .$$

The area under the minimum of the two bounding curves is $\frac{2}{\pi}\sqrt{\int |\phi| \int |\phi''|}$.

Proof of Theorem 3.1.

Since ϕ is absolutely integrable, f can be computed as follows from ϕ:

$$f(x) = \frac{1}{2\pi} \int \phi(t) e^{-itx} \, dt .$$

Furthermore, because the first absolute moment is finite, ϕ' exists and

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

Because the second moment is finite, ϕ'' exists and

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

(Loeve, 1963, p. 199). From this, all the inequalities follow trivially. ■

The integrability condition on ϕ implies that f is bounded and continuous. The integrability condition on ϕ'' translates into a strong tail condition: the tail of f can be tucked under a quickly decreasing curve. This explains why f can globally be tucked under a bounded integrable curve. Based upon Theorem 3.1, we can now formulate a first general rejection algorithm for characteristic functions satisfying the conditions of the Theorem.

General rejection algorithm for characteristic functions

[SET-UP]

$a \leftarrow \frac{1}{2\pi} \int |\phi|$, $b \leftarrow \frac{1}{2\pi} \int |\phi''|$

[GENERATOR]

REPEAT

Generate two iid uniform [-1,1] random variates U, V.

IF $U < 0$

THEN $X \leftarrow \sqrt{\frac{b}{a}} V$, $T \leftarrow |U| a$

ELSE $X \leftarrow \sqrt{\frac{b}{a}} \frac{1}{V}$, $T \leftarrow \frac{|U| b}{X^2}$

(Note that this is $|U| a V^2$.)

UNTIL $T \leq f(X)$

RETURN X

Various simplifications are possible in this rudimentary algorithm. What matters is that f is still required in the acceptance step.

Remark 3.1.

The expected number of iterations is $\frac{2}{\pi}\sqrt{\int |\phi| \int |\phi''|}$. This is a scale invariant quantity: indeed, let X have characteristic function ϕ. Then, under the conditions of Theorem 3.1, $\phi(t)=E(e^{itX})$, $\phi''(t)=E(-X^2e^{itX})$. For the scaled random variable aX, we obtain respectively $\phi(at)$ and $a^2\phi''(at)$. The product of the integrals of the last two functions does not depend upon a. Unfortunately, the product is not translation invariant. Noting that $X+c$ has characteristic function $\phi(t)e^{itc}$, we see that $\int |\phi|$ is translation invariant. However,

$$\int |\phi''| = \int |E(-(X-c)^2e^{itX})|$$

is not. From the quadratic form of the integrand, one deduces quickly that the integral is approximately minimal when $c=E(X)$, i.e. when the distribution is centered at the mean. This is a common sense observation, reinforced by the symmetric form of the dominating curve. Let us finally note that in Theorem 3.1 we have implicitly proved the inequality

$$\int |\phi| \int |\phi''| \geq \frac{\pi^2}{4},$$

which is of independent interest in mathematical statistics. ■

If the evaluation of f is to be avoided, then we must find at the very least a converging series for f. Assume first that ϕ is absolutely integrable, symmetric and nonnegative. Then $f(x)$ is sandwiched between consecutive partial sums in the series

$$f(0)-\frac{x^2}{2!}f'(0)+\frac{x^4}{4!}f''(0)- \cdots .$$

This can be seen as follows: since $\cos(tx)$ is sandwiched between consecutive partial sums in its Taylor series expansion, and since

$$f(x) = \frac{1}{2\pi}\int \phi(t)\cos(tx)\, dt ,$$

we see that by our assumptions on ϕ, $f(x)$ is sandwiched between consecutive partial sums in

$$\nu_0-\frac{x^2}{2!}\nu_2+\frac{x^4}{4!}\nu_4- \cdots ,$$

where

$$\nu_{2n} = \frac{1}{2\pi} \int t^{2n} \phi(t)\, dt \ .$$

If $\int t^{2n} \phi(t)\, dt$ is finite, then $f^{(2n)}$ exists, and its value at 0 is equal to it. This gives the desired collection of inequalities. Note thus that for an inequality involving $f^{(2n)}$ to be valid, we need to ask that

$$\int t^{2n} \phi(t)\, dt < \infty \ .$$

This moment condition on ϕ is a smoothness condition on f. For extremely smooth f, all moments can be finite. Examples include the normal density, the Cauchy density and all symmetric stable densities with parameter at least equal to one. Also, all characteristic functions with compact support are included, such as the triangular characteristic function. If furthermore the series $x^{2n} \nu_{2n} /(2n)!$ is summable for all $x > 0$, we see that f is determined by all its derivatives at 0. A sufficient condition is

$$\nu_{2n}^{\frac{1}{2n}} = o(n) \ .$$

This class of densities is enormously smooth. In addition, these densities are unimodal with a unique mode at 0 (see exercises). Random variate generation can thus be based upon the alternating series method. As dominating curve, we can use any curve available to us. If Theorem 3.1 is used, note that $\int |\phi| = \int \phi = f(0)$.

Series method for very smooth densities

[NOTE: This algorithm is valid for densities with a symmetric real nonnegative characteristic function for which the value of f is uniquely determined by the Taylor series expansion of f about 0.]

[SET-UP]

$a \leftarrow \frac{1}{2\pi}\int |\phi| \ (=f(0))$, $b \leftarrow \frac{1}{2\pi}\int |\phi''|$.

[GENERATOR]

REPEAT

Generate a uniform [0,1] random variate U, and a random variate X with density proportional to $g(x)=\min(a, b/x^2)$.

$T \leftarrow Ug(X)$

$S \leftarrow f(0)$, $n \leftarrow 0$, $Q \leftarrow 1$ (prepare for series method)

WHILE $T \le S$ DO

$n \leftarrow n+1$, $Q \leftarrow -QX^2/(2n(2n-1))$

$S \leftarrow S + Qf^{(n)}(0)$

IF $T \le S$ THEN RETURN X

$n \leftarrow n+1$, $Q \leftarrow -QX^2/(2n(2n-1))$, $S \leftarrow S+Qf^{(n)}(0)$

UNTIL False

This algorithm could have been presented in the section on the series method, or in the section on universal algorithms. It has a place in this section because it shows how one can avoid inverting the characteristic function in a general rejection method for characteristic functions.

3.3. A black box method.

When ϕ is absolutely integrable, the value of the density f can be computed by the inversion formula

$$f(x) = \frac{1}{2\pi}\int \phi(t)e^{-itx}\ dt = \int \psi(t)\ dt \ .$$

This integral can be approximated in a number of ways, by using well-known techniques from numerical integration. If such approximations are to be useful, it is essential that we have good explicit estimates of the error. The approximations include the **rectangular rule**

$$r_n(x) = \frac{b-a}{n}\sum_{j=0}^{n-1}\psi(a+(b-a)\frac{j}{n}) \ ,$$

where $[a,b]$ is a finite interval. Other popular rules are the **trapezoidal rule**

$$t_n(x) = \frac{b-a}{n}\sum_{j=1}^{n}\left(\frac{1}{2}\psi\left(a+\frac{(j-1)(b-a)}{n}\right)+\frac{1}{2}\psi\left(a+\frac{j(b-a)}{n}\right)\right),$$

and **Simpson's rule**

$$s_n(x) = \frac{b-a}{n}\sum_{j=1}^{n}\left(\frac{1}{6}\psi\left(a+\frac{(j-1)(b-a)}{n}\right)\right.$$

$$\left.+\frac{4}{6}\psi\left(a+\frac{(j-\frac{1}{2})(b-a)}{n}\right)+\frac{1}{6}\psi\left(a+\frac{j(b-a)}{n}\right)\right).$$

These are the first few rules in an infinite sequence of rules called the **Newton-Cotes integration formulas.** The simple trapezoidal rule integrates linear functions on $[a,b]$ exactly, and Simpson's rule integrates cubics exactly. The next few rules, listed for example in Davis and Rabinowitz (1975, p. 63-64), integrate higher degree polynomials exactly. For example, **Boole's rule** is

$$b_n(x) = \frac{b-a}{n}\sum_{j=1}^{n}\left(\frac{7}{90}\psi\left(a+\frac{(j-1)(b-a)}{n}\right)+\frac{32}{90}\psi\left(a+\frac{(j-\frac{3}{4})(b-a)}{n}\right)\right.$$

$$+\frac{12}{90}\psi\left(a+\frac{(j-\frac{1}{2})(b-a)}{n}\right)+\frac{32}{90}\psi\left(a+\frac{(j-\frac{1}{4})(b-a)}{n}\right)$$

$$\left.+\frac{7}{90}\psi\left(a+\frac{j(b-a)}{n}\right)\right).$$

The error committed by these rules is very important to us. In general ψ is a complex-valued function; and so are the estimates r_n, t_n, etcetera. A little care should be taken when we use only the real parts of these estimates. The main tools are collected in Theorem 3.2:

Theorem 3.2.

Let $[-a, a]$ be a finite interval on the real line, let n be an arbitrary integer, and let the density $f(x)$ be approximated by $f_n(x)$ where $f_n(x)$ is $\mathrm{Re}(r_n(x))$, $\mathrm{Re}(t_n(x))$, $\mathrm{Re}(s_n(x))$, or $\mathrm{Re}(b_n(x))$. Let X be a random variable with density f and j-th absolute moment μ_j. Define the absolute difference $E_n = | f(x) - f_n(x) |$, and the tail integral

$$T_n = \frac{1}{2\pi}\left(\int_{-\infty}^{-a} |\phi| + \int_{a}^{\infty} |\phi|\right).$$

Then:

A. If r_n is used and $\mu_1 < \infty$, then

$$E_n \le T_n + \frac{(2a)^2}{4\pi n}\left(|x| + \mu_1 \right).$$

B. If t_n is used and $\mu_2 < \infty$, then

$$E_n \le T_n + \frac{(2a)^3}{24\pi n^2}\left(|x| + \mu_2^{\frac{1}{2}} \right)^2.$$

C. If s_n is used and $\mu_4 < \infty$, then

$$E_n \le T_n + \frac{(2a)^5}{360\pi n^4}\left(|x| + \mu_4^{\frac{1}{4}} \right)^4.$$

D. If b_n is used and $\mu_6 < \infty$, then

$$E_n \le T_n + \frac{(2a)^7}{3870720\pi n^6}\left(|x| + \mu_6^{\frac{1}{6}} \right)^6.$$

Before proving Theorem 3.2, it is helpful to point out the following inequalities:

Lemma 3.1.

Let ϕ be a characteristic function, and let ψ be defined by

$$\psi(t) = \phi(t)e^{-itx} .$$

Assume that the absolute moments for the distribution corresponding to ϕ are denoted by μ_j. Then, if the j-th absolute moment is finite,

$$\sup_t |\psi^{(j)}(t)| \le \left(|x| + \mu_j^{\frac{1}{j}} \right)^j ,$$

where $j=0,1,2,\ldots$.

Proof of Lemma 3.1.

Note that $\psi^{(j)} = g_j e^{-itx}$ for some function g_j. It can be verified by induction that

$$g_j = \sum_{k=0}^{j} \binom{j}{k} (-ix)^k \phi^{(j-k)} .$$

When $\mu_j < \infty$, $\phi^{(j)}$ is a bounded continuous function given by $\int (ix)^j e^{itx} f(x)\, dx$. In particular, $|\phi^{(j)}| \le \mu_j$. If we also use the inequalities

$$\mu_k \le \mu_j^{\frac{k}{j}} \quad (k \le j),$$

then we obtain

$$|\psi^{(j)}| \le |g_j| \le \sum_{k=0}^{j} \binom{j}{k} |x|^k \mu_{j-k}$$

$$\le \sum_{k=0}^{j} \binom{j}{k} |x|^k \mu_j^{\frac{j-k}{j}}$$

$$= \left(|x| + \mu_j^{\frac{1}{j}} \right)^j . \blacksquare$$

Proof of Theorem 3.2.

Let us define $\psi(t) = \frac{1}{2\pi}\phi(t)e^{-itx}$. Then by Lemma 3.1,

$$2\pi |\psi^{(j)}| \le \left(|x| + \mu_j^{\frac{1}{j}} \right)^j ,$$

where μ_j is the finite j-th absolute moment of the distribution. Next, we need some estimates from numerical analysis. In particular,

$$| f(x)-f_n(x) | \leq T_n + | \int_{-a}^{a} \mathrm{Re}(\psi(t))\, dt - f_n(x) | .$$

To the last term, which is an error term in the estimation of the integral of $\mathrm{Re}(\psi)$ over a finite interval, we can apply estimates such as those given in Davis and Rabinowitz (1975, pp. 40-64). To apply these estimates, we recall that, when $\mu_j < \infty$, ψ is a bounded continuous function on the real line. If r_n is used and $\mu_1 < \infty$, then the last term does not exceed

$$\frac{(2a)^2}{2n} \sup | \mathrm{Re}(\psi)' | \leq \frac{(2a)^2}{2n} \sup | \psi^{(1)} |$$
$$\leq \frac{(2a)^2}{4\pi n} (| x | + \mu_1) .$$

If t_n is used and $\mu_2 < \infty$, then the last term does not exceed

$$\frac{(2a)^3}{12n^2} \sup | \psi^{(2)} | \leq \frac{(2a)^3}{24\pi n^2} \left(| x | + \mu_2^{\frac{1}{2}} \right)^2 .$$

If s_n is used and $\mu_4 < \infty$, then the last term does not exceed

$$\frac{(2a)^5}{180n^4} \sup | \psi^{(4)} | \leq \frac{(2a)^5}{360\pi n^4} \left(| x | + \mu_4^{\frac{1}{4}} \right)^4 .$$

If b_n is used and $\mu_6 < \infty$, then the last term does not exceed

$$\frac{(2a)^7}{1935360n^6} \sup | \psi^{(6)} | \leq \frac{(2a)^7}{3870720\pi n^6} \left(| x | + \mu_6^{\frac{1}{6}} \right)^6 . \blacksquare$$

The bounds of Theorem 3.2 allow us to apply the series method. There are two key problems left to solve:

A. The choice of a as a function of n.

B. The selection of a dominating curve g for rejection.

It is wasteful to compute $t_n, t_{n+1}, t_{n+2}, \ldots$ when trying to make an acceptance or rejection decision. Because the error decreases at a polynomial rate with n, it seems better to evaluate t_{c^k} for some $c > 1$ and $k = 1,2,\ldots$. Additionally, it is advantageous to use the standard dyadic "trick" of computing only t_2, t_4, t_8, etcetera. When computing t_{2n}, the computations made for t_n can be reused provided that we align the cutpoints. In other words, if a_n is the constant a with the dependence upon n made explicit, it is necessary to demand that

$$\frac{a_{2^k}}{2^k}$$

be equal to

$$\frac{a_{2^{k+1}}}{2^{k+1}}$$

or to

$$\frac{a_{2^{k+1}}}{2^k} .$$

Thus, $a_{2^{k+1}}$ is equal to a_{2^k} or to twice that value. Note that for the estimates f_n in Theorem 3.2 to tend to $f(x)$, it is necessary that $a_n \to \infty$ (unless the characteristic function has compact support), and that $a_n = o(n^{\frac{j}{j+1}})$ where j is 1,2,4 or 6 depending upon the estimator used. Thus, it does not hurt to choose a_n monotone and of the form

$$a_{2^k} = a_0 2^{c_k}$$

where c_k is a positive integer sequence satisfying $c_{k+1} - c_k \in \{0,1\}$, and a_0 is a constant.

The problem of the selection of a dominating curve has a simple solution in many cases. To be able to use Theorem 3.2, we need upper bounds for μ_j and $\int_a^\infty |\phi|$. Luckily, this is also sufficient for the design of good upper bounds. To make this point, we consider several examples, after an auxiliary lemma.

Lemma 3.2.

Let ϕ be a characteristic function with continuous absolutely integrable n-th derivative $\phi^{(n)}$ where n is a nonnegative integer. Then ϕ has a density f where

$$f(x) \le \frac{\int |\phi^{(n)}|}{2\pi |x|^n} .$$

If $\int |t| |\phi(t)| dt < \infty$, then ϕ has a Lipschitz density f with Lipschitz constant not exceeding

$$\frac{\int |t| |\phi(t)| dt}{2\pi} .$$

Proof of Lemma 3.2.

When ϕ has a continuous absolutely integrable n-th derivative $\phi^{(n)}$, then a density f exists, and the following inversion formula is valid:

$$(ix)^n f(x) = \frac{1}{2\pi} \int \phi^{(n)}(t) e^{-tx} \, dt .$$

The first inequality follows directly from this. Next, assume that $\int |t| \, |\phi(t)| \, dt < \infty$. Once again, a density f exists, and because f can be computed by the standard inversion formula, we have

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

$$\leq \frac{1}{2\pi} \int |e^{-it(y-x)} - 1| \, |\phi(t)| \, dt$$

$$\leq \frac{1}{2\pi} |y-x| \int |t| \, |\phi(t)| \, dt . \blacksquare$$

Example 3.1. Characteristic functions with compact support.

Assume that ϕ is known to vanish outside $[-A, A]$ for some finite value A. It should be stressed that this is a very strong condition of smoothness for the density f of this distribution. From Lemma 3.2, we know that f is a bounded density:

$$f(x) \leq \frac{A}{\pi} .$$

Furthermore, f is Lipschitz with Lipschitz constant C not exceeding $A^2/(2\pi)$. The densities in this class can have arbitrarily large tails, and can not be uniformly bounded without imposing some sort of tail condition. For a detailed discussion of this, we refer to section VII.3.3, and in particular to Example VII.3.4, where a dominating curve for a Lipschitz (C) density on the positive real line with absolute moment μ_j $(j>2)$ is given. The area under that dominating curve is

$$2\sqrt{8C} \frac{j}{j-2} \mu_j^{\frac{1}{j}} .$$

Here the factor 2 allows for the extension of the bound to the entire real line. Note that with $C = A^2/(2\pi)$, the rejection constant becomes

$$\frac{4A}{\sqrt{\pi}} \frac{j}{j-2} \mu_j^{\frac{1}{j}} ,$$

which is scale invariant.

We suggest that a be picked constant and equal to A, since $T_n=0$ in Theorem 3.2 when $a \geq A$. ■

Example 3.2. Unimodal densities.

For unimodal densities with mode at 0, a variety of good dominating curves were given in section VII.3.2. These required a bound on the value of $f(0)$ and one additional piece of information, such as an upper bound for μ_j. For the bound at the mode, we can use

$$f(x) \leq \frac{\int |\phi|}{2\pi} .$$

It is difficult to verify the unimodality of a density from a characteristic function, so this example is not as strong as Example 3.1. Also, the choice of a causes a few extra problems. See Example 3.3 below. ■

Example 3.3. Optimization of parameter a.

Using a Chebyshev type inequality applied to characteristic functions,

$$\int_a^{\infty} |\phi| \leq \frac{\int_0^{\infty} |t|^r \, |\phi(t)| \, dt}{a^r} ,$$

we can obtain upper bounds of the form $ca^k + da^{-r}$ for the error E_n in Theorem 3.2, where c, d, k, r are positive constants, and c depends upon n. Considered as a function of a, this has one minimum at

$$a = \left(\frac{dr}{ck} \right)^{\frac{1}{k+r}} .$$

The minimal value is

$$c^{\frac{r}{k+r}} d^{\frac{k}{k+r}} \left(\left(\frac{r}{k}\right)^{\frac{k}{k+r}} + \left(\frac{k}{r}\right)^{\frac{r}{k+r}} \right) .$$

What matters here is that the only factor depending upon n is the first one, and that it tends to 0 at the rate $c^{r/(k+r)}$. Since c varies typically as $n^{-(k-1)}$ for the estimators given in Theorem 3.2, we obtain the rate

$$n^{-\frac{r(k-1)}{k+r}} .$$

This rate is necessarily sublinear when $r=1$, regardless of how large k is. Note that it decreases quickly when $r \geq 2$ for all usual values of k. For example, with $r=2$ and Simpson's rule ($k=5$), our rate is $n^{-8/7}$. With $r=3$ and the trapezoidal rule ($k=3$), our rate is $n^{-3/2}$. ■

Example 3.4. Sums of iid uniform random variables.

The uniform density on $[-1,1]$ has characteristic function $\phi(t) = \sin(t)/t$. The sum of m iid uniform $[-1,1]$ random variables has characteristic function

$$\phi_m(t) = \left(\frac{\sin(t)}{t} \right)^m .$$

The corresponding density is unimodal, which should be of help in the derivation of bounds for the density. By taking consecutive derivatives of ϕ_m, it is easily established that the second moment μ_2 is $\frac{m}{3}$, and that the fourth moment μ_4 is $\frac{m^2}{3}-\frac{2m}{15}$. Furthermore, the mode, which occurs at zero, has value

$$\begin{aligned}
&\frac{1}{2\pi} \int \phi_m(t)\, dt \\
&\leq \frac{1}{2\pi} \int \min((1-\frac{t^2}{6}+\frac{t^4}{120})^m , | t |^{-m})\, dt \\
&\leq \frac{1}{2\pi} \int \min(e^{-\frac{m}{6}t^2(1-\frac{t^2}{20})} , | t |^{-m})\, dt \\
&\leq \frac{2}{2\pi(m-1)} + \int \frac{1}{2\pi} e^{-\frac{m}{6}\frac{19}{20}t^2}\, dt \\
&= \frac{1}{\pi(m-1)} + \sqrt{\frac{60}{19m}} = M ,
\end{aligned}$$

where we split the integral over the intervals $[-1,1]$ and its complement. We now refer to Theorem VII.3.2 for symmetric unimodal densities bounded by M and having r-th absolute moment μ_r. Such densities are bounded by $\min(M,(r+1)\mu_r / | x |^{r+1})$, and the dominating curve has integral

$$\frac{r+1}{r}((r+1)\mu_r)^{\frac{1}{r+1}} M^{\frac{r}{r+1}} .$$

For example, for $r=4$, we obtain in our example

$$\frac{5}{4}(5\mu_4)^{\frac{1}{5}} M^{\frac{4}{5}} \sim \frac{5}{4}(\frac{5}{3})^{\frac{1}{5}}(\frac{60}{19})^{\frac{2}{5}} .$$

as $m \rightarrow \infty$. In other words, as $m \rightarrow \infty$, the rejection constant tends to a fixed value. One can verify that this same property holds true for all values of $r>0$. This example is continued in Example 3.6. ■

This leaves us with the black box algorithm and its analysis. We assume that a dominating curve cg is known, where g is a density, that another function h is known having the property that

$$\frac{1}{2\pi}\left(\int_{-\infty}^{-a} |\phi| + \int_{a}^{\infty} |\phi|\right) \leq h(a) \quad (a>0) ,$$

and that integrals will be evaluated only for the subsequence $a_0 2^k$ $,k \geq 0$, where a_0 is a given integer. Let f_n denote a numerical integral estimating ψ such as r_n, s_n, t_n or b_n. This estimate uses as interval of integration $[-l(n,x),l(n,x)]$ for some function l which normally diverges as n tends to ∞.

Series method based upon numerical integration

REPEAT

- Generate a random variate X with density g.
- Generate a uniform [0,1] random variate U.
- Compute $T \leftarrow Ucg(X)$ (recall that $f \leq cg$).
- $n \leftarrow a_0 2^k, a \leftarrow l(n,X)$ (prepare for integration)
- REPEAT
  - $W \leftarrow f_n(X)$ (f_n is an integral estimate of $f=\int\psi$ with parameter n on interval $[-a,a]$; the number of evaluations of ϕ required is proportional to n)
  - Compute an upper bound on the error, E. (Use the bounds of Theorem 3.2 plus $h(a)$.)
  - $n \leftarrow 2n$
- UNTIL $|T-W|>E$

UNTIL $T<W$

RETURN X

The first issue is that of correctness of the algorithm. This boils down to verifying whether the algorithm halts with probability one. We have:

Theorem 3.3.

The algorithm based upon the series method given above is correct, i.e. halts with probability one, when

$$\lim_{n \to \infty} l(n,x) = \infty \quad (\text{all } x),$$

$$\lim_{a \to \infty} h(a) = \infty$$

(this forces ϕ to be absolutely integrable), and one of the following conditions holds:

A. r_n is used, $\mu_1 < \infty$, and $l(n,x) = o(n^{1/2})$ for all x.

B. t_n is used, $\mu_2 < \infty$, and $l(n,x) = o(n^{2/3})$ for all x.

C. s_n is used, $\mu_4 < \infty$, and $l(n,x) = o(n^{4/5})$ for all x.

D. b_n is used, $\mu_6 < \infty$, and $l(n,x) = o(n^{6/7})$ for all x.

Here μ_j is the j-th absolute moment for f.

Proof of Theorem 3.3.

We need only verify that the error bound used in the algorithm tends to 0 as $n \to \infty$ for all x. Theorem 3.3 is a direct corollary of Theorem 3.2. ∎

Theorem 3.3 is reassuring. Under very mild conditions on the density, a valid algorithm indeed exists. We have to know μ_j for some j and we need also an explicit expression for the tail bound $h(a)$. The theorem just states that given this information, we can choose a function $l(n,x)$ and an estimator f_n which guarantee the validity. Unfortunately, there is a snake in the grass. The function $l(n,x)$ has a profound impact on the time before halting. In many examples, the expected time is ∞. Thus, let us consider the expected number of evaluations of ψ (or ϕ) before halting. This can't possibly be given without discussing how large $h(.)$ is, and which function $l(.,.)$ is picked. Perhaps the best thing to do at this stage is to offer a helpful lemma, and then to illustrate it on a few examples.

Lemma 3.3.

Consider the series method given above, and assume that for the given functions h and l, we have an inequality of the type

$$| f(x)-f_n(x) | \leq C(x)n^{-\alpha} \quad (n \geq 1 \text{ , all } x) ,$$

where C is a positive function and $\alpha>1$ is a constant. If $a_0=1$ and f_n requires $\beta n+1$ evaluations of ψ for some constant β (for t_n, $\beta=1$, and for s_n, $\beta=2$), then the expected number of evaluations of ψ before halting does not exceed

$$\leq c(\beta+1) + 2^\gamma c^{1-\gamma} \int C^\gamma g^{1-\gamma} \frac{2\beta+1}{1-2^{1-\gamma\alpha}}$$

$$\leq c(\beta+1) + 2^\gamma c^{1-\gamma} \frac{2\beta+1}{1-2^{1-\gamma\alpha}} \left(\int Cg\right)^{1-\gamma} \left(\int C^{2-\frac{1}{\gamma}}\right)^{\gamma} ,$$

where γ is a number satisfying

$$\alpha\gamma > 1 , \gamma \leq 1 .$$

Proof of Lemma 3.3.

By Wald's equation, our expected number is equal to c times the expected number of evaluations in the first iteration (regardless of acceptance or rejection). Let us first condition on $X=x$ with density g. For f_1, we use up $\beta+1$ evaluations in all cases. The probability of having to evaluate f_2 does not exceed $2C(x)1^{-\alpha}/cg(x)$. Continuing in this fashion, it is easily seen that the expected number of evaluations of ψ is not greater than

$$\sum_{k=0}^{\infty} \left((\beta 2^{k+1}+1) \min\left(\frac{2C(x)(2^k)^{-\alpha}}{cg(x)}, 1\right) \right) + \beta+1 .$$

Taking expectations with respect to $g(x)\,dx$ and multiplying with c gives the unconditional upper bound

$$c(\beta+1) + \sum_{k=0}^{\infty} \left((\beta 2^{k+1}+1) \int \min(2C(x)(2^k)^{-\alpha}, cg(x))\, dx \right)$$

$$\leq c(\beta+1) + \sum_{k=0}^{\infty} \left((\beta 2^{k+1}+1) \int \min(2C(x)(2^k)^{-\alpha}, cg(x))\, dx \right)$$

$$\leq c(\beta+1) + \int (2C(x))^\gamma (cg(x))^{1-\gamma}\, dx \sum_{k=0}^{\infty} 2^{-k\gamma\alpha}(\beta 2^{k+1}+1)$$

$$= c(\beta+1) + 2^\gamma c^{1-\gamma} \int C^\gamma g^{1-\gamma} \left(\frac{2\beta}{1-2^{1-\gamma\alpha}} + \frac{1}{1-2^{-\gamma\alpha}}\right)$$

$$\leq c(\beta+1) + 2^\gamma c^{1-\gamma} \int C^\gamma g^{1-\gamma} \frac{2\beta+1}{1-2^{1-\gamma\alpha}} ,$$

where γ is a number satisfying

$$\alpha\gamma > 1 \ , \ \gamma \leq 1 \ .$$

By Holder's inequality, the integral in the last expression does not exceed

$$\left(\int Cg\right)^{1-\gamma} \left(\int C^{2-\frac{1}{\gamma}}\right)^{\gamma} . \blacksquare$$

Lemma 3.3 reveals the extent to which the efficiency of the algorithm is affected by c ,$C(x)$,$g(x)$ and μ_j.

Example 3.5. Characteristic functions with compact support.

Assume that the characteristic function vanishes outside $[-A,A]$. If we take $l(n,x)=A$, then $h \equiv 0$ in the algorithm. Note that this choice violates the consistency conditions of Theorem 3.3, but leads nevertheless to a consistent procedure. With t_n , we have $\beta=1,\alpha=2$ and an error

$$E_n \leq C(x)n^{-\alpha}$$

where

$$C(x) = \frac{(2A)^3}{24\pi}(\,|\,x\,|+\sqrt{\mu_2})^2 \ .$$

With s_n , we have $\beta=2$, $\alpha=4$ and

$$C(x) = \frac{(2A)^5}{360\pi}(\,|\,x\,|+\mu_4^{\frac{1}{4}})^4 \ .$$

'With both error bounds, $\int C=\infty$, so we can't take $\gamma=1$ in Lemma 3.3. Also,

$$\int C^{2-\frac{1}{\gamma}} < \infty$$

when $\frac{1}{\gamma}>2+\frac{1}{\alpha}$. Thus, for the bound of Lemma 3.3 to be useful, we need to choose

$$\frac{1}{\alpha} < \gamma < \frac{\alpha}{2\alpha+1} \ .$$

This yields the intervals $(\frac{1}{2},\frac{2}{5})$ and $(\frac{1}{4},\frac{4}{9})$ respectively. Of course, the former interval is empty. This is due to the fact that the last inequality in Lemma 3.3 (combined with Theorem 3.2) never leads to a finite upper bound for the trapezoidal rule. Let us further concentrate therefore on s_n . Note that

$$\int Cg \leq \frac{(2A)^5}{360\pi} \int g(x)(|x| + \mu_4^{\frac{1}{4}})^4 \, dx$$
$$\leq \frac{(2A)^5}{360\pi} 8 \int g(x)(|x|^4 + \mu_4) \, dx$$
$$= \frac{32A^5(\mu_4^* + \mu_4)}{45\pi},$$

where μ_4^* is the fourth absolute moment for g. Typically, when g is close to f, the fourth moment is close to that of f. We won't proceed here with the explicit computation of the full bound of Lemma 3.3. It suffices to note that the bound is large when either A or μ_4 is large. In other words, it is large when the support of ϕ is large (the density is less smooth) and/or the tail of the density is large. Let us conclude this section by repeating the algorithm:

Series method based upon numerical integration

[NOTE: The characteristic function ϕ vanishes off $[-A,A]$, and the fourth absolute moment does not exceed μ_4.]

REPEAT

  Generate a random variate X with density g.

  Generate a uniform [0,1] random variate U.

  Compute $T \leftarrow Ucg(X)$ (recall that $f \leq cg$).

  $n \leftarrow a_0$ (prepare for integration)

  REPEAT

    $W \leftarrow \mathrm{Re}(s_n(X))$ (s_n is Simpson's integral estimate of $f = \int \psi$ with parameter n on interval $[-A,A]$; the number of evaluations of ϕ required is $2n+1$)

    $E \leftarrow \frac{(2A)^5}{360\pi}(|X| + \mu_4^{\frac{1}{4}})^4 n^{-4}$

    $n \leftarrow 2n$

  UNTIL $|T-W| > E$

UNTIL $T < W$

RETURN X

For dominating curves cg, there are numerous possibilities. See for example Lemma 3.2. In Example 3.1, a dominating curve based upon an inequality for Lipschitz densities (section VII.3.4) was developed. The rejection constant c for that example is

$$\frac{8}{\sqrt{\pi}} A \mu_4^{\frac{1}{4}} . \blacksquare$$

Example 3.6. Sums of iid uniform random variables.

This is a continuation of Example 3.4, where a good dominating density was found for use in the rejection algorithm. What is left here is mainly the choice of h and l for use in the algorithm. Let us start with the decision to estimate f by Simpson's rule s_n. This is based upon a quick preliminary analysis which shows that the trapezoidal rule for example just isn't good enough to obtain finite expected time.

The function $h(a)$ can be chosen as

$$h(a) = \frac{1}{\pi a^{m-1}(m-1)}$$

where m is the number of uniform [-1,1] random variables that are summed. To see this, note that

$$2\int_a^\infty \frac{1}{2\pi} \left| \frac{\sin(t)}{t} \right|^m dt \leq \frac{1}{\pi}\int_a^\infty |t|^{-m} dt = h(a) .$$

Given $X=x$ in the algorithm, we see that with s_n, the error E_n is not greater than

$$E_n \leq h(a)+\frac{(2a)^5(|x|+\mu_4^{1/4})^4}{360 \pi n^4} ,$$

where a determines the integration interval (Theorem 3.2). Optimization of the upper bound with respect to a is simple and leads to the value

$$a = \left(\frac{9n^4}{4(|x|+\mu_4^{1/4})} \right)^{\frac{1}{m+4}} .$$

With this value for a (or $l(n,x)$), we obtain

$$E_n \leq C(x)n^{-\alpha}$$

for $\alpha=4(m-1)/(m+4)$ and

$$C(x) = \frac{m}{m-1}\frac{1}{\pi}\left(\frac{4}{9}\right)^{\frac{m-1}{m+4}} (|x|+\mu_4^{1/4})^\alpha .$$

This is all the users need to implement the algorithm. We can now apply Lemma 3.3 to obtain an idea of the expected complexity of the algorithm. We will show that the expected time is better than $O(m^{(5+\epsilon)/8})$ for all $\epsilon>0$. A brief outline of the proof should suffice at this point. In Lemma 3.3, we need to pick a constant γ. The conditions $\alpha\gamma>1$ and $\int C^{2-\frac{1}{\gamma}}<\infty$ force us to impose the conditions

$$\frac{m+4}{4m-4} < \gamma < \frac{4m-4}{9m-4} .$$