Chapter Nine
CONTINUOUS UNIVARIATE DENSITIES

Chapters IX and X are included for the convenience of a large subpopulation of users, the statisticians. The main principles in random variate generation were developed in the first eight chapters. Most particular distributions found here are members of special classes of densities for which universal methods are available. For example, a short algorithm for log-concave densities was developed in section VII.2. When speed is at a premium, then one of the table methods of the previous chapter could be used. This chapter is purely complementary. We are not in the least interested in a historical review of the different methods proposed over the years for the popular densities. Some interesting developments which give us new insight or illustrate certain general principles will be reported. The list of distributions corresponds roughly speaking to the list of distributions in the three volumes of Johnson and Kotz.

1. THE NORMAL DENSITY.

1.1. Definition.

A random variable X is **normally distributed** if it has density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

When X is normally distributed, then $\mu+\sigma X$ is said to be normal (μ,σ^2). The mean μ and the variance σ^2 are uninteresting from a random variate generation point of view.

Comparative studies of normal generators were published by Muller (1959), Ahrens and Dieter (1972), Atkinson and Pearce (1976), Kinderman and Ramage (1976), Payne (1979) and Best (1979). In the table below, we give a general outline of how the available algorithms are related.

Method	References	Speed	Size of code	Section
Inversion	Muller (1959)	Slow	Moderate	II.2.3
Polar method	Box and Muller (1958)	Moderate	Small	V.4.4
	Bell (1968)			
Rejection	Von Neumann (1951)	Moderate	Small	II.3.2
	Sibuya (1962)			
Ratio-of-uniforms	Kinderman and Monahan (1977)	Fast	Small to moderate	IV.7.2
Composition/rejection	Marsaglia and Bray (1964)	Fast	Small to moderate	
	Ahrens and Dieter (1972)			
	Kinderman and Ramage (1976)			
	Sakasegawa (1978)			
Series method		Fast	Small to moderate	IV.5.3
Almost-exact inversion	Wallace (1976)	Moderate	Small	IV.3.3
Table methods	Marsaglia, Maclaren and Bray (1964)	Very fast	Large	
Forsythe's method	Forsythe (1972)	Fast	Moderate	IV.2.1
	Ahrens and Dieter (1973)			
	Brent (1974)			

The list given here is not exhaustive. Many references are missing. What matters are the general trends. We know that table methods are fast, and the rectangle-wedge-tail method of Marsaglia, Maclaren and Bray (1964) is no exception. At the other end of the scale are the small programs of moderate speed, such as the programs for the polar method and some rejection methods. In between are moderate-sized programs that have good speed, such as the ratio-of-uniforms method, the series method, Forsythe's method and the composition/rejection method. Only the inversion method is inadmissible because it is slower and less space efficient than all of the other methods, the table methods excepted. Below, we will mainly focus on the composition/rejection methods which have not been described in earlier chapters. Because we will cut off the tail of the normal density, it seems important to show how random variates with a density proportional to the tail can be generated.

1.2. The tail of the normal density.

In this section, we consider generators for the family of tail densities

$$f(x) = \frac{e^{-\frac{x^2}{2}}}{\Phi(a)} \quad (x > a),$$

where $\Phi(a) = \int_a^\infty e^{-\frac{x^2}{2}}$ is a normalization constant and $a > 0$ is a parameter. Two algorithms will be described:

Marsaglia's method for the tail-of-the-normal density (Marsaglia, 1964)

REPEAT
Generate iid uniform [0,1] random variates U, V.
$X \leftarrow \sqrt{a^2 - 2\log(U)}$
UNTIL $VX \leq a$
RETURN X

Marsaglia's method is based upon the trivial inequality

$$e^{-\frac{x^2}{2}} \leq \frac{x}{a} e^{-\frac{x^2}{2}} \quad (x \geq a) .$$

But $xe^{\frac{a^2-x^2}{2}}$ $(x \geq a)$ is a density having distribution function

$$F(x) = 1 - e^{\frac{a^2-x^2}{2}} \quad (x \geq a) ,$$

which is the tail part of the Rayleigh distribution function. Thus, by inversion, $\sqrt{a^2-2\log(U)}$ has distribution function F, which explains the algorithm. The probability of acceptance in the rejection algorithm is

$$P(VX \leq a) = E\left(\frac{a}{X}\right) = \int_a^\infty a e^{\frac{a^2-x^2}{2}} \, dx = a e^{\frac{a^2}{2}} \Phi(a) \rightarrow 1$$

as $a \rightarrow \infty$. Thus, the rejection algorithm is asymptotically optimal. Even for small values of a, the probability of acceptance is quite high: it is about 66% for $a = 1$ and about 88% for $a = 3$. Note that Marsaglia's method can be sped up somewhat by postponing the square root until after the acceptance:

REPEAT
Generate iid uniform [0,1] random variates U, V.
$X \leftarrow c - \log(U)$ (where $c = a^2/2$)
UNTIL $V^2 X \leq c$
RETURN $\sqrt{2X}$

An algorithm which does not require any square roots can be obtained by rejection from an exponential density. We begin with the inequality

$$e^{-\frac{x^2}{2}} \leq e^{\frac{a^2}{2}-ax} \qquad (x \geq a) ,$$

which follows from the observation that $(x-a)^2 \geq 0$. The upper bound is proportional to the density of $a+\frac{E}{a}$ where E is exponentially distributed. This yields without further work the following algorithm:

REPEAT

Generate iid exponential random variates $E, E*$.

UNTIL $E^2 \leq 2a^2 E*$

RETURN $X \leftarrow a+\frac{E}{a}$

The probability of acceptance is precisely as for Marsaglia's method:

$$P(E* \geq E^2/(2a^2)) = \int_0^\infty e^{-\frac{x^2}{2a^2}} \, dx = ae^{\frac{a^2}{2}} \Phi(a) \rightarrow 1 \qquad (a \rightarrow \infty) .$$

If a fast exponential random variate generator is available, the second rejection algorithm is probably faster than Marsaglia's.

1.3. Composition/rejection methods.

The principle underlying all good composition/rejection methods is the following: decompose the density of f into two parts, $f(x)=pg(x)+(1-p)h(x)$ where $p \in (0,1)$ is a mixture parameter, g is an easy density, and h is a residual density not very often needed when p is close to 1. We rarely stumble upon a good choice for g by accident. But we can always find the optimal g_θ in a family of suitable candidates parametrized by θ. The weight of g_θ in the mixture is denoted by $p(\theta)$:

$$p(\theta) = \inf_x \frac{f(x)}{g_\theta(x)} .$$

The candidates g_θ should preferably be densities of simple transformations of independent uniform [0,1] random variables. Among the simple transformations one might consider, we cite:

(1) $\theta(V_1 + \cdots + V_n)$;
(2) θ median$(V_1, \ldots, V_n)$;
(3) $\theta_1 V_1 + \theta_2 V_2$;
(4) $\theta_1 V_1 + \theta_2 (V_1)^3$.

Here $V_1, V_2, \ldots$ are iid uniform $[-1,1]$ random variates, and $\theta, \theta_1, \theta_2$ are parameters to be selected. Marsaglia and Bray (1964) used the first choice with $n=3$ and with the deliberately suboptimal value $\theta=1$ (because a time-consuming multiplication is avoided for this value). Kinderman and Ramage (1976) optimized θ for choice (1) when $n=2$. And Ahrens and Dieter (1972) proposed to use choice (3). Because the shape of g_θ is trapezoidal, this method is known as the **trapezoidal method.** All three approaches lead to algorithms of about equal length and speed. We will look at choices (1) and (2) in more detail below, and provide enough detail for the reader to be able to reconstruct the algorithms of Marsaglia and Bray (1964) and Kinderman and Ramage (1976).

Theorem 1.1.
The density of θ median$(V_1, \ldots, V_{2n+1})$ for n positive and $\theta > 0$ is

$$c\left(1-\frac{x^2}{\theta^2}\right)^n \qquad (|x| \leq \theta)$$

where $c = \frac{(2n+1)!}{2^{2n+1} n!^2 \theta}$. The maximal value of $p(\theta)$ is reached for $\theta = \sqrt{2n+1}$, and takes the value

$$p = \frac{2^{2n+1} n!^2 \sqrt{n}}{\sqrt{\pi e}\,(2n+1)!}\left(1+\frac{1}{2n}\right)^{n+\frac{1}{2}} .$$

We have

(i) $p = \sqrt{\frac{6}{\pi e}} \approx 0.8382112 \quad (n=1)$;

(ii) $p = \sqrt{\frac{125}{18\pi e}} \approx 0.9017717 \quad (n=2)$;

(iii) $\lim_{n \to \infty} p = 1$.

Proof of Theorem 1.1.

The density can be derived very easily after recalling that the median of $2n+1$ iid uniform $[0,1]$ random variables has a symmetric beta density given by

$$\frac{(2n+1)!}{n!^2}(x(1-x))^n \qquad (0 \leq x \leq 1) .$$

Define $g_\theta(x) = c(1-(x^2/\theta^2))^n \quad (|x| \leq \theta)$, and note that $\log(f/g_\theta)$ attains an extremum at some point x for which the derivative of the logarithm is 0. This yields the equation

$$-x+\frac{2x}{\theta^2}\frac{n}{1-\frac{x^2}{\theta^2}}=0 ,$$

or,

$$x=0 \ ; \ x^2=\theta^2-2n .$$

When $\theta^2<2n$, f/g_θ attains only one minimum, at $x=0$. When $\theta^2>2n$, the function f/g_θ is symmetric around 0: it has a local peak at 0, dips to a minimum, and increases monotonically again to ∞ as $x\uparrow\theta$. Thus, we have

$$p(\theta)=\inf_x \frac{f(x)}{g_\theta(x)}=\begin{cases}\frac{1}{\sqrt{2\pi}c}=\frac{2^{2n+1}n!^2\theta}{(2n+1)!\sqrt{2\pi}} & (\theta^2<2n)\\ \frac{1}{\sqrt{2\pi}c}(\frac{e}{2n})^n\theta^{2n}e^{-\frac{\theta^2}{2}} & (\theta^2>2n) .\end{cases}$$

We still have to maximize this function with respect to θ. The function $p(\theta)$ increases linearly from 0 up to $\theta=\sqrt{2n}$. Then, it increases some more , peaks, and decreases in a bell-shaped fashion. The maximum is attained for some value $\theta>\sqrt{2n}$. Since in that region, $p(\theta)$ is a constant times $\theta^{2n+1}e^{-\theta^2/2}$, the maximum is attained for $\theta=\sqrt{2n+1}$. This gives the desired result. ■

Had we considered the Taylor series expansion of f about 0, given by

$$f(x)=\frac{1}{\sqrt{2\pi}}(1-\frac{x^2}{2}+\frac{x^4}{8}-\frac{x^6}{48}+\cdots) ,$$

which is known to give partial sums that alternately overestimate and underestimate f , then we would have been tempted to choose $g(x)=\frac{3}{4\sqrt{2}}(1-\frac{x^2}{2})$, because of

$$f(x)\geq\frac{1}{\sqrt{2\pi}}(1-\frac{x^2}{2})=pg(x) \quad (|x|\leq\sqrt{2})$$

where $p=\frac{4}{3\sqrt{\pi}}\approx0.7522528$ is the weight of g in the mixture. This illustrates the usefulness and the shortcomings of Taylor's series. Simple polynomial bounds are very easy to obtain, but the choice could be suboptimal. From Theorem 1.1 for example, we recall that the optimal g of the inverted parabolic form is a constant times $(1-\frac{x^2}{3})$ $(|x|\leq\sqrt{3})$. Sometimes a suboptimal choice of θ is preferable because the residual density h is easier to handle. This is the case for $n=1$ in Theorem 1.1. The suboptimal choice $\theta=\sqrt{2n}$, which is the choice implicit in

Taylor's series expansion, yields a much cleaner residual density. For $n=2$, we need 5 random variates instead of 3, an increase of 66%, while the gain in efficiency (in value of p) is only of the order of 10%. For this reason, the case $n>1$ is less important in practice. Let us briefly describe the entire algorithm for the case $n=1$, $\theta=\sqrt{2}$. We can decompose f as follows:

$$f(x) = pg(x)+qh(x)+rt(x)$$

where

$$\text{(i)} \quad g(x) = \frac{3}{4\sqrt{2}}(1-\frac{x^2}{2});$$

$$p = \frac{4}{3\sqrt{\pi}} \approx 0.7522528;$$

$$\text{(ii)} \quad t(x) = \frac{1}{r}\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \quad (|x|>\sqrt{2});$$

$$r = \int_{|x|>\sqrt{2}} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx \approx 0.15729921;$$

$$\text{(iii)} \quad h(x) = \frac{1}{q}\frac{1}{\sqrt{2\pi}}(e^{-x^2/2}-(1-\frac{x^2}{2})) \quad (|x|\le\sqrt{2});$$

$$q = \int_{|x|\le\sqrt{2}} \frac{1}{\sqrt{2\pi}}(e^{-x^2/2}-(1-\frac{x^2}{2})\,dx \approx 0.09044801 .$$

Sampling from the tail density t has been discussed in the previous sub-section. Sampling from g is simple: just generate three iid uniform $[-1,1]$ random variates, and take $\sqrt{2}$ times the median. Sampling from the residual density h can be done as follows:

REPEAT
  Generate V uniformly on $[-1,1]$, and U uniformly on $[0,6]$.
  $X \leftarrow \sqrt{2}V/|V|^{4/5}$
  Accept $\leftarrow [U>X^2]$
  IF NOT Accept THEN
    IF $U \ge X^2(1-\frac{X^2}{8})$ THEN
      Accept $\leftarrow [(1-\frac{U}{6})X^4 \le 8(e^{-\frac{X^2}{2}}-(1-\frac{X^2}{2}))]$
UNTIL Accept
RETURN X

This is a simple rejection algorithm with squeezing based upon the inequalities

$$e^{-\frac{x^2}{2}}-(1-\frac{x^2}{2}) \le \frac{x^4}{8} \quad (|x|\le\sqrt{2});$$

$$\frac{x^4}{8}-\frac{x^6}{48} \leq e^{-\frac{x^2}{2}}-(1-\frac{x^2}{2}) \leq \frac{x^4}{8}-\frac{x^6}{48}+\frac{x^8}{384} .$$

The reader can easily work out the details. The probability of immediate acceptance in the first iteration is

$$P(X^2<U) = \int_0^1 P(X^2<6x)\,dx = \int_0^1 P(|X|<\sqrt{6x})\,dx$$

$$= \int_0^{\frac{1}{3}} \frac{5}{4\sqrt{2}} \frac{(\sqrt{6x})^5}{5}\,dx + \int_{\frac{1}{3}}^1 dx$$

$$= \frac{2}{3}+\frac{2}{7}\frac{6^{\frac{5}{2}}}{4\sqrt{2}3^{\frac{7}{2}}} = \frac{6}{7} .$$

The same smooth performance for a residual density could not have been obtained had we not based our decomposition upon the Taylor series expansion.

Let us next look at the density g_θ of $\theta(V_1+V_2+V_3)$ where the V_i's are iid uniform $[-1,1]$ random variables. For the density of $\theta(V_1+V_2)$, the triangular density, we refer to the exercises where among other things it is shown that the optimal θ is 1.1080179... , and that the corresponding value $p(\theta)$ is 0.8840704... .

Theorem 1.2.

The optimal value for θ in the decomposition of the normal density into $p(\theta)g_\theta(x)$ plus a residual density (where g_θ is the density of $\theta(V_1+V_2+V_3)$ and the V_i's are iid uniform $[-1,1]$ random variables), is

$$\theta = 0.956668451229... .$$

The corresponding optimal value for $p(\theta)$ is 0.962365327... .

Proof of Theorem 1.2.

The density g_θ of $\theta(V_1+V_2+V_3)$ is

$$g_\theta(x) = \begin{cases} \frac{1}{8\theta}(3-(\frac{x}{\theta})^2) & (|x| \leq \theta) \\ \frac{1}{16\theta}(3-|\frac{x}{\theta}|)^2 & (\theta \leq |x| \leq 3\theta) \\ 0 & (|x| > 3\theta) . \end{cases}$$

The function $h_\theta = f / g_\theta$ can be written as

$$h_\theta(x) = \begin{cases} \frac{8\theta}{\sqrt{2\pi}} \frac{e^{-\frac{x^2}{2}}}{3-\frac{x^2}{\theta^2}} & (0 < x \le \theta) \\ \frac{16\theta}{\sqrt{2\pi}} \frac{e^{-\frac{x^2}{2}}}{(3-\frac{x}{\theta})^2} & (\theta \le x \le 3\theta)\,, \end{cases}$$

when $x > 0$. We need to find the value of θ for which $\min_{0<x\le 3\theta} h_\theta(x)$ is maximal. By setting the derivative of $\log(h_\theta)$ with respect to x equal to 0, and by analyzing the shape of h_θ, we see that the minimum of h_θ belongs to the following set of values: 0, θ, b, c, where

$$b = \theta\sqrt{3-\frac{2}{\theta^2}}\;;$$

$$c = \frac{3\theta}{2}+\frac{\theta}{2}\sqrt{9-\frac{8}{\theta^2}}\;.$$

The following table gives all the local minima together with the values for h_θ.

Local minimum	Value of h_θ at minimum	Local minimum exists when:
0	$\eta = \frac{8\theta}{3\sqrt{2\pi}}$	$\theta^2 \le \frac{2}{3}$
b	$\psi = \frac{4\theta^3}{\sqrt{2\pi}} e^{-\frac{b^2}{2}}$	$1 \ge \theta^2 \ge \frac{2}{3}$
θ	$\xi = \frac{4\theta}{\sqrt{2\pi}} e^{-\frac{a^2}{2}}$	$\theta = 1$
c	$\phi = \frac{16\theta}{\sqrt{2\pi}} \frac{e^{-\frac{c^2}{2}}}{(3-\frac{c}{\theta})^2}$	$\theta^2 \ge \frac{8}{9}$

The general shape of h_θ is as follows: when $\theta^2 \ge 1$, there is no local minimum on $(0,\theta)$, and h_θ decreases monotonically to reach a global minimum at $x = c$ equal to ϕ, after which it increases again. When $\theta^2 = 1$, the same shape is observed, but a zero derivative occurs at $x = \theta$, although this does not correspond to a local minimum. When $\frac{8}{9} < \theta^2 < 1$, there are two local minima, one on $(0,\theta)$ (at b, of value ψ), and one on $(\theta, 3\theta)$ (at c, of value ϕ). For $\frac{2}{3} < \theta^2 < \frac{8}{9}$, the local minimum at c ceases to exist. We have again a function with one minimum, this time at $b < \theta$, of value ψ. Finally, for $\theta^2 \le \frac{2}{3}$, the function increases monotonically, and its global minimum occurs at $x = 0$ and has value η.

Consider now the behavior of η and ψ as a function of θ. Clearly, η increases linearly with θ. Furthermore, ψ is gamma shaped with global peak at $\theta=1$, and $\eta=\psi$ for $\theta^2=\frac{2}{3}$. The value of ϕ on the other hand decreases monotonically on the set $\theta^2 \geq \frac{8}{9}$. We verify easily that ϕ and ψ cross each other on the segment $\frac{8}{9}<\theta^2<1$. It is at this point that $\min_{0<x<3\theta} h_\theta(x)$ is maximal. This cross-over point is precisely the value given in the statement of the theorem. ■

Theorem 1.2 can be used in the design of a fast composition/rejection algorithm. In particular, the tail beyond the optimal 3θ is very small, having probability 0.004104648... . The residual density on $[-3\theta,3\theta]$ has probability 0.033530022... , but has unfortunately enough five peaks, the largest of which occurs at the origin. It is clear once again that the maximization criterion does not take the complexity of the residual density into account. A suboptimal value for θ sometimes leads to better residual densities. For example, when $\theta=1$, we save one multiplication and end up with a more manageable residual density. This choice was first suggested by Marsaglia and Bray (1964). We conclude this section by giving their algorithm in its entirety.

From the proof of Theorem 1.2, we see that (in the notation of that proof),

$$p(\theta) = \phi = \frac{16}{\sqrt{2\pi e}} = 0.86385546\ldots .$$

The normal density f can be decomposed as follows:

$$f(x) = \sum_{i=1}^{4} p_i f_i(x) ,$$

where (p_1,p_2,p_3,p_4) is a probability vector, and the f_i's are densities defined as follows:

(i) $p_1=0.86385546\ldots$, f_1 is the density of $V_1+V_2+V_3$, where the V_i's are iid uniform $[-1,1]$ random variables.

(ii) $p_4=0.002699796063\ldots= \int_{|x|\geq 3} f$; f_4 is the tail-of-the-normal density restricted to $|x| \geq 3$.

(iii) $f_2(x) = \frac{1}{9}(6-4|x|)$ $(|x| \leq \frac{3}{2})$; $p_2=0.1108179673\ldots$.

(iv) $p_3=1-p_1-p_4-p_2=0.02262677245\ldots$; $f_3=\frac{1}{p_3}(f-p_1f_1-p_2f_2-p_4f_4)$.

In the design, Marsaglia and Bray decided upon the triangular form of f_2 first, because random variates with this density can be generated simply as $\frac{3}{4}(V_4+V_5)$ where the V_i's are again iid uniform $[-1,1]$ random variates. After having picked this simple f_2, it is necessary to choose the best (largest) weight p_2, given by

$$p_2 = \inf_x \frac{f(x) - p_1 f_1(x)}{f_2(x)} .$$

This infimum is found as follows. The derivative of the ratio is 0 at $|x|=2$ and at $|x|=0.87386312884...$. Only the latter $|x|$ corresponds to a minimum, and the corresponding value for p_2 is $p_2=0.1108179673...$. Having determined random variate generation methods for all parts except f_3, it remains to establish just this for f_3. First, note that f_3 has supremum 0.3181471173... . If we use rejection from a rectangular density with support on [-3,3], then the expected number of iterations is

$$\frac{6\times 0.3181471173...}{p_3} = 1.9088827038... .$$

Combining all of this into one algorithm, we have:

Normal generator of Marsaglia and Bray (1964)

[NOTE: This algorithm follows the implementation suggested by Kinderman and Ramage (1977).]

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

CASE

$0 \le U \le 0.8638$:

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

RETURN $X \leftarrow 2.3153508\ldots U - 1 + V + W$

$0.8638 < U \le 0.9745$:

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

RETURN $X \leftarrow \frac{3}{2}(V - 1 + 9.0334237\ldots(U - 0.8638))$

$0.9973002\ldots < U \le 1$:

REPEAT

Generate iid uniform [0,1] random variates V,W.

$X \leftarrow \frac{9}{2} - \log(W)$

UNTIL $XV^2 \le \frac{9}{2}$

RETURN $X \leftarrow \sqrt{2X}\ \mathrm{sign}(U - 0.9986501\ldots)$

$0.9745 < U \le 0.9973002\ldots$:

REPEAT

Generate a uniform [-3,3] random variate X and a uniform [0,1] random variate U.

$V \leftarrow |X|$

$W \leftarrow 6.6313339\ldots(3-V)^2$

Sum $\leftarrow 0$

IF $V < \frac{3}{2}$ THEN Sum $\leftarrow 6.0432809\ldots(\frac{3}{2} - V)$

IF $V < 1$ THEN Sum $\leftarrow$ Sum $+13.2626678\ldots(3-V^2) - W$

UNTIL $U \le 49.0024445\ldots e^{-\frac{V^2}{2}} - \mathrm{Sum} - W$

RETURN X

1.4. Exercises.

1. In the trapezoidal method of Ahrens and Dieter (1972), the largest symmetric trapezoid under the normal density is used as the main component in the mixture. Show that this trapezoid is defined by the vertices $(-\xi,0),(\xi,0),(\eta,\rho),(-\eta,\rho)$ where $\xi=2.1140280833...$, $\eta=0.2897295736...$, $\rho=0.3825445560...$. (Note: the area under the trapezoid is 0.9195444057... .) A random variate with such a trapezoidal density can be generated as aV_1+bV_2 for some constants $a,b>0$ where V_1,V_2 are iid uniform $[-1,1]$ random variates. Determine a,b in this case.
2. Show that as $a \uparrow \infty$,
$$\int_a^\infty e^{-\frac{x^2}{2}} \sim \frac{1}{a} e^{-\frac{a^2}{2}} .$$
3. The optimal probability p in Theorem 1.1 depends upon n. Use Stirling's formula to determine a constant c such that $p \geq 1-\frac{c}{n}$, valid for all $n \geq 3$.
4. If we want to generate a normal random variate by rejection from the exponential density $\frac{\lambda}{2}e^{-\lambda|x|}$, the smallest rejection constant is obtained when $\lambda=1$. The constant is $\sqrt{\frac{2e}{\pi}}$. Show this. Note that the corresponding rejection algorithm is:

   REPEAT

   Generate two iid exponential random variates, X,E.

   UNTIL $2E \leq (X-1)^2$

   RETURN SX where S is a random sign.

   This algorithm is mentioned in Abramowitz and Stegun (1970), where von Neumann is credited. Butcher (1961) attributes it to Kahn. Others have rediscovered it later.
5. **Teichroew's distribution.** Teichroew (1957) has shown that the functions $\phi(t) = \frac{1}{(1+t^2)^a}$ are valid characteristic functions for all values $a>0$ of the parameter. Show that random variates from this family can be generated as
   (i) G_1-G_2, where the G_i's are iid gamma (a) random variables;
   (ii) $N\sqrt{2G}$ where N,G are independent random variables with a normal and gamma (a) distribution respectively.
6. This question is related to the algorithm of Kinderman and Ramage (1976) (programs given in Kinderman and Ramage (1977)). Consider the isosceles

triangular density g_θ of the random variable $\theta(V_1+V_2)$ where V_1,V_2 are iid uniform [-1,1] random variates. Show that the largest triangle to fit under the normal density f touches f at the origin. Show next that the sides of the largest triangle touch f somewhere else. Conclude that the optimal θ is given by θ=1.1080179... , and that the corresponding optimal weight of the triangle is p =0.88407040... .

7. **The lognormal density.** When N is normally distributed, then $\theta+e^{\varsigma+\sigma N}$ is lognormal with parameters θ,ς,σ, all real numbers. The lognormal distribution has a density with support on (θ,∞) given by

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

Random variate generation requires the exponentiation of a normal random variate, and can be beaten speedwise by the judicious use of a composition/rejection algorithm, or a rejection algorithm with a good squeeze step. Develop just such an algorithm. To help you find a solution, it is instructive to draw several lognormal densities. Consider only the case θ=0 since θ is a translation parameter. Show also that in that case, the mode is at $e^{\varsigma-\sigma^2}$, the median is at e^{ς}, and that the r-th moment is $e^{r\varsigma+r^2\sigma^2/2}$ when $r>0$.

8. In the composition/rejection algorithm of Marsaglia and Bray (1964), we return the sum of three independent uniform [-1,1] random variates about 86% of the time. Schuster (1983) has shown that by considering sums of the form $a_1V_1+a_2V_2+a_3V_3$, where the V_i's are iid uniform [-1,1] random variates, it is possible to find coefficients a_1,a_2,a_3 such that we can return the said sum about 97% of the time (note however that the multiplications could actually cause a slowdown). Find these coefficients, and give the entire algorithm.

2. THE EXPONENTIAL DENSITY.

2.1. Overview.

We hardly have to convince the reader of the crucial role played by the exponential distribution in probability and statistics and in random variate generation. We have discussed various generators in the early chapters of this book. No method is shorter than the inversion method, which returns $-\log(U)$ where U is a uniform [0,1] random variate. For most users, this method is satisfactory for their needs. In a high level language, the inversion method is difficult to beat. A variety of algorithms should be considered when the computer does not have a log operation in hardware and one wants to obtain a faster method. These include:

1. The uniform spacings method (section V.3.5).
2. von Neumann's method (section IV.2.2).
3. Marsaglia's exponential generator, or its modifications (discussed below).
4. The ratio-of-uniforms method (section IV.7.2).
5. The series method (section IV.5.3).
6. Table methods.

The methods listed under points 4 and 5 will not be discussed again in this chapter. Methods 2, 3 and 6 are all based upon the memoryless property of the exponential distribution, which states that given that an exponential random variable E exceeds $x>0$, $E-x$ is again exponentially distributed. This is at the basis of Lemma IV.2.1, repeated here for the sake of readability:

Lemma IV.2.1.

An exponential random variable E is distributed as $(Z-1)\mu+Y$ where Z,Y are independent random variables and $\mu>0$ is an arbitrary positive number: Z is geometrically distributed with

$$P(Z=i)=\int_{(i-1)\mu}^{i\mu} e^{-x}\,dx = e^{-(i-1)\mu}-e^{-i\mu} \quad (i\geq 1),$$

and Y is a truncated exponential random variable with density

$$f(x)=\frac{e^{-x}}{1-e^{-\mu}} \quad (0\leq x\leq \mu).$$

Since Z,Y are independent, exponential random variate generation can truly be considered as the problem of the generation of a discrete random variate plus a continuous random variate with compact support. And because the continuous random variate has compact support, any fast table method can be used .

The uniform spacings method is based upon the fact that $GS_1, \ldots, GS_n$ are iid exponential random variables when G is gamma (n), and $S_1, \ldots, S_n$ are spacings defined by a uniform sample of size $n-1$. For $n=2$ this is sometimes faster than straightforward inversion:

Generate iid uniform [0,1] random variates U, V, W.
$Y \leftarrow -\log(UV)$
RETURN $WY, (1-W)Y$

Notice that three uniform random variates and one logarithm are needed per couple of exponential random variates. The overhead for the case $n=3$ is sometimes a drawback. We summarize nevertheless:

Generate iid uniform [0,1] random variates U_1, U_2, U_3, U_4, U_5.
$Y \leftarrow -\log(U_1 U_2 U_3)$
$V \leftarrow \min(U_4, U_5), W \leftarrow \max(U_4, U_5)$
RETURN $VY, (W-V)Y, (1-W)Y$

2.2. Marsaglia's exponential generator.

Marsaglia (1961) proved the following theorem:

Theorem 2.1. (Marsaglia, 1961)

Let $U_1, U_2, \ldots$ be iid uniform [0,1] random variables. Let Z be a truncated Poisson random variate with probability vector

$$P(Z=i) = \frac{1}{e^{\mu}-1} \frac{\mu^i}{i!} \quad (i \geq 1),$$

where $\mu > 0$ is a constant. Let M be a geometric random vector with probability vector

$$P(M=i) = (1-e^{-\mu})e^{-\mu i} \quad (i \geq 0).$$

Then $X \leftarrow \mu(M + \min(U_1, \ldots, U_Z))$ is exponentially distributed. Also,

$$E(M) = \frac{1}{e^{\mu}-1},$$

$$E(Z) = \frac{\mu e^{\mu}}{e^{\mu}-1}.$$

Proof of Theorem 2.1.

We note that for $\mu \geq x > 0$,

$$P(\mu \min(U_1, \ldots, U_Z) \leq x) = \sum_{i=1}^{\infty} P(Z=i) P(\mu \min(U_1, \ldots, U_i) \leq x)$$

$$= \sum_{i=1}^{\infty} \frac{1}{e^{\mu}-1} \frac{\mu^i}{i!} (1-(1-\frac{x}{\mu})^i)$$

$$= 1 - \sum_{i=1}^{\infty} \frac{1}{e^{\mu}-1} \frac{(\mu(1-\frac{x}{\mu}))^i}{i!}$$

$$= 1 - \frac{e^{\mu-x}-1}{e^{\mu}-1}$$

$$= \frac{1-e^{-x}}{1-e^{-\mu}}.$$

Thus, $\mu \min(U_1, \ldots, U_Z)$ has the exponential distribution truncated to $[0,\mu]$. The first part of the theorem now follows directly from Lemma IV.2.1. For the second part, use the fact that $M+1$ is geometrically distributed, so that $E(M+1) = \frac{1}{1-e^{-\mu}}$. Furthermore,

$$E(Z) = \frac{1}{e^{\mu}-1} (\frac{\mu^1}{0!} + \frac{\mu^2}{1!} + \frac{\mu^3}{2!} + \cdots)$$

$$= \frac{\mu e^{\mu}}{e^{\mu}-1}. \blacksquare$$

We can now suggest an algorithm based upon Theorem 2.1:

Marsaglia's exponential generator

Generate a geometric random variate M defined by $P(M=i)=(1-e^{-\mu})e^{-\mu i}$ $(i \geq 0)$.
$Z \leftarrow 1$
Generate iid uniform [0,1] random variates U,V.
$Y \leftarrow V$
WHILE True Do
  IF $U \leq F(Z)$ (Note: $F(i)=\frac{1}{e^{\mu}-1}\sum_{j=1}^{i}\frac{\mu^j}{j!}$.)
    THEN RETURN $X \leftarrow \mu(M+Y)$
    ELSE
      $Z \leftarrow Z+1$
      Generate a uniform [0,1] random variate V.
      $Y \leftarrow \min(Y,V)$

For the geometric random variate, the inversion method based upon sequential search seems the obvious choice. This can be sped up by storing the cumulative probabilities, or by mixing sequential search with the alias method. Similarly, the cumulative distribution function F of Z can be partially stored to speed up the second part of the algorithm. The design parameter μ must be found by compromise. Note that if sequential search based inversion is used for the geometric random variate M, then $\frac{1}{1-e^{-\mu}}$ comparisons are needed on the average: this decreases from ∞ to 1 as μ varies from 0 to ∞. Also, the expected number of accesses of F in the second part of the algorithm is equal to $E(Z)=\frac{\mu}{1-e^{-\mu}}$, and this increases from 1 to ∞ as μ varies from 0 to ∞. Furthermore, the algorithm in its entirety requires on the average $2+E(Z)$ uniform [0,1] random variates. The two effects have to be properly balanced. For most implementations, a value μ in the range 0.40...0.80 seems to be optimal. This point was addressed in more detail by Sibuya (1961). Special advantages are offered by the choices $\mu=1$ and $\mu=\log(2)$.

The special case $\mu=\log(2)$ allows one to generate the desired geometric random variate by analyzing the random bits in a uniform [0,1] random variate, which can be done conveniently in assembly language by the logical shift operation. This algorithm was proposed by Ahrens and Dieter (1972), where the reader can also find an excellent survey of exponential random variate generation. Again, a table of $F(i)$ values is needed.

Exponential generator of Ahrens and Dieter (1972)

[NOTE: a table of values $F(i)=\sum_{j=1}^{i} \frac{(\log(2))^j}{j!}$ is required.]

$M \leftarrow 0$

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

WHILE $U < \frac{1}{2}$ DO $U \leftarrow 2U$, $M \leftarrow M+\log(2)$

> (M is now correctly distributed. It is equal to the number of 0's before the first 1 in the binary expansion of U. Note that $U \leftarrow 2U$ is implementable by a shift operation.)

$U \leftarrow 2U-1$ (U is again uniform [0,1] and independent of M.)

IF $U < \log(2)$

    THEN RETURN $X \leftarrow M+U$

    ELSE

        $Z \leftarrow 2$

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

        $Y \leftarrow V$

        WHILE True Do

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

            $Y \leftarrow \min(Y,V)$

            IF $U \le F(Z)$

                THEN RETURN $X \leftarrow M+Y\log(2)$

                ELSE $Z \leftarrow Z+1$

Ahrens and Dieter squeeze the first uniform [0,1] random variate U dry. Because of this, the algorithm requires very few uniform random variates on the average: the expected number is 1+log(2), which is about 1.69315.

2.3. The rectangle-wedge-tail method.

One of the fastest table methods for the exponential distribution was first published by Maclaren, Marsaglia and Bray (1964). It is ideally suited for implementation in machine language, but even in a high level language it is faster than most other methods described in this section. The extra speed is obtained by principles related to the table method. First, the tail of the density is cut off at some point $n\mu$ where n is a design integer and $\mu>0$ is a small design constant. The remainder of the graph of f is then divided into n equal strips of width μ. And on interval $[(i-1)\mu, i\mu]$, we divide the graph into a rectangular piece of height $e^{-i\mu}$, and a wedge $f(x)-e^{-i\mu}$. Thus, the density is decomposed into

$2n+1$ pieces of the following weights:

one tail of weight $e^{-n\mu}$;
n rectangles with weights $\mu e^{-i\mu}$ $, 1 \le i \le n$;
n wedges of weights $e^{-i\mu}(e^{\mu}-1-\mu)$ $, 1 \le i \le n$.

These numbers can be used to set up a table for discrete random variate generation. The algorithm then proceeds as follows:

The rectangle-wedge-tail method

[NOTE: we refer to the $2n+1$ probabilities defined above.]
$X \leftarrow 0$
REPEAT
Generate a random integer Z with values in $1, \ldots, 2n+1$ having the given probability vector.
CASE
Rectangle i chosen: RETURN $X \leftarrow X+(i-1+U)\mu$ where U is a uniform $[0,1]$ random variate.

Wedge i chosen: RETURN $X \leftarrow X+(i-1)\mu+Y$ where Y is a random variate having the wedge density $g(x)=\frac{e^{\mu-x}-1}{e^{\mu}-1-\mu}$ $, 0 \le x \le \mu$.

Tail is chosen: $X \leftarrow X+n\mu$
UNTIL False

Note that when the tail is picked, we do in fact reject the choice, but keep at the same time track of the number of rejections. Equivalently, we could have returned $n\mu-\log(U)$ but this would have been less elegant since we would in effect rely on a logarithm. The recursive approach followed here seems cleaner. Random variates from the wedge density can be obtained in a number of ways. We could proceed by rejection from the triangular density: note that

$$g(x) = \frac{e^{\mu-x}-1}{e^{\mu}-1-\mu} \le \frac{\mu-x}{\mu} \frac{e^{\mu}-1}{e^{\mu}-1-\mu}$$

and

$$g(x) \ge \frac{e^{\mu}-xe^{\mu}-1}{e^{\mu}-1-\mu},$$

so that the following rejection algorithm is valid:

Wedge generator

REPEAT

Generate two iid uniform [0,1] random variates X,U.

IF $X>U$ THEN $(X,U)\leftarrow(U,X)$ ((X,U) is now uniformly distributed under the triangle with unit sides.)

IF $U \le 1-X\frac{\mu e^{\mu}}{e^{\mu}-1}$

THEN RETURN μX

ELSE IF $U \le \frac{e^{\mu-\mu X}-1}{e^{\mu}-1}$ THEN RETURN μX

UNTIL False

The wedge generator requires on the average

$$\frac{1}{2}\frac{\mu(e^{\mu}-1)}{e^{\mu}-1-\mu}$$

iterations. It is easy to see that this tends to 1 as $\mu\downarrow 0$. The expected number of uniform random variates needed is thus twice this number. But note that this can be bounded as follows:

$$\mu\frac{e^{\mu}-1}{e^{\mu}-1-\mu} = \mu(1+\frac{\mu}{e^{\mu}-1-\mu}) \le \mu(1+\frac{2}{\mu}) = \mu+2 .$$

Here we used an inequality based upon the truncated Taylor series expansion. In view of the squeeze step, the expected number of evaluations of the exponential function is of course much less than the expected number of iterations. Having established this, we can summarize the performance of the algorithm by repeated use of Wald's equation:

Theorem 2.2.

This theorem is about the analysis of the rectangle-wedge-tail algorithm shown above.

(i) The expected number of global iterations is $A = \frac{1}{1-e^{-n\mu}}$.

(ii) The expected number of uniform [0,1] random variates needed (excluding the discrete random variate generation portion) is $\frac{\mu}{1-e^{-\mu}}$.

Proof of Theorem 2.2.

Theorem 2.2 is established as follows: we have 1 uniform random variate per rectangle (the probability of this is $\sum_{i=1}^{n} \mu e^{-i\mu} = \mu \frac{e^{-\mu}-e^{-(n+1)\mu}}{1-e^{-\mu}}$ in the first iteration). We have $\mu \frac{e^{\mu}-1}{e^{\mu}-1-\mu}$ per wedge (the probability of this is $\sum_{i=1}^{n} e^{-i\mu}(e^{\mu}-1-\mu) = \frac{e^{-\mu}-e^{-(n+1)\mu}}{1-e^{-\mu}}(e^{\mu}-1-\mu)$ in the first iteration). Thus, by establishing the correctness of statement (1), and applying Wald's equation, we observe that the expected number of uniform random variates needed is

$$
\begin{aligned}
& A\left(\mu \frac{e^{-\mu}-e^{-(n+1)\mu}}{1-e^{-\mu}} + \mu \frac{e^{\mu}-1}{e^{\mu}-1-\mu} \frac{e^{-\mu}-e^{-(n+1)\mu}}{1-e^{-\mu}}(e^{\mu}-1-\mu)\right) \\
&= A\left(\mu e^{\mu} \frac{e^{-\mu}-e^{-(n+1)\mu}}{1-e^{-\mu}}\right) \\
&= A\left(\mu \frac{1-e^{-n\mu}}{1-e^{-\mu}}\right) \\
&= \frac{\mu}{1-e^{-\mu}} . \blacksquare
\end{aligned}
$$

The number of intervals n does not affect the expected number of uniform random variates needed in the algorithm. Of course, the expected number of discrete random variates needed depends very much on n, since it is $\frac{1}{1-e^{-n\mu}}$. It is clear that μ should be made very small because as $\mu \downarrow 0$, the expected number of uniform random variates is $1+\frac{\mu}{2}+o(\mu)$. But when μ is small, we have to choose n large to keep the expected number of iterations down. For example, if we want the expected number of iterations to be $\frac{1}{1-e^{-4}}$, which is entirely reasonable, then we should choose $n = \frac{4}{\mu}$. When $\mu = \frac{1}{20}$, the table size is $2n+1=161$.

The algorithm given here may differ slightly from the algorithms found elsewhere. The idea remains basically the same: by picking certain design constants, we can practically guarantee that one exponential random variate can be obtained at the expense of one discrete random variate and one uniform random variate. The discrete random variate in turn can be obtained extremely quickly by the alias method or the alias-urn method at the cost of one other uniform random variate and either one or two table look-ups.

2.4. Exercises.

1. It is important to have a fast generator for the truncated exponential density $f(x)=e^{-x}/(1-e^{-\mu})$, $0 \le x \le \mu$. From Theorem 2.1, we recall that a random variate with this density can be generated as $\mu \min(U_1, \ldots, U_Z)$ where the U_i's are iid uniform [0,1] random variates and Z is a truncated Poisson variate with probability vector

$$P(Z=i) = \frac{1}{e^{\mu}-1} \frac{\mu^i}{i!} \quad (i \ge 1) .$$

The purpose of this exercise is to explore alternative methods. In particular, compare with a strip table method based upon n equi-sized intervals and with a grid table method based upon n equi-sized intervals. Compare also with rejection from a trapezoidal dominating function, combined with clever squeeze steps.

2. **The Laplace density.** The Laplace density is $f(x)=\frac{1}{2}e^{-|x|}$. Show that a random variate X with this density can be generated as SE or as E_1-E_2 where E, E_1, E_2 are iid exponential random variates, and S is a random sign.

3. Find the density of the sum of two iid Laplace random variables, and verify its bell shape. Prove that such a random variate can be generated as $\log(\frac{U_1U_2}{U_3U_4})$ where the U_i's are iid uniform [0,1] random variates. Develop a rejection algorithm for normal random variates with quick acceptance and rejection steps based upon the inequalities:

$$1-\frac{x^3}{3} \le \frac{e^{-\frac{x^2}{2}}}{(1+x)e^{-x}} \le \begin{cases} 1 & ,x>0 \\ 1-\frac{x^3}{3}(\frac{23}{27})(1-\frac{x^2}{6}) & ,x>0 \end{cases}$$

Prove these inequalities by using Taylor's series expansion truncated at the third term.

3. THE GAMMA DENSITY.

3.1. The gamma family.

A random variable X is **gamma** (a,b) distributed when it has density

$$f(x) = \frac{x^{a-1}e^{-\frac{x}{b}}}{\Gamma(a)b^a} \quad (x \ge 0) .$$

Here $a>0$ is the shape parameter and $b>0$ is the scale parameter. We say that X is gamma (a) distributed when it is gamma $(a,1)$. Before reviewing random variate generation techniques for this family, we will look at some key properties that are relevant to us and that could aid in the design of an algorithm.

The density is unimodal with mode at $(a-1)b$ when $a \geq 1$. When $a<1$, it is monotone with an infinite peak at 0. The moments are easily computed. For example, we have

$$E(X)=\int_0^\infty xf(x)\,dx = \frac{\Gamma(a+1)b^{a+1}}{\Gamma(a)b^a} = ab \ ;$$

$$E(X^2)=\int_0^\infty x^2 f(x)\,dx = \frac{\Gamma(a+2)b^{a+2}}{\Gamma(a)b^a} = a(a+1)b^2 \ .$$

Thus, $Var(X)=ab^2$.

The gamma family is closed under many operations. For example, when X is gamma (a,b), then cX is gamma (a,bc) when $c>0$. Also, summing gamma random variables yields another gamma random variable. This is perhaps best seen by considering the characteristic function $\phi(t)$ of a gamma (a,b) random variable:

$$\phi(t)=E(e^{itX})=\int_0^\infty \frac{x^{a-1}e^{-x(\frac{1}{b}-it)}}{\Gamma(a)b^a}\,dx$$

$$=\frac{(\frac{b}{1-itb})^a}{b^a}\int_0^\infty \frac{x^{a-1}e^{-x(\frac{1}{b}-it)}}{\Gamma(a)(\frac{b}{1-itb})^a}\,dx$$

$$=\frac{1}{(1-itb)^a} \ .$$

Thus, if $X_1, \ldots, X_n$ are independent gamma $(a_1), \ldots$, gamma (a_n) random variables, then $X=\sum_{i=1}^n X_i$ has characteristic function

$$\phi(t)=\prod_{j=1}^n \frac{1}{(1-it)^{a_j}} = \frac{1}{(1-it)^{\sum_{j=1}^n a_j}} \ ,$$

and is therefore gamma $(\sum_{j=1}^n a_j, 1)$ distributed. The family is also closed under more complicated transformations. To illustrate this, we consider Kullback's result (Kullback, 1934) which states that when X_1,X_2 are independent gamma (a) and gamma $(a+\frac{1}{2})$ random variables, then $2\sqrt{X_1X_2}$ is gamma $(2a)$.

The gamma distribution is related in innumerable ways to other well-known distributions. The exponential density is a gamma density with parameters (1,1). And when X is normally distributed, then X^2 is gamma $(\frac{1}{2},2)$ distributed. This is called the **chi-square** distribution with one degree of freedom. In general, a gamma $(\frac{r}{2},2)$ random variable is called a chi-square random variable with r degrees of freedom. We will not use the chi-square terminology in this section. Perhaps the most important property of the gamma density is its relationship with the beta density. This is summarized in the following theorem:

Theorem 3.1.

If X_1,X_2 are independent gamma (a_1) and gamma (a_2) random variables, then $\frac{X_1}{X_1+X_2}$ and X_1+X_2 are independent beta (a_1,a_2) and gamma (a_1+a_2) random variables. Furthermore, if Y is gamma (a) and Z is beta $(b,a-b)$ for some $b>a>0$, then YZ and $Y(1-Z)$ are independent gamma (b) and gamma $(a-b)$ random variables.

Proof of Theorem 3.1.

We will only prove the first part of the theorem, and leave the second part to the reader (see exercises). Consider first the transformation $y=x_1/(x_1+x_2)$, $z=x_1+x_2$, which has an inverse $x_1=yz,x_2=(1-y)z$. The Jacobian of the transformation is

$$\begin{vmatrix} \frac{\partial x_1}{\partial y} & \frac{\partial x_1}{\partial z} \\ \frac{\partial x_2}{\partial y} & \frac{\partial x_2}{\partial z} \end{vmatrix} = \begin{vmatrix} z & y \\ -z & 1-y \end{vmatrix} = |z| .$$

Thus, the density $f(y,z)$ of $(Y,Z)=(\frac{X_1}{X_1+X_2},X_1+X_2)$ is

$$\frac{(yz)^{a_1-1}e^{-yz}}{\Gamma(a_1)} \frac{((1-y)z)^{a_2-1}e^{-(1-y)z}}{\Gamma(a_2)} z$$

$$= \frac{\Gamma(a_1+a_2)y^{a_1-1}(1-y)^{a_2-1}}{\Gamma(a_1)\Gamma(a_2)} \frac{z^{a_1+a_2-1}e^{-z}}{\Gamma(a_1+a_2)} ,$$

which was to be shown. ∎

The observation that for large values of a, the gamma density is close to the normal density could aid in the choice of a dominating curve for the rejection method. This fact follows of course from the observation that sums of gamma random variables are again gamma random variables, and from the central limit theorem. However, since the central limit theorem is concerned with the convergence of distribution functions, and since we are interested in a local central limit theorem, convergence of a density to a density, it is perhaps instructive to give a direct proof of this result. We have:

Theorem 3.2.

If X_a is gamma (a) distributed and if f_a is the density of the normalized gamma random variable $(X_a - a)/\sqrt{a}$, then

$$\lim_{a \uparrow \infty} f_a(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (x \in R).$$

Proof of Theorem 3.2.

The density of $(X_a - a)/\sqrt{a}$ evaluated at x is

$$\sqrt{a}\,\frac{(x\sqrt{a}+a)^{a-1}e^{-(x\sqrt{a}+a)}}{\Gamma(a)} \sim \frac{\sqrt{a}\,a^{a-1}(1+\frac{x}{\sqrt{a}})^{a-1}e^{-a}\,e^{-x\sqrt{a}}}{(\frac{a-1}{e})^{a-1}\sqrt{2\pi(a-1)}}$$

$$\sim \frac{1}{\sqrt{2\pi}}\frac{1}{e}(1+\frac{1}{a-1})^{a-1} e^{x\sqrt{a}+\frac{(a-1)x}{\sqrt{a}}-\frac{(a-1)x^2}{2a}+O(\frac{1}{\sqrt{a}})}$$

$$= \frac{1}{\sqrt{2\pi}}(1+o(1))e^{-\frac{x^2}{2}+O(\frac{1}{\sqrt{a}})}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}(1+o(1)).$$

Here we used Stirling's approximation, and the Taylor series expansion for $\log(1+u)$ when $0<u<1$. ■

3.2. Gamma variate generators.

Features we could appreciate in good gamma generators include

(i) Uniform speed: the expected time is uniformly bounded over all values of a, the shape parameter.

(ii) Simplicity: short easy programs are more likely to become widely used.

(iii) Small or nonexistent set-up times: design parameters which depend upon a need to be recalculated every time a changes. These recalculations take often more time than the generator.

No family has received more attention in the literature than the gamma family. Many experimental comparisons are available in the general literature: see e.g. Atkinson and Pearce (1976), Vaduva (1977), or Tadikamalla and Johnson

(1980,1981).

For special cases, there are some good recipes: for example, when $a=1$, we return an exponential random variate. When a is a small integer, we can return either

$$\sum_{i=1}^{a} E_i$$

where the E_i's are iid exponential random variates, or

$$-\log(\prod_{i=1}^{a} U_i)$$

where the U_i's are iid uniform [0,1] random variates. When a equals $\frac{1}{2}+k$ for some small integer k, it is possible to return

$$\frac{1}{2}N^2+\sum_{i=1}^{k} E_i$$

where N is a normal random variate independent of the E_i's. In older texts one will often find the recommendation that a gamma (a) random variate should be generated as the sum of a gamma ($\lfloor a \rfloor$) and a gamma ($a-\lfloor a \rfloor$) random variate. The former random variate is to be obtained as a sum of independent exponential random variates. The parameter of the second gamma variate is less than 1. All these strategies take time linearly increasing with a; none lead to good gamma generators in general.

There are several successful approaches in the design of good gamma generators: first and foremost are the rejection algorithms. The rejection algorithms can be classified according to the family of dominating curves used. The differences in timings are usually minor: they often depend upon the efficiency of some quick acceptance step, and upon the way the rejection constant varies with a as $a \uparrow \infty$. Because of Theorem 3.2, we see that for the rejection constant to converge to 1 as $a \uparrow \infty$ it is necessary for the dominating curve to approach the normal density. Thus, some rejection algorithms are suboptimal from the start. Curiously, this is sometimes not a big drawback provided that the rejection constant remains reasonably close to 1. To discuss algorithms, we will inherit the names available in the literature for otherwise our discussion would be too verbose. Some successful rejection algorithms include:

GB. (Cheng, 1977): rejection from the Burr XII distribution. To be discussed below.

GO. (Ahrens and Dieter, 1974): rejection from a combination of normal and exponential densities.

GC. (Ahrens and Dieter, 1974): rejection from the Cauchy density.

XG. (Best, 1978): rejection from the t distribution with 2 degrees of freedom.

TAD2.
(Tadikamalla, 1978): rejection from the Laplace density.

Of these approaches, algorithm GO has the best asymptotic value for the rejection constant. This by itself does not make it the fastest and certainly not the shortest algorithm. The real reason why there are so many rejection algorithms around is that the normalized gamma density cannot be fitted under the normal density because its tail decreases much slower than the tail of the normal density. We can of course apply the almost exact inversion principle and find a nonlinear transformation which would transform the gamma density into a density which is very nearly normal, and which among other things would enable us to tuck the new density under a normal curve. Such normalizing transformations include a quadratic transformation (Fisher's transformation) and a cubic transformation (the Wilson-Hilferty transformation): the resulting algorithms are extremely fast because of the good fit. A prototype algorithm of this kind was developed and analyzed in detail in section IV.3.4, Marsaglia's algorithm RGAMA (Marsaglia (1977), Greenwood (1974)). In section IV.7.2, we presented some gamma generators based upon the ratio-of-uniforms method, which improve slightly over similar algorithms published by Kinderman and Monahan (1977, 1978, 1979) (algorithm GRUB) and Cheng and Feast (1979, 1979) (algorithm GBH). Despite the fact that no ratio-of-uniforms algorithm can have an asymptotically optimal rejection constant, they are typically comparable to the best rejection algorithms because of the simplicity of the dominating density. Most useful algorithms fall into one of the categories described above. The universal method for log-concave densities (section VII.2.3) (Devroye, 1984) is of course not competitive with specially designed algorithms.

There are no algorithms of the types described above which are uniformly fast for all a because the design is usually geared towards good performance for large values of a. Thus, for most algorithms, we have uniform speed on some interval $[a^*,\infty)$ where a^* is typically near 1. For small values of a, the algorithms are often not valid - this is due to the fact that the gamma density has an infinite peak at 0 when $a<1$, while dominating curves are often taken from a family of bounded densities. We will devote a special section to the problem of gamma generators for values $a<1$.

Sometimes, there is a need for a very fast algorithm which would be applied for a fixed value of a. What one should do in such case is cut off the tail, and use a strip-based table method (section VIII.2) on the body. Since these table methods can be automated, it is not worth spending extra time on this issue. It is nevertheless worth noting that some automated table methods have table sizes that in the case of the gamma density increase unboundedly as $a\to\infty$ if the expected time per random variate is to remain bounded, unless one applies a specially designed technique similar to what was done for the exponential density in the rectangle-wedge-tail method. In an interesting paper, Schmeiser and Lal (1980) have developed a semi-table method: the graph of the density is partitioned into about 10 pieces, all rectangular, triangular or exponential in shape, and the set-up time, about five times the time needed to generate one random variate, is reasonable. Moreover, the table size (number of pieces) remains fixed for all values of a. When speed per random variate is at a premium, one should certainly use some sort of table method. When speed is important, and a varies

Theorem 3.3.

A. The density g has distribution function

$$G(x) = \frac{1}{2}\left(1+\frac{\frac{x}{\sqrt{2}}}{\sqrt{1+\frac{x^2}{2}}}\right).$$

A random variate with this distribution can be generated as

$$\frac{\sqrt{2}(U-\frac{1}{2})}{\sqrt{U(1-U)}}$$

where U is a uniform $[0,1]$ random variate.

B. Let f be the gamma (a) density, and let g_a be the density of $(a-1)+Y\sqrt{\frac{3a}{2}-\frac{3}{8}}$ where Y has density g. Then

$$f(x) \leq c_a g_a(x) = \frac{1}{\Gamma(a)\left(1+\frac{1}{2}\left(\frac{x-(a-1)}{\sqrt{\frac{3a}{2}-\frac{3}{8}}}\right)^2\right)^{\frac{3}{2}}},$$

where the rejection constant is given by

$$c_a = \frac{2\sqrt{3a-\frac{3}{4}}}{\Gamma(a)}\left(\frac{a-1}{e}\right)^{a-1}.$$

C. We have $\sup_{a\geq 1} c_a \leq e\sqrt{\frac{6}{\pi}}$, and $\lim_{a\uparrow\infty} c_a = \sqrt{\frac{6}{\pi}}$.

Proof of Theorem 3.3.

The claim about the distribution function G is quickly verified. When U is uniformly distributed on $[0,1]$, then the solution X of $G(X)=U$ is precisely $X=\frac{\sqrt{2}(U-\frac{1}{2})}{\sqrt{U(1-U)}}$. This proves part A.

Let Y have density g. Then $(a-1)+Y\sqrt{\frac{3a}{2}-\frac{3}{8}}$ has density

$$\frac{1}{2\sqrt{2}\sqrt{\frac{3a}{2}-\frac{3}{8}}}\left(1+\frac{1}{2}\left(\frac{x-(a-1)}{\sqrt{\frac{3a}{2}-\frac{3}{8}}}\right)^2\right)^{-\frac{3}{2}}$$

with each call, the almost-exact-inversion method seems to be the winner in most experimental comparisons, and certainly when fast exponential and normal random variate generators are available. The best ratio-of-uniforms methods and the best rejection methods (XG,GO,GB) are next in line, well ahead of all table methods.

Finally, we will discuss random variate generation for closely related distributions such as the Weibull distribution and the exponential power distribution.

3.3. Uniformly fast rejection algorithms for $a \geq 1$.

We begin with one of the shortest algorithms for the gamma density, which is based upon rejection from the t density with 2 degrees of freedom:

$$g(x) = \frac{1}{2\sqrt{2}}\left(1+\frac{x^2}{2}\right)^{-\frac{3}{2}}$$

This density decreases as x^{-3}, and is symmetric bout 0. Thus, it can be used as a dominating curve of a properly rescaled and translated gamma density. Best's algorithm XG (Best, 1978) is based upon the following facts:

To prove statement B, we need only show that for $x > 0$,

$$x^{a-1}e^{-x} \le \left(\frac{a-1}{e}\right)^{a-1} \frac{1}{\left(1+\frac{1}{2}\left(\frac{x-(a-1)}{\sqrt{\frac{3a}{2}-\frac{3}{8}}}\right)^2\right)^{\frac{3}{2}}} ,$$

or, after resubstitution $y = x-(a-1)$, that for $y \ge -(a-1)$,

$$e^{-y}\left(1+\frac{y}{a-1}\right)^{a-1} \le \left(1+\frac{y^2}{3a-\frac{3}{4}}\right)^{-\frac{3}{2}} .$$

Taking logarithms, we see that we must show that

$$h(y) = -y+(a-1)\log(1+\frac{y}{a-1})+\frac{3}{2}\log\left(1+\frac{y^2}{3a-\frac{3}{4}}\right) \le 0 .$$

Clearly, $h(0)=0$. It suffices to show that $h'(y) \ge 0$ for $y \le 0$ and that $h'(y) \le 0$ for $y \ge 0$. But

$$h'(y) = -1+\frac{a-1}{(a-1)(1+\frac{y}{a-1})}+\frac{3}{2}\frac{2y}{3a-\frac{3}{4}}\frac{1}{1+\frac{y^2}{3a-\frac{3}{4}}}$$

$$= -\frac{y}{a-1+y}+\frac{y}{a-\frac{1}{4}+\frac{y^2}{3}}$$

$$= \frac{y(y-\frac{3}{4}-\frac{y^2}{3})}{(a-1+y)(a-\frac{1}{4}+\frac{y^2}{3})} .$$

The denominator is ≥ 0 for $a \ge \frac{1}{4}$. The numerator is ≥ 0 for $y \le 0$, and is ≤ 0 for $y \ge 0$ (this can be seen by rewriting it as $-\frac{y}{3}(y-\frac{3}{2})^2$. This concludes the proof of part B.

For part C, we apply Stirling's approximation, and observe that

$$c_a \sim \frac{2\sqrt{3a}}{(\frac{a}{e})^a\sqrt{\frac{2\pi}{a}}}\left(\frac{a-1}{e}\right)^{a-1}$$

$$= \frac{2e\sqrt{3a}}{\sqrt{2\pi a}}\left(1-\frac{1}{a}\right)^{a-1}$$

$$\sim \sqrt{\frac{6}{\pi}} .$$

The first $\sim$ is also an upper bound, so that

$$c_a \leq \sqrt{\frac{6}{\pi}} e^{\frac{1}{a}}$$

when $a \geq 1$. This proves part C. ■

Based upon Theorem 3.3, we can now state Best's rejection algorithm:

Best's rejection algorithm XG for gamma random variates (Best, 1978)

[SET-UP]

$b \leftarrow a-1, c \leftarrow 3a-\frac{3}{4}$

[GENERATOR]

REPEAT

Generate iid uniform [0,1] random variates U, V.

$W \leftarrow U(1-U), Y \leftarrow \sqrt{\frac{c}{W}}(U-\frac{1}{2}), X \leftarrow b+Y$

IF $X \geq 0$

THEN

$Z \leftarrow 64 W^3 V^2$

Accept $\leftarrow [Z \leq 1-\frac{2Y^2}{X}]$

IF NOT Accept

THEN Accept $\leftarrow [\log(Z) \leq 2(b \log(\frac{X}{b})-Y)]$

UNTIL Accept

RETURN X

The random variate X generated at the outset of the REPEAT loop has density g_a. The acceptance condition is

$$e^{-Y}(1+\frac{Y}{a-1})^{a-1} \geq V\left(1+\frac{Y^2}{3a-\frac{3}{4}}\right)^{-\frac{3}{2}} .$$

This can be rewritten in a number of ways: for example, in the notation of the algorithm,

$$e^{-Y}(\frac{X}{b})^{b} \geq V(4W)^{\frac{3}{2}} ;$$

$$-Y+b\log(\frac{X}{b}) \geq \frac{1}{2}\log(4^3V^2W^3) ;$$

$$2(-Y+b\log(\frac{X}{b})) \geq \log(Z) .$$

This explains the acceptance condition used in the algorithm. The squeeze step is derived from the acceptance condition, by noting that

(i) $\log(Z) \leq Z-1$;

(ii) $2(b\log(1+\frac{Y}{b})-Y) \geq 2Y(-\frac{Y}{b+Y}) = -\frac{2Y^2}{X}$.

The last inequality is obtained by noting that the left hand side as a function of Y is 0 at $Y=0$, and has derivative $-\frac{Y}{b+Y}$. Therefore, by the Taylor series expansion truncated at the first term, we see that for $Y \geq 0$, the left hand side is at least equal to $2(0+Y(-\frac{Y}{b+Y}))$. For $Y \leq 0$, the same bound is valid. Thus, when $Z-1 \leq -2Y^2/X$, we are able to conclude that the acceptance condition is satisfied. It should be noted that in view of the rather large rejection constant, the squeeze step is probably not very effective, and could be omitted without a big time penalty.

We will now move on to Cheng's algorithm GB which is based upon rejection from the Burr XII density

$$g(x) = \lambda\mu\frac{x^{\lambda-1}}{(\mu+x^\lambda)^2}$$

for parameters $\mu,\lambda>0$ to be determined as a function of a. Random variates with this density can be obtained as

$$(\frac{\mu U}{1-U})^{\frac{1}{\lambda}}$$

where U is uniformly distributed on [0,1]. This follows from the fact that the distribution function corresponding to g is $x^\lambda/(\mu+x^\lambda), x \geq 0$. We have to choose λ and μ. Unfortunately, minimization of the area under the dominating curve does not give explicitly solvable equations. It is useful to match the curves of f and g, which are both unimodal. Since f peaks at $a-1$, it makes sense to match this peak. The peak of g occurs at

$$x = (\frac{(\lambda-1)\mu}{\lambda+1})^{\frac{1}{\lambda}} .$$

If we choose λ large, i.e. increasing with a, then this peak will approximately match the other peak when $\mu=a^\lambda$. Consider now $\log(\frac{f}{g})$. The derivative of this function is

$$\frac{a-\lambda-x}{x}+\frac{2\lambda x^{\lambda-1}}{a^\lambda+x^\lambda} .$$

This derivative attains the value 0 when $(a+\lambda-x)x^\lambda+(a-\lambda-x)a^\lambda=0$. By analyzing the derivative, we can see that it has a unique solution at $x=0$ when $\lambda=\sqrt{2a-1}$. Thus, we have

$$f(x) \le cg(x)$$

where

$$\begin{aligned} c &= \frac{a^{a-1}e^{-a}(2a^\lambda)^2}{\Gamma(a)\lambda a^\lambda a^{\lambda-1}} \\ &= \frac{a^a e^{-a} 4}{\Gamma(a)\lambda} \\ &\sim \frac{4\sqrt{a}}{\sqrt{2\pi}\lambda} \quad (a \uparrow \infty) . \end{aligned}$$

Resubstitution of the value of λ yields the asymptotic value of $\sqrt{\frac{4}{\pi}}\approx 1.13$. In fact, we have

$$c \le \frac{4\sqrt{a}}{\sqrt{2\pi}\lambda} = \sqrt{\frac{4}{\pi}}\sqrt{a/(a-\frac{1}{2})} \le \sqrt{\frac{8}{\pi}} ,$$

uniformly over $a \ge 1$. Thus, the rejection algorithm suggested by Cheng has a good rejection constant. In the design, we notice that if X is a random variate with density g, and U is a uniform [0,1] random variate, then the acceptance condition is

$$4(\frac{a}{e})^a(\frac{a^\lambda}{X^{\lambda+1}})\frac{X^{2\lambda}}{(a^\lambda+X^\lambda)^2}U \le X^{a-1}e^{-X} .$$

Equivalently, since $V=X^\lambda/(a^\lambda+X^\lambda)$ is uniformly distributed on [0,1], the acceptance condition can be rewritten as

$$4(\frac{a}{e})^a a^\lambda V^2 U \le X^{\lambda+a}e^{-X} ,$$

or

$$\log(4)+(\lambda+a)\log(a)-a+\log(UV^2) \le (\lambda+a)\log(X)-X ,$$

or

$$\log(UV^2) \le a-\log(4)+(\lambda+a)\log(\frac{X}{a})-X .$$

A quick acceptance step can be introduced which uses the inequality

$$\log(UV^2) \le d(UV^2)-\log(d)-1$$

which is valid for all d. The value $d=\frac{9}{2}$ was suggested by Cheng. Combining all of this, we obtain:

Cheng's rejection algorithm GB for gamma random variates (Cheng, 1977)

[SET-UP]

$b \leftarrow a - \log(4)$, $c \leftarrow a + \sqrt{2a-1}$

[GENERATOR]

REPEAT

Generate iid uniform [0,1] random variates U, V.

$Y \leftarrow a \log(\frac{V}{1-V})$, $X \leftarrow ae^{V}$

$Z \leftarrow UV^2$

$R \leftarrow b + cY - X$

Accept $\leftarrow [R \geq \frac{9}{2}Z - (1+\log(\frac{9}{2}))]$ (note that $(1+\log(\frac{9}{2}))=2.5040774\ldots$)

IF NOT Accept THEN Accept $\leftarrow [R \geq \log(Z)]$

UNTIL Accept

RETURN X

We will close this section with a word about the historically important algorithm GO of Ahrens and Dieter (1974), which was the first uniformly fast gamma generator. It also has a very good asymptotic rejection constant, slightly larger than 1. The authors got around the problem of the tail of the gamma density by noting that most of the gamma density can be tucked under a normal curve, and that the right tail can be tucked under an exponential curve. The breakpoint must of course be to the right of the peak $a-1$. Ahrens and Dieter suggest the value $(a-1)+\sqrt{6(a+\sqrt{\frac{8a}{3}})}$. We recall that if X is gamma (a) distributed, then $\frac{(X-a)}{\sqrt{a}}$ tends in distribution to a normal density. Thus, with the breakpoint of Ahrens and Dieter, we cannot hope to construct a dominating curve with integral tending to 1 as $a \uparrow \infty$ (for this, the breakpoint must be at $a-1$ plus a term increasing faster than $\sqrt{a}$). It is true however that we are in practice very close. The almost-exact inversion method for normal random variates yields asymptotically optimal rejection constants without great difficulty. For this reason, we will delegate the treatment of algorithm GO to the exercises.

3.4. The Weibull density.

A random variable has the standard **Weibull density** with parameter $a>0$ when it has density

$$f(x)=ax^{a-1}e^{-x^a} \quad (x\geq 0).$$

In this, we recognize the density of $E^{\frac{1}{a}}$ where E is an exponential random variable. This fact can also be deduced from the form of its distribution function,

$$F(x)=1-e^{-x^a} \quad (x\geq 0).$$

Because of this, it seems hardly worthwhile to design rejection algorithms for this density. But, turning the tables around for the moment, the Weibull density is very useful as an auxiliary density in generators for other densities.

Example 3.1. Gumbel's extreme value distribution.

When X is Weibull (a), then $Y=-a\log(X)$ has the extreme value density

$$f(x)=e^{-x}e^{-e^{-x}} \quad (x\in R).$$

By the fact that X is distributed as $E^{\frac{1}{a}}$, we see of course that the parameter a plays no special role: thus, $-\log(E)$ and $-\log(\log(\frac{1}{U}))$ are both extreme value random variables when E is exponentially distributed, and E is exponentially distributed. ■

Example 3.2. A compound Weibull distribution.

Dubey (1968) has pointed out that the ratio $W_a/G_b^{\frac{1}{a}}$ has the Pareto-like density

$$f(x)=\frac{abx^{a-1}}{(1+x^a)^{b+1}} \quad (x\geq 0).$$

Here W_a is a Weibull (a) random variable, and G_b is a gamma (b) random variable. As a special case, we note that the ratio of two independent exponential random variables has density $\frac{1}{(1+x)^2}$ on $[0,\infty)$. ■

Example 3.3. Gamma variates by rejection from the Weibull density.

Consider the gamma (a) density f with parameter $0<a\leq 1$. For this density, random variates can be generated by rejection from the Weibull (a) density (which will be called g). This is based upon the inequality

$$\frac{f(x)}{g(x)}=\frac{e^{x^a-x}}{a\Gamma(a)}\leq\frac{e^{b-b^{\frac{1}{a}}}}{\Gamma(a+1)}$$

where

$$b=a^{\frac{a}{1-a}} .$$

A rejection algorithm based upon this inequality has rejection constant

$$\frac{e^{(1-a)a^{\frac{a}{1-a}}}}{\Gamma(1+a)} .$$

The rejection constant has the following properties:

1. It tends to 1 as $a\downarrow 0$, or $a\uparrow 1$.
2. It is not greater than $\frac{e}{0.88560}$ for any value of $a\in(0,1]$. This can be seen by noting that $(1-a)b\leq 1-a\leq 1$ and that $\Gamma(1+a)\geq 0.8856031944\ldots$ (the gamma function at $1+a$ is absolutely bounded from below by its value at $1+a=1.4616321449\ldots$; see e.g. Abramowitz and Stegun (1970, pp. 259)).

This leads to a modified version of an algorithm of Vaduva's (1977):

Gamma generator for parameter smaller than 1

[SET-UP]

$c\leftarrow\frac{1}{a}$, $d\leftarrow a^{\frac{a}{1-a}}(1-a)$

[GENERATOR]

REPEAT

Generate iid exponential random variates Z,E. Set $X\leftarrow Z^c$ (X is Weibull (a)).

UNTIL $Z+E\leq d+X$

RETURN X ■

3.5. Johnk's theorem and its implications.

Random variate generation for the case $a < 1$ can be based upon a special property of the beta and gamma distributions. This property is usually attributed to Johnk (1964), and has later been rediscovered by others (Newman and Odell, 1971; Whittaker, 1974). We have:

Theorem 3.4. (Johnk, 1964)

Let $a, b > 0$ be given constants, and let U, V be iid uniform [0,1] random variables. Then, conditioned on $U^{\frac{1}{a}} + V^{\frac{1}{b}} \leq 1$, the random variable

$$\frac{U^{\frac{1}{a}}}{U^{\frac{1}{a}} + V^{\frac{1}{b}}}$$

is beta (a, b) distributed.

Theorem 3.5. (Berman, 1971)

Let $a, b > 0$ be given constants, and let U, V be iid uniform [0,1] random variables. Then, conditioned on $U^{\frac{1}{a}} + V^{\frac{1}{b}} \leq 1$, the random variable

$$U^{\frac{1}{a}}$$

is beta $(a, b+1)$ distributed.

Proof of Theorems 3.4 and 3.5.

Note that $X = U^{\frac{1}{a}}$ has distribution function x^a on [0,1]. The density is ax^{a-1}. Thus, the joint density of X and $Y = V^{\frac{1}{b}}$ is

$$f(x, y) = bx^{a-1}y^{b-1} \quad (0 \leq x, y \leq 1) .$$

Consider the transformation $z = x + y, t = \frac{x}{x+y}$ with inverse $x = tz, y = (1-t)z$. This transformation has Jacobian

$$\begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} z & t \\ -z & 1-t \end{vmatrix} = |z| .$$

The joint density of $(Z,T)=(X+Y,\frac{X}{X+Y})$ is

$$|z| f(tz,(1-t)z) = zab(tz)^{a-1}((1-t)z)^{b-1} \quad (0\leq tz,(1-t)z\leq 1)$$
$$= abt^{a-1}(1-t)^{b-1}z^{a+b-1} \quad (0\leq tz,(1-t)z\leq 1) .$$

The region in the (z,t) plane on which this density is nonzero is $A=\{(z,t):t>0,0<z<\min(\frac{1}{t},\frac{1}{1-t})\}$. Let A_t be the collection of values z for which $0<z<\min(\frac{1}{t},\frac{1}{1-t})$. Then, writing $g(z,t)$ for the joint density of (Z,T) at (z,t), we see that the density of T conditional on $Z\leq 1$ is given by

$$\frac{\int\limits_{z\leq 1,z\in A_t} g(z,t)dz}{\int\limits_{A} g(z,t)dz\ dt}$$
$$= \frac{1}{c}\frac{ab}{a+b}t^{a-1}(1-t)^{b-1}$$

where $c=\int\limits_{A} g(z,t)dz\ dt$ is a normalization constant. Clearly,

$$c = P(X+Y\leq 1) = \frac{ab}{a+b}\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{\Gamma(a+1)\Gamma(b+1)}{\Gamma(a+b+1)} .$$

This concludes the proof of Theorem 3.4.

For Berman's theorem, consider the transformation $x=x,z=x+y$ with inverse $x=x,y=z-x$. The joint density of (X,Z) is $f(x,z-x)=abx^{a-1}(z-x)^{b-1}I_B(x,z)$ where B is the set of (z,x) satisfying $0<x<1,0<x<z<x+1$. This is a parallelepid in the (z,x) plane. The density of X conditional on $Z<1$ is equal to a constant times

$$\int\limits_{x<z<1} abx^{a-1}(z-x)^{b-1}\,dz = ax^{a-1}(1-x)^{b} .$$

This concludes the proof of Theorem 3.5. ■

These theorems provide us with recipes for generating gamma and beta variates. For gamma random variates , we observe that YZ is gamma (a) distributed when Y is beta $(a,1-a)$ and Z is gamma (1) (i.e. exponential), or when Y is beta $(a,2-a)$ and Z is gamma (2). Summarizing all of this, we have:

Johnk's beta generator

REPEAT

Generate iid uniform [0,1] random variates U, V.

$X \leftarrow U^{\frac{1}{a}}, Y \leftarrow V^{\frac{1}{b}}$

UNTIL $X+Y \leq 1$

RETURN $\frac{X}{X+Y}$ (X is beta (a, b) distributed)

Berman's beta generator

REPEAT

Generate iid uniform [0,1] random variates U, V.

$X \leftarrow U^{\frac{1}{a}}, Y \leftarrow V^{\frac{1}{b}}$

UNTIL $X+Y \leq 1$

RETURN X (X is beta ($a, b+1$) distributed)

Johnk's gamma generator

REPEAT

Generate iid uniform [0,1] random variates U, V.

$X \leftarrow U^{\frac{1}{a}}, Y \leftarrow V^{\frac{1}{1-a}}$

UNTIL $X+Y \leq 1$

Generate an exponential random variate E.

RETURN $\frac{EX}{X+Y}$ (X is gamma (a) distributed)

Berman's gamma generator

REPEAT

Generate iid uniform [0,1] random variates U,V.

$X \leftarrow U^{\frac{1}{a}}, Y \leftarrow V^{\frac{1}{1-a}}$

UNTIL $X+Y \leq 1$

Generate a gamma (2) random variate Z (either as the sum of two iid exponential random variates or as $-\log(U^*V^*)$ where U^*, V^* are iid uniform [0,1] random variates).

RETURN ZX (X is gamma (a) distributed)

Both beta generators require on the average

$$\frac{1}{P(X+Y \leq 1)} = \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b+1)}$$

iterations, and this increases rapidly with a and b. It is however uniformly bounded over all a,b with $0<a,b \leq 1$. The two gamma generators should only be used for $a \leq 1$. The expected number of iterations is in both cases

$$\frac{1}{\Gamma(1+a)\Gamma(2-a)} .$$

It is known that $\Gamma(a)\Gamma(1-a) = \pi/\sin(\pi a)$. Thus, the expected number of iterations is

$$\frac{\sin \pi a}{\pi a(1-a)} ,$$

which is a symmetric function of a around $\frac{1}{2}$ taking the value 1 near both endpoints ($a \downarrow 0$, $a=1$), and peaking at the point $a=\frac{1}{2}$: thus, the rejection constant does not exceed $\frac{4}{\pi}$ for any $a \in (0,1]$.

3.6. Gamma variate generators when $a \leq 1$.

We can now summarize the available algorithms for gamma (a) random variate generation when the parameter is less than one. The fact that there is an infinite peak eliminates other time-honored approaches (such as the ratio-of-uniforms method) from contention. We have:

1. Rejection from the Weibull density (Vaduva, 1977): see section IX.3.7.

2. The Johnk and Berman algorithms (Johnk, 1971; Berman, 1971): see section IX.3.8.
3. The generator based upon Stuart's theorem (see section IV.6.4): $G_{a+1}U^{\frac{1}{a}}$ is gamma (a) distributed when G_{a+1} is gamma ($a+1$) distributed, and U is uniformly distributed on [0,1]. For G_{a+1} use an efficient gamma generator with parameter greater than unity.
4. The Forsythe-von Neumann method (see section IV.2.4).
5. The composition/rejection method, with rejection from an exponential density on $[1,\infty)$, and from a polynomial density on [0,1]. See sections IV.2.5 and II.3.3 for various pieces of the algorithm mainly due to Vaduva (1977). See also algorithm GS of Ahrens and Dieter (1974) and its modification by Best (1983) developed in the exercise section.
6. The transformation of an EPD variate obtained by the rejection method of section VII.2.6.

All of these algorithms are uniformly fast over the parameter range. Comparative timings vary from experiment to experiment. Tadikamalla and Johnson (1981) report good results with algorithm GS but fail to include some of the other algorithms in their comparison. The algorithms of Johnk and Berman are probably better suited for beta random variate generation because two expensive powers of uniform random variates are needed in every iteration. The Forsythe-von Neumann method seems also less efficient time-wise. This leaves us with approaches 1,3,5 and 6. If a very efficient gamma generator is available for $a>1$, then method 3 could be as fast as algorithm GS, or Vaduva's Weibull-based rejection method. Methods 1 and 6 are probably comparable in all respects, although the rejection constant of method 6 certainly is superior.

3.7. The tail of the gamma density.

As for the normal density, it is worthwhile to have a good generator for the tail gamma (a) density truncated at t. It is only natural to look at dominating densities of the form $be^{b(t-x)}$ ($x \geq t$). The parameter b has to be picked as a function of a and t. Note that a random variate with this density can be generated as $t+\frac{E}{b}$ where E is an exponential random variate. We consider the cases $a<1$ and $a \geq 1$ separately. We can take $b=1$ because the gamma density decreases faster than e^{-x}. Therefore, rejection can be based upon the inequality

$$x^{a-1}e^{-x} \leq t^{a-1}e^{-x} \quad (x \geq t) .$$

It is easily seen that the corresponding algorithm is

REPEAT

Generate a uniform random variate U and an exponential random variate E. Set $X \leftarrow t+E$

UNTIL $XU^{\frac{1}{1-a}} \leq a$

RETURN X (X has the gamma density restricted to $[t,\infty)$)

The efficiency of the algorithm is given by the ratio of the integrals of the two functions. This gives

$$\frac{t^{a-1}e^{-t}}{\int_t^\infty x^{a-1}e^{-x}\,dx}$$

$$= \frac{1}{\int_t^\infty \left(\frac{x}{t}\right)^{a-1} e^{t-x}\,dx}$$

$$= \frac{1}{\int_0^\infty \left(1+\frac{x}{t}\right)^{a-1} e^{-x}\,dx}$$

$$\leq \frac{1}{\int_0^\infty e^{x\left(\frac{a-1}{t}-1\right)}\,dx}$$

$$= 1+\frac{1-a}{t}$$

$$\rightarrow 1 \text{ as } t \rightarrow \infty .$$

When $a \geq 1$, the exponential with parameter 1 does not suffice because of the polynomial portion in the gamma density. It is necessary to take a slightly slower decreasing exponential density. The inequality that we will use is

$$\left(\frac{x}{t}\right)^{a-1} \leq e^{(a-1)\left(\frac{x}{t}-1\right)}$$

which is easily established by standard optimization methods. This suggests the choice $b=1-\frac{a-1}{t}$ in the exponential curve. Thus, we have

$$x^{a-1}e^{-x} \leq t^{a-1}e^{(a-1)\left(\frac{x}{t}-1\right)-x} .$$

Based on this, the rejection algorithm becomes

REPEAT

Generate two iid exponential random variates E, E^*.

$X \leftarrow t + \frac{E}{1-\frac{a-1}{t}}$

UNTIL $\frac{X}{t}-1+\log(\frac{t}{X}) \leq \frac{E^*}{a-1}$

RETURN X (X has the gamma (a) density restricted to $[t,\infty)$.)

The algorithm is valid for all $a>1$ and all $t>a-1$ (the latter condition states that the tail should not include the mode of the gamma density). A squeeze step can be included by noting that $\log(\frac{X}{t})=\log(1+\frac{X-t}{t}) \geq 2\frac{X-t}{X+t}=\frac{2E}{(1-\frac{a-1}{t})(X+t)}$. Here we used the inequality $\log(1+u) \geq 2u/(u+2)$. Thus, the quick acceptance step to be inserted in the algorithm is

IF $\frac{E^2}{(1-\frac{a-1}{t})^2 t(X+t)} \leq \frac{E^*}{a-1}$ THEN RETURN X

We conclude this section by showing that the rejection constant is asymptotically optimal as $t \uparrow \infty$: the ratio of the integrals of the two functions involved is

$$\frac{t^{a-1}e^{-t}}{(1-\frac{a-1}{t})\int_t^\infty x^{a-1}e^{-x}\,dx}$$

$$=\frac{1}{(1-\frac{a-1}{t})\int_0^\infty (1+\frac{x}{t})^{a-1} e^{-x}\,dx}$$

which once again tends to 1 as $t \rightarrow \infty$. We note here that the algorithms given in this section are due to Devroye (1980). The algorithm for the case $a>1$ can be slightly improved at the expense of more complicated design parameters. This possibility is explored in the exercises.

3.8. Stacy's generalized gamma distribution.

Stacy (1962) introduced the generalized gamma distribution with two shape parameters, $c, a > 0$: the density is

$$f(x) = \frac{c}{\Gamma(a)} x^{ca-1} e^{-x^c} \quad (x \geq 0) .$$

This family of densities includes the gamma densities ($c=1$), the halfnormal density ($a=\frac{1}{2}, c=2$) and the Weibull densities ($a=1$). Because of the flexibility of having two shape parameters, this distribution has been used quite often in modeling stochastic inputs. Random variate generation is no problem because we observe that $G_a^{\frac{1}{c}}$ has the said distribution where G_a is a gamma (a) random variable.

Tadikamalla (1979) has developed a rejection algorithm for the case $a > 1$ which uses as a dominating density the Burr XII density used by Cheng in his algorithm GB. The parameters μ,λ of the Burr XII density are $\lambda = c\sqrt{2a-1}$, $\mu = a^{\sqrt{2a-1}}$. The rejection constant is a function of a only. The algorithm is virtually equivalent to generating G_a by Cheng's algorithm GB and returning $G_a^{\frac{1}{c}}$ (which explains why the rejection constant does not depend upon c).

3.9. Exercises.

1. Show Kullback's result (Kullback, 1934) which states that when X_1,X_2 are independent gamma (a) and gamma ($a+\frac{1}{2}$) random variables, then $2\sqrt{X_1 X_2}$ is gamma ($2a$).
2. Prove Stuart's theorem (the second statement of Theorem 3.1): if Y is gamma (a) and Z is beta ($b, a-b$) for some $b > a > 0$, then YZ and $Y(1-Z)$ are independent gamma (b) and gamma ($a-b$) random variables.
3. **Algorithm GO (Ahrens and Dieter, 1974).** Define the breakpoint $b = a-1+\sqrt{6(a+\sqrt{\frac{8a}{3}})}$. Find the smallest exponentially decreasing function dominating the gamma (a) density to the right of b. Find a normal curve centered at $a-1$ dominating the gamma density to the left of b, which has the property that the area under the dominating curve divided by the area under the leftmost piece of the gamma density tends to a constant as $a \uparrow \infty$. Also, find the similarly defined asymptotic ratio for the rightmost

piece, and establish that it is greater than 1. By combining this, obtain an expression for the limit value of the rejection constant. Having established the bounds, give a rejection method for generating a random variate with the gamma density. Find efficient squeeze steps if possible.

4. **The Weibull density.** Prove the following properties of the Weibull (a) distribution:

   A. For $a \geq 1$, the density is unimodal with mode at $(1-\frac{1}{a})^{\frac{1}{a}}$. The position of the mode tends to 1 as $a \uparrow \infty$.

   B. The value of the distribution function at $x=1$ is $1-\frac{1}{e}$ for all values of a.

   C. The r-th moment is $\Gamma(1+\frac{r}{a})$.

   D. The minimum of n iid Weibull random variables is distributed as a constant times a Weibull random variable. Determine the constant and the parameter of the latter random variable.

   E. As $a \uparrow \infty$, the first moment of the Weibull distribution varies as $1-\frac{\gamma}{a}+o(\frac{1}{a})$ where γ=0.57722... is Euler's constant. Also, the variance $\sim \pi^2/6a^2$.

5. Obtain a good uniform upper bound for the rejection constant in Vaduva's algorithm for gamma random variates when $a \leq 1$ which is based upon rejection from the Weibull density.

6. **Algorithm GS (Ahrens and Dieter, 1974).** The following algorithm was proposed by Ahrens and Dieter (1974) for generating gamma (a) random variates when the parameter a is ≤ 1:

Rejection algorithm GS for gamma variates (Ahrens and Dieter, 1974)

[SET-UP]

$b \leftarrow \frac{e+a}{e}$, $c \leftarrow \frac{1}{a}$

[GENERATOR]

REPEAT

Generate iid uniform [0,1] random variates U,W. Set $V \leftarrow bU$.

IF $V \leq 1$

THEN

$X \leftarrow V^c$

Accept $\leftarrow [W \leq e^{-X}]$

ELSE

$X \leftarrow -\log(c(b-V))$

Accept $\leftarrow [W \leq X^{a-1}]$

UNTIL Accept

RETURN X

The algorithm is based upon the inequalities: $f(x) \leq \frac{a}{\Gamma(1+a)} x^{a-1}$ $(0 \leq x \leq 1)$ and $f(x) \leq \frac{a}{\Gamma(1+a)} e^{-x}$ $(x \geq 1)$. Show that the rejection constant is $\frac{e+a}{e\,\Gamma(1+a)}$. Show that the rejection constant approaches 1 as $a \downarrow 0$, that it is $1+\frac{1}{e}$ at $a=1$, and that it is uniformly bounded over $a \in (0,1]$ by a number not exceeding $\frac{3}{2}$. Show that in sampling from the composite dominating density, we have probability weights $\frac{e}{e+a}$ for ax^{a-1} $(0<x \leq 1)$, and $\frac{a}{e+a}$ for e^{1-x} $(x \geq 1)$ respectively.

7. Show that the exponential function of the form ce^{-bx} $(x \geq t)$ of smallest integral dominating the gamma (a) density on $[t,\infty)$ (for $a>1$, $t>0$) has parameter b given by

$$b = \frac{t-a+\sqrt{(t-a)^2+4t}}{2t} .$$

Hint: show first that the ratio of the gamma density over e^{-bx} reaches a peak at $x=\frac{a-1}{1-b}$ (which is to the right of t when $b \geq 1-\frac{a-1}{t}$). Then compute the optimal b and verify that $b \geq 1-\frac{a-1}{t}$. Give the algorithm for the tail of the gamma density that corresponds to this optimal inequality. Show furthermore that as $t \uparrow \infty$, $b=1-\frac{a-1}{t}+o\left(\frac{1}{t}\right)$, which proves that the choice of b in the text is asymptotically optimal (Dagpunar, 1978).

8. **Algorithm RGS (Best, 1983).** Algorithm GS (of exercise 6) can be optimized by two devices: first, the gamma density f with parameter a can be maximized by a function which is $x^{a-1}/\Gamma(a)$ on $[0,t]$ and $t^{a-1}e^{-x}/\Gamma(a)$ on $[t,\infty)$, where t is a breakpoint. In algorithm GS, the breakpoint was chosen as $t=1$. Secondly, a squeeze step can be added.

   A. Show that the optimal breakpoint (in terms of minimization of the area under the dominating curve) is given by the solution of the transcendental equation $t=e^{-t}(1-a+t)$. (Best approximates this solution by $0.07+0.75\sqrt{1-a}$.)

   B. Prove the inequalities $e^{-x}\geq(2-x)/(2+x)$ $(x\geq 0)$ and $(1+x)^{-c}\geq 1/(1+cx)$ $(x\geq 0, 1\geq c\geq 0)$. (These are needed for the squeeze steps.)

   C. Show that the algorithm given below is valid:

Algorithm RGS for gamma variates (Best, 1983)

[SET-UP]

$t\leftarrow 0.07+0.75\sqrt{1-a}$, $b\leftarrow 1+\frac{e^{-t}a}{t}$, $c\leftarrow\frac{1}{a}$

[GENERATOR]

REPEAT

  Generate iid uniform [0,1] random variates U,W. Set $V\leftarrow bU$.

  IF $V\leq 1$

    THEN

      $X\leftarrow tV^c$

      Accept $\leftarrow[W\leq\frac{2-X}{2+X}]$

      IF NOT Accept THEN Accept $\leftarrow[W\leq e^{-X}]$

    ELSE

      $X\leftarrow -\log(ct(b-V))$, $Y\leftarrow\frac{X}{t}$

      Accept $\leftarrow[W(a+Y-aY)\leq 1]$

      IF NOT Accept THEN Accept $\leftarrow[W\leq Y^{a-1}]$

UNTIL Accept

RETURN X

9. **Algorithm G4PE (Schmeiser and Lal, 1980).** The graph of the gamma density can be covered by a collection of rectangles, triangles and exponential curves having the properties that (i) all parameters involved are easy to compute; and (ii) the total area under the dominating curve is uniformly bounded over $a\geq 1$. One such proposal is due to Schmeiser and Lal (1980): define five breakpoints,

$$t_3 = a-1$$
$$t_4 = t_3+\sqrt{t_3}$$
$$t_5 = t_4(1+1/(t_4-t_3))$$
$$t_2 = \max(0, t_3-\sqrt{t_3})$$
$$t_1 = t_2(1-1/(t_3-t_2))$$

where t_3 is the mode, and t_2, t_4 are the points of inflection of the gamma density. Furthermore, t_1, t_5 are the points at which the tangents of f at t_2 and t_4 cross the x-axis. The dominating curve has five pieces: an exponential tail on $(-\infty, t_1]$ with parameter $1-t_3/t_1$ and touching f at t_1. On $[t_5, \infty)$ we have a similar exponential dominating curve with parameter $1-t_3/t_5$. On $[t_1, t_2]$ and $[t_4, t_5]$, we have a linear dominating curve touching the density at the breakpoints. Finally, we have a constant piece of height $f(t_3)$ on $[t_2, t_4]$. All the strips except the two tail sections are partitioned into a rectangle (the largest rectangle fitted under the curve of f) and a leftover piece. This gives ten pieces, of which four are rectangles totally tucked under the gamma density. For the six remaining pieces, we can construct very simple linear acceptance steps.

A. Develop the algorithm.

B. Compute the area under the dominating curve, and determine its asymptotic value.

C. Determine the asymptotic probability that we need only one uniform random variate (the random variate needed to select one of the four rectangles is recycled). This is equivalent to computing the asymptotic area under the four rectangles.

D. With all the squeeze steps defined above in place, compute the asymptotic value of the expected number of evaluations of f .

Hint: obtain the values for an appropriately transformed normal density and use the convergence of the gamma density to the normal density.

10. **The t-distribution.** Show that when $G_{1/2}, G_{a/2}$ are independent gamma random variables, then $\sqrt{aG_{1/2}/G_{a/2}}$ is distributed as the absolute value of a random variable having the t distribution with a degrees of freedom. (Recall that the t density is

$$f(x) = \frac{\Gamma(\frac{a+1}{2})}{\sqrt{\pi a}\,\Gamma(\frac{a}{2})(1+\frac{x^2}{2})^{\frac{a+1}{2}}} \text{ .)}$$

In particular, if G, G^* are iid gamma $(\frac{1}{2})$ random variables, then $\sqrt{G/G^*}$ is Cauchy distributed.

11. **The Pearson VI distribution.** Show that G_a/G_b has density

$$f(x) = \frac{x^{a-1}}{B_{a,b}(1+x)^{b-1}} \qquad (x \geq 0)$$

when G_a, G_b are independent gamma random variables with parameters a and b respectively. Here $B_{a,b} = \Gamma(a)\,\Gamma(b)/\Gamma(a+b)$ is a normalization constant. The density in question is the Pearson VI density. It is also called the beta density of the second kind with parameters a and b. b/a times the random variable in question is also called an F distributed random variable with $2a$ and $2b$ degrees of freedom.

4. THE BETA DENSITY.

4.1. Properties of the beta density.

We say that a random variable X on [0,1] is **beta** (a,b) distributed when it has density

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B_{a,b}} \qquad (0 \le x \le 1)$$

where $a,b > 0$ are shape parameters, and

$$B_{a,b} = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

is a normalization constant. The density can take a number of interesting shapes:

1. When $0 < a,b < 1$, the density is U-shaped with infinite peaks at 0 and 1.
2. When $0 < a < 1 \le b$, the density is said to be J-shaped: it has an infinite peak at 0 and decreases monotonically to a positive constant (when $b=1$) or to 0 (when $b>1$).
3. When $a = 1 < b$, the density is bounded and decreases monotonically to 0.
4. When $a = b = 1$, we have the uniform [0,1] density.
5. When $1 < a,b$, the density is unimodal, and takes the value 0 at the endpoints.

The fact that there are two shape parameters makes the beta density a solid candidate for illustrating the various techniques of nonuniform random variate generation. It is important for the design to understand the basic properties. For example, when $a,b > 1$, the mode is located at $\frac{a-1}{a+b-2}$. It is also quite trivial to show that for $r > -a$,

$$E(X^r) = \frac{B_{a+r,b}}{B_{a,b}} .$$

In particular, $E(X) = \frac{a}{a+b}$ and $Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$. There are a number of relationships with other distributions. These are summarized in Theorem 4.1:

Theorem 4.1.

This is about the relationships between the beta (a,b) density and other densities.

A. Relationship with the **gamma** density: if G_a, G_b are independent gamma (a), gamma (b) random variables, then $\frac{G_a}{G_a+G_b}$ is beta (a,b) distributed.

B. Relationship with the **Pearson VI** (or β_2) density: if X is beta (a,b), then $Y=\frac{X}{1-X}$ is $\beta_2(a,b)$, that is, Y is a beta of the second kind, with density $f(x)=\frac{x^{a-1}}{B_{a,b}(1+x)^{a+b}}$ $(x \geq 0)$.

C. Relationship with the **(Student's) t** distribution: if X is beta $(\frac{1}{2},\frac{a}{2})$, and S is a random sign, then $S\sqrt{\frac{aX}{1-X}}$ is t-distributed with a degrees of freedom, i.e. it has density

$$f(x)=\frac{\Gamma(\frac{a+1}{2})}{\sqrt{\pi a}\,\Gamma(\frac{a}{2})(1+\frac{x^2}{a})^{\frac{a+1}{2}}} .$$

By the previous property, note that $\sqrt{aY}$ is t-distributed with parameter a when Y is $\beta_2(a,b)$. Furthermore, if X denotes a beta (a,a) random variable, and T denotes a t random variable with $2a$ degrees of freedom, then we have the following equality in distribution: $X=\frac{1}{2}+\frac{1}{2}\frac{T}{\sqrt{2b+T^2}}$, or $T=\frac{\sqrt{2a}\,(2X-1)}{2\sqrt{X-X^2}}$. In particular, when U is uniform on $[0,1]$, then $\frac{\sqrt{2}(U-\frac{1}{2})}{\sqrt{U-U^2}}$ is t with 2 degrees of freedom.

D. Relationship with the **F (Snedecor)** distribution: when X is beta (a,b), then $\frac{bX}{a(1-X)}$ is F-distributed with a and b degrees of freedom, i.e. it has density $\frac{a}{b}f(\frac{ax}{b})$ $(x>0)$, where f is the $\beta_2(\frac{a}{2},\frac{b}{2})$ density.

E. Relationship with the **Cauchy** density: when X is beta $(\frac{1}{2},\frac{1}{2})$ distributed (this is called the arc sine distribution), then $\sqrt{\frac{X}{1-X}}$ is distributed as the absolute value of a Cauchy random variable.

Proof of Theorem 4.1.

All the properties can be obtained by applying the methods for computing densities of transformed random variables explained for example in section I.4.1. ■

We should also mention the important connection between the beta distribution and order statistics. When $0 < U_{(1)} < \cdots < U_{(n)}$ are the order statistics of a uniform [0,1] random sample, then $U_{(k)}$ is beta $(k, n-k+1)$ distributed. See section I.4.3.

4.2. Overview of beta generators.

Beta variates can be generated by exploiting special properties of the distribution. The order statistics method, applicable only when both a and b are integer, proceeds as follows:

Order statistics method for beta variates

Generate $a+b-1$ iid uniform [0,1] random variates.
Find the a-th order statistic X (a-th smallest) among these variates.
RETURN X

This method, mentioned as early as 1963 by Fox, requires time at least proportional to $a+b-1$. If standard sorting routines are used to obtain the a-th smallest element, then the time complexity is even worse, possibly $\Omega((a+b-1)\log(a+b-1))$. There are obvious improvements: it is wasteful to sort a sample just to obtain the a-th smallest number. First of all, via linear selection algorithms we can find the a-th smallest in worst case time $O(a+b-1)$ (see e.g. Blum, Floyd, Pratt, Rivest and Tarjan (1973) or Schonhage, Paterson and Pippenger (1976)). But in fact, there is no need to generate the entire sample. The uniform sample can be generated directly from left to right or right to left, as shown in section V.3. This would reduce the time to $O(\min(a,b))$. Except in special applications, not requiring non-integer or large parameters, this method is not recommended.

When property A of Theorem 4.1 is used, the time needed for one beta variate is about equal to the time required to generate two gamma variates. This method is usually very competitive because there are many fast gamma generators. In any case, if the gamma generator is uniformly fast, so will be the beta generator. Formally we have:

Beta variates via gamma variates

Generate two independent gamma random variates, G_a and G_b.

RETURN $\frac{G_a}{G_a+G_b}$

Roughly speaking, we will be able to improve over this generator by at most 50%. There is no need to discuss beta variate generators which are not time efficient. A survey of pre-1972 methods can be found in Arnason (1972). None of the methods given there has uniformly bounded expected time. Among the competitive approaches, we mention:

A. Standard rejection methods. For example, we have:

 Rejection from the Burr XII density (Cheng, 1978).

 Rejection from the normal density (Ahrens and Dieter, 1974).

 Rejection from polynomial densities (Atkinson and Whittaker, 1976, 1979; Atkinson, 1979).

 Rejection and composition with triangles, rectangles, and exponential curves (Schmeiser and Babu, 1980).

 The best of these methods will be developed below. In particular, we will highlight Cheng's uniformly fast algorithms. The algorithm of Schmeiser and Babu (1980), which is uniformly fast over $a,b \geq 1$, is discussed in section VII.2.6.

B. Forsythe's method, as applied for example by Atkinson and Pearce (1976). This method requires a lot of code and the set-up time is considerable. In comparison with this investment, the speed obtainable via this approach is disappointing.

C. Johnk's method (Johnk, 1964) and its modifications. This too should be considered as a method based upon special properties of the beta density. The expected time is not uniformly bounded in the parameters. It should be used only when both parameters are less than one. See section IX.3.5.

D. Universal algorithms. The beta density is unimodal when both parameters are at least one, and it is monotone when one parameter is less than one and one is at least equal to one. Thus, the universal methods of section VII.3.2 are applicable. At the very least, the inequalities derived in that section can be used to design good (albeit not superb) bounds for the beta density. In any case, the expected time is provably uniform over all parameters a,b with $\max(a,b) \geq 1$.

E. Strip table methods, as developed in section VIII.2.2. We will study below how many strips should be selected as a function of a and b in order to have uniformly bounded expected generation times.

The bottom line is that the choice of a method depends upon the user: if he is not willing to invest a lot of time, he should use the ratio of gamma variates. If he does not mind coding short programs, and a and/or b vary frequently, one of the rejection methods based upon analysis of the beta density or upon universal inequalities can be used. The method of Cheng is very robust. For special cases, such as symmetric beta densities, rejection from the normal density is very competitive. If the user does not foresee frequent changes in a and b, a strip table method or the algorithm of Schmeiser and Babu (1980) are recommended. Finally, when both parameters are smaller than one, it is possible to use rejection from polynomial densities or to apply Johnk's method.

4.3. The symmetric beta density.

In this section, we will take a close look at one of the simplest special cases, the symmetric beta density with parameter a:

$$f(x) = \frac{\Gamma(2a)}{\Gamma^2(a)}(x(1-x))^{a-1} = C(x(1-x))^{a-1} \quad (0 \le x \le 1) .$$

For large values of a, this density is quite close to the normal density. To see this, consider $y = x - \frac{1}{2}$, and

$$\log(f(x)) = \log(C) + (a-1)\log(1+2y) + (a-1)\log(1-2y) - (a-1)\log 4$$
$$= \log(C) - (a-1)\log 4 + (a-1)\log(1-4y^2) .$$

The last term on the right hand side is not greater than $-4(a-1)y^2$, and it is at least equal to $-4(a-1)y^2 - 16(a-1)y^4/(1-4y^2)$. Thus, $\log(f(\frac{1}{2}+\frac{x}{\sqrt{8(a-1)}}))$ tends to $-\log(\sqrt{2\pi}) - \frac{x^2}{2}$ as $a \to \infty$ for all $x \in R$. Here we used Stirling's formula to prove that $\log(C) - (a-1)\log 4$ tends to $-\log(\sqrt{2\pi})$. Thus, if X is beta (a,a), then the density of $\sqrt{8(a-1)}(X - \frac{1}{2})$ tends to the standard normal density as $a \to \infty$. The only hope for an asymptotically optimal rejection constant in a rejection algorithm is to use a dominating density which is either normal or tends pointwise to the normal density as $a \to \infty$. The question is whether we should use the normalization suggested by the limit theorem stated above. It turns out that the best rejection constant is obtained not by taking $8(a-1)$ in the formula for the normal density, but $8(a-\frac{1}{2})$. We state the algorithm first, then announce its properties in a theorem:

Symmetric beta generator via rejection from the normal density

[NOTE: $b=(a-1)\log(1+\frac{1}{2a-2})-\frac{1}{2}$.]

[GENERATOR]

REPEAT

  REPEAT

    Generate a normal random variate N and an exponential random variate E.

    $X \leftarrow \frac{1}{2}+\frac{N}{\sqrt{8a-4}}, Z \leftarrow N^2$

  UNTIL $Z<2a-1$ (now, $X \in [0,1]$)

  Accept $\leftarrow [E+\frac{Z}{2}-\frac{(a-1)Z}{2a-1-Z}+b \geq 0]$

  IF NOT Accept THEN Accept $\leftarrow [E+\frac{Z}{2}+(a-1)\log(1-\frac{Z}{2a-1})+b \geq 0]$

UNTIL Accept

RETURN X

Theorem 4.2.

Let f be the beta (a) density with parameter $a \geq 1$. Then let $\sigma>0$ be a constant and let c_σ be the smallest constant such that for all x,

$$f(x) \leq c_\sigma \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\frac{1}{2})^2}{2\sigma^2}} .$$

Then c_σ is minimal for $\sigma^2=\frac{1}{8a-4}$, and the minimal value is

$$c_\sigma = \left(\frac{8(a-1)}{4e\,(8a-4)} \right)^{a-1} \frac{\sqrt{2\pi}}{\sqrt{8a-4}B_{a,a}} e^{\frac{8a-4}{8}} .$$

In the rejection algorithm shown above, the rejection constant is c_σ. The rejection constant is uniformly bounded for $a \in [1,\infty)$: selected values are $\sqrt{\frac{\pi e}{6}}$ at $a=2$, $\sqrt{36\pi e}$ at $a=3$. We have

$$\lim_{a \to \infty} c_\sigma = 1 .$$

and in fact, $c_\sigma \leq e^{\frac{1}{24a}+\frac{1}{2a-1}}$.

Proof of Theorem 4.2.

Let us write $g(x)$ for the normal density with mean $\frac{1}{2}$ and variance σ^2. We first determine the supremum of f/g by setting the derivative of $\log(\frac{f}{g})$ equal to zero. This yields the equation

$$(x-\frac{1}{2})(\sigma^{-2}-\frac{2(a-1)}{x(1-x)}) = 0 .$$

One can easily see from this that f/g has a local minimum at $x=\frac{1}{2}$ and two local maxima symmetrically located on either side of $\frac{1}{2}$ at $\frac{1}{2}\pm\frac{1}{2}\sqrt{1-8(a-1)\sigma^2}$. The value of f/g at the maxima is

$$c_\sigma = \left(\frac{8(a-1)\sigma^2}{4e}\right)^{a-1} \frac{\sqrt{2\pi}\sigma}{B_{a,a}} e^{\frac{1}{8\sigma^2}} .$$

This depends upon σ as follows: $\sigma^{2a-1}e^{\frac{1}{8\sigma^2}}$. This has a unique minimum at $\sigma=1/\sqrt{8a-4}$. Resubstitution of this value gives

$$c_\sigma = \left(\frac{a-1}{4a-2}\right)^{a-1} \frac{\sqrt{2\pi}}{\sqrt{8a-4}B_{a,a}} e^{\frac{1}{2}} .$$

By well-known bounds on the gamma function (Whittaker ans Watson, 1927, p. 253), we have

$$\frac{1}{B_{a,a}} \leq 4^{a-\frac{1}{2}}\sqrt{\frac{a}{\pi}}e^{\frac{1}{24a}} ,$$

$$\frac{1}{B_{a,a}} \sim 4^{a-\frac{1}{2}}\sqrt{\frac{a}{\pi}}$$

as $a\to\infty$. Thus,

$$c_\sigma \leq \left(\frac{a-1}{4a-2}\right)^{a-1} \frac{\sqrt{2\pi}}{\sqrt{8a-4}} 4^{a-\frac{1}{2}}\sqrt{\frac{a}{\pi}}e^{\frac{1}{24a}}e^{\frac{1}{2}}$$

$$= \sqrt{ae/(a-\frac{1}{2})}e^{\frac{1}{24a}}(1-\frac{1}{2a-1})^{a-1}$$

$$\leq \sqrt{ae/(a-\frac{1}{2})}e^{\frac{1}{24a}}e^{-\frac{a-1}{2a-1}}$$

$$= \sqrt{1+\frac{1}{2a-1}}e^{\frac{1}{24a}+\frac{1}{4a-2}}$$

$$\leq e^{\frac{1}{24a}+\frac{1}{2a-1}} . \blacksquare$$

The algorithm shown above is applicable for all $a \geq 1$. For large values of a, we need about one normal random variate per beta random variate, and the probability that the long acceptance condition has to be verified at all tends to 0 as $a \to \infty$ (exercise 4.1). There is another school of thought, in which normal random variates are avoided altogether, and the algorithms are phrased in terms of uniform random variates. After all, normal random variates are also built from uniform random variates. In the search for a good dominating curve, help can be obtained from other symmetric unimodal long-tailed distributions. There are two examples that have been explicitly mentioned in the literature, one by Best (1978), and one by Ulrich (1984):

Theorem 4.3.

When Y is a t distributed random variable with parameter $2a$, then $X \leftarrow \frac{1}{2}+\frac{1}{2}\frac{Y}{\sqrt{2a+Y^2}}$ is beta (a,a) distributed (Best, 1978).

When U,V are independent uniform [0,1] random variables, then

$$X \leftarrow \frac{1}{2}+\frac{1}{2}\sin(2\pi V)\sqrt{1-U^{\frac{2}{2a-1}}}$$

is beta (a,a) distributed (Ulrich, 1984).

Proof of Theorem 4.3.

The proof is left as an exercise on transformations of random variables. ∎

If we follow Best, then we need a fast t generator, and we refer to section IX.5 for such algorithms. Ulrich's suggestion is intriguing because it is reminiscent of the polar method. Recall that when X,Y is uniformly distributed in the unit circle with $S=X^2+Y^2$, then $(\frac{X}{\sqrt{S}},\frac{Y}{\sqrt{S}})$ and S are independent, and S is uniformly distributed on [0,1]. Also, switching to polar coordinates (R,Θ), we see that $XY/S=\cos(\Theta)\sin(\Theta)=2\sin(2\Theta)$. Thus, since 2Θ is uniformly distributed on $[0,4\pi]$, we see that the random variable

$$\frac{1}{2}+\frac{XY}{S}\sqrt{1-S^{\frac{2}{2a-1}}}$$

has a beta (a,a) distribution. We summarize:

Ulrich's polar method for symmetric beta random variates

REPEAT

Generate U uniformly on $[0,1]$ and V uniformly on $[-1,1]$.

$S \leftarrow U^2+V^2$

UNTIL $S \leq 1$

RETURN $X \leftarrow \frac{1}{2}+\frac{UV}{S}\sqrt{1-S^{\frac{2}{2a-1}}}$

It should be stressed that Ulrich's method is valid for all $a>0$, provided that for the case $a=1/2$, we obtain X as $1/2+UV/S$, that is, X is distributed as a linearly transformed arc sin random variable. Despite the power and the square root needed in the algorithm for general a, its elegance and generality make it a formidable candidate for inclusion in computer libraries.

4.4. Uniformly fast rejection algorithms.

The beta (a,b) density has two shape parameters. If we are to construct a uniformly fast rejection algorithm, it seems unlikely that we can just consider rejection from a density with no shape parameter such as the normal density. This is generally speaking only feasible when there is one shape parameter as in the case of the gamma or symmetric beta families. The trick will then be to find a flexible family of easy dominating densities. In his work, Cheng has repeatedly used the Burr XII density with one scale parameter and one shape parameter with a great deal of success. This density is constructed as follows. If U is uniformly distributed on $[0,1]$, then $\frac{U}{1-U}$ has density $(1+x)^{-2}$ on $[0,\infty)$. For $\mu,\lambda>0$, the density of

$$(\mu\frac{U}{1-U})^{\frac{1}{\lambda}}$$

is

$$g(x)=\frac{\lambda\mu x^{\lambda-1}}{(\mu+x^\lambda)^2} \quad (x>0) .$$

This is an infinite-tailed density, of little direct use for the beta density. Fortunately, beta and β_2 random variables are closely related (see Theorem 4.1), so that we need only consider the infinite-tailed β_2 density with parameters (a,b):

$$f(x)=\frac{x^{a-1}}{B_{a,b}(1+x)^{a+b}} \quad (x\geq 0) .$$

The values of μ and λ suggested by Cheng (1978) for good rejection constants are

$$\mu = \left(\frac{a}{b}\right)^{\lambda} ;$$

$$\lambda = \begin{cases} \min(a,b) & (\min(a,b) \leq 1) \\ \sqrt{\frac{2ab-(a+b)}{a+b-2}} & (\min(a,b) > 1) \end{cases} .$$

With these choices, it is not difficult to verify that f/g is maximal at $x = a/b$, and that $f \leq cg$ where

$$c = \frac{4a^a b^b}{\lambda B_{a,b} (a+b)^{a+b}} .$$

Note that $cg(x)/f(x)$ can be simplified quite a bit. The unadorned algorithm is:

Cheng's rejection algorithm BA for beta random variates (Cheng, 1978)

[SET-UP]

$s \leftarrow a+b$

IF $\min(a,b) \leq 1$

THEN $\lambda \leftarrow \min(a,b)$

ELSE $\lambda \leftarrow \sqrt{\frac{2ab-s}{s-2}}$

$u \leftarrow a+\lambda$

[GENERATOR]

REPEAT

Generate two iid uniform [0,1] random variates U_1, U_2.

$V \leftarrow \frac{1}{\lambda} \frac{U_1}{1-U_1}$, $Y \leftarrow ae^V$

UNTIL $s \log(\frac{s}{b+Y}) + uV - \log(4) \geq \log(U_1^2 U_2)$

RETURN $X \leftarrow \frac{Y}{b+Y}$

The fundamental property of Cheng's algorithm is that

$$\sup_{a,b>0} c = 4 ; \sup_{a,b \geq 1} c = \frac{4}{e} \approx 1.47 .$$

For fixed a, c is minimal when $b = a$ and increases when $b \downarrow 0$ or $b \uparrow \infty$. The details of the proofs of the various statements about this algorithm are left as an exercise. There exists an improved version of the algorithm for the case that both parameters are greater than 1 which is based upon the squeeze method (Cheng's algorithm BB). Cheng's algorithm is slowest when $\min(a,b) < 1$. In that region of the parameter space, it is worthwhile to design special algorithms that may or may not be uniformly fast over the entire parameter space.

4.5. Generators when min(a,b)$\leq$1.

Cheng's algorithm BA is robust and can be used for all values of a,b. However, when both a,b are smaller than one, and $a+b \leq 1.5$, Johnk's method is typically more efficient. When $\min(a,b)$ is very small, and $\max(a,b)$ is rather large, neither Johnk's method nor algorithm BA are particularly fast. To fill this gap, several algorithms were proposed by Atkinson and Whittaker (1976, 1979) and Atkinson (1979). In addition, Cheng (1977) developed an algorithm of his own, called algorithm BC.

Atkinson and Whittaker (1976,1979) split [0,1] into $[0,t]$ and $[t,1]$, and construct a dominating curve for use in the rejection method based upon the inequalities:

$$x^{a-1}(1-x)^{b-1} \leq \begin{cases} x^{a-1}(1-t)^{b-1} & (x \leq t) \\ t^{a-1}(1-x)^{b-1} & (x > t) \end{cases} .$$

The areas under the two pieces of the dominating curve are, respectively, $(1-t)^{b-1}\frac{t^a}{a}$ and $t^{a-1}\frac{(1-t)^b}{b}$. Thus, the following rejection algorithm can be used:

First algorithm of Atkinson and Whittaker (1976, 1979)

[SET-UP]

Choose $t \in [0,1]$.

$p \leftarrow \frac{bt}{bt+a(1-t)}$

[GENERATOR]

REPEAT

Generate a uniform [0,1] random variate U and an exponential random variate E.

IF $U \leq p$

THEN

$X \leftarrow t\left(\frac{U}{p}\right)^{\frac{1}{a}}$

Accept $\leftarrow [(1-b)\log(\frac{1-X}{1-t}) \leq E]$

ELSE

$X \leftarrow 1-(1-t)\left(\frac{1-U}{1-p}\right)^{\frac{1}{b}}$

Accept $\leftarrow [(1-a)\log(\frac{X}{t}) \leq E]$

UNTIL Accept

RETURN X

Despite its simplicity, this algorithm performs remarkably well when both parameters are less than one, although for $a+b<1$, Johnk's algorithm is still to be preferred. The explanation for this is given in the next theorem. At the same time, the best choice for t is derived in the theorem.

Theorem 4.4.

Assume that $a \leq 1, b \leq 1$. The expected number of iterations in Johnk's algorithm is

$$c = \frac{\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b+1)} .$$

The expected number of iterations ($E(N)$) in the first algorithm of Atkinson and Whittaker is

$$c \frac{bt+a(1-t)}{(a+b)t^{1-a}(1-t)^{1-b}} .$$

When $a+b \leq 1$, then for all values of t, $E(N) \geq c$. In any case, $E(N)$ is minimized for the value

$$t_{opt} = \frac{\sqrt{a(1-a)}}{\sqrt{a(1-a)}+\sqrt{b(1-b)}} .$$

With $t=t_{opt}$, we have $E(N)<c$ whenever $a+b>1$. For $a+b>1$, $t=\frac{1}{2}$, it is also true that $E(N)<c$.

Finally, $E(N)$ is uniformly bounded over $a,b \leq 1$ when $t=\frac{1}{2}$ (and it is therefore uniformly bounded when $t=t_{opt}$).

Proof of Theorem 4.4.

We begin with the fundamental inequality:

$$x^{a-1}(1-x)^{b-1} \leq \begin{cases} x^{a-1}(1-t)^{b-1} & (x \leq t) \\ t^{a-1}(1-x)^{b-1} & (x>t) \end{cases} .$$

The area under the top curve is $(1-t)^{b-1}\frac{t^a}{a} + t^{a-1}\frac{(1-t)^b}{b}$. The area under the bottom curve is of course $\Gamma(a)\Gamma(b)/\Gamma(a+b)$. The ratio gives us the expression for $E(N)$. $E(N)$ is minimal for the solution t of

$$(1-t)^2 a(a-1)-t^2 b(b-1) = 0 ,$$

which gives us $t=t_{opt}$. For the performance of Johnk's algorithm, we refer to Theorem 3.4. To compare performances for $a+b \leq 1$, we have to show that for all t,

$$\left(\frac{1}{t}\right)^a \left(\frac{1}{1-t}\right)^b \leq \frac{1}{a+b}\left(\frac{b}{1-t}+\frac{a}{t}\right) .$$

By the arithmetic-geometric mean inequality, the left hand side is in fact not greater than

$$\left(\frac{1}{a+b}\left(\frac{b}{1-t}+\frac{a}{t}\right)\right)^{a+b}$$

$$\leq \frac{1}{a+b}\left(\frac{b}{1-t}+\frac{a}{t}\right)$$

because $a+b \leq 1$, and the argument of the power is a number at least equal to 1. When $a+b>1$, it is easy to check that $E(N)<c$ for $t=\frac{1}{2}$. The statement about the uniform boundedness of $E(N)$ when $t=\frac{1}{2}$ follows simply from

$$E(N) = 2^{1-a-b}\, c$$

and the fact that c is uniformly bounded over $a,b \leq 1$. ∎

Generally speaking, the first algorithm of Atkinson and Whittaker should be used instead of Johnk's when $a,b \leq 1$ and $a+b \geq 1$. The computation of t_{opt}, which involves one square root, is only justified when many random variates are needed for the same values of a and b. Otherwise, one should choose $t=\frac{1}{2}$.

When $a \leq 1$ and $b \geq 1$, the performance of the first algorithm of Atkinson and Whittaker deteriorates with increasing values of b: for fixed $a<1$, $\lim_{b \to \infty} E(N)=\infty$. The inequalities used to develop the algorithm are altered slightly:

$$x^{a-1}(1-x)^{b-1} \leq \begin{cases} x^{a-1} & (x \leq t) \\ t^{a-1}(1-x)^{b-1} & (x>t) \end{cases}.$$

The areas under the two pieces of the dominating curve are, respectively, $\frac{t^a}{a}$ and $t^{a-1}\frac{(1-t)^b}{b}$. The following rejection algorithm can be used:

Second algorithm of Atkinson and Whittaker (1976, 1979)

[SET-UP]

Choose $t \in [0,1]$.

$p \leftarrow \frac{bt}{bt+a(1-t)^b}$

[GENERATOR]

REPEAT

  Generate a uniform [0,1] random variate U and an exponential random variate E.

  IF $U \leq p$

    THEN

      $X \leftarrow t\left(\frac{U}{p}\right)^{\frac{1}{a}}$

      Accept $\leftarrow [(1-b)\log(1-X) \leq E]$

    ELSE

      $X \leftarrow 1-(1-t)\left(\frac{1-U}{1-p}\right)^{\frac{1}{b}}$

      Accept $\leftarrow [(1-a)\log(\frac{X}{t}) \leq E]$

UNTIL Accept

RETURN X

Simple calculations show that

$$E(N) = c\frac{bt^a + a(1-t)^b t^{a-1}}{a+b}$$

where c is the expected number of iterations in Johnk's algorithm (see Theorems 3.4 and 4.4). The optimum value of t is the solution of

$$bt + (a-1)(1-t)^b - bt(1-t)^{b-1} = 0 .$$

Although this can be solved numerically, most of the time we can not afford a numerical solution just to generate one random variate. We have, however, the following reassuring performance analysis for a choice for t suggested by Atkinson and Whittaker (1976):

Theorem 4.5.

For the second algorithm of Atkinson and Whittaker with $t=\frac{1-a}{b+1-a}$,

$$\sup_{a \le 1, b \ge 1} E(N) < \infty ,$$

$$\lim_{b \to \infty} E(N) = \infty \quad (\text{all } a > 1) .$$

4.6. Exercises.

1. For the symmetric beta algorithm studied in Theorem 4.2, show that the quick acceptance step is valid, and that with the quick acceptance step in place, the expected number of evaluations of the full acceptance step tends to 0 as $a \to \infty$.
2. Prove Ulrich's part of Theorem 4.3.
3. Let X be a $\beta_2(a,b)$ random variable. Show that $\frac{1}{Y}$ is $\beta_2(b,a)$, and that $E(Y)=\frac{a}{b-1}$ $(b>1)$, and $Var(Y)=\frac{a(a+b-1)}{(b-1)^2(b-2)}$ $(b>2)$.
4. In the table below, some densities are listed with one parameter $a>0$ or two parameters $a,b>0$. Let c be the shorthand notation for $1/B(a,b)$. Show for each density how a random variate can be generated by a suitable transformation of a beta random variate.

$2cx^{2a-1}(1-x^2)^{b-1}$ $(0 \le x \le 1)$
$2c \sin^{2a-1}(x)\cos^{2b-1}(x)$ $(0 \le x \le \frac{\pi}{2})$
$\frac{cx^{a-1}}{(1+x)^{a+b}}$ $(x \ge 0)$
$\frac{2cx^{2a-1}}{(1+x^2)^{a+b}}$ $(x \ge 0)$
$c\frac{x^{a-1}+x^{b-1}}{(1+x)^{a+b}}$ $(0 \le x \le 1)$
$\frac{(1-x)^{a-1}}{2^{2a-1}B(a,a)\sqrt{x}}$ $(0 \le x \le 1)$
$\frac{(1-x^2)^{a-1}}{2^{2a-2}B(a,a)}$ $(0 \le x \le 1)$

5. Prove Theorem 4.5.
6. **Grassia's distribution.** Grassia (1977) introduced a distribution which is close to the beta distribution, and can be considered to be as flexible, if not more flexible, than the beta distribution. When X is gamma (a,b), then e^{-X} is Grassia I, and $1-e^{-X}$ is Grassia II. Prove that for every possible combination of skewness and kurtosis achievable by the beta density, there

exists a Grassia distribution with the same skewness and kurtosis (Tadikamalla, 1981).

7. A continuation of exercise 6. Use the Grassia distribution to obtain an efficient algorithm for the generation of random variates with density

$$f(x) = \frac{8a^2 x^{a-1} \log(\frac{1}{x})}{\pi^2 (1-x^{2a})} \quad (0<x<1) ,$$

where $a>0$ is a parameter.

5. THE t DISTRIBUTION.

5.1. Overview.

The **t distribution** plays a key role in statistics. The distribution has a symmetric density with one shape parameter $a>0$:

$$f(x) = \frac{\Gamma(\frac{a+1}{2})}{\sqrt{\pi a}\, \Gamma(\frac{a}{2})(1+\frac{x^2}{a})^{\frac{a+1}{2}}}$$

This is a bell-shaped density which can be dealt with in a number of ways. As special members, we note the **Cauchy density** ($a=1$), and the $\mathbf{t_3}$ density ($a=3$). When a is integer-valued, it is sometimes referred to as the number of degrees of freedom of the distribution. Random variate generation methods for this distribution include:

1. The inversion method. Explicit forms of the distribution function are only available in special cases: for the Cauchy density ($a=1$), see section II.2.1. For the t_2 density ($a=2$), see Theorem IX.3.3 in section IX.3.3. For the t_3 density ($a=3$), see exercise II.2.4. In general, the inversion method is not competitive because the distribution function is only available as an integral, and not as a simple explicit function of its argument.
2. Transformation of gamma variates. When N is a normal random variate, and $G_{a/2}$ is a gamma ($\frac{a}{2}$) random variate independent of N,

$$\frac{\sqrt{2a}\, N}{\sqrt{G_{a/2}}}$$

is t_a distributed. Equivalently, if $G_{1/2}, G_{a/2}$ are independent gamma random variables, then

$$S\sqrt{a} \sqrt{\frac{G_{1/2}}{G_{a/2}}}$$

is t_a distributed where S is a random sign. See example I.4.6 for the derivation of this property. Somewhat less useful, but still noteworthy, is the property that if $G_{a/2}, G*_{a/2}$ are iid gamma random variates, then

$$\frac{\sqrt{a}}{2} \frac{G_{a/2}-G*_{a/2}}{\sqrt{G_{a/2}G*_{a/2}}}$$

is t_a distributed (Cacoullos, 1965).

3. Transformation of a symmetric beta random variate. It is known that if X is symmetric beta $(\frac{a}{2},\frac{a}{2})$, then

$$\sqrt{a} \frac{X-\frac{1}{2}}{\sqrt{X(1-X)}}$$

is t_a distributed. Symmetric beta random variate generation was studied in section IX.4.3. The combination of a normal rejection method for symmetric random variates, and the present transformation was proposed by Marsaglia (1980).

4. Transformation of an F random variate. When S is a random sign and X is $F(1,a)$distributed, then $S\sqrt{X}$ is t_a distributed (see exercise I.4.6). Also, when X is symmetric F with parameters a and a, then

$$\frac{\sqrt{a}}{2} \frac{1-X}{\sqrt{X}}$$

is t_a distributed.

5. The ratio-of-uniforms method. See section IV.7.2.

6. The ordinary rejection method. Since the t density cannot be dominated by densities with exponentially decreasing tails, one needs to find a polynomially decreasing dominating function. Typical candidates for the dominating curve include the Cauchy density and the t_3 density. The corresponding algorithms are quite short, and do not rely on fast normal or exponential generators. See below for more details.

7. The composition/rejection method, similar to the method used for normal random variate generation. The algorithms are generally speaking longer, more design constants need to be computed for each choice of a, and the speed is usually a bit better than for the ordinary rejection method. See for example Kinderman, Monahan and Ramage (1977) for such methods.

8. The acceptance-complement method (Stadlober, 1981).

9. Table methods.

One of the transformations of gamma or beta random variates is recommended if one wants to save time writing programs. It is rare that additional speed is required beyond these transformation methods. For direct methods, good speed can be obtained with the ratio-of-uniforms method and with the ordinary rejection methods. Typically, the expected time per random variate is uniformly bounded over a subset of the parameter range, such as $[1,\infty)$ or $[3,\infty)$. Not unexpectedly, the small values of a are the troublemakers, because these densities decrease as $x^{-(a+1)}$, so that no fixed exponent polynomial dominating density exists. The large values of a give least problems because it is easy to see that for every x,

$$\lim_{a \to \infty} f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} .$$

The problem of small a is not important enough to warrant a special section. See however the exercises.

5.2. Ordinary rejection methods.

Let us first start with the development of simple upper bounds for f. For example, when $a \geq 1$, the following inequality is trivially true:

$$\frac{1}{(1+\frac{x^2}{a})^{\frac{a+1}{2}}} \leq \frac{1}{1+\frac{a+1}{2a}x^2} .$$

The top bound is proportional to the density of $\sqrt{\frac{2a}{a+1}}C$ where C is a Cauchy random variate. If we want to verify just how good this inequality is, we note that the area under the dominating curve is $\pi\sqrt{\frac{2a}{a+1}}$. The area under the curve on the left hand side of the inequality is $\frac{\sqrt{\pi a}\,\Gamma(\frac{a}{2})}{\Gamma(\frac{a+1}{2})}$. By the convergence to the normal density, we deduce without computations that this quantity tends to $\sqrt{2\pi}$. Thus, the ratio of the areas, our rejection constant, tends to $\sqrt{\pi}$ as $a \to \infty$. The fit is not very good, except perhaps for a close to 1: for $a=1$, the rejection constant is obviously 1. The details of the rejection algorithm are left to the reader.

Consider next rejection from the t_3 density

$$g(x) = \frac{1}{\sqrt{3}B(\frac{1}{2},\frac{3}{2})(1+\frac{x^2}{3})^2} .$$

Best (1978) has shown the following:

Theorem 5.1.

Let f be the t_a density with $a \geq 3$, and let g be the t_3 density. Then:

$$f(x) \leq cg(x)$$

where

$$c = \frac{8\pi\sqrt{3}}{9\sqrt{a}\,B(\frac{1}{2},\frac{a}{2})\left(1+\frac{1}{a}\right)^{\frac{a+1}{2}}} .$$

Also, if

$$T(x) = \frac{f(x)}{cg(x)} = \frac{9}{16}\frac{\left(1+\frac{x^2}{3}\right)^2}{\left(\frac{1+\frac{x^2}{a}}{1+\frac{1}{a}}\right)^{\frac{a+1}{2}}} ,$$

then

$$T(x) \geq \frac{9}{16}e^{\frac{1}{2}-\frac{x^2}{2}}\left(1+\frac{x^2}{3}\right)^2 .$$

Finally,

$$c \leq \sqrt{\frac{32\pi}{27e}}\sqrt{\frac{a}{a+1}}\,e^{\frac{1}{6a+1}}$$

and

$$\lim_{a \to \infty} c = \sqrt{\frac{32\pi}{27e}} .$$

Proof of Theorem 5.1.

Verify that f/g is maximal for $x = \pm 1$. The lower bound for $T(x)$ follows from the inequality

$$\left(\frac{1+\frac{x^2}{a}}{1+\frac{1}{a}}\right)^{a+1} = \left(1+\frac{x^2-1}{a+1}\right)^{a+1} \leq e^{1-x^2} .$$

Finally, the statement about c follows from Stirling's formula and bounds related to Stirling's formula. For example, the upper bound is obtained as follows:

$$
\begin{aligned}
c &= \frac{8\pi\sqrt{3}}{9\sqrt{a}\,B(\frac{1}{2},\frac{a}{2})\left(1+\frac{1}{a}\right)^{\frac{a+1}{2}}} \\
&\leq \frac{8\sqrt{3\pi}}{9\sqrt{a}}\left(\frac{a+1}{2e}\right)^{\frac{a+1}{2}}\left(\frac{2e}{a}\right)^{\frac{a}{2}}\sqrt{\frac{a}{a+1}}\,e^{\frac{1}{6(a+1)}}\left(\frac{a}{a+1}\right)^{\frac{a+1}{2}} \\
&= \frac{8\sqrt{3\pi}}{9\sqrt{2e}}\sqrt{\frac{a}{a+1}}\,e^{\frac{1}{6(a+1)}} \\
&\to \sqrt{\frac{32\pi}{27e}} .
\end{aligned}
$$

A similar lower bound is valid, which establishes the asymptotic result. ■

The fit with the t_3 dominating density is much better than with the Cauchy density. Also, recalling the ratio-of-uniforms method for generating t_3 random variates in a form convenient to us (see section IV.7.2),

t3 generator based upon the ratio-of-uniforms method

REPEAT

Generate iid uniform [0,1] random variates U, V. Set $V \leftarrow V - \frac{1}{2}$.

UNTIL $U^2+V^2 \leq U$

RETURN $X \leftarrow \sqrt{3}\frac{V}{U}$

We can summarize Best's algorithm as follows:

t generator based upon rejection from a t3 density (Best, 1978)

REPEAT

Generate a t_3 random variate X by the ratio-of-uniforms method (see above).

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

$Z \leftarrow X^2$, $W \leftarrow 1+\frac{Z}{3}$

$Y \leftarrow 2 \log\left(\frac{\frac{9}{16}W^2}{U}\right)$

Accept $\leftarrow [Y \geq 1-Z]$

IF NOT Accept THEN Accept $\leftarrow [Y \geq (a+1)\log(\frac{a+1}{a+Z})]$

UNTIL Accept

RETURN X

The algorithm given above differs slightly from that given in Best (1978). Best adds another squeeze step before the first logarithm.

5.3. The Cauchy density.

The **Cauchy density**

$$f(x) = \frac{1}{\pi(1+x^2)}$$

plays another key role in statistics. It has no shape parameters, and the mean does not exist. Just as for the exponential distribution, it is easily seen that this density causes no problems whatsoever. To start with, the inversion method is applicable because the distribution function is

$$F(x) = \frac{1}{2}+\frac{1}{\pi}\text{arc tan } x \ .$$

This leads to the generator $\tan(\pi U)$ where U is a uniform random variate. The tangent being a relatively slow operation, there is hope for improvement. The main property of the Cauchy density is that whenever (X,Y) is a radially distributed random vector in R^2 without an atom at the origin, then $\frac{X}{Y}$ is Cauchy distributed. The proof uses the fact that if (R,Θ) are the polar coordinates for (X,Y), then $\frac{Y}{X}=\tan(\Theta)$, and Θ is distributed as $2\pi U$ where U is a uniform [0,1] random variate. This leads to two straightforward algorithms for generating

Cauchy random variates:

Polar method I for Cauchy random variates

Generate iid normal random variates N_1,N_2.

RETURN $X \leftarrow \frac{N_1}{N_2}$

Polar method II for Cauchy random variates

REPEAT

Generate iid uniform $[-1,1]$ random variates V_1,V_2.

UNTIL $V_1^2+V_2^2 \leq 1$

RETURN X

Even though the expected number of uniform random variates needed in the second algorithm is $\frac{8}{\pi}$, it seems unlikely that the expected time of the second algorithm will be smaller than the expected time of the algorithm based upon the ratio of two normal random variates. Other algorithms have been proposed in the literature, see for example the acceptance-complement method (section II.5.4 and exercise II.5.1) and the article by Kronmal and Peterson (1981).

5.4. Exercises.

1. **Laha's density (Laha, 1958).** The ratio of two independent normal random variates is Cauchy distributed. This property is shared by other densities as well, in the sense that the term "normal" can be replaced by the name of some other distributions. Show first that the ratio of two independent random variables with Laha's density

$$f(x) = \frac{\sqrt{2}}{\pi(1+x^4)}$$

is Cauchy distributed. Give a good algorithm for generating random variates with Laha's density.

2. Let (X,Y) be uniformly distributed on the circle with center (a,b). Describe the density of $\frac{X}{Y}$. Note that when $(a,b)=(0,0)$, you should obtain the

Cauchy density.

3. Consider the class of generalized Cauchy densities

$$f(x) = \frac{a \sin(\frac{\pi}{a})}{2\pi(1+|x|^a)} ,$$

where $a > 1$ is a parameter. The densities in this class are dominated by the Cauchy density times a constant when $a \geq 2$. Use this fact to develop a generator which is uniformly fast on $[2,\infty)$. Can you also suggest an algorithm which is uniformly fast on $(1,\infty)$?

4. The density

$$f(x) = \frac{1}{\pi(1+x)\sqrt{x}} \quad (x > 0)$$

possesses both a heavy tail and a sharp peak at 0. Suggest a good and short algorithm for the generation of random variates with this density.

5. **Cacoullos's theorem (Cacoullos, 1965).** Prove that when $G, G*$ are iid gamma $(\frac{a}{2})$ random variates, then

$$X \leftarrow \frac{\sqrt{a}}{2} \frac{G-G*}{\sqrt{GG*}}$$

is t_a distributed. In particular, note that when N_1, N_2 are iid normal random variates, then $(N_1-N_2)/(2\sqrt{N_1N_2})$ is Cauchy distributed.

6. The following family of densities has heavier tails than any member of the t family:

$$f(x) = \frac{a-1}{x(\log(x))^a} \quad (x > e) .$$

Here $a > 1$ is a parameter. Propose a simple algorithm for generating random variates from this family, and verify that it is uniformly fast over all values $a > 1$.

7. In this exercise, let C_1, C_2, C_3 be iid Cauchy random variables, and let U be a uniform $[0,1]$ random variable. Prove the following distributional properties:

   A. C_1C_2 has density $(\log(x^2))/(\pi^2(x^2-1))$ (Feller, 1971, p. 64).

   B. $C_1C_2C_3$ has density $(\pi^2+(\log(x^2))^2)/(2\pi^3(1+x^2))$.

   C. UC_1 has density $\log(\frac{1+x^2}{x^2})/(2\pi)$.

8. Show that when X, Y are iid random variables with density $\frac{2}{\pi(e^x+e^{-x})}$, then $X+Y$ has density

$$g(x) = \frac{4x}{\pi^2(e^x-e^{-x})} = \frac{2}{\pi^2(1+\frac{x^2}{3!}+\frac{x^4}{5!}+\cdots)} .$$

Hint: find the density of $\log(|C|)$ first, where C is a Cauchy random variate, and use the previous exercise. Show how you can generate random variates with density g directly and efficiently by the rejection method (Feller, 1971, p. 64).

9. Develop a composition-rejection algorithm for the t distribution which is based on the inequality

$$\frac{1}{(1+\frac{x^2}{a})^{\frac{a+1}{2}}} \geq e^{-\frac{(a+1)x^2}{2a}}$$

which for large a is close to $e^{-\frac{x^2}{2}}$. Make sure that if the remainder term is majorized for use in the rejection algorithm, that the area under the remainder term is $o(1)$ as $a \to \infty$. Note: the remainder term must have tails which increase at least as $|x|^{-(a+1)}$. Note also that the ratio of the areas under the normal lower bound and the area under the t density tends to 1 as $a \to \infty$.

10. **The tail of the Cauchy density.** We consider the family of tail densities of the Cauchy, with the tail being defined as the interval $[t,\infty)$, where $t>0$ is a parameter. Show first that

$$X \leftarrow \tan\left(\arctan(t)(1-U)+\frac{\pi U}{2}\right)$$

has such a tail density. (This is the inversion method.) By using the polar properties of the Cauchy density, show that the following rejection method is also valid, and that the rejection constant tends to 1 as $t \to \infty$:

REPEAT

  Generate iid uniform [0,1] random variates U,V.

  $X \leftarrow \frac{t}{U}$

UNTIL $V(1+\frac{1}{X^2}) \leq 1$

RETURN X

11. This exercise is about inequalities for the function

$$f_a(x) = (1+\frac{x^2}{a})^{-\frac{a+1}{2}}$$

which is proportional to the t density with parameter $a \geq 1$. The inequalities have been used by Kinderman, Monahan and Ramage (1977) in the development of several rejection algorithms with squeeze steps:

A. $f_a(x) \leq \min(1, \frac{1}{x^2})$. Using this inequality in the rejection method corresponds to using the ratio-of-uniforms method.

B. $f_a(x) \geq 1 - \frac{|x|}{2}$. The triangular lower bound is the largest such lower bound not depending upon a that is valid for all $a \geq 1$.

C. $f_a(x) \leq \frac{c}{1+x^2}$ where $c = 2(1+\frac{1}{a})^{-\frac{a+1}{2}} \leq \frac{2}{\sqrt{e}}$. If this inequality is used in the rejection method, then the rejection constant tends to $\sqrt{\frac{2\pi}{e}}$ as $a \to \infty$. The bound can also be used as a quick rejection step.

12. A uniformly fast rejection method for the t family can be obtained by using a combination of a constant bound ($f(0)$) and a polynomial tail bound: for the function $(1+\frac{x^2}{a})^{-\frac{a+1}{2}}$, find an upper bound of the form $\frac{c}{x^b}$ where c, b are chosen to keep the area under the combined upper bound uniformly bounded over $a > 0$.

6. THE STABLE DISTRIBUTION.

6.1. Definition and properties.

It is well known that the sum of iid random variables with finite variance tends in distribution to the normal law. When the variance is not finite, the sum tends in distribution to one of the stable laws, see e.g. Feller (1971). Stable laws have thicker tails than the normal distribution, and are well suited for modeling economic data, see e.g. Mandelbrot (1963), Press (1975). Unfortunately, stable laws are not easy to work with because with a few exceptions no simple expressions are known for the density or distribution function of the stable distributions. The stable distributions are most easily defined in terms of their characteristic functions. Without translation and scale parameters, the characteristic function ϕ is usually defined by

$$\log(\phi(t)) = \begin{cases} -|t|^{\alpha}(1-i\beta\,\mathrm{sgn}(t)\tan(\frac{\alpha\pi}{2})) & (\alpha \neq 1) \\ -|t|(1+i\beta\frac{2}{\pi}\mathrm{sgn}(t)\log(|t|)) & (\alpha = 1) \end{cases},$$

where $-1 \leq \beta \leq 1$ and $0 < \alpha \leq 2$ are the parameters of the distribution, and $\mathrm{sgn}(t)$ is the sign of t. This will be called Levy's representation. There is another parametrization and representation, which we will call the polar form (Zolotarev, 1959; Feller, 1971):

$$\log(\phi(t)) = -|t|^{\alpha} e^{-i\gamma\,\mathrm{sgn}(t)} .$$

Here, $0<\alpha\leq 2$ and $|\gamma| \leq \frac{\pi}{2}\min(\alpha,2-\alpha)$ are the parameters. Note however that one should not equate the two forms to deduce the relationship between the parameters because the representations have different scale factors. After throwing in a scale factor, one quickly notices that the α's are identical, and that β and γ are related via the equation $\beta=\tan(\gamma)/\tan(\alpha\pi/2)$. Because γ has a range which depends upon α, it is more convenient to replace γ by $\frac{\pi}{2}\min(\alpha,2-\alpha)\delta$, where δ is now allowed to vary in $[-1,1]$. Thus, we rewrite the polar form as follows:

$$\log(\phi(t)) = -|t|^{\alpha} e^{-i\frac{\pi}{2}\min(\alpha,2-\alpha)\delta\,\mathrm{sgn}(t)} .$$

When we say that a random variable is stable (1.3,0.4), we are referring to the last polar form with $\alpha=1.3$ and $\delta=0.4$. The parameters β, γ and δ are called the skewness parameters. For $\beta=0$ ($\gamma=0$, $\delta=0$), we obtain the symmetric stable distribution, which is by far the most important sub-class of stable distributions. For all forms, the symmetric stable characteristic function is

$$\phi(t) = e^{-|t|^{\alpha}}$$

By using the product of characteristic functions, it is easy to see that if $X_1, \ldots, X_n$ are iid symmetric stable (α), then

$$n^{-\frac{1}{\alpha}} \sum_{i=1}^{n} X_i$$

is again symmetric stable (α). The following particular cases are important: the symmetric stable (1) law coincides with the Cauchy law, and the symmetric stable (2) distribution is normal with zero mean and variance 2. These two representatives are typical: all symmetric stable densities are unimodal (Ibragimov and Chernin, 1959; Kanter, 1975) and in fact bell-shaped with two infinite tails. All moments exist when $\alpha=2$. For $\alpha<2$, all moments of order $<\alpha$ exist, and the α-th moment is ∞.

The asymmetric stable laws have a nonzero skewness parameter, but in all cases, α is indicative of the size of the tail(s) of the density. Roughly speaking, the tail or tails drop off as $|x|^{-(1+\alpha)}$ as $|x| \rightarrow \infty$. All densities are unimodal, and the existence or nonexistence of moments is as for the symmetric stable densities with the same value of α. There are two infinite tails when $|\delta| \neq 1$ or when $\alpha \geq 1$, and there is one infinite tail otherwise. When $0<\alpha<1$, the mode has the same sign as δ. Thus, for $\alpha<1$, a stable $(\alpha,1)$ random variable is positive, and a stable $(\alpha,-1)$ random variable is negative. Both are shaped as the gamma density.

There are a few relationships between stable random variates that will be useful in the sequel. It is not necessary to treat negative-valued skewness parameters since minus a stable (α,δ) random variable is stable $(\alpha,-\delta)$ distributed. Next, we have the following basic relationship:

Lemma 6.1.

Let Y be a stable $(\alpha',1)$ random variable with $\alpha'<1$, and let X be an independent stable (α,δ) random variable with $\alpha\neq 1$. Then $XY^{1/\alpha}$ is stable $(\alpha\alpha',\delta\frac{\alpha'\min(\alpha,2-\alpha)}{\min(\alpha\alpha',2-\alpha\alpha')})$. Furthermore, the following is true:

A. If N is a normal random variable, and Y is an independent stable $(\alpha',1)$ random variable with $\alpha'<1$, then $N\sqrt{2Y}$ is stable $(2\alpha',0)$.

B. A stable $(\frac{1}{2},1)$ random variable is distributed as $1/(2N^2)$ where N is a normal random variable. In other words, it is Pearson V distributed.

C. If $N_1,N_2,\ldots$ are iid normal random variables, then for integer $k\geq 1$,

$$\prod_{j=0}^{k-1}\frac{1}{(2N_j{}^2)^{2^j}}$$
$$=2^{-(2^k-1)}\prod_{j=1}^{k}\frac{1}{N_j{}^{2^j}}$$

is stable $(2^{-k},1)$.

D. For $N_1,N_2,\ldots$, iid normal random variables, and integer $k\geq 1$,

$$N_{k+1}2^{-(2^{k-1}-1)}\prod_{j=1}^{k}\frac{1}{N_j{}^{2^j-1}}$$

is stable $(2^{1-k},0)$.

E. For $N_1,N_2,\ldots$, iid normal random variables , and integer $k\geq 0$,

$$\frac{N_{k+1}}{N_{k+2}}\prod_{j=0}^{k}\left(\frac{1}{2N_j{}^2}\right)^{2^j}$$
$$=\frac{N_{k+1}}{N_{k+2}}2^{-(2^{k+1}-1)}\prod_{j=0}^{k}\left(\frac{1}{N_j{}^2}\right)^{2^j}$$

is stable $(\frac{1}{2^{k+1}},0)$.

Proof of Lemma 6.1.

The first statement is left as an exercise. If in it, we take $\alpha=2$, $\delta=0$, we obtain part A. It is also seen that a symmetric stable (1) is distributed as a symmetric stable (2) random variable times $\sqrt{X}$ where X is stable $(\frac{1}{2},1)$. But by the property that stable (1) random variables are nothing but Cauchy random variables, i.e. ratios of two independent normal random variables, we conclude that

X must be distributed as $1/(2N^2)$ where N is normally distributed. This proves part B. Next, again by the main property, if X is as above, and Y is stable $(\alpha',1)$, then XY^2 is stable $(\frac{\alpha'}{2},1)$, at least when $\alpha'<1$. If this is applied successively for $\alpha'=\frac{1}{2},\frac{1}{4},\frac{1}{8},\ldots$, we obtain statement C. Statement D follows from statements A and C. Finally, using the fact that a symmetric stable $(1/2^{k+1})$ is distributed as a symmetric stable $(1/2^k)$ times X^{2^k}, where X is stable $(\frac{1}{2},1)$, we see that a stable $(1/2^{k+1},0)$ is distributed as a Cauchy random variable times

$$\prod_{j=0}^{k}\left(\frac{1}{2N_j^{\,2}}\right)^{2^j}.$$

This concludes the proof of part E. ■

Properties A-E in Lemma 6.1 are all corollaries of the main property given there. The main property is due to Feller (1971). Property A tells us that all symmetric stable random variables can be obtained if we can obtain all positive ($\delta=1$) stable random variables with parameter $\alpha<1$. Property B is due to Levy (1940). Property C goes back to Brown and Tukey (1946). Property D is but a simple corollary of property C, and finally, property E is a representation of Mitra's (1981). For other similar representations, see Mitra (1982).

There is another property worthy of mention. It states that all stable (α,δ) random variables can be written as weighted sums of two iid stable $(\alpha,1)$ random variables. It was mentioned in chapter IV (Lemma 6.1), but we reproduce it here for the sake of completeness.

Lemma 6.2.

If X and Y are iid stable$(\alpha,1)$, then $Z \leftarrow pX-qY$ is stable(α,δ) where

$$p^{\alpha}=\frac{\sin(\frac{\pi\min(\alpha,2-\alpha)(1+\delta)}{2})}{\sin(\pi\min(\alpha,2-\alpha))},$$

$$q^{\alpha}=\frac{\sin(\frac{\pi\min(\alpha,2-\alpha)(1-\delta)}{2})}{\sin(\pi\min(\alpha,2-\alpha))}.$$

Proof of Lemma 6.2.

The characteristic function of Z is

$$\phi(t) = E(e^{itpX})E(e^{-itqY})$$
$$= \psi(pt)\psi(-qt)$$

where ψ is the characteristic function of the stable $(\alpha,1)$ law:

$$\psi(t) = e^{-|t|^\alpha e^{-i\frac{\pi}{2}\min(\alpha,2-\alpha)\,\mathrm{sgn}(t)}} .$$

Note next that for $u > 0$, $p^\alpha e^{-iu} + q^\alpha e^{iu}$ is equal to

$$\cos(u)(p^\alpha + q^\alpha) - i\sin(u)(p^\alpha - q^\alpha)$$
$$= \frac{1}{\sin(\pi \min(\alpha,2-\alpha))} 2(\cos(u)\sin(\frac{\pi}{2}\min(\alpha,2-\alpha))\cos((\frac{\pi}{2}\delta\min(\alpha,2-\alpha))) -$$
$$i\sin(u)\cos(\frac{\pi}{2}\min(\alpha,2-\alpha))\sin((\frac{\pi}{2}\delta\min(\alpha,2-\alpha)))) .$$

After replacing u by its value, $\frac{\pi}{2}\min(\alpha,2-\alpha)$, we see that we have

$$\frac{2\cos(u)\sin u}{\sin(2u)}(\cos(\delta u) - i\sin(\delta u)) = e^{-i\delta u} .$$

Resubstitution gives us our result. ∎

6.2. Overview of generators.

The difficulty with most stable densities and distribution functions is that no simple analytical expression for its computation is available. The exceptions are spelled out in the previous section. Basically, stable random variates with parameter α equal to 2^{-k} for $k \geq 0$, and with arbitrary value for δ, can be generated quite easily by the methods outlined in Lemmas 6.1 and 6.2. One just needs to combine an appropriate number of iid normal random variates. For general α,δ, methods requiring accurate values of the density or distribution function are thus doomed, because these cannot be obtained in finite time. Approximate inversions of the distribution function are reported in Fama and Roll (1968), Dumouchel (1971) and Paulson, Holcomb and Leitch (1975). Paulauskas (1982) suggests another approximate method in which enough iid random variables are summed. Candidates for summing include the Pareto densities. For symmetric stable densities, Bartels (1978) also presents approximate methods. Bondesson (1982) proposes yet another approximate method in which a stable random variable is written as an infinite sum of powers of the event times in a homogeneous Poisson process on $[0,\infty)$. The sum is truncated, and the tail sum is replaced by an appropriately picked normal random variate.

Fortunately, exact methods do exist. First of all, the stable density can be written as an integral which in turn leads to a simple formula for generating stable random variates as a combination of one uniform and one exponential random variate. These generators were developed in section IV.6.6, and are based upon integral representations of Ibragimov and Chernin (1959) and Zolotarev (1966). The generators themselves were proposed by Kanter (1975) and Chambers, Mallows and Stuck (1976), and are all of the form $g(U)E^{-\frac{1-\alpha}{\alpha}}$ where E is exponentially distributed, and $g(U)$ is a function of a uniform [0,1] random variate U. The sheer simplicity of the representation makes this method very attractive, even though g is a rather complicated function of its argument involving several trigonometric and exponential/logarithmic operations. Unless speed is absolutely at a premium, this method is highly recommended.

For symmetric stable random variates with $\alpha \leq 1$, there is another representation: such random variates are distributed as

$$\frac{Y}{(E_1+E_2 I_{[U<\alpha]})^{\frac{1}{\alpha}}}$$

where Y has the Fejer-de la Vallee Poussin density, and E_1, E_2 are iid exponential random variates. This representation is based upon properties of Polya characteristic functions, see section IV.6.7, Theorems IV.6.8, IV.6.9, and Example IV.6.7. Since the Fejer-de la Vallee Poussin density does not vary with α, random variates with this density can be generated quite quickly (remark IV.6.1). This can lead to speeds which are superior to the speed of the method of Kanter and Chambers, Mallows and Stuck.

In the rest of this section we outline how the series method (section IV.5) can be used to generate stable random variates. Recall that the series method is based upon rejection, and that it is designed for densities that are given as a convergent series. For stable densities, such convergent series were obtained by Bergstrom (1952) and Feller (1971). In addition, we will need good dominating curves for the stable densities, and sharp estimates for the tail sums of the convergent series. In the next section, the Bergstrom-Feller series will be presented, together with estimates of the tail sums due to Bartels (1981). Inequalities for the stable distribution which lead to practical implementations of the series method are obtained in the last section. At the same time, we will obtain estimates of the expected time performance as a function of the parameters of the distribution.

6.3. The Bergstrom-Feller series.

The purpose of this section is to get ready for the next section, where the series method for stable random variates is developed. The form of the characteristic function most convenient to us is the first polar form, with parameters α and γ. To obtain series expansions for the stable density function, we consider the Fourier inverse of ϕ, which takes a simple form since $|\phi|$ is absolutely integrable:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} e^{-|t|^{\alpha} e^{-i\gamma \operatorname{sgn}(t)}} dt$$

$$= \mathrm{Re}\left\{ \frac{1}{\pi}\int_0^{\infty} e^{-itx} e^{-t^{\alpha} e^{-i\gamma}} dt \right\}$$

$$= \mathrm{Re}\left\{ \frac{1}{\pi}\int_0^{\infty} e^{-txe^{i(\frac{\pi}{2}+\psi)}} e^{-t^{\alpha} e^{i\psi} e^{i(\alpha\psi-\gamma)}} dt \right\}$$

provided that $|\alpha\psi-\gamma| \le \frac{\pi}{2}$ and that $|\frac{\pi}{2}+\psi| \le \frac{\pi}{2}$ with at least one of these being a strict inequality. We have used the fact that changing the sign of γ is equivalent to mirroring the density about the origin, and we have considered a contour in the complex plane. The last expression for f will be our starting point. Recall that we need not only a convergent series, but also good bounds for f and for the tail sums. Bergstrom (1952) replaces each of the exponents in the last expression in turn by its Maclaurin series, and integrates (see also Feller (1971)). Bartels (1981) uses Darboux's formula (1876) for the remainder term in the series expansion to obtain good truncation bounds. In Theorem 6.1 below, we present the two Bergstrom-Feller series together with Bartels's bounds. The proof follows Bartels (1981).

Theorem 6.1.

The stable (α,γ) density f can be expanded for values $x \geq 0$ as follows:

$$f(x) = \sum_{j=1}^{n} a_n(x) + A_{n+1}^*(x)$$

where

$$a_j(x) = \frac{1}{\alpha\pi}(-1)^{j-1}\frac{\Gamma(\frac{j}{\alpha})x^{j-1}\sin(j(\frac{\pi}{2}+\frac{\gamma}{\alpha}))}{j-1!},$$

$$|A_{n+1}^*(x)| \leq A_{n+1}(x) = \frac{1}{\alpha\pi}\frac{\Gamma(\frac{n+1}{\alpha})x^n}{n!(\cos(\theta))^{\frac{n+1}{\alpha}}},$$

where $\theta=0$ if $\gamma\leq 0$ and $\theta=\gamma$ if $\gamma>0$. For $x<0$, note that the value of the density is equal to $f(-x)$ provided that γ is replaced by $-\gamma$. The expansion converges for $1<\alpha\leq 2$. For $0<\alpha<1$, we have a divergent asymptotic series for small $|x|$, i.e., for fixed n, $A_n(x)\to 0$ as $|x|\to 0$. Note also that

$$f(x) \leq \frac{\Gamma(\frac{1}{\alpha})}{\alpha\pi(\cos(\theta))^{\frac{1}{\alpha}}}.$$

A second expansion for $f(x)$ when $x>0$ is given by

$$f(x) = \sum_{j=1}^{n} b_n(x) + B_{n+1}^*(x),$$

where

$$b_j(x) = \frac{(-1)^{j-1}\Gamma(\alpha j+1)\sin(j(\frac{\alpha\pi}{2}+\gamma))}{\pi j! x^{\alpha j+1}},$$

$$|B_{n+1}^*(x)| \leq B_{n+1}(x) = \frac{\Gamma(\alpha(n+1)+1)}{\pi(n+1)!(x\cos(\theta))^{\alpha(n+1)+1}},$$

with $\theta=\max(0,\frac{\pi}{2}+\frac{1}{\alpha}(\gamma-\frac{\pi}{2}))$. The expansion is convergent for $0<\alpha<1$, and is a divergent asymptotic expansion at $|x|\to\infty$ when $\alpha>1$, i.e. for fixed n, $B_n(x)\to 0$ as $|x|\to\infty$. Furthermore, for all α,

$$f(x) \leq \frac{\Gamma(\alpha+1)}{\pi(x\cos(\theta))^{\alpha+1}}.$$

Proof of Theorem 6.1.

The proof is based upon a formula of Darboux (1876), which when applied to e^z with complex z leads to

$$e^z = \sum_{j=0}^{n-1} \frac{z^j}{j!} + \frac{z^n}{n!} M_n ,$$

where $M_n = \lambda e^{\theta z}$, λ being a complex constant with $|\lambda| \leq 1$, and θ being a real constant in the range $0 \leq \theta < 1$. In particular, for $\mathrm{Re}(z) > 0$, $|M_n| \leq |e^z|$. For $\mathrm{Re}(z) \leq 0$, $|M_n| \leq 1$. Apply this result with $z = -txe^{j(\frac{\pi}{2}+\psi)}$ in the inversion formula for f, and note that $\mathrm{Re}(z) \leq 0$. Take the integrals, and observe that the remainder term can be bounded as follows:

$$|A^*_{n+1}(x)| \leq \frac{x^n}{\pi n!} \int_0^\infty t^n \, | e^{-t^\alpha e^{j(\alpha\psi-\gamma)}} | \, dt$$

$$= \frac{x^n}{\pi n!} \int_0^\infty t^n \, e^{-t^\alpha \cos(\alpha\psi-\gamma)} dt$$

$$= \frac{1}{\alpha\pi} \frac{\Gamma(\frac{n+1}{\alpha}) x^n}{n!(\cos(\alpha\psi-\gamma))^{\frac{n+1}{\alpha}}} .$$

The angle ψ can be chosen within the restrictions put on it, to make the upper bound as small as possible. This leads to the choice $\frac{\gamma}{\alpha}$ when $\gamma \leq 0$, and 0 when $\gamma > 0$. It is easy to verify that for $1 < \alpha \leq 2$, the expansion is convergent. Finally, the upper bound is obtained by noting that $f(x) \leq A_1(x)$.

The second expansion is obtained by applying Darboux's formula to $e^{-t^\alpha e^{j(\alpha\psi-\gamma)}}$ and integrating. Repeating the arguments used for the first expansion, we obtain the second expansion. Using Stirling's formula, it is easy to verify that for $0 < \alpha < 1$, the expansion is convergent. Furthermore, for fixed n, $B_n(x) \to 0$ as $|x| \to \infty$, and $f(x) \leq B_1(x)$. ■

The convergent series expansion for $\alpha > 1$ requires an increasing number of terms to reach a given truncation error as $|x|$ increases. The asymptotic series increases in accuracy and needs fewer terms as $|x|$ increases. As pointed out by Bartels (1981), the convergent series generally tends to increase first, before converging, and the intermediate values may become so large that the final answer no longer has sufficient significant digits. This drawback occurs mainly for values of α near 1, and large values of $|\gamma|$.

6.4. The series method for stable random variates.

From Theorem 6.1, we deduce the following useful bound for the stable (α,γ) density when $\gamma \geq 0$:

$$f(x) \leq \begin{cases} \dfrac{\Gamma(\frac{1}{\alpha})}{\alpha\pi(\cos(\gamma))^{\frac{1}{\alpha}}} & (x \geq 0) \\ \dfrac{\Gamma(\alpha+1)}{\pi(x\cos(\eta))^{\alpha+1}} & (x \geq 0) \\ \dfrac{\Gamma(\frac{1}{\alpha})}{\alpha\pi} & (x<0) \\ \dfrac{\Gamma(\alpha+1)}{\pi(-x\cos(\theta))^{\alpha+1}} & (x<0) \end{cases}$$

where $\theta=\max(0,\frac{\pi}{2}+\frac{1}{\alpha}(-\gamma-\frac{\pi}{2}))$ and $\eta=\max(0,\frac{\pi}{2}+\frac{1}{\alpha}(\gamma-\frac{\pi}{2}))$. The bounds are valid for all values of α. The dominating curve will be used in the rejection algorithm to be presented below. Taking the minimum of the bounds gives basically two constant pieces near the center and two polynomially decreasing tails. There is no problem whatsoever with the generation of random variates with density proportional to the dominating curve. Unfortunately, the bounds provided by Theorem 6.1 are not very useful for asymmetric stable random variates because the mode is located away from the origin. For example, for the positive stable density, we even have $f(0)=0$. Thus, a constant/polynomial dominating curve does not cap the density very well in the region between the origin and the mode. For a good fit, we would have needed an expansion around the mode instead of two expansions, one around the origin, and one around ∞. The inefficiency of the bound is easily born out in the integral under the dominating curve. We will consider four cases:

$\gamma=0, \alpha>1$ (symmetric stable).
$\gamma=0, \alpha\leq 1$ (symmetric stable).
$\gamma=(2-\alpha)\frac{\pi}{2}, \alpha>1$ (positive stable).
$\gamma=\alpha\frac{\pi}{2}, \alpha\leq 1$ (positive stable).

The upper bound given to us is of the form $\min(A, Bx^{-(1+\alpha)})$ for $x>0$. For the symmetric stable density, the dominating curve can be mirrored around the origin, while for the asymmetric cases, we need to replace A, B by values A^*, B^*, and x by $-x$. Recalling that

$$\int_0^\infty \min(A, Bx^{-(1+\alpha)})\,dx = \frac{1+\alpha}{\alpha} A^{\frac{\alpha}{1+\alpha}} B^{\frac{1}{1+\alpha}},$$

It is easy to compute the areas under the various dominating curves. We offer the following table for A ,B :

CASE	A	B
1	$\frac{\Gamma(\frac{1}{\alpha})}{\pi\alpha}$	$\frac{\Gamma(1+\alpha)}{\pi(\sin(\frac{\pi}{2\alpha}))^{\alpha+1}}$
2	$\frac{\Gamma(\frac{1}{\alpha})}{\pi\alpha}$	$\frac{\Gamma(1+\alpha)}{\pi}$
3	$\frac{\Gamma(\frac{1}{\alpha})}{\alpha\pi(\sin((\alpha-1)\frac{\pi}{2}))^{\frac{1}{\alpha}}}$	$\frac{\Gamma(\alpha+1)}{\pi(\cos(\frac{\pi}{2\alpha}))^{\alpha+1}}$
4	$\frac{\Gamma(\frac{1}{\alpha})}{\alpha\pi(\cos(\frac{\alpha\pi}{2}))^{\frac{1}{\alpha}}}$	$\frac{\Gamma(\alpha+1)}{\pi(-\cos(\frac{\pi}{2\alpha}))^{\alpha+1}}$

For example, in case 1, we see that the area under the dominating curve is

$$2\frac{\alpha+1}{\alpha}\left(\frac{\Gamma(\frac{1}{\alpha})}{\pi\alpha}\right)^{\frac{\alpha}{1+\alpha}}\left(\frac{\Gamma(1+\alpha)}{\pi(\sin(\frac{\pi}{2\alpha}))^{\alpha+1}}\right)^{\frac{1}{\alpha+1}}$$

$$\leq \frac{4}{\pi}\left(\Gamma(\frac{1}{\alpha})\right)^{\frac{\alpha}{1+\alpha}}\left(\Gamma(1+\alpha)\right)^{\frac{1}{\alpha+1}}$$

$$\leq \frac{4}{\pi}\pi^{\frac{1}{3}}\sqrt{2}$$

where we used the following inequalities: (i) $(\alpha+1)/\alpha^{\alpha/(1+\alpha)}\leq 2$ $(\alpha\geq 1)$; (ii) $\sin(\pi/(2\alpha))\geq 1/\alpha$; (iii) $\Gamma(u)\leq 2$ $(2\leq u\leq 3)$; (iv) $\Gamma(u)\leq\Gamma(\frac{1}{2})=\sqrt{\pi}$ $(\frac{1}{2}\leq u\leq 1)$. Some of the inequalities are rather loose, so that the actual fit is probably much better than what is predicted by the upper bound. For $\alpha=2$, the normal density, we obtain $32^{1/6}\pi^{-2/3}$. The importance of the good fit is clear: we can now use the dominating curve quite confidently in any rejection type algorithm for symmetric stable random variate generation when $\alpha\geq 1$. The story is not so rosy for the three other cases, because the integral of the dominating curve is not uniformly bounded over the specified parameter ranges. The actual verification of this statement is left as an exercise, but we conclude that it is not worth to use the Bergstrom-Feller series for asymmetric stable random variates. For this reason, we will just concentrate on the symmetric case. The notation a_n ,b_n , A_n , B_n is taken from Theorem 6.1. Furthermore, we define a density g and a normalization constant c by

$$cg(x) = \min \begin{cases} \frac{\Gamma(\frac{1}{\alpha})}{\alpha\pi} \\ \frac{\Gamma(\alpha+1)}{\pi(\mid x \mid \sin(\varsigma))^{\alpha+1}} \end{cases}$$

where $\varsigma=0$ for $\alpha<1$, and $\varsigma=\pi/(2\alpha)$ otherwise. The algorithm is of the following form:

Series method for symmetric stable density; case of parameter > 1

REPEAT
    Generate X with density g.
    Generate a uniform [0,1] random variate U.
    $T \leftarrow Ucg(X)$
    $S \leftarrow 0$, $n \leftarrow 0$ (Get ready for series method.)
    REPEAT
        $n \leftarrow n+1, S \leftarrow S+a_n(X)$
    UNTIL $\mid S-T \mid \geq A_{n+1}(X)$
UNTIL $T \leq S$
RETURN X

Because of the convergent nature of the series $\sum a_n$, this algorithm stops with probability one. Note that the divergent asymptotic expansion is only used in the definition of cg. It could of course also be used for introducing quick acceptance and rejection steps. But because of the divergent nature of the expansion it is useless in the definition of a stopping rule. One possible use is as indicated in the modified algorithm shown below.

Series method for symmetric stable density; case of parameter > 1

REPEAT
    Generate X with density g.
    Generate a uniform [0,1] random variate U.
    $T \leftarrow Ucg(X)$
    $S \leftarrow 0$, $n \leftarrow 0$ (Get ready for series method.)
    $V \leftarrow B_2(X)$, $W \leftarrow b_1(X)$
    IF $T \leq W-V$
        THEN RETURN X
        ELSE IF $W-V < T \leq W+V$
            THEN
                REPEAT
                    $n \leftarrow n+1, S \leftarrow S+a_n(X)$
                UNTIL $|S-T| \geq A_{n+1}(X)$
UNTIL $T \leq S$ AND $T \leq W+V$
RETURN X

Good speed is obtainable if we can set up some constants for a fixed value of α. In particular, an array of the first m coefficients of x^{j-1} in the series expansion can be computed beforehand. Note that for $\alpha < 1$, both algorithms shown above can be used again, provided that the roles of a_n and b_n are interchanged. For the modified version, we have:

Series method for symmetric stable density; case of parameter less than or equal to one

REPEAT

Generate X with density g.

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

$T \leftarrow Ucg(X)$

$S \leftarrow 0$, $n \leftarrow 0$ (Get ready for series method.)

$V \leftarrow A_2(X)$, $W \leftarrow a_1(X)$

IF $T \leq W-V$

THEN RETURN X

ELSE IF $W-V<T \leq W+V$

THEN

REPEAT

$n \leftarrow n+1, S \leftarrow S+b_n(X)$

UNTIL $|S-T| \geq B_{n+1}(X)$

UNTIL $T \leq S$ AND $T \leq W+V$

RETURN X

6.5. Exercises.

1. Prove that a symmetric stable random variate with parameter $\frac{1}{2}$ can be obtained as $c(N_1^{-2}-N_2^{-2})$ where N_1,N_2 are iid normal random variates, and $c>0$ is a constant. Determine c too.
2. The expected number of iterations in the series method for symmetric stable random variates with parameter α ,based upon the inequalities given in the text (based upon the Bergstrom-Feller series), is asymptotic to

$$\frac{2}{\pi e \alpha^2}$$

as $\alpha \downarrow 0$.
3. Consider the series method for stable random variates given in the text, without quick acceptance and rejection steps. For all values of α, determine $E(N)$, where N is the number of computations of some term a_n or b_n (note that since a_n or b_n are computed in the inner loop of two nested loops, it is an appropriate measure of the time needed to generate a random variate). For which values, if any, is $E(N)$ finite ?

4. Some approximate methods for stable random variate generation are based upon the following limit law, which you are asked to prove. Assume that $X_1, \ldots$ are iid random variables with common distribution function F satisfying

$$1-F(x) \sim \left(\frac{b}{x}\right)^\alpha \quad (x \to \infty),$$

$$F(-x) \sim \left(\frac{cb*}{|x|}\right)^\alpha \quad (x \to -\infty),$$

for some constants $0<\alpha<2$, $b, b* \geq 0$, $b+b*>0$. Show that there exist normalizing constants c_n such that

$$\frac{1}{n^{\frac{1}{\alpha}}} \sum_{j=1}^{n} X_j - c_n$$

tends in distribution to the stable (α,β) distribution with parameter

$$\beta = \frac{b^\alpha - b*^\alpha}{b^\alpha + b*^\alpha}.$$

(Feller, 1971).

5. This is a continuation of the previous exercise. Give an example of a distribution with a density satisfying the tail conditions mentioned in the exercise, and show how you can generate a random variate. Furthermore, suggest for your example how c_n can be chosen.
6. Prove the first statement of Lemma 6.1.
7. Find a simple dominating curve with uniformly bounded integral for all positive stable densities with parameter $\alpha \geq 1$. Mention how you would proceed with the generation of a random variate with density proportional to this curve.
8. In the spirit of the previous exercise, find a simple dominating curve with uniformly bounded integral for all symmetric stable densities; α can take all values in $(0,2]$.

7. NONSTANDARD DISTRIBUTIONS.

7.1. Bessel function distributions.

The **Polya-Aeppli** distribution is a three-parameter distribution with density

$$f(x) = Cx^{\frac{\lambda-1}{2}} e^{-\theta x} I_{\lambda-1}(\beta\sqrt{x}) \quad (x \geq 0)$$

where $\theta>0$, $\lambda>0$, $\beta\geq 0$ are the parameters and $I_a(x)$ is the modified Bessel function of the first kind, formally defined by

$$I_a(x) = \sum_{j=0}^{\infty} \frac{1}{j!\Gamma(j+a+1)} (\frac{x}{2})^{2j+a} .$$

The normalization constant C is given by

$$C = (\frac{2}{\beta})^{\lambda-1} \theta^\lambda e^{-\frac{\beta^2}{4\theta}} .$$

The name Polya-Aeppli is used in many texts such as Ord (1972, p. 125-126). Others prefer the name "type I Bessel function distribution" (Feller, 1971, p. 57). By using the expansion of the Bessel function, it is not difficult to see that if Z is Poisson $(\frac{\beta^2}{4\theta})$ distributed, and G is gamma $(\lambda+Z)$ distributed, then $\frac{G}{\theta}$ has the Polya-Aeppli distribution. We summarize:

Polya-Aeppli random variate generator

Generate a Poisson $(\frac{\beta^2}{4\theta})$ random variate Z.

Generate a gamma $(\lambda+Z)$ random variate G.

RETURN $X \leftarrow \frac{G}{\theta}$

The Polya-Aeppli family contains as a special case the gamma family (set $\beta=0$, $\theta=1$). Other distributions can be derived from it without much trouble: for example, if X is Polya-Aeppli $(\beta,\lambda,\frac{\theta}{2})$, then X^2 is a type II Bessel function distribution with parameters (β,λ,θ), i.e. X^2 has density

$$f(x) = Dx^\lambda e^{-\theta\frac{x^2}{2}} I_{\lambda-1}(\beta x) \quad (x \geq 0) ,$$

where $D=\theta^\lambda \beta^{1-\lambda} e^{-\beta^2/(2\theta)}$. Special cases here include the folded normal distribution and the Rayleigh distribution. For more about the properties of type I and II Bessel function distributions, see for example Kotz and Srinivasan (1969), Lukacs and Laha (1964) and Laha (1954).

Bessel functions of the second kind appear in other contexts. For example, the product of two iid normal random variables has density

$$\frac{1}{\pi} K_0(x)$$

where K_0 is the Bessel function of the second kind with purely imaginary argument of order 0 (Springer, 1979, p. 160).

In the study of random walks, the following density appears naturally:

$$f(x) = \frac{r}{x} e^{-x} I_r(x) \quad (x > 0) ,$$

where $r > 0$ is a parameter (see Feller (1971, pp. 59-60,476)). For integer r, this is the density of the time before level r is crossed for the first time in a symmetric random walk, when the time between epochs is exponentially distributed:

```
X←0,L←0
REPEAT
    Generate a uniform [-1,1] random variate U.
    L←L+sign(U)
    X←X-log(|U|)
UNTIL L=r
RETURN X
```

Unfortunately, the expected number of iterations is ∞, and the number of iterations is bounded from below by r, so this algorithm is not uniformly fast in any sense. We have however:

Theorem 7.1.

Let $r > 0$ be a real number. If G, B are independent gamma (r) and beta $(\frac{1}{2}, r+\frac{1}{2})$ random variables, then

$$\frac{G}{2B}$$

has density

$$f(x) = \frac{r}{x} e^{-x} I_r(x) \quad (x > 0) .$$

Proof of Theorem 7.1.

We use an integral representation of the Bessel function I_r which can be found for example in Magnus et al. (1966, p. 84):

$$f(x) = \frac{r}{x} e^{-x} I_r(x)$$

$$= \frac{1}{\Gamma(r+\frac{1}{2})} \frac{r}{x} e^{-x} \frac{1}{\sqrt{\pi}} (\frac{x}{2})^r \int_{-1}^{1} e^{-zx} (1-z^2)^{r-\frac{1}{2}} dz$$

$$= \frac{1}{\Gamma(r+\frac{1}{2})} \frac{r}{x} \frac{1}{\sqrt{\pi}} (\frac{x}{2})^r 2^{2r} \int_0^1 e^{-2yx} (y(1-y))^{r-\frac{1}{2}} dy .$$

The result follows directly from this. ■

The algorithm suggested by Theorem 7.1 is uniformly fast over all $r>0$ if uniformly fast gamma and beta generators are used. Of course, we can also use direct rejection. Bounds for f can for example be obtained starting from the integral representation for f given in the proof of Theorem 7.1. The acceptance or rejection has to be decided based upon the series method in that case.

7.2. The logistic and hyperbolic secant distributions.

A random variable has the **logistic distribution** when it has distribution function

$$F(x) = \frac{1}{1+e^{-x}}$$

on the real line. The corresponding density is

$$f(x) = \frac{1}{2+e^x+e^{-x}} .$$

For random variate generation, we can obviously proceed by inversion: when U is uniformly distributed on [0,1], then $X \leftarrow \log(\frac{U}{1-U})$ is logistic. To beat this method, one needs either an extremely efficient rejection or acceptance-complement algorithm, or a table method. Rejection could be based upon one of the following inequalities:

A. $f(x) \le e^{-|x|}$: this is rejection from the Laplace density. The rejection constant is 2.

B. $f(x) \le \frac{1}{4+x^2}$: this is rejection from the density of $2C$ where C is a Cauchy random variate. The rejection constant is $\frac{\pi}{2} \approx 1.57$.

A distribution related to the logistic distribution is the **hyperbolic secant** distribution (Talacko, 1956). The density is given by

$$f(x) = \frac{2}{\pi(e^x+e^{-x})} .$$

Both the logistic and hyperbolic secant distributions are members of the family of Perks distributions (Talacko, 1956), with densities of the form $c/(a+e^x+e^{-x})$, where $a \geq 0$ is a parameter and c is a normalization constant. For this family, rejection from the Cauchy density can always be used since the density is bounded from above by $c/(a+2+x^2)$, and the resulting rejection algorithm has uniformly bounded rejection constant for $a \geq 0$. For the hyperbolic secant distribution in particular, there are other possibilities. One can easily see that it has distribution function

$$F(x) = \frac{2}{\pi} \text{arc} \tan(e^x) .$$

Thus, $X \leftarrow \log(\tan(\frac{\pi}{2}U))$ is a hyperbolic secant random variate whenever U is a uniform [0,1] random variate. We can also use rejection from the Laplace density, based upon the inequality $f(x) \leq \frac{2}{\pi}e^{-|x|}$. This yields a quite acceptable rejection constant of $\frac{4}{\pi}$. The rejection condition can be considerably simplified:

Rejection algorithm for the hyperbolic secant distribution

REPEAT
  Generate U uniformly on [0,1] and V uniformly on [-1,1].
  $X \leftarrow \text{sign}(V)\log(|V|)$
UNTIL $U(|V|+1) \leq 1$
RETURN X

Both the logistic and hyperbolic secant distributions are intimately related to a host of other distributions. Most of the relations can be deduced from the inversion method. For example, by the properties of uniform spacings, we observe that $\frac{U}{1-U}$ is distributed as E_1/E_2, the ratio of two independent exponential random variates. Thus, $\log(E_1)-\log(E_2)$ is logistic. This in turn implies that the difference between two iid extreme-value random variables (i.e., random variables with distribution function $e^{-e^{-x}}$) is logistic. Also, $\tan(\frac{\pi}{2}U)$ is distributed as the absolute value of a Cauchy random variable. Thus, if C is a Cauchy random variable, and N_1, N_2 are iid normal random variables, then $\log(|C|)$ and $\log(|N_1|)-\log(|N_2|)$ are both hyperbolic secant.

Many properties of the logistic distribution are reviewed in Olusegun George and Mudholkar (1981).

7.3. The von Mises distribution.

The **von Mises distribution** for points on a circle has become important in the statistical theory of directional data. For its properties, see for example the survey paper by Mardia (1975). The distribution is completely determined by the distribution of the random angle Θ on $[-\pi,\pi]$. There is one shape parameter, $\kappa>0$, and the density is given by

$$f(\theta) = \frac{e^{\kappa\cos(\theta)}}{2\pi I_0(\kappa)} \quad (|\theta| \leq \pi) .$$

Here I_0 is the modified Bessel function of the first kind of order 0:

$$I_0(x) = \sum_{j=0}^{\infty} \frac{1}{j!^2}\left(\frac{x}{2}\right)^{2j} .$$

Unfortunately, the distribution function does not have a simple closed form, and there is no simple relationship between von Mises (κ) random variables and von Mises (1) random variables which would have allowed us to eliminate in effect the shape parameter. Also, no useful characterizations are as yet available. It seems that the only viable method is the rejection method. Several rejection methods have been suggested in the literature, e.g. the method of Seigerstetter (1974) (see also Ripley (1983)), based upon the obvious inequality

$$f(\theta) \leq f(0)$$

which leads to a rejection constant $2\pi f(0)$ which tends quickly to ∞ as $\kappa\to\infty$. We could use the universal bounding methods of chapter 7 for bounded monotone densities since f is bounded, U-shaped (with modes at π and $-\pi$) and symmetric about 0. Fortunately, there are much better alternatives. The leading work on this subject is by Best and Fisher (1979), who, after considering a variety of dominating curves, suggest using the wrapped Cauchy density as a dominating curve. We will just content ourselves with a reproduction of the Best-Fisher algorithm.

We begin with the wrapped Cauchy distribution function with parameter ρ:

$$G(x) = \frac{1}{2\pi}\arccos\left(\frac{(1+\rho^2)\cos(x)-2\rho}{1+\rho^2-2\rho\cos(x)}\right) \quad (|x| \leq \pi) .$$

For later reference, the density g for G is:

$$g(x) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(x)} \quad (|x| \leq \pi) .$$

A random variate with this distribution can easily be generated via the inversion method:

Wrapped Cauchy generator; inversion method

[SET-UP]

$s \leftarrow \frac{1+\rho^2}{2\rho}$

[GENERATOR]

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

$Z \leftarrow \cos(\pi U)$

RETURN $\Theta \leftarrow \frac{\text{sign}(U)}{\cos(\frac{1+sZ}{s+Z})}$

If the wrapped Cauchy distribution is to be used for rejection, we need to fine tune the distribution, i.e. choose ρ as a function of κ.

Theorem 7.2. (Best and Fisher, 1979)

Let f be the von Mises density with parameter $\kappa > 0$, and let g be the wrapped Cauchy density with parameter $\rho > 0$. Then

$$f(x) \leq c g(x) \quad (|x| \leq \pi)$$

where c is a constant depending upon κ and ρ only. The constant is minimized with respect to ρ for the value

$$\rho = \frac{r - \sqrt{2r}}{2\kappa}$$

where

$$r = 1 + \sqrt{1+4\kappa^2} .$$

The expected number of iterations in the rejection algorithm is

$$c = \frac{\frac{2\rho}{\kappa} e^{\kappa \frac{1+\rho^2}{2\rho} - 1}}{(1-\rho^2) I_0(\kappa)} .$$

Furthermore, $\lim_{\kappa \downarrow 0} c = \infty$ and $\lim_{\kappa \to \infty} c = \sqrt{\frac{2\pi}{e}}$.

Proof of Theorem 7.2.

Consider the ratio

$$h(x) = \frac{f(x)}{g(x)} = \frac{(1+\rho^2-2\rho\cos(x))e^{\kappa\cos(x)}}{I_0(\kappa)(1-\rho^2)} .$$

The derivative of h is zero for $\sin(x)=0$ and for $\cos(x)=(1+\rho^2-\frac{2\rho}{\kappa})/(2\rho)$. By verifying the second derivative of h, we find a local maximum value

$$M_1 = (1-\rho)^2 e^{\kappa}$$

at $\sin(x)=0$ when

$$\frac{2\rho}{(1-\rho)^2} < \kappa ,$$

and a local maximum value

$$M_2 = \frac{2\rho}{\kappa} e^{\kappa\frac{1+\rho^2}{2\rho}-1}$$

at $\cos(x)=(1+\rho^2-\frac{2\rho}{\kappa})/(2\rho)$ when

$$\frac{2\rho}{(1+\rho)^2} < \kappa < \frac{2\rho}{(1-\rho)^2} .$$

Let ρ_0 and ρ_1 be the roots in (0,1) of $\frac{2\rho}{(1-\rho)^2}=\kappa$ and $\frac{2\rho}{(1+\rho)^2}=\kappa$ respectively. The two intervals for ρ defined by the the two sets of inequalities are nonoverlapping. The two intervals are $(0,\rho_0)$ and $(\rho_0,\min(1,\rho_1))$ respectively. The maximum M is defined as M_1 on $(0,\rho_0)$ and as M_2 on $(\rho_0,\min(1,\rho_1))$.

To find the best value of ρ, it suffices to find ρ for which M as a function of ρ is minimal. First, M_1 considered as a function of ρ is minimal for $\rho=\rho_0$. Next, M_2 considered as a function of ρ is minimal at the solution of

$$-\kappa\rho^4+2\rho^3+2\kappa\rho^2+2\rho-\kappa=0 ,$$

i.e. at $\rho=\rho^*=(r-\sqrt{2r})/(2r)$ where $r=1+\sqrt{1+4\kappa^2}$. It can be verified that $\rho^*\in(\rho_0,\min(1,\rho_1))$. But because $M_1(\rho_0)=M_2(\rho_0)\geq M_2(\rho^*)$, it is clear that the overall minimum is attained at ρ^*. The remainder of the statements of Theorem 7.2 are left as an exercise. ■

The rejection algorithm based upon the inequality of Theorem 7.2 is given below:

von Mises generator (Best and Fisher, 1979)

[SET-UP]

$s \leftarrow \frac{1+\rho^2}{2\rho}$

[GENERATOR]

REPEAT

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

$Z \leftarrow \cos(\pi U)$

$W \leftarrow \frac{1+sZ}{s+Z}$

$Y \leftarrow \kappa(s-W)$

Accept $\leftarrow [W(2-W)-V \geq 0]$ (Quick acceptance step)

IF NOT Accept THEN Accept $\leftarrow [\log(\frac{W}{V})+1-W \geq 0]$

UNTIL Accept

RETURN $\Theta \leftarrow \frac{\mathrm{sign}(U)}{\cos(W)}$

Two final computational remarks. The cosine in the definition of Z can be avoided by using an appropriate polar method. The cosine in the last statement of the algorithm cannot be avoided.

7.4. The Burr distribution.

In a series of papers, Burr (1942, 1968, 1973) has proposed a versatile family of densities. For the sake of completeness, his original list is reproduced here. The parameters r,k,c are positive real numbers. The fact that k could take non-integer values is bound to be confusing, but at this point it is undoubtedly better to stick to the standard notation. Note that a list of distribution functions, not densities, is provided in the table.

NAME	$F(x)$	RANGE FOR x
Burr I	x	$[0,1]$
Burr II	$(1+e^{-x})^{-r}$	$(-\infty,\infty)$
Burr III	$(1+x^{-k})^{-r}$	$[0,\infty)$
Burr IV	$(1+(\frac{c-x}{x})^{\frac{1}{c}})^{-r}$	$[0,c]$
Burr V	$(1+ke^{-\tan(x)})^{-r}$	$[-\frac{\pi}{2},\frac{\pi}{2}]$
Burr VI	$(1+ke^{-\sinh(x)})^{-r}$	$(-\infty,\infty)$
Burr VII	$2^{-r}(1+\tanh(x))^r$	$(-\infty,\infty)$
Burr VIII	$(\frac{2}{\pi}\arctan(e^x))^r$	$(-\infty,\infty)$
Burr IX	$1-\frac{2}{2+k((1+e^x)^r-1)}$	$(-\infty,\infty)$
Burr X	$(1+e^{-x^2})^r$	$[0,\infty)$
Burr XI	$(x-\frac{1}{2\pi}\sin(2\pi x))^r$	$[0,1]$
Burr XII	$1-(1+x^c)^{-k}$	$[0,\infty)$

Most of the densities in the Burr family are unimodal. In all cases, we can generate random variates directly via the inversion method. By far the most important of these distributions is the Burr XII distribution. The corresponding density,

$$f(x) = \frac{kcx^{c-1}}{(1+x^c)^k} \quad (x \geq 0)$$

with parameters $c,k>0$ can take a variety of shapes. Thus, f is particularly useful as a flexible dominating curve in random variate generation (see e.g. Cheng (1977)). As pointed out by Tadikamalla (1980), the Burr III density is even more flexible. It is called the reciprocal Burr distribution because the reciprocal of a Burr XII with parameters c,k has the Burr III distribution function

$$F(x) = \frac{1}{(1+x^c)^k} .$$

The density is

$$f(x) = \frac{kcx^{ck-1}}{(1+x^c)^{k+1}} .$$

It should be noted that a myriad of relationships exist between all the Burr distributions, because of the fact that all are directly related to the uniform distribution via the probability integral transform.

7.5. The generalized inverse gaussian distribution.

The **generalized inverse gaussian**, or GIG, distribution is a three-parameter distribution with density

$$f(x) = \frac{(\frac{\psi}{\chi})^{\frac{\lambda}{2}}}{2K_\lambda(\sqrt{\psi\chi})} x^{\lambda-1} e^{-\frac{1}{2}(\frac{\chi}{x}+\psi x)} \qquad (x>0) .$$

Here $\lambda \in R$, $\chi>0$, and $\psi>0$ are the parameters of the distribution, and K_λ is the modified Bessel function of the third kind, defined by

$$K_\lambda(u) = \frac{1}{2}\int_{-\infty}^{\infty} \cosh(\lambda u) e^{-z\cosh(u)}\, du \ .$$

A random variable with the density given above will be called a GIG (λ,ψ,χ) random variable. The GIG family was introduced by Barndorff-Nielsen and Halgreen (1977), and its properties are reviewed by Blaesild (1978) and Jorgensen (1982). The individual densities are gamma-shaped, and the family has had quite a bit of success recently because of its applicability in modeling. Furthermore, many well-known distributions are but special cases of GIG distributions. To cite a few:

A. $\chi=0$: the gamma density.

B. $\psi=0$: the density of the inverse of a gamma random variable.

C. $\lambda=-\frac{1}{2}$: the inverse gaussian distribution (see section IV.4.3).

Furthermore, the GIG distribution is closely related to the generalized hyperbolic distribution (Barndorff-Nielsen (1977, 1978), Blaesild (1978), Barndorff-Nielsen and Blaesild (1980)), which is of interest in itself. For the relationship, we refer to the exercises.

We begin with a partial list of properties, which show that there are really only two shape parameters, and that for random variate generation purposes, we need only consider the cases of $\chi=\psi$ and $\lambda>0$.

Lemma 7.4.

Let GIG (.,.,.) and Gamma (.) denote GIG and gamma distributed random variables with the given parameters, and let all random variables be independent. Then, we have the following distributional equivalences:

A. $\mathrm{GIG}\,(\lambda,\psi,\chi) = \frac{1}{c}\mathrm{GIG}\,(\lambda,\frac{\psi}{c},\chi c\,)$ for all $c>0$. In particular,

$$\mathrm{GIG}(\lambda,\psi,\chi) = \sqrt{\frac{\chi}{\psi}}\mathrm{GIG}(\lambda,\sqrt{\psi\chi},\sqrt{\psi\chi}) \ .$$

B.

$$\mathrm{GIG}(\lambda,\psi,\psi) = \mathrm{GIG}(-\lambda,\psi,\psi) + \frac{2}{\psi}\mathrm{Gamma}(\lambda) \ .$$

C.

$$\mathrm{GIG}(\lambda,\psi,\chi) = \frac{1}{\mathrm{GIG}(-\lambda,\chi,\psi)} \ .$$

For random variate generation purposes, we will thus assume that $\chi=\psi$ and that $\lambda>0$. All the other cases can be taken care of via the equivalences shown in Lemma 7.4. By considering $\log(f\,)$, it is not hard to verify that the distribution is unimodal with mode m at

$$m = \frac{1}{\sqrt{(\frac{\lambda-1}{\psi})^2+1}} - \frac{\lambda-1}{\psi}$$

In addition, the density is log concave for $\lambda\geq 1$. In view of the analysis of section VII.2, we know that this is good news. Log concave densities can be dealt with quite efficiently in a number of ways. First of all, one could employ the universal algorithm for log concave densities given in section VII.2. This has two disadvantages: first, the value of $f\,(m\,)$ has to be computed at least once for every choice of the parameters (recall that this involves computing the modified Bessel function of the third kind); second, the expected number of iterations in the rejection algorithm is large (but not more than 4). The advantages are that the user does not have to do any error-prone computations, and that he has the guarantee that the expected time is uniformly bounded over all $\psi>0$, $\lambda\geq 1$. The expected number of iterations can further be reduced by using the non-universal rejection method of section VII.2.6, which uses rejection from a density with a flat part around m, and two exponential tails. In Theorem 2.6, a simple formula is given for the location of the points where the exponential tails should touch f : place these points such that the value of f at the points is $\frac{1}{e}f\,(m\,)$. Note that to solve this equation, the normalization constant in f cancels out conveniently.

Because $f'(0)=0$, the equation has two well-defined solutions, one on each side of the mode. In some cases, the numerical solution of the equation is well worth the trouble. If one just cannot afford the time to solve the equation numerically, there is always the possibility of placing the points symmetrically at distance $e/((e-1)f(m))$ from m (see section VII.2.6), but this would again involve computing $f(m)$. Atkinson (1979,1982) also uses two exponential tails, both with and without flat center parts, and to optimize the dominating curve, he suggests a crude step search. In any case, the generation process for f can be automated for the case $\lambda \geq 1$.

When $0<\lambda<1$, f is log concave for $x \leq \psi/(1-\lambda)$, and is log convex otherwise. Note that this cut-off point is always greater than the mode m, so that for the part of the density to the left of m, we can use the standard exponential/constant dominating curve as described above for the case $\lambda \geq 1$. The right tail of the GIG density can be bounded by the gamma density (by omitting the $1/x$ term in the exponent). For most choices of $\lambda<1$ and $\psi>0$, this is satisfactory.

7.6. Exercises.

1. **The generalized logistic distribution.** When X is beta (a,b), then $\log(\frac{X}{1-X})$ is generalized logistic with parameters (a,b) (Johnson and Kotz, 1970; Olusegun George and Ojo, 1980). Give a uniformly fast rejection algorithm for the generation of such random variates when $a=b \geq 1$. Do not use the transformation of a beta method given above.
2. Show that if $L_1, L_2, \ldots$ are iid Laplace random variates, then $\sum_{j=1}^{\infty} \frac{L_j}{j^2}$ is logistic. Hint: show first that the logistic distribution has characteristic function $\frac{\pi i t}{\sin(\pi i t)} = \Gamma(1-it)\Gamma(1+it)$. Then use a key property of the gamma function.
3. Complete the proof of Theorem 7.2 by proving that for the von Mises generator of Best and Fisher, $\lim_{\kappa \to \infty} c = \sqrt{\frac{2\pi}{e}}$.
4. **The Pearson system.** In the beginning of this century, Karl Pearson developed his well-known family of distributions. The Pearson system was, and still is, very popular because the family encompasses nearly all well-known distributions, and because every allowable combination of skewness and kurtosis is covered by at least one member of the family. The family has 12 member distributions, and is described in great detail in Johnson and Kotz (1970). In 1973, McGrath and Irving pointed out that random variates for 11 member distributions can be generated by simple transformations of one or two beta or gamma random variates. The exception is the Pearson IV distribution. Fortunately, the Pearson IV density is log-concave, and can be dealt with quite efficiently using the methods of section VII.2 (see exercise

VII.2.1). The Pearson densities are listed in the table below. In the table, a, b, c, d are shape parameters, and C is a normalization constant. Verify the correctness of the generators, and in doing so, determine the normalization constants C.

PEARSON DENSITIES				
Pearson	$f(x)$	PARAMETERS	SUPPORT	GENERATOR
I	$C(1+\frac{x}{a})^b(1-\frac{x}{c})^d$	$b, d > -1; a, c > 0$	$[-a, c]$	$\frac{(a+c)X}{X+Y}-a$ X gamma(b) Y gamma(d)
II	$C(1-(\frac{x}{a})^2)^b$	$b > -1; a > 0$	$[-a, a]$	$\frac{a(X-Y)}{X+Y}$ X gamma($b+1$) Y gamma($b+1$)
III	$C(1+\frac{x}{a})^{ba}e^{-bx}$	$ba > -1; b > 0$	$[-a, \infty]$	$\frac{X}{b}-a$ X gamma($ba+1$)
IV	$C(1+(\frac{x}{a})^2)^{-b}e^{-c\arctan(\frac{x}{a})}$	$a > 0; b > \frac{1}{2}$		
V	$Cx^{-b}e^{-\frac{c}{x}}$	$b > 1; c > 0$	$[0, \infty)$	$\frac{1}{cX}$ X gamma($b-1$)
VI	$C(x-a)^b x^{-c}$	$c > b+1 > 0; a > 0$	$[a, \infty)$	$a\frac{X+Y}{X}$ X gamma($c-b-1$) Y gamma($b+1$)
VII	$C(1+(\frac{x}{a})^2)^{-b}$	$b > \frac{1}{2}; a > 0$		$\frac{aN}{\sqrt{X/2}}$ N normal X gamma($b-\frac{1}{2}$)
VIII	$C(1+\frac{x}{a})^{-b}$	$0 \le b \le 1; a > 0$	$[-a, 0]$	$a(U^{-\frac{1}{b-1}}-1)$ U uniform[0,1]
IX	$C(1+\frac{x}{a})^b$	$b > 0; a > 0$	$[-a, 0]$	$a(U^{\frac{1}{b+1}}-1)$ U uniform[0,1]
X	$\frac{1}{a}e^{-\frac{x}{a}}$	$a > 0$	$[0, \infty)$	aE E exponential
XI	$C(\frac{a}{x})^b$	$a > 0; b > 1$	$[a, \infty)$	$aU^{-\frac{1}{b-1}}$ U uniform[0,1]
XII	$C(\frac{a+x}{b-x})^c$	$0 < b < a; 0 \le c < 1$	$[-a, b]$	$(a+b)X-a$ X beta($c+1, 1-c$)

5. **The arcsine distribution.** A random variable X on $[-1,1]$ is said to have an arcsine distribution if its density is of the form $f(x) = (\pi\sqrt{1-x^2})^{-1}$. Show first that when U, V are iid uniform $[0,1]$ random variables, then $\sin(\pi U)$, $\sin(2\pi U)$, $-\cos(2\pi U)$, $\sin(\pi(U+V))$, and $\sin(\pi(U-V))$ are all have

the arcsine distribution. This immediately suggests several polar methods for generating such random variates: prove, for example, that if (X,Y) is uniformly distributed in C_2, then $(X^2-Y^2)/(X^2+Y^2)$ has the arcsine distribution. Using the polar method, show further the following properties for iid arcsine random variables X,Y:

(i) XY is distributed as $\frac{1}{2}(X+Y)$ (Norton, 1978).

(ii) $\frac{1+X}{2}$ is distributed as X^2 (Arnold and Groeneveld, 1980).

(iii) X is distributed as $2X\sqrt{1-X^2}$ (Arnold and Groeneveld, 1980).

(iv) X^2-Y^2 is distributed as XY (Arnold and Groeneveld, 1980).

6. **Ferreri's system.** Ferreri (1964) suggests the following family of densities:

$$f(x) = \frac{\sqrt{b}}{C(c+e^{a+b(x-\mu)^2})},$$

where a,b,c,μ are parameters, and

$$C = \Gamma(\frac{1}{2}) \sum_{j=1}^{\infty} (-c)^{j-1} e^{-ja} j^{-\frac{1}{2}}$$

is a normalization constant. The parameter c takes only the values ± 1. As $a \rightarrow \infty$, the density approaches the normal density. Develop an efficient uniformly fast generator for this family.

7. The family of distributions of the form $aX+bY$ where $a,b \in R$ are parameters, and X,Y are iid gamma random variables was proposed by McKay (1932) and studied by Bhattacharyya (1942). This family has basically two shape parameters. Derive its density, and note that its form is a product of a gamma density multiplied with a modified Bessel function of the second kind when $a,b>0$.

8. **Toranzos's system.** Show how you can generate random variates from Toranzos's class (Toranzos, 1952) of bell-shaped densities of the form $Cx^c e^{-(a+bx)^2}$ $(x>0)$ (C is a normalization constant) in expected time uniformly bounded over all allowable values of the parameters. Do not use C in the generator, and do not compute C for the proof of the uniform boundedness of the expected time.

9. **Tukey's lambda distribution.** In 1960, Tukey proposed a versatile family of symmetric densities in terms of the inverse distribution function:

$$F^{-1}(U) = \frac{1}{\lambda}(U^{\lambda}-(1-U)^{\lambda}),$$

where $\lambda \in R$ is a shape parameter. Clearly, if U is a uniform [0,1] random variate, then $F^{-1}(U)$ has the given distribution. Note that the density is not known in closed form. Tukey's distribution was later generalized in several directions, first by Ramberg and Schmeiser (1972) who added a location and a scale parameter. The most significant generalization was by Ramberg and Schmeiser (1974), who defined

$$F^{-1}(U) = \lambda_1 + \frac{1}{\lambda_2}(U^{\lambda_3} - (1-U)^{\lambda_4}) .$$

For yet another generalization, see Ramberg (1975). In the Ramberg-Schmeiser form, λ_1 is a location parameter, and λ_2 is a scale parameter. The merit of this family of distributions is its versatility with respect to its use in modeling data. Furthermore, random variate generation is trivial. It is therefore important to understand which shapes the density can take. Prove all the statements given below.

A. As $\lambda_3=\lambda_4 \to 0$, the density tends to the logistic density.

B. The density is J-shaped when $\lambda_3=0$.

C. When $\lambda_1=\lambda_3=0$, and $\lambda_2=\lambda_4 \to 0$, the density tends to the exponential density.

D. The density is U-shaped when $1 \leq \lambda_3, \lambda_4 \leq 2$.

E. Give necessary and sufficient conditions for the distribution to be truncated on the left (right).

F. No positive moments exist when $\lambda_3 < -1$ and $\lambda_4 > 1$, or vice versa.

G. The density $f(x)$ can be found by computing $1/F^{-1\prime}(u)$, where u is related to x via the equality $x = F^{-1}(u)$. Thus, by letting u vary between 0 and 1, we can compute pairs $(x, f(x))$, and thus plot the density.

H. Show that for $\lambda_1=0$, $\lambda_2=0.1975$, $\lambda_3=\lambda_4=0.1349$, the distribution function thus obtained differs from the normal distribution function by at most 0.002.

For a general description of the family, and a more complete bibliography, see Ramberg, Tadikamalla, Dudewicz and Mykytka (1979).

10. **The hyperbolic distribution.** The hyperbolic distribution, introduced by Barndorff-Nielsen (1977, 1978) has density

$$f(x) = \frac{\varsigma}{2\alpha K_1(\varsigma)} e^{-\alpha\sqrt{1+x^2}+\beta x} .$$

Here $\alpha > |\beta|$ are the parameters, $\varsigma = \sqrt{\alpha^2-\beta^2}$, and K_1 is the modified Bessel function of the third kind. For $\beta=0$, the density is symmetric. Show the following:

A. The distribution is log-concave.

B. If N is normally distributed, and X is GIG $(1, \alpha^2-\beta^2, 1)$, then $\beta X + N\sqrt{X}$ has the given density.

C. The parameters for the optimal non-universal rejection algorithm for log-concave densities are explicitly computable. (Compute them, and obtain an expression for the expected number of iterations. Hint: apply Theorem VII.2.6.)

11. **The hyperbola distribution.** The hyperbola distribution, introduced by Barndorff-Nielsen (1978) has density

$$f(x) = \frac{1}{2K_0(\varsigma)\sqrt{1+x^2}} e^{-\alpha\sqrt{1+x^2}+\beta x} .$$

Here $\alpha > |\beta|$ are the parameters, $\varsigma=\sqrt{\alpha^2-\beta^2}$, and K_0 is the modified Bessel function of the third kind. For $\beta=0$, the density is symmetric. Show the following:

A. The distribution is not log-concave.

B. If N is normally distributed, and X is GIG $(0,\alpha^2-\beta^2,1)$, then $\beta X+N\sqrt{X}$ has the given density.

12. **Johnson's system.** Every possible combination of skewness and kurtosis corresponds to one and only one distribution in the Pearson system. Other systems have been designed to have the same property too. For example, Johnson (1949) introduced a system defined by the densities of suitably transformed normal (μ,σ) random variables N: his system consists of the S_L, or lognormal, densities (of e^N), of the S_B densities (of $e^N/(1+e^N)$), and the S_U densities (of $\sinh(N)=\frac{1}{2}(e^N-e^{-N})$). This system has the advantage that fitting of parameters by the method of percentiles is simple. Also, random variate generation is simple. In Johnson (1954), a similar system in which N is replaced by a Laplace random variate with center at μ and variance σ^2 is described. Give an algorithm for the generation of a Johnson system random variable when the skewness and kurtosis are given (recall that after normalization to zero mean and unit variance, the skewness is the third moment, and kurtosis is the fourth moment). Note that this forces you in effect to determine the different regions in the skewness-kurtosis plane. You should be able to test very quickly which region you are in. However, your main problem is that the equations linking μ and σ to the skewness and kurtosis are not easily solved. Provide fast-convergent algorithms for their numerical solution.