Chapter Seven
UNIVERSAL METHODS

1. BLACK BOX PHILOSOPHY.

In the next two chapters we will apply the tools of the previous chapters in the design of algorithms that are applicable to large families of distributions. Described in terms of a common property, such as the family of all unimodal densities with mode at 0, these families are generally speaking nonparametric in nature. A method that is applicable to such a large family is called a **universal method.** For example, the rejection method can be used for all bounded densities on [0,1], and is thus a universal method. But to actually apply the rejection method correctly and efficiently would require knowledge of the supremum of the density. This value cannot be estimated in a finite amount of time unless we have more information about the density in question, usually in the form of an explicit analytic definition. Universal methods which do not require anything beyond what is given in the definition of the family are called black box methods.

Consider for example all discrete distributions on the positive integers. Assume only that for each i we can evaluate p_i (consider this evaluation as being performed by a black box). Then the sequential inversion method (section III.2) can be used to generate a random variate with this distribution, and can thus be called a black box method for this family. The inversion method for distributions with a continuous distribution function is not a black box method because finite time generation is only possible in special cases (e.g., the distribution function is piecewise linear).

The larger the family for which we design a black box method, the less we should expect from the algorithm timewise: a case in point is the sequential inversion method for discrete random variates. The undeniable advantage of having a few black box methods in one's computer library is that one can always fall back on these when everything else fails. Comparative timings with algorithms specially designed for particular distributions are not fair.

In chapters IX and X we will mainly be concerned with fast algorithms for parametric families that are widely used by the statistical community. In this chapter too, we will be concerned with speed, but it is by no means the driving force. Because continuous distributions are more difficult to handle in general, we will only focus on families with densities. In section 2, we present a case study for the class of log-concave densities, to wet the appetite. Since the whole story in black box methods is told in terms of inequalities when the rejection method is involved, it is important to show how standard probability theoretical inequalities can aid in the design of black box algorithms. This is done in section 3. In section 4, the inversion-rejection principle is presented, which combines the sequential inversion method for discrete random variates with the rejection method. It is demonstrated there that this method can be used for the generation of random variables with a unimodal or monotone density.

2. LOG-CONCAVE DENSITIES.

2.1. Definition.

A density f on R^d is called **log-concave** when $\log f$ is concave on its support. In this section we will obtain universal methods for this class of densities when $d=1$. The class of densities is very important in statistics. A partial list of member densities is given in the table below.

Name of density	Density	Parameter(s)
Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	
Gamma (a)	$\frac{x^{a-1}e^{-x}}{\Gamma(a)}$ $(x>0)$	$a>1$
Weibull (a)	$ax^{a-1}e^{-x^a}$ $(x>0)$	$a \geq 1$
Beta (a,b)	$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}$ $(0 \leq x \leq 1)$	$a,b \geq 1$
Exponential power (a)	$\frac{e^{-\|x\|^a}}{2\Gamma(1+\frac{1}{a})}$	$a \geq 1$
Perks (a)	$\frac{c}{e^x+e^{-x}+a}$	$a>-2$
Logistic	same as above, $a=2$	
Hyperbolic secant	same as above, $a=0$	
Extreme value (k)	$\frac{k^k}{(k-1)!} e^{-kx-ke^{-x}}$	$k \geq 1$,integer
Generalized inverse gaussian	$cx^{a-1}e^{-bx-\frac{b^*}{x}}$ $(x \geq 0)$	$a \geq 1, b, b^* > 0$

Important individual members of this family also include the uniform density (as a special case of the beta family), and the exponential density (as a special case of the gamma family). For studies on the less known members, see for example Perks (1932) (for the Perks densities), Talacko (1956) (for the hyperbolic secant density), Gumbel (1958) (for the extreme value distributions) and Jorgensen (1982) (for the generalized inverse gaussian densities).

The family of log-concave densities on R is also important to the mathematical statistician because of a few key properties involving closedness under certain operations: for example, the class is closed under convolutions (Ibragimov (1956), Lekkerkerker (1953)).

The algorithms of this section are based upon rejection. They are of the black box type for all log-concave densities with mode at 0 (note that all log-concave densities are bounded and have a mode, that is, a point x such that f is nonincreasing on $[x,\infty)$ and nondecreasing on $(-\infty,x]$). Thus, the mode must be given to us beforehand. Because of this, we will mainly concentrate on the class $LC_{0,1}$, the class of all log-concave densities with a mode at 0 and $f(0)=1$. The restriction $f(0)=1$ is not crucial: since $f(0)$ can be computed at run-time, we can always rescale the axis after having computed $f(0)$ so that the value of $f(0)$ after rescaling is 1. We define LC_0 as the class of all log-concave densities with a mode at 0.

The bottom line of this section is that there is a rejection-based black box method for LC_0 which takes expected time uniformly bounded over this class if the computation of f at any point and for any f takes one unit of time. The algorithm can be implemented in about ten lines of FORTRAN or PASCAL code. The fundamental inequality needed to achieve this is developed in the next sub-section. All of the results in this section were first published in Devroye (1984).

2.2. Inequalities for log-concave densities.

Theorem 2.1.

Assume that f is a log-concave density on $[0,\infty)$ with a mode at 0, and that $f(0)=1$. Then $f(x)\leq g(x)$ where

$$g(x)=\begin{cases}1 & (0\leq x\leq 1)\\ \text{the unique solution } t<1 \text{ of } t=e^{-x(1-t)} & (x>1)\end{cases}.$$

The inequality cannot be improved because g is the supremum of all densities in the family.

Furthermore, for any log-concave density f on $[0,\infty)$ with mode at 0,

$$\int_x^\infty f \leq e^{-xf(0)} \quad (x\geq 0).$$

Proof of Theorem 2.1.

We need only consider the case $x>1$. The density f in the given class which yields the maximal value of $f(x)$ when $x>1$ is fixed is given by

$$\log f(u)=\begin{cases}-au & (0\le u\le x)\\ -\infty & (x<u)\end{cases}$$

for some $a>0$. Thus, $f(u)=e^{-au}$, $0\le u\le x$. Here a is chosen for the sake of normalization. We must have

$$1=\frac{1-e^{-ax}}{a}.$$

Replace $1-a$ by t.

The second part of the theorem follows by a similar geometrical argument. First fix $x>0$. Then notice that the tail probability beyond x is maximal for the exponential density, which because of normalization must be of the form $f(0)e^{-yf(0)}$, $y\ge 0$. The tail probability is $e^{-xf(0)}$. ∎

Theorem 2.2.

The function g of Theorem 2.1 can be bounded by two sequences of functions $y_n(x), z_n(x)$ for $x>1$, where

(i) $0=z_0(x)\le z_1(x)\le\cdots\le g(x)$;

(ii) $g(x)\le\cdots\le y_1(x)\le y_0(x)=\frac{1}{x}$;

(iii) $\lim_{n\to\infty} y_n(x)=g(x)$;

(iv) $\lim_{n\to\infty} z_n(x)=g(x)$;

(v) $y_{n+1}(x)=e^{-x(1-y_n(x))}$;

(vi) $z_{n+1}(x)=e^{-x(1-z_n(x))}$.

Proof of Theorem 2.2.

Fix $x>1$. Consider the functions $f_1(u)=u$ and $f_2(u)=e^{-x(1-u)}$ for $0\le u\le 1$. We have $f_1(1)=f_2(1)=1$, $f'_2(1)=x>1=f'_1(1)$, $f'_2(0)=xe^{-x}<1=f'_1(0)$. Also, f_2 is convex and increases from e^{-x} at $u=0$ to 1 at $u=1$. Thus, there exists precisely one solution in (0,1) for the equation $f_1(u)=f_2(u)$. This solution can be obtained by ordinary functional iteration: if one starts with $z_0(x)=0$, and uses $z_{n+1}(x)=f_2(z_n(x))$, then the unique solution is approached from below in a monotone manner. If we start with $y_0(x)$ at least equal to the value of the solution, then the functional iteration $y_{n+1}(x)=f_2(y_n(x))$ can be used to approach the solution from above in a monotone way. Since $f(x) \le \frac{1}{x}$ for all monotone densities f on $[0,\infty)$, we have $g(x) \le \frac{1}{x}$, and thus, we can take $y_0(x) = \frac{1}{x}$. ■

When f is a log-concave density on $[m,\infty)$ with mode at m, then

$$\frac{f\left(m+\frac{x}{f(m)}\right)}{f(m)} \le \min(1, e^{1-x}) \qquad (x \ge 0) .$$

The area under the bounding curve is exactly 2. The inequality applies to all log-concave densities with mode at m (in which case the condition $x>0$ must be dropped and $1-x$ is replaced by $1-|x|$). But unfortunately, the area under the dominating curve becomes 4. The two features that make the inequality useful for us are

(i) The fact that the area under the curve does not depend upon f. (This gives us a uniform guarantee about its performance.)

(ii) The fact that the top curve itself does not depend upon f. (This is a necessary condition for a true black box method.)

2.3. A black box algorithm.

Let us start with the rejection algorithm based upon the inequality

$$\frac{f\left(m+\frac{x}{f(m)}\right)}{f(m)} \le \min(1, e^{1-x}) \qquad (x \ge 0)$$

valid for log-concave densities on $[m,\infty)$ with mode at m:

Rejection algorithm for log-concave densities

[SET-UP](can be omitted)

$c \leftarrow f(m)$

[GENERATOR]

REPEAT

  Generate U uniformly on [0,2] and V uniformly on [0,1].

  IF $U \leq 1$

    THEN $(X,Z) \leftarrow (U,V)$

    ELSE $(X,Z) \leftarrow (1-\log(U-1), V(U-1))$

  $X \leftarrow m + \frac{X}{c}$

UNTIL $Z \leq \frac{f(X)}{c}$

RETURN X

The validity of this algorithm is quickly verified: just note that the random vector (X,Z) generated in the middle section of the algorithm is uniformly distributed under the curve $\min(1,e^{1-x})$ $(x \geq 0)$. Because of the excellent properties of the algorithm, it is worth pointing out how we can proceed when f is log-concave with support on both sides of the mode m. It suffices to add a random sign to X just after (X,Z) is generated. We should note here that we pay rather heavily for the presence of two tails because the rejection constant becomes 4. A quick fix-up is not possible because of the fact that the sum of two log-concave functions is not necessarily log-concave. Thus, we cannot "add" the left portion of f to the right portion suitably mirrored and apply the given algorithm to the sum. However, when f is symmetric about the mode m, it is possible to keep the rejection constant at 2 by replacing the statement $X \leftarrow m + \frac{X}{c}$ by $X \leftarrow m + \frac{SX}{2c}$ where S is a random sign.

Let us conclude this section of algorithms with an exponential version of the previous method which should be fast when exponential random variates can be generated cheaply and if the computation of $\log(f)$ can be done efficiently (in most cases, $\log(f)$ can be computed faster than f).

Rejection method for log-concave densities. Exponential version

[SET-UP](can be omitted)

$c \leftarrow f(m)$, $r \leftarrow \log c$

[GENERATOR]

REPEAT

Generate U uniformly on [0,2]. Generate an exponential random variate E.

IF $U \leq 1$

THEN $(X,Z) \leftarrow (U,-E)$

ELSE $(X,Z) \leftarrow (1+E^*,-E-E^*)$ (E^* is a new exponential random variate)

CASE

f log-concave on $[m,\infty)$: $X \leftarrow m+\frac{X}{c}$

f log-concave on $(-\infty,\infty)$:

Generate a random sign S.

CASE

f symmetric: $X \leftarrow m+\frac{SX}{2c}$

f not known to be symmetric: $X \leftarrow m+\frac{SX}{c}$

UNTIL $Z \leq \log f(X)-r$

RETURN X

One of the practical stumbling blocks is that often most of the time spent in the computation of $f(X)$ is spent computing a complicated normalization factor. When f is given analytically, it can be sidestepped by setting up a subprogram for the computation of the ratio $f(x)/f(m)$ since this is all that is needed in the algorithms. For example, for the generalized inverse gaussian distribution, the normalization constant has several factors including the value of the Bessel function of the third kind. The factors cancel out in $f(x)/f(m)$. Note however that we cannot entirely ignore the issue since $f(m)$ is needed in the computation of X. Because m is fixed, we call this a set-up step.

2.4. The optimal rejection algorithm.

In this section, we assume that f is in $LC_{0,1}$. The optimal rejection algorithm uses the best possible uniform bounding curve, that is, the function g of Theorem 2.1. The problem is that g is only defined implicitly. Nevertheless, it is possible to generate random variates with density $g / \int g$ without great difficulty:

Theorem 2.3.

Let E_1, E_2, U, D be independent random variables with the following distributions: E_1, E_2 are exponentially distributed, U is uniformly distributed on [0,1] and D is integer-valued with $P(D=n)=6/(\pi^2 n^2)$, $n \geq 1$. Then

$$(X,Y) = (U\frac{(E_1+E_2)/D}{1-e^{-(E_1+E_2)/D}}, e^{-(E_1+E_2)/D})$$

is uniformly distributed in $\{(x,y) : x \geq 0, 0 \leq y \leq g(x)\}$ where g is defined in Theorem 2.1. In particular, X has density $g / \int g$ and Y is distributed as $Vg(X)$ where V is a uniform [0,1] random variable independent of X.

Proof of Theorem 2.3.

Flip the axes around, and observe that the desired Y should have density proportional to $-\log(y)/(1-y)$, $0 \leq y \leq 1$, and that X should be distributed as $U(-\log(Y)/(1-Y))$ where U is independent of Y. By the transformation $y=e^{-z}$, $Y=e^{-Z}$, we see that Z has density proportional to

$$\frac{ze^{-z}}{1-e^{-z}} = \sum_{n=0}^{\infty} ze^{-(n+1)z}$$

$$= \frac{\pi^2}{6}(\sum_{n=1}^{\infty}(n^2 ze^{-nz})(\frac{6}{\pi^2 n^2})) \quad (z \geq 0),$$

i.e., Z is distributed as $(E_1+E_2)/D$ (since E_1+E_2 has density ze^{-z}, $z \geq 0$). Thus, the couple $(UZ/(1-e^{-Z})), e^{-Z})$ has the correct uniform distribution. ■

In the proof of Theorem 2.3, we have also shown that

$$\int g = \frac{\pi^2}{6} \approx 1.6433.$$

This is about 18% better than for the algorithms of the previous section. The algorithm based upon Theorem 2.3 is as follows:

Optimal rejection algorithm for log-concave densities

```
[NOTE: f ∈LC_{0,1}]
REPEAT
    Generate a uniform [0,1] random variate U.
    Generate iid exponential random variates E_1,E_2. Set E←E_1+E_2.
    Generate a discrete random variate D with P(D=n)=6/(π^2 n^2), n≥1.
    Z←E/D
    Y←e^{-Z}, X←UZ/(1-Y)
UNTIL Y≤f(X)
RETURN X
```

For the generation of D, we could use yet another rejection method such as:

```
REPEAT
    Generate iid uniform [0,1] random variates U,V.
    IF U≤1/2
        THEN D←1
        ELSE D←⌈1/(2(1-U))⌉
UNTIL DV≥1
RETURN D
```

If D is generated as suggested, we have a rejection constant of $\frac{12}{\pi^2}$. When used in the former algorithm, this will offset the 18% gain so painstakingly obtained. Since the D generator does not vary with f, it should preferably be implemented based upon a combination of the alias method and a rejection method for the tail of the distribution.

2.5. The mirror principle.

Consider now a normalized log-concave f with two tails, $m=0$, and $f(0)=1$. In this case, the original algorithms have a rejection constant equal to 4. However, there are two observations of Richard Brent which will considerably improve the performance. The first observation is that if $p=F(m)$ is known (F is the distribution function), then the rejection constant can be reduced to 2 again. This is based upon the following inequality:

Theorem 2.4.

If f is a log-concave density with mode $m=0$ and $f(0)=1$, then, writing p for $F(0)$, we have

$$f(x) \leq \begin{cases} \min(1, e^{1-\frac{|x|}{1-p}}) & (x \geq 0) \\ \min(1, e^{1-\frac{|x|}{p}}) & (x < 0) \end{cases} .$$

The area under the bounding curve is 2.

Proof of Theorem 2.4.

Note that $\frac{f(x)}{1-p}$ is a log-concave density on $(0,\infty)$, and that $\frac{f(x)}{p}$ is a log-concave density on $(-\infty,0)$. Since $f(x(1-p))$ is log-concave on $(0,\infty)$, we have

$$f(x(1-p)) \leq \min(1, e^{1-x}) \quad (x \geq 0) .$$

The inequality and the statement about the area follow without further work. ∎

The details of the rejection algorithm based upon Theorem 2.4 are left as an exercise. Brent's second observation applies to the case that $F(m)$ is not available. The expected number of iterations in the rejection algorithm can be reduced to between 2.64 and 2.75 at the expense of an increased number of computations of f.

Theorem 2.5.

Let f be a log-concave density on R with mode at 0 and $f(0)=1$. Then, for $x>0$,

$$f(x)+f(-x) \le g(x) = \sup_{p \in (0,1)} \left(\min(1,e^{1-\frac{x}{1-p}})+\min(1,e^{1-\frac{x}{p}})\right)$$

$$= \begin{cases} 2 & (0 \le x \le \frac{1}{2}) \\ 1+e^{2-\frac{1}{1-x}} & (\frac{1}{2} \le x \le 1) \\ e^{1-x} & (x \ge 1) \end{cases} .$$

Furthermore,

$$\int g = \frac{5}{2}+\frac{1}{4}\int_0^\infty \frac{e^{-u}}{(1+\frac{u}{2})^2}\,du < \frac{5}{2}+\frac{1}{4}\int_0^\infty \frac{e^{-u}}{1+u}\,du \approx 2.6491 .$$

Define another function $g*$ where $g*=g$ except on $(\frac{1}{2},1)$, where $g*$ is linear with values $g*(\frac{1}{2})=2, g*(1)=1$. Then $g* \ge g$ and $\int g* = \frac{11}{4}$.

Proof of Theorem 2.5.

Let us write once again $p=F(0)$. The first inequality follows directly from Theorem 2.4. We will first rewrite g as $\sup_{0 \le p \le \frac{1}{2}} h_p(x)$ where $h_p(x)$ is defined by

$$\begin{cases} 2 & (x \le p) \\ 1+e^{1-\frac{x}{p}} & (p \le x \le 1-p) \\ e^{1-\frac{x}{p}}+e^{1-\frac{x}{1-p}} & (1-p \le x < \infty) \end{cases} .$$

To prove the main statement of Theorem 2.5, we first show that g is at least equal to the right-hand-side of the main equation. For $x \le \frac{1}{2}$, we have $h_{1/2}(x)=2$. For $\frac{1}{2} \le x \le 1$, observe that $h_{1-x}(x)=1+e^{2-1/(1-x)}$. Finally, for $x \ge 1$, we have $h_0(x)=e^{1-x}$. We now show that g is at most equal to the right-hand-side of the main equation. To do this, decompose h_p as $h_{p1}+h_{p2}+h_{p3}$ where $h_{p1}=h_p I_{[0,p]}$, $h_{p2}=h_p I_{(p,1-p)}$, $h_{p3}=h_p I_{[1-p,\infty)}$. Clearly, $h_{p1} \le g$ for all $p \le \frac{1}{2}, x \ge 0$. Since $(p,1-p) \subseteq [0,1]$, we have $h_{p2} \le g$ for all $p \le \frac{1}{2}, x \ge 0$. It suffices to show that $h_{p3} \leq e^{1-x}$ for all $x \geq 1, p \leq \frac{1}{2}$. This follows if for all such p,

$$e^{-\frac{1}{p}} + e^{-\frac{1}{1-p}} \leq \frac{1}{e}$$

because this would imply, for $x \geq 1$,

$$e\left(\left(e^{-\frac{1}{p}}\right)^x + \left(e^{-\frac{1}{1-p}}\right)^x\right)$$

$$\leq e\left(e^{-\frac{1}{p}} + e^{-\frac{1}{1-p}}\right)^x$$

$$\leq e^{1-x} .$$

Putting $u = \frac{1-p}{p}$, we have

$$e\left(e^{-\frac{1}{p}} + e^{-\frac{1}{1-p}}\right) = e^{-u} + e^{-\frac{1}{u}} .$$

The last function has equal maxima at $u=0$ and $u \uparrow \infty$, and a minimum at $u=1$. The maximal value is 1 and the minimal value is $\frac{2}{e}$. This concludes the proof of the main equation in the theorem.

Next, $\int g$ is

$$\frac{5}{2} + e^2 \int_{\frac{1}{2}}^{1} e^{-\frac{1}{1-x}} \, dx = \frac{5}{2} + \frac{1}{4} \int_0^\infty \left(1+\frac{u}{2}\right)^{-2} e^{-u} \, du$$

where we used the transformation $u = \frac{1}{1-x} - 2$. The rest follows easily. For example, a formula for the exponential integral is used at one point (Abramowitz and Stegun, 1970, p. 231). The last statement of the theorem is a direct consequence of the fact that h_{p2} is convex on $[\frac{1}{2}, 1]$. ■

We conclude this section by mentioning the algorithm derived from Theorem 2.5. It requires on the average 2.75 iterations and 5.5 evaluations of f per random variate. It should be used only when the number of uniform random variates per generated random variate must be kept reasonable.

Rejection method for log-concave densities on the real line

[NOTE]

We assume that f has a mode at 0 and that $f(0)=1$. Otherwise, use a linear transformation to enforce this condition.

[GENERATOR]

REPEAT

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

IF $U \leq \frac{4}{11}$

THEN $(X,Y) \leftarrow (\frac{W}{2}, 2V)$

ELSE IF $U \leq \frac{7}{11}$

THEN

Generate a uniform [0,1] random variate $W*$.

$(X,Y) \leftarrow (\frac{1}{2}+\frac{1}{2}\min(W, 2W*), V(1+2(1-X)))$

ELSE $(X,Y) \leftarrow (1-\log(W), VW)$

UNTIL $Y \leq f(X)+f(-X)$

Generate a uniform [0,1] random variate Z (this can be done by reuse of the unused portion of U).

IF $Z \leq \frac{f(X)}{f(X)+f(-X)}$

THEN RETURN X

ELSE RETURN $-X$

2.6. Non-universal rejection methods.

The universal rejection algorithm developed in the previous sections is suboptimal for individual log-concave densities in the following sense: one can find dominating curves which consist of a constant function around the mode and two exponential tails and have at the same time a smaller integral than that of the dominating curves for the universal method. The improvements are individual, because for each density we require additional information about the density not normally available in the black box model. The resulting algorithms are comparable with the ratio-of-uniforms method, where the exponential tails are replaced with quadratic tails. Since log-concave densities have sub-exponential tails, the fit will often be much better than with the ratio-of-uniforms method. More importantly, we can give a very elegant recipe for finding the optimal dominating curve which is valid for all log-concave densities.

By log-concavity, we know that $h = \log(f)$ can be majorized by the derivative of h at any point (the derivative being considered as a line). This corresponds to fitting an exponential curve over f. The problem we have is that of finding points $m+a \geq m$ and $m-b \leq m$ (where m is the mode of f) such that the area under

$$g(x) = \min(f(m), f(m+a)e^{(x-(m+a))h'(m+a)}, f(m-b)e^{(x-(m-b))h'(m-b)})$$

is minimal. We will formally allow $h'(m+a)=-\infty$ and $h'(m-b)=+\infty$. In those cases, the corresponding terms in the definition of g are either ∞ or 0. This distinction is important for compact support densities where a or b point at the extremal point in the support of f. We can offer the following general principle for finding a and b.

Theorem 2.6.

Let f be decomposed as $f_r + f_l$ where f_r, f_l refer to the parts of of f to the right and left of the mode respectively. The inverses of f_r and f_l are well-defined when evaluated at a point strictly between 0 and $f(m)$. (In case of a continuous f_r, there is no problem. If f_r has a discontinuity at y, then we know that $f_r(x)>0$ for $x<y$ and $f_r(x)=0$ for $x>y$. In that case, the inverse, if necessary, is forced to be y.)

The area under g is minimal when

$$m+a = f_r^{-1}\left(\frac{f(m)}{e}\right),$$

$$m-b = f_l^{-1}\left(\frac{f(m)}{e}\right).$$

The minimal area is given by

$$f(m)(a+b).$$

Furthermore, the minimal area does not exceed $\frac{2e}{e-1}$, and can be as small as 1. When in g we use values of $m+a$ and $m-b$ further away from the mode than those given above, the area under g is bounded from above by $f(m)(a+b)$.

Proof of Theorem 2.6.

We will prove the theorem for a monotone density f on $[m,\infty)$ only. The full theorem then follows by a simple combination of antisymmetric results. We begin thus with the inequality

$$g(x) = \min(f(m), f(m+a)e^{(x-(m+a))h'(m+a)}).$$

The cross-over point between the top curves is at a point z between m and $m+a$:

$$z = m+a+\frac{1}{h'(m+a)}\log(\frac{f(m)}{f(m+a)}) .$$

The area under the curve g to the right of m is given by

$$f(m)(z-m)+\int_z^\infty f(m+a)e^{(x-a)h'(m+a)}\,dx$$

$$= f(m)(z-m)+\frac{f(m+a)}{-h'(m+a)}e^{(z-(m+a))h'(m+a)}$$

$$= f(m)(z-m-\frac{1}{h'(m+a)})$$

$$= f(m)(a+\frac{1}{h'(m+a)}(h(m)-h(m+a)-1)) .$$

The derivative of this expression with respect to a is

$$\frac{f(m)h''(m+a)(1+h(m+a)-h(m))}{h'^2(m+a)}$$

which is zero for $h(m+a)=h(m)-1$, i.e. $f(m+a)=\frac{f(m)}{e}$. Note also that $h''(m+a)\leq 0$, and thus that the derivative is nonpositive for values of $m+a$ smaller than this threshold value, and that it is nonnegative for larger values of $m+a$, so that we do indeed have a global minimum for the area under g. At the suggested value of $m+a$, the area is given by $af(m)$. For $m+a$ larger than the suggested value, the area is bounded from above by $af(m)$, since $h'(m+a)\leq 0$, $h(m)-h(m+a)-1\geq 0$.

To obtain a distribution-free upper bound for the area $af(m)$ when a is optimally chosen, we use the inequality of Theorem 2.1. If we use the upper bound on f given there, and set it equal to $\frac{1}{e}$, then the solution is a number greater than $af(m)$. But that solution is $\frac{e}{e-1}$. Thus, for the optimal a, $af(m)\leq\frac{e}{e-1}$. ■

Theorem 2.6 is important. If a lot is known about the density in question, good rejection algorithms can be obtained. Several examples will be given below. If we want to bound f from above by a combination of pieces of exponential functions, then the area can be reduced even further although, as we will see from the examples given below, the reduction is often hardly worth the extra effort since the rejection constant is already good to begin with.

The formal algorithm is as follows:

Rejection with two exponential tails touching at m-b and m+a

[SET-UP]

m is the mode; $a, b \geq 0$ are assumed given.

$\lambda_r \leftarrow -1/h'(m+a), \lambda_l \leftarrow 1/h'(m-b)$ (where $h = \log(f)$).

$f_m \leftarrow f(m)$

$a* \leftarrow a + \lambda_r \log(\frac{f(m+a)}{f_m})$, $b* \leftarrow b + \lambda_l \log(\frac{f(m-b)}{f_m})$. ($m+a*$ and $m-b*$ are the thresholds.)

Compute the mixture probabilities: $s \leftarrow \lambda_l + \lambda_r + a* + b*$, $p_l \leftarrow \lambda_l / s$, $p_r \leftarrow \lambda_r / s$, $p_m \leftarrow (a*+b*)/s$.

[GENERATOR]

REPEAT

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

  IF $U \leq p_m$ THEN

    Generate a uniform [0,1] random variate Y (which can be done as $Y \leftarrow U/p_m$).

    $X \leftarrow m - b* + Y(a*+b*)$

    Accept $\leftarrow [V f_m \leq f(X)]$

  ELSE IF $p_m < U \leq p_m + p_r$ THEN

    Generate an exponential random variate E (which can be done as $E \leftarrow -\log(\frac{U-p_m}{p_r})$).

    $X \leftarrow m + a* + \lambda_r E$

    Accept $\leftarrow [V f_m e^{-(X-(m+a*))/\lambda_r} \leq f(X)]$ (which is equivalent to Accept $\leftarrow [V f_m e^{-E} \leq f(X)]$, or to Accept $\leftarrow [V f_m \frac{U-p_m}{p_r} \leq f(X)]$)

  ELSE

    Generate an exponential random variate E (which can be done as $E \leftarrow -\log(\frac{U-(p_m+p_r)}{1-p_m-p_r})$).

    $X \leftarrow m - b* - \lambda_l E$

    Accept $\leftarrow [V f_m e^{(X-(m-b*))/\lambda_l} \leq f(X)]$ (which is equivalent to Accept $\leftarrow [V f_m e^{-E} \leq f(X)]$, or to Accept $\leftarrow [V f_m \frac{U-(p_m+p_r)}{1-p_m-p_r} \leq f(X)]$)

UNTIL Accept

RETURN X

In most implementations, this algorithm can be considerably simplified. For one thing, the set-up step can be integrated in the algorithm. When the density is monotone or symmetric unimodal, other obvious simplifications are possible.

Example 2.1. The exponential power distribution (EPD).

The EPD density with parameter $\tau>0$ is

$$f(x) = \left(2\Gamma(1+\frac{1}{\tau})\right)^{-1} e^{-|x|^\tau} .$$

Generation for this density has been dealt with in Example IV.6.1, by transformations of gamma random variables. For $\tau \geq 1$, the density is log-concave. The values of a, b in the optimal rejection algorithm are easily found in this case: $a=b=1$. Before giving the details of the algorithm, observe that the rejection constant, the area under the dominating curve, is $f(0)(a+b)$, which is equal to $1/\Gamma(1+\frac{1}{\tau})$. As a function of τ, the rejection constant is a unimodal function with value 1 at the extremes $\tau=1$ (the Laplace density) and $\tau \uparrow \infty$ (the uniform $[-1,1]$ density), and peak at $\tau = \frac{1}{0.4616321449...}$. At the peak, the value is $\frac{1}{0.8856031944...}$ (see e.g. Abramowitz and Stegun (1970, p. 259)). Thus, uniformly over all $\tau \geq 1$, the rejection rate is extremely good. For the important case of the normal density ($\tau=2$) we obtain a value of $1/\Gamma(\frac{3}{2}) = \sqrt{\frac{4}{\pi}}$. The algorithm can be summarized as follows:

REPEAT

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

IF $U \leq 1-\frac{1}{\tau}$

THEN

$X \leftarrow U$ (note that X is uniform on $[0,1-\frac{1}{\tau}]$)

Accept $\leftarrow [\,|X|^{\tau} \leq E*]$

ELSE

Generate an exponential random variate E (which can be done as $E \leftarrow -\log(\tau(1-U))$).

$X \leftarrow 1-\frac{1}{\tau}+\frac{1}{\tau}E$

Accept $\leftarrow [\,|X|^{\tau} \leq E+E*]$

UNTIL Accept

RETURN SX where S is a random sign.

The reader will have little difficulty verifying the validity of the algorithm. Consider the monotone density on $[0,\infty)$ given by $(\Gamma(1+\frac{1}{\tau}))^{-1} e^{-x^{\tau}}$. Thus, with $m=0, a=1, h'(1)=-\tau$, we obtain $a*=1-\frac{1}{\tau}$. Since we know that $|X|^{\tau}$ is distributed as a gamma $(\frac{1}{\tau})$ random variable, it is easily seen that we have at the same time a good generator for gamma random variates with parameter less than one. For the sake of easy reference, we give the algorithm in full:

Gamma generator with parameter a less than one

REPEAT

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

IF $U \leq 1-a$

THEN

$X \leftarrow U^{\frac{1}{a}}$ (note that U is uniform on $[0,1-a]$)

Accept $\leftarrow [\,|X| \leq E*]$

ELSE

Generate an exponential random variate E (which can be done as $E \leftarrow -\log(\frac{1-U}{a})$).

$X \leftarrow (1-a+aE)^{\frac{1}{a}}$

Accept $\leftarrow [\,|X| \leq E+E*]$

UNTIL Accept

RETURN X ■

Example 2.2. Complicated densities.

For more complicated densities, the equation $f(x)=f(m)/e$ can be difficult to solve explicitly. It is always possible to take the pessimistic, or minimax, approach, by setting a and b both equal to $\frac{e}{(e-1)f(m)}$. In some cases, b can be set equal to 0. In the set-up of the algorithm, it is still necessary to evaluate the derivative of $\log(f)$ at the points $m+a$, $m-b$, but this can be done explicitly when f is given in analytic form. This approach can be automated for the beta and generalized inverse gaussian distributions, for example. When $m+a$ or $m-b$ fall outside the support of f, one should consider one-tailed dominating curves with the constant section truncated at the relevant extremal point of the support. For the beta density for example, this leads to an algorithm which resembles in many respects algorithm B2PE of Schmeiser and Babu (1980). ■

Example 2.3. Algorithm B2PE (Schmeiser and Babu, 1980) for beta random variates.

In 1980, Schmeiser and Babu proposed a highly efficient algorithm for generating beta random variates with parameters a and b when both parameters are at least one. Recall that for these values of the parameters, the beta density is log-concave. Schmeiser and Babu partition the interval [0,1] into three intervals: in the center interval, around the mode $m = \frac{a-1}{a+b-2}$, they use as dominating function a constant function $f(m)$. In the tail intervals, they use exponential dominating curves that touch the graph of f at the breakpoints. At the breakpoints, Schmeiser and Babu have a discontinuity. Nevertheless, analysis similar to that carried out in Theorem 2.6 can be used to obtain the optimal placement of the breakpoints. Schmeiser and Babu suggest placing the breakpoints at the inflection points of the density, if they exist. The inflection points are at

$$\max(m-\sigma,0)$$

and

$$\min(m+\sigma,1)$$

where $\sigma = \sqrt{\frac{m(1-m)}{a+b-3}}$ if $a+b>3$ and $\sigma=\infty$ otherwise. Two inflection points exist on [0,1] when $m-\sigma$ and $m+\sigma$ both take values in [0,1]. In that case, the area under the dominating curve is easily seen to be equal to

$$2\sigma f(m)+f(m)\left(\frac{1}{|h'(m-\sigma)|}+\frac{1}{|h'(m+\sigma)|}\right)$$
$$= f(m)\left(2\sigma+\frac{1}{\sigma(a+b-2)}((m+\sigma)(1-m-\sigma)+(m-\sigma)(1-m+\sigma))\right)$$
$$= f(m)\left(2\sigma+\frac{1}{\sigma(a+b-2)}2m(1-m)\left(1-\frac{1}{a+b-3}\right)\right)$$
$$= f(m)\left(2\sigma+2\sqrt{\frac{m(1-m)}{a+b-3}}\right)$$
$$= 4f(m)\sigma .$$

Thus, we have the interesting result that the probability mass under the exponential tails equals that under the constant center piece. One or both of the tails could be missing. In those cases, one or both of the contributions $f(m)\sigma$ needs to be replaced by $f(m)m$ or $f(m)(1-m)$. Thus, $4f(m)\sigma$ is a conservative upper bound which can be used in all cases. It can be shown (see exercises) that as $a,b\to\infty$, $4f(m)\sigma\to\sqrt{\frac{8}{\pi}}$. Furthermore, a little additional analysis shows that the expected area under the dominating curve is uniformly bounded over all values of $a,b\geq 1$. Even though the fit is far from perfect, the algorithm can be made very fast by the judicious use of the squeeze principle. Another acceleration trick proposed by Schmeiser and Babu (algorithm B4PE) consists of partitioning [0,1] into 5 intervals instead of 3, with a linear dominating curve added in the new intervals.

Algorithm B2PE for beta (a,b) random variates

[SET-UP]

$m \leftarrow \frac{a-1}{a+b-2}$

IF $a+b>3$ THEN $\sigma \leftarrow \sqrt{\frac{m(1-m)}{a+b-3}}$

IF $a<2$

THEN $x \leftarrow 0, p \leftarrow 0$

ELSE

$x \leftarrow m-\sigma$

$\lambda \leftarrow \frac{a-1}{x} - \frac{b-1}{1-x}$

$v \leftarrow e^{(a-1)\log(\frac{x}{a-1})+(b-1)\log(\frac{1-x}{b-1})+(a+b-2)\log(a+b-2)}$

$p \leftarrow \frac{v}{\lambda}$

Now, x is the left breakpoint, p the probability under the left exponential tail, λ the exponential parameter, and v the value of the normalized density f at x.

IF $b<2$

THEN $y \leftarrow 1, q \leftarrow 0$

ELSE

$y \leftarrow m+\sigma$

$\mu \leftarrow -\frac{a-1}{y} + \frac{b-1}{1-y}$

$w \leftarrow e^{(a-1)\log(\frac{y}{a-1})+(b-1)\log(\frac{1-y}{b-1})+(a+b-2)\log(a+b-2)}$

$q \leftarrow \frac{w}{\mu}$

Now, y is the left breakpoint, q the probability under the left exponential tail, μ the exponential parameter, and w the value of the normalized density f at y.

```
[GENERATOR]
REPEAT
    Generate iid uniform [0,1] random variates U,V. Set U←U(p+q+y-x).
    CASE
        U≤y-x:
            X←x+U (X is uniformly distributed on [x,y])
            IF X<m
                THEN Accept ←[V≤v+(X-x)(1-v)/(m-x)]
                ELSE Accept ←[V≤w+(y-X)(1-w)/(y-m)]
        y-x<U≤y-x+p:
            U←(U-(y-x))/p (create a new uniform random variate)
            X←x+(1/λ)log(U) (X is exponentially distributed)
            Accept ←[V≤(λ(X-x)+1)/U]
            V←VUv (create a new uniform random variate)
        y-x+p≤U:
            U←(U-(y-x+p))/q (create a new uniform random variate)
            X←y-(1/μ)log(U) (X is exponentially distributed)
            Accept ←[V≤(μ(y-X)+1)/U]
            V←VUw (create a new uniform random variate)
    IF NOT Accept THEN
        T←log(V)
        IF T>-2(a+b-2)(X-m)^2
            THEN
            Accept ←[T≤(a-1)log(X/(a-1))+(b-1)log((1-X)/(b-1))+(a+b-2)log(a+b-2)]
UNTIL Accept
RETURN X
```

The algorithm can be improved in many ways. For example, many constants can be computed in the set-up step, and quick rejection steps can be added when X falls outside [0,1]. Note also the presence of another quick rejection step, based upon the following inequality:

$$\log\left(\frac{f(x)}{f(m)}\right) \le -2(a+b-2)(x-m)^2 .$$

The quick rejection step is useful in situations just like this, i.e. when the fit is not very good. ■

Example 2.4. Tails of log-concave densities.

When f is log-concave, and a random variate from the right tail of f, truncated at $t > m$ where m is the mode of f, is needed, one can always use the exponential majorizing function:

$$f(x) \leq f(t) e^{\frac{f'(t)}{f(t)}(x-t)} \qquad (x \geq t) .$$

The first systematic use of these exponential tails can be found in Schmeiser (1980). The expected number of iterations in the rejection algorithm is

$$\frac{f^2(t)}{|f'(t)| \int_t^\infty f} . \blacksquare$$

2.7. Exercises.

1. **The Pearson IV density.** The Pearson IV density on R has two parameters, $m > \frac{1}{2}$ and $s \in R$, and is given by

$$f(x) = \frac{c}{(1+x^2)^m} e^{-s \arctan x} .$$

Here c is a normalization constant. For $s = 0$ we obtain the t density. Show the following:

A. If X is Pearson IV (m, s), and $m \geq 1$, then $\arctan(X)$ has a log-concave density

$$g(x) = c \cos^{2(m-1)}(x) e^{-sx} \qquad (|x| \leq \frac{\pi}{2}) .$$

B. The mode of g occurs at $\arctan(-\frac{s}{2(m-1)})$.

C. Give the complete rejection algorithm (exponential version) for the distribution. For the symmetric case of the t density, give the details of the rejection algorithm with rejection constant 2.

D. Find a formula for the computation of c .

2. Prove that a mixture of two log-concave densities is not necessarily log-concave.

3. Give the details of the rejection algorithm that is based upon the inequality of Theorem 2.4.

4. Log-concave densities can also occur in R^d. For example, the multivariate normal density is log-concave. The closure under convolutions also holds in R^d (Davidovic et al., 1969), and marginals of log-concave densities are again log-concave (Prekopa, 1973). Unfortunately, it is useless to try to look for a generalization of the inequalities of this section to R^d with $d \geq 2$ because of the following fact which you are asked to show: the supremum over all log-concave densities with mode at 0 and $f(0)=1$ is the constant function 1.

5. To speed up the algorithms of this section at the expense of preprocessing, we can compute the normalized log-concave density at $n > 1$ carefully selected points, and use rejection (perhaps combined with squeezing) with a dominating curve consisting of several pieces. Can you give a universal recipe for locating the points of measurement so that the rejection constant is guaranteed to be smaller than a function of n only, and this function of n tends to 1 as $n \rightarrow \infty$? Make sure that random variate generation from the dominating density is not difficult, and provide the details of your algorithm.

6. This is about the area under the dominating curve in algorithm B2PE (Schmeiser and Babu, 1980) for beta random variate generation (Example 2.3). Assume throughout that $a, b \geq 1$.

   (i) $\sigma \leq m$ if and only if $a \geq 2$, $\sigma \leq 1-m$ if and only if $b \geq 2$. (Thus, for $a, b \geq 2$, the area under the dominating curve is precisely $4f(m)\sigma$.)

   (ii) $\lim_{a,b \rightarrow \infty} 4f(m)\sigma = \sqrt{\frac{8}{\pi}}$. Use Stirling's approximation.

   (iii) The area under the dominating curve is uniformly bounded over all $a, b \geq 1$. Use sharp inequalities for the gamma function to bound $f(m)$. Consider 3 cases: both $a, b \geq 2$, one of a, b is ≥ 2, and one is < 2, and both a, b are < 2. Try to obtain as good a uniform bound as possible.

   (iv) Prove the quick rejection inequality used in the algorithm:

$$\log\left(\frac{f(x)}{f(m)}\right) \leq -2(a+b-2)(x-m)^2 .$$

3. INEQUALITIES FOR DENSITIES.

3.1. Motivation.

The previous section has shown us the utility of upper bounds in the development of universal methods or black box methods. The strategy is to obtain upper bounds for densities in a large class which

(i) have a small integral;

(ii) are defined in terms of quantities that are either computable or present in the definition of the class.

For the log-concave densities with mode at 0 we have for example obtained an upper bound in section VII.2 with integral 4, which requires knowledge of the position of the mode (this is in the definition of the class), and of the value of $f(0)$ (this can be computed). In general, quantities that are known could include:

A. A uniform upper bound for f (called M);
B. The r-th moment μ_r;
C. The value of a functional $\int f^{\alpha}$;
D. A Lipschitz constant;
E. A uniform bound for the s-th derivative;
F. The entire moment generating function $M(t)$, $t \in R$;
G. The entire distribution function $F(x)$, $x \in R$;
H. The support of f.

When this information is combined in various ways, a multitude of useful dominating curves can be obtained. The goodness of a dominating curve is measured in terms of its integral and the ease with which random variates with a density proportional to the dominating curve can be generated. We show by example how some inequalities can be obtained.

3.2. Bounds for unimodal densities.

Let us start with the class of monotone densities on [0,1] which are bounded by M. Note that if M is unknown, it can easily be computed as $f(0)$. Thus, the only true restriction is that we must know that f vanishes off [0,1]. The trivial inequality

$$f(x) \leq M I_{[0,1]}(x)$$

is not very useful, since the integral under the dominating curve is M. There are several ways to increase the efficiency:

1. Use a table method by evaluating in a set-up step the value of f at many points. Basically, the dominating curve is piecewise constant and hugs the curve of f much better. These methods are very fast but the need for extra storage (usually growing with M) and an additional preprocessing step makes this approach somehow different. It should not be compared with

methods not requiring these extra costs. It will be developed systematically in chapter VIII.

2. Use as much information as possible to improve the bound. For example, in the inequality $f(x) \leq M$, the monotonicity is not used.
3. Ask the user if he has additional knowledge in the form of moments, quantiles, functionals and the like. Then construct good dominating curves.

We will illustrate approaches 2 and 3. For all monotone densities, the following is true:

Theorem 3.1.

For all monotone densities f on $[0,\infty)$,

$$f(x) \leq \frac{1}{x} .$$

If f is also convex, then

$$f(x) \leq \frac{1}{2x} .$$

Proof of Theorem 3.1.

Fix $x > 0$. Then, by monotonicity,

$$xf(x) \leq \int_0^x f(y)\,dy \leq 1 .$$

When f is also convex, we can in fact use a geometrical argument: if we wish to find the convex f for which $f(x)$ is maximal, it suffices to consider only triangles. This class is the class of all densities $2a(1-ax)_+$, $0 \leq x \leq \frac{1}{a}$. Thus, we find a for which $f(x)$ is maximal. Setting the derivative with respect to a equal to 0 gives the equation $1-ax-ax=0$, i.e. $a=\frac{1}{2x}$. Resubstitution gives the bound. ■

The bounds of Theorem 3.1 cannot be improved in the sense that for every x, there exists a monotone (or monotone and convex) f for which the upper bound is attained. If we return now to the class of monotone densities on [0,1] bounded by M, we see that the following inequality can be used:

$$f(x) \leq \min(M,\frac{1}{x}) I_{[0,1]}(x) .$$

The area under the dominating curve is $1+\log(M)$. Clearly, this is always less than M. In most applications the improvement in computer time obtainable by using the last inequality is noticeable if not spectacular. Let us therefore take a moment to give the details of the corresponding rejection algorithm. The dominating density for rejection is

$$g(x) = \frac{1}{1+\log(M)} \min(M, \frac{1}{x}) I_{[0,1]}(x) .$$

It has distribution function

$$\begin{cases} \frac{Mx}{1+\log(M)} & , 0 \le x \le \frac{1}{M} \\ \frac{1+\log(Mx)}{1+\log(M)} & , \frac{1}{M} \le x \le 1 . \end{cases}$$

Using inversion for generation from g, we obtain

Rejection algorithm for monotone densities on [0,1] bounded by M

REPEAT

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

IF $U \le \frac{1}{1+\log(M)}$

THEN

$X \leftarrow \frac{U}{M}(1+\log(M))$

IF $VM \le f(X)$ THEN RETURN X

ELSE

$X \leftarrow \frac{1}{M} e^{U(1+\log(M))-1}$

IF $V \le Xf(X)$ THEN RETURN X

UNTIL False

When f is also convex, we can use the inequality

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

where

$$g(x) = \frac{2}{1+\log(2M)} \min(M, \frac{1}{2x}) I_{[0,1]}(x) .$$

It has distribution function

$$\begin{cases} \frac{2Mx}{1+\log(2M)} & , 0 \le x \le \frac{1}{2M} \\ \frac{1+\log(2Mx)}{1+\log(2M)} & , \frac{1}{2M} \le x \le 1 . \end{cases}$$

Using inversion for generation from g, we obtain

Rejection algorithm for monotone convex densities on [0,1] bounded by M

REPEAT

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

IF $U \leq \frac{1}{1+\log(2M)}$

THEN

$X \leftarrow \frac{U}{2M}(1+\log(2M))$

IF $VM \leq f(X)$ THEN RETURN X

ELSE

$X \leftarrow \frac{1}{2M} e^{U(1+\log(2M))-1}$

IF $V \leq 2Xf(X)$ THEN RETURN X

UNTIL False

The expected number of iterations now is $\frac{1+\log(2M)}{2}$, which is for large M roughly speaking half of the expected number of iterations for the nonconvex cases.

The function $\frac{1}{x}$ is not integrable on $[1,\infty)$, so that Theorem 3.1 is useless for handling infinite tails of monotone densities. We have to tuck the tails under some integrable function, yet uniformly over all monotone densities we cannot get anything better than $\frac{1}{x}$. Thus, additional information is required.

Theorem 3.2.

Let f be a monotone density on $[0,\infty)$.

A. If $\int x^r f(x)\,dx \leq \mu_r < \infty$ where $r > 0$, then

$$f(x) \leq \frac{(r+1)\mu_r}{x^{r+1}} \quad (x>0) .$$

B. In any case, for all $0<\alpha\leq 1$,

$$f(x) \leq \frac{(\int f^{\alpha})^{\frac{1}{\alpha}}}{x^{\frac{1}{\alpha}}} \quad (x>0) .$$

Proof of Theorem 3.2.

For part A we proceed as follows:

$$\mu_r \geq \int_0^x y^r f(y)\, dy \geq \frac{f(x) x^{r+1}}{r+1} .$$

For part B, we use the trivial observation

$$x f^{\alpha}(x) \leq \int f^{\alpha} . \blacksquare$$

For monotone densities on $[0,\infty)$, bounded by $M = f(0)$, Theorem 3.2 provides us with bounds of the form

$$f(x) \leq \min(M, \frac{A}{x^a}) \quad (x > 0)$$

where we can take (A, a) as follows:

Information	A	a
$\int x^r f(x) dx \leq \mu_r < \infty$	$(r+1)\mu_r$	$r+1$
$(\int f^{\alpha})^{\frac{1}{\alpha}} \leq \nu_\alpha < \infty$	ν_α	$\frac{1}{\alpha}$

In all cases, the area under the dominating curve is

$$\frac{a}{a-1} A^{\frac{1}{a}} M^{\frac{a-1}{a}} .$$

Furthermore, random variate generation for the dominating density can be done quite easily via the inversion method or the inverse-of-f method (section IV.6.3):

Theorem 3.3.

Let g be the density on $[0,\infty)$ proportional to $\min(M,\frac{A}{x^a})$ where $M>0,A>0,a>1$ are parameters. Then the following random variables X have density g:

A. $X=(\frac{A}{M})^{\frac{1}{a}}\frac{U}{V^{a-1}}$ where U,V are iid uniform $[0,1]$ random variates.

B. Let $x*$ be $(\frac{A}{M})^{\frac{1}{a}}$ and let U be uniform on $[0,1]$. Then $X\leftarrow\frac{a}{a-1}Ux*$ if $U\leq\frac{a-1}{a}$, and $X\leftarrow\frac{x*}{(aU-(a-1))^{\frac{1}{a-1}}}$ else.

Proof of Theorem 3.3.

By the inverse-of-f method (section IV.6.3), it suffices to note that a random variate with monotone density f can be obtained as $Uf^{-1}(Y)$ where Y has density f^{-1}. It is easy to see that for monotone g not necessarily integrating to one, $Ug^{-1}(Y)$ has density proportional to g if Y has density proportional to g^{-1}. In our case, $g^{-1}(y)=(\frac{A}{y})^{\frac{1}{a}}$ $,0\leq y\leq M$. To generate Y with density proportional to this, we apply the inversion method. Verify that $MV^{\frac{a}{a-1}}$ has distribution function $(\frac{y}{M})^{1-\frac{1}{a}}$ on $[0,M]$, which yields a density proportional to g^{-1}. Plugging this Y back into $Ug^{-1}(Y)$ proves part A.

Part B is obtainable by straightforward inversion. Note that $x*$ is the breakpoint where $M=\frac{A}{x^a}$, that $\int_0^{x*}g=Mx*$, and that $\int_{x*}^{\infty}g=\frac{A}{a-1}x*^{-(a-1)}$. The sum of the two areas is

$$A^{\frac{1}{a}}M^{1-\frac{1}{a}}(1+\frac{1}{a-1}) .$$

Thus, with probability $\frac{a-1}{a}$, X is distributed uniformly on $[0,x*]$, and with the complementary probability, X is distributed as $\frac{x*}{V^{\frac{1}{a-1}}}$ where V is uniformly distributed on $[0,1]$ (the latter random variable has density decreasing as x^{-a} on $[x*,\infty)$). The uniform random variates needed here can be recovered from the uniform random variate U used in the comparison with $\frac{a-1}{a}$: given that $U\leq\frac{a-1}{a}$, $U\frac{a}{a-1}$ is again uniform. Given that $U>\frac{a-1}{a}$, $aU-(a-1)$ is in turn uniformly distributed on [0,1]. ■

For the sake of completeness, we will now give the rejection algorithm for generating random variates with density f based upon the inequality

$$f(x) \leq \min(M, \frac{A}{x^a}) \quad (x \geq 0) .$$

Rejection method based upon part A of Theorem 3.3

REPEAT

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

$Y \leftarrow MV^{\frac{a}{a-1}}$

$X \leftarrow U(\frac{A}{Y})^{\frac{1}{a}}$

UNTIL $Y \leq f(X)$

RETURN X

The validity of this algorithm is based upon the fact that $(Y, Ug^{-1}(Y))=(Y, X)$ is uniformly distributed under the curve of g^{-1}. By swapping coordinate axes, we see that (X, Y) is uniformly distributed under g, and can thus be used in the rejection method. Note that the power operation is unavoidable. Based upon part B, we can use rejection with fewer powers.

Rejection method based upon part B of Theorem 3.3.

REPEAT

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

IF $U \leq \frac{a-1}{a}$

THEN

$X \leftarrow \frac{a}{a-1} U x*$

IF $VM \leq f(X)$ THEN RETURN X

ELSE

$X \leftarrow x*(aU-(a-1))^{-\frac{1}{a-1}}$

IF $VA \leq X^a f(X)$ THEN RETURN X

UNTIL False

For both implementations, the expected number of computations of f is equal to the expected number of iterations,

$$E(N) = \frac{a}{a-1} A^{\frac{1}{a}} M^{\frac{a-1}{a}} .$$

It is instructive to analyze this measure of the performance in more detail. Consider the moment version for example , where $A=(r+1)\mu_r$, $a=r+1$ and μ_r is the r-th moment of the monotone density. We have

Theorem 3.4.

Let $E(N), M, r, A, a, \mu_r$ be as defined above. Then for all monotone densities on $[0,\infty)$,

$$E(N) \geq 1+\frac{1}{r} .$$

For all monotone densities that are concave on their support,

$$E(N) \leq 2(1+\frac{1}{r})(r+2)^{-\frac{1}{r+1}} \leq 2(1+\frac{1}{r}) .$$

Finally, for all monotone log-concave densities,

$$E(N) \leq (1+\frac{1}{r})(\Gamma(r+2))^{\frac{1}{r+1}} \sim \frac{r+1}{e} (\text{as } r \rightarrow \infty) .$$

Proof of Theorem 3.4.

We start from the expression

$$E(N) = (1+\frac{1}{r})\,((r+1)M^r \mu_r)^{\frac{1}{r+1}} .$$

The product $M\mu_r$ is scale invariant, so that we can take $M=1$ without loss of generality. For all such bounded densities, we have $1-F(x) \geq (1-x)_+$. Thus,

$$\mu_r = \int_0^\infty x^r f(x)\,dx = \int_0^\infty r x^{r-1}(1-F(x))\,dx$$

$$\geq \int_0^1 r x^{r-1}(1-x)\,dx$$

$$= 1-\frac{r}{r+1} = \frac{1}{r+1} .$$

This proves the first part of the theorem. Note that we have implicitly used the fact that every random variable with a density bounded by 1 on $[0,\infty)$ is stochastically larger than a uniform $[0,1]$ random variate.

For the second part, we use the fact that all random variables with a monotone concave density satisfying $f(0)=M=1$ are stochastically smaller than a random variable with density $(1-\frac{x}{2})_+$ (exercise 3.1). Thus, for this density,

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

Resubstitution gives us part B for concave densities. Finally, for log-concave densities we need the fact that $f(0)X$ is stochastically smaller than an exponential random variate. Thus, in particular,

$$M^r \mu_r \leq \int_0^\infty y^r e^{-y}\,dy = \Gamma(r+1) .$$

This proves the last part of the theorem. ∎

A brief discussion of Theorem 3.4 is in order here. First of all, the inequalities are quite inefficient when r is near 0 in view of the lower bound $E(N) \geq 1+\frac{1}{r}$. What is important here is that for important subclasses of monotone densities, the performance is uniformly bounded provided that we know the r-th moment of the density in case. For example, for the log-concave densities, we have the following values for the upper bound for $E(N)$:

r	$E(N) \le$	Approximate value
1	$\frac{4}{\sqrt{3}}$	2.3094...
2	$\frac{3}{4^{\frac{1}{3}}}$	1.88988...
3	$\frac{8}{3\,5^{\frac{1}{4}}}$	1.7833...
4	$\frac{5}{2\,6^{\frac{1}{5}}}$	1.7470...
5	$\frac{12}{5\,7^{\frac{1}{6}}}$	1.7352...
6	$\frac{7}{3\,8^{\frac{1}{7}}}$	1.73366...
7	$\frac{16}{7\,9^{\frac{1}{8}}}$	1.7367...
$\uparrow\infty$	$\uparrow 2$	

The upper bound is minimal for r near 6. The algorithm is guaranteed to perform at its best when the sixth moment is known. In the exercises, we will develop a slightly better inequality for concave monotone densities. One of the features of the present method is that we do not need any information about the support of f - such information would be required if ordinary rejection from a uniform density is used. Unfortunately, very few important densities are concave on their support, and often we do not know whether a density is concave or not.

The family of log-concave densities is more important. The upper bound for $E(N)$ in Theorem 3.4 has acceptable values for the usual values of r:

r	$E(N) \le$	Approximate value
1	$\sqrt{8}$	2.82...
2	$\frac{3}{2}6^{\frac{1}{3}}$	2.7256...
3	$\frac{4}{3}24^{\frac{1}{4}}$	2.9511...
$\uparrow\infty$	$\uparrow\infty$	

In this case, the optimal integer value of r is 2. Note that if μ_r is not known, but is replaced in the algorithm and the analysis by its upper bound $\frac{\Gamma(r+1)}{M^r}$, then both the algorithm and the performance analysis of Theorem 3.4 remain valid. In that case, we obtain a black box method for all log-concave densities on $[0,\infty)$ with mode at 0, as in the previous section. For $r=2$, the expected number of iterations (about 2.72) is about 36% larger than the algorithm of the previous section which was specially developed for log-concave densities only.

3.3. Densities satisfying a Lipschitz condition.

We say that a function f is **Lipschitz** (C) when

$$\sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|} \leq C .$$

When f is absolutely continuous with a.e. derivative f', then we can take $C = \sup |f'|$. Unfortunately, some important functions are not Lipschitz, such as $\sqrt{x}$. However, many of these functions are Lipschitz of order α: formally, we say that f is Lipschitz of order α with constant C (and we write $f \in Lip_{\alpha}(C)$) when

$$\sup_{x \neq y} \frac{|f(x)-f(y)|}{|x-y|^{\alpha}} \leq C .$$

Here $\alpha \in (0,1]$ is a constant. It can be shown (exercise 3.6) that the classes $Lip_{\alpha}(C)$ for $\alpha > 1$ contain no densities. The fundamental inequality for the Lipschitz classes is given below:

Theorem 3.5.

When f is a density in $Lip_{\alpha}(C)$ for some $C > 0$, $\alpha \in (0,1]$, then

$$f(x) \leq \left(\min(F(x),1-F(x)) \frac{\alpha+1}{\alpha} C^{\frac{1}{\alpha}}\right)^{\frac{\alpha}{\alpha+1}} .$$

Here F is the distribution function for f. In particular, for $\alpha=1$, we have

$$f(x) \leq \sqrt{2C \min(F(x),1-F(x))} .$$

Proof of Theorem 3.5.

Fix x, and define $y = f(x)$. Then fix $z > x$. We clearly have

$$f(z) \geq f(x) - C(z-x)^{\alpha} .$$

The density f which yields the maximal value for $f(x)$ is equal to the lower bound for $f(z)$ given above. It vanishes beyond

$$z* = x + \left(\frac{f(x)}{C}\right)^{\frac{1}{\alpha}} .$$

By integration of the previous inequality we have

$$1-F(x) \geq \int_x^{z*} (f(x) - C(z-x)^{\alpha}) \, dz$$

$$= f(x)(z*-x) - \frac{C(z*-x)^{\alpha+1}}{\alpha+1}$$

$$= f(x)\left(\frac{f(x)}{C}\right)^{\frac{1}{\alpha}} - \frac{C}{\alpha+1}\left(\frac{f(x)}{C}\right)^{\frac{\alpha+1}{\alpha}}$$

$$= f(x)^{\frac{\alpha+1}{\alpha}} \frac{\alpha}{\alpha+1} C^{-\frac{1}{\alpha}} .$$

By symmetry, the same lower bound is valid for $F(x)$. Rearranging the terms gives us our result. ∎

Theorem 3.5 provides us with an important bridging device. For many distributions, tail inequalities are readily available: standard textbooks usually give Markov's and Chebyshev's inequalities, and these are sometimes supplemented by various exponential inequalities. If f is in $Lip_\alpha(C)$ on $(0,\infty)$ (thus, a discontinuity could occur at 0), then we still have

$$f(x) \le \left((1-F(x)) \frac{\alpha+1}{\alpha} C^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{\alpha+1}} .$$

Before we proceed with some examples of the use of Theorem 3.5, we collect some of the best known tail inequalities in a lemma:

Lemma 3.1.

Let F be a distribution function of a random variable X. Then the following inequalities are valid:

A. $P(|X| \ge x) \le \frac{E(|X|^r)}{|x|^r}$, $r>0$ (Chebyshev's inequality) .

B. $1-F(x) \le M(t)e^{-tx}$, $t>0$ where $M(t)=E(e^{tX})$ is the moment generating function (Markov's inequality); note that by symmetry, $F(x) \le M(-t)e^{tx}$, $t>0$.

C. For log-concave f with mode at 0 and support on $[0,\infty)$, $1-F(x) \le e^{-f(0)x}$.

D. For monotone f on $[0,\infty)$, $1-F(x) \le (\frac{r}{r+1})^r \frac{E(|X|^r)}{|x|^r}$, $x,r>0$ (Narumi's inequality).

Proof of Lemma 3.1.

Parts A and B are but special cases of a more general inequality: assume that ψ is a nonnegative function at least equal to one on a set A. Then

$$P(X\in A) = \int_A dF(x) \le \int_A \psi(x)\, dF(x) \le E(\psi(X)) .$$

For part A, take $A=[x,\infty)\cup(-\infty,x]$ and $\psi(y)=\frac{|y|^r}{|x|^r}$. For part B, take

$A=[x,\infty)$ and $\psi(y)=e^{t(y-x)}$ for some $t>0$. Part C follows simply from the fact that for log-concave densities on $[0,\infty)$ with mode at 0, $f(0)X$ is stochastically smaller than an exponential random variable. Thus, only part D seems non-trivial; see exercise 3.7. ■

If inequalities other than those given here are needed, the reader may want to consult the survey article of Savage (1961) or the specialized text by Godwin (1964).

Example 3.1. Convex densities.

When a convex density f on $[0,\infty)$ is in $Lip_1(C)$, we can take $C=f'(0)$. By Narumi's inequality for monotone densities,

$$f(x) \le \min\left(f(0), \frac{\sqrt{2f'(0)(\frac{r}{r+1})^r \mu_r}}{x^{\frac{r}{2}}}\right),$$

where $\mu_r=E(|X|^r)$. This is of the general form dealt with in Theorem 3.3. It should be noted that for this inequality to be useful, we need $r>2$. ■

Example 3.2. Densities with known moment generating function.

Patel, Kapadia and Owen (1976) give several examples of the use of moment generating functions $M(t)$ in statistics. Using the exponential version of Markov's inequality, we can bound any $Lip_1(C)$ density as follows:

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

Here $t>0$ is a constant. There is nothing that keeps us from making t depend upon x except perhaps the simplicity of the bound. If we do not wish to upset this simplicity, we have to take one t for all x. When f is also symmetric about the origin, then the bound can be written as follows:

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

where $g(x)=\frac{t}{4}e^{-\frac{t}{2}|x|}$ is the Laplace density with parameter $\frac{t}{2}$, and $c=\sqrt{32\ C\ M(t)/t^2}$ is a constant which depends upon t only. If this bound is used in a rejection algorithm, the expected number of iterations is c. Thus, the best value for t is the value that minimizes $M(t)/t^2$. Note that c increases with C (decreasing smoothness) and with $M(t)$ (increasing size of the tail). Having picked t, the following rejection algorithm can be used:

Rejection method for symmetric Lipschitz densities with known moment generating function

[SET-UP]

$b \leftarrow \sqrt{2CM(t)}$

[GENERATOR]

REPEAT

Generate E, U, independent exponential and uniform [0,1] random variates.

$X \leftarrow \frac{2}{t} E$

UNTIL $Ube^{-E} \leq f(X)$

RETURN SX where S is a random sign. ■

Example 3.3. The generalized gaussian family.

The generalized gaussian family of distributions contains all distributions for which for some constant $s \geq 0$, $M(t) \leq e^{s^2t^2/2}$ for all t (Chow, 1966). The mean of these distributions exists and is 0. Also, as shown by Chow (1966), both $1-F(x)$ and $F(-x)$ do not exceed $e^{-x^2/(2s^2)}$ for all $x > 0$. Thus, by Theorem 3.5, when $f \in Lip_1(C)$,

$$f(x) \leq s\sqrt{8C\pi}\left(\frac{1}{s\sqrt{4\pi}} e^{-\frac{x^2}{4s^2}}\right) .$$

The function in parentheses is a normal $(0, s\sqrt{2})$ density. The rejection constant is $s\sqrt{8C\pi}$. In its crudest form the algorithm can be summarized as follows:

Rejection algorithm for generalized gaussian distributions with a Lipschitz density

REPEAT

Generate N, E, independent normal and exponential random variates.

$X \leftarrow Ns\sqrt{2}$

UNTIL $-\frac{N^2}{2}-E \leq \log(\frac{f(X)}{\sqrt{2C}})$

RETURN X ■

Example 3.4. Densities with known moments.

The previous three examples apply to rather small families of distributions. If only the r-th absolute moment μ_r is known, the we have by Chebyshev's inequality,

$$1-F(x) \leq \frac{\mu_r}{|x|^r}$$

for all $x, r>0$. This leads to the inequality

$$f(x) \leq \sqrt{2C} \min\left(1, \frac{\sqrt{\mu_r}}{|x|^{\frac{r}{2}}}\right),$$

which is only useful to us for $r>2$ (otherwise, the dominating function is not integrable). The integral of the dominating curve is $\sqrt{8C}\frac{r}{r-2}\mu_r^{\frac{1}{r}}$. Just which r is best depends upon the distribution: $\frac{r}{r-2}$ decreases monotonically with r whereas $\mu_r^{\frac{1}{r}}$ is nondecreasing in r (this is known as Lyapunov's inequality, which can be obtained in one line from Jensen's inequality). ■

Example 3.5. Log-concave densities.

Assume that f is log-concave with mode at 0 and support contained in $[0,\infty)$. Using $1-F(x) \leq e^{-xf(0)}$, we observe that

$$f(x) \leq \frac{\sqrt{8C}}{f(0)}\left(\frac{f(0)}{2}e^{-\frac{xf(0)}{2}}\right) \quad (x>0) .$$

The top bound is $\frac{\sqrt{8C}}{f(0)}$ times a Laplace density. It is thus not difficult to see that the following algorithm is useful:

Rejection method for log-concave Lipschitz densities

REPEAT

Generate iid exponential random variates E_1,E_2

$X \leftarrow \frac{2}{f(0)}E_1$

UNTIL $-E_2-E_1 \leq \log(\frac{f(X)}{\sqrt{2C}})$

RETURN X ■

3.4. Normal scale mixtures.

Many distributions in statistics can be written as mixtures of normal densities in which the variance is the mixture parameter. These normal scale mixtures have far-reaching applications ranging from modeling to mathematical statistics. The corresponding random variables X are thus distributed as NY, where N is normal, and Y is a positive-valued random variable. The class of normal scale mixtures is selected here to be contrasted against the class of log-concave densities. It should be clear that we could have picked other classes of mixture distributions.

There are two situations that should be clearly distinguished: in the first case, the distribution of Y is known. In the second case, the distribution of Y is not explicitly given, but it is known nevertheless that X is a normal scale mixture. The first case is trivial: one just generates N and Y and exits with NY. In the table below, some examples are given:

DENSITY OF X	DENSITY OF Y
Cauchy	Density of $1/N$ where N is normal
Laplace	Density of $1/\sqrt{2E}$ where E is exponential
Logistic	Density of $2K$ where K has the Kolmogorov-Smirnov distribution
t_a	Density of $\sqrt{\frac{2a}{G}}$ where G is gamma $(\frac{a}{2})$
Symmetric stable (α)	Density of $\sqrt{S}$ where S is positive stable $(\frac{\alpha}{2})$

This table is far from complete, and all the representations have been known for quite some time. For the inclusion of the symmetric stable, see e.g. Feller (1971), and for the inclusion of the logistic, see e.g. Andrews and Mallows (1974). In fact, it is known that an even density f is a normal scale mixture if and only if the derivatives of $f(\sqrt{x})$ are of alternating sign for all $x>0$ (Kelker, 1971). Unfortunately, for all the densities given in the table, efficient direct methods of generation are known, so there is no reason why one should use the decomposition.

The more interesting case is the one in which we just know that the distribution is a normal scale mixture. To develop universal rejection methods for this class of distributions, general inequalities are needed. The following inequalities are useful for this purpose:

Theorem 3.6.

Let f be the density of a normal scale mixture, and let X be a random variable with density f. Then f is symmetric and unimodal, $f(x) \leq f(0)$, and for all $a \geq -1$,

$$f(x) \leq C_a \frac{\mu_a}{|x|^{1+a}}$$

where

$$\mu_a = E(|X|^a)$$

is the a-th absolute moment of X, and

$$C_a = \left(\frac{1+a}{e}\right)^{\frac{1+a}{2}} \frac{1}{2^{\frac{1+a}{2}} \Gamma(\frac{1+a}{2})} .$$

For $a=1$ and $a=2$, we have respectively,

$$f(x) \leq \min(f(0), \frac{E(|X|)}{e|x|^2}) ,$$

$$f(x) \leq \min(f(0), (\frac{3}{e})^{\frac{3}{2}} \frac{E(X^2)}{\sqrt{2\pi}|x|^3}) .$$

The areas under the dominating curves are respectively, $\frac{4}{\sqrt{e}}\sqrt{f(0)\mu_1}$, and $C(\mu_2 f(0)^2)^{1/3}$ where $C=3(3/e)^{1/2}(2\pi)^{-1/6}$.

Proof of Theorem 3.6.

The unimodality is obvious. The upper bounds for f follow directly from similar upper bounds for the normal density. Note that we have, for all $x, \sigma>0$,

$$e^{-\frac{x^2}{2\sigma^2}} \leq (\frac{\sigma}{|x|})^{1+a} (\frac{1+a}{e})^{\frac{1+a}{2}} .$$

Observe that

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

where Y is a random variable used in the mixture (recall that $X=NY$). Using the normal-polynomial bound mentioned above, this leads to the inequality

$$f(x) \leq E(Y^a) \frac{1}{\sqrt{2\pi}|x|^{1+a}} (\frac{1+a}{e})^{\frac{1+a}{2}} .$$

But in view of the relationship $X = NY$, we have $E(Y^a) = E(|X|^a)/E(|N|^a)$. Now, use the fact that $E(|N|^a)\sqrt{2\pi} = 2^{\frac{1+a}{2}}\Gamma(\frac{1+a}{2})$ (which follows by definition of the gamma integral). This gives the main inequality. The special cases are easily obtained from the main inequality, as are the areas under the dominating curves. ■

The algorithms of section 3.2 are once again applicable. However, we are in much better shape now. If we had just used the unimodality, we would have obtained the inequality

$$f(x) \leq \min(f(0), \frac{a+1}{2}\frac{\mu_a}{|x|^{a+1}}),$$

which is useful for $a > 0$. See the proof of Theorem 3.2. The area under this dominating curve is larger than the corresponding area for Theorem 3.6, which should come as no surprise because we are using more information in Theorem 3.6. Notice that, just as in section 3.2, the areas under the dominating curves are scale invariant. The choice of a depends of course upon f. Because the class of normal mixtures contains densities with arbitrarily large tails, we may be forced to choose a very close to 0 in order to make μ_a finite. Such a strategy is appropriate for the symmetric stable density.

3.5. Exercises.

1. Prove the following fact needed in Theorem 3.4: all monotone densities on $[0,\infty)$ with value 1 at 0 and concave on their support are stochastically smaller than the triangular density $f(x) = (1-\frac{x}{2})_+$, i.e. their distribution functions all dominate the distribution function of the triangular density.
2. In the rejection algorithm immediately preceding Theorem 3.4, we exit some of the time with $X \leftarrow \frac{x^*}{\sqrt{aU-(a-1)}}$. The square root is costly. The special case $a = 3$ is very important. Show that $\sqrt{3U-2}$ is distributed as $\max(3U-2, W)$ where W is another uniform [0,1] random variate.
3. **Concave monotone densities.** In this exercise, we consider densities f which are concave on their support and monotone on $[0,\infty)$. Let us use $M = f(0)$, $\mu_r = \int x^r f(x)\,dx$.

   A. Show that $f(x) \leq \min(M, (\frac{2\mu_r(r+1)}{x^{r+1}} - M)_+)$.

   B. Show that the area under the dominating curve is $2-2^{\frac{1}{r+1}}$ times the area under the dominating curve shown in Theorem 3.4. That is, the area is

$$(2-2^{\frac{1}{r+1}})(1+\frac{1}{r})M^{\frac{r}{r+1}}((r+1)\mu_r)^{\frac{1}{r+1}} .$$

C. Noting that the improvement is most outspoken for $r=1$ $(2-\sqrt{2}\approx 0.59)$ and $r=2$ and that it is negligible when r is very large, give the details of the rejection algorithm for these two cases.

4. Give the strongest counterparts of Theorems 3.1-3.4 you can find for unimodal densities on the real line with a mode at 0. Because this class contains the class dealt with in the section, all the bounds given in the section remain valid for $f(|x|)$, and this leads to performances that are precisely double those of the various theorems. Mimicking the development of section VII.2 for log-concave densities, this can be improved if we know $F(0)$, the value of the distribution function at 0, or are willing to apply Brent's mirror principle (generate a random variate X with density $f(x)+f(-x)$,$x>0$, and exit with X or $-X$ with probabilities $\frac{f(x)}{f(x)+f(-x)}$ and $\frac{f(-x)}{f(x)+f(-x)}$ respectively). Work out the details.

5. Compare the rejection constant of Example 3.5 (log-concave densities on $[0,\infty)$) with 2, the rejection constant obtained for the algorithm of section VII.2. Show that it is always at least 2, that is, show that for all log-concave densities on $[0,\infty)$ belonging to $Lip_1(C)$,

$$\frac{\sqrt{8C}}{f(0)} \geq 2 .$$

Hint: fix C, and try to find the density in the class under consideration for which $f(0)$ is maximal. Conclude that one should never use the algorithm of Example 3.5.

6. Show that the class $Lip_\alpha(C)$ has no densities whenever $\alpha>1$.

7. Prove Narumi's inequalities (Lemma 3.1, part D).

8. When f is a normal scale mixture, show that for all $a>0$, the bound of Theorem 3.6 is at least as good as the corresponding bound of Theorem 3.2.

9. Show that f is an exponential scale mixture if and only if for all $x>0$, the derivatives of f are of alternating sign (see e.g. Feller (1971), Keilson and Steutel (1974)). These mixtures consist of convex densities densities on $[0,\infty)$. Derive useful bounds similar to those of Theorem 3.6.

10. **The z-distribution.** Barndorff-Nielsen, Kent and Sorensen (1982) introduced the class of z-distributions with two shape parameters. The symmetric members of this family have density

$$f(x)=\frac{1}{4^a B_{a,a}\cosh^{2a}(\frac{x}{2})} \qquad (x\in R) ,$$

where $a>0$ is a parameter. The translation and scale parameters are omitted. For $a=1/2$, this gives the hyperbolic cosine distribution. For $a=1$ we have the logistic distribution. For integer a it is also called the generalized logistic distribution (Gumbel, 1944). Show the following:

A. The symmetric z-distributions are normal scale mixtures (Barndorff-Nielsen, Kent and Sorensen, 1982).

B. A random variate can be generated as $\log(\frac{Y}{1-Y})$ where Y is symmetric beta distributed with parameter a.

C. If a random variate is generated by rejection based upon the inequalities of Theorem 3.6, the expected time stays uniformly bounded over all values of a.

Additional note: the general z distribution with parameters $a,b>0$ is defined as the distribution of $\log(\frac{Y}{1-Y})$ where Y is beta (a,b).

11. **The residual life density.** In renewal theory and the study of Poisson processes, one can associate with every distribution function F on $[0,\infty)$ the residual life density

$$f(x)=\frac{1-F(x)}{\mu},$$

where $\mu=\int(1-F)$ is the mean for F. Assume that besides the mean we also know the second moment μ_2. This is the second moment of F, not f. Show the following:

A. $f(x)\leq\mu_2/(\mu(x^2+\mu_2))$

B. The black box algorithm shown below is valid and has rejection constant $\pi\sqrt{\mu_2}/\mu$. The rejection constant is at least equal to π, and can be arbitrarily large.

REPEAT

Generate a Cauchy random variate Y, and a uniform $[0,1]$ random variate U.

$X\leftarrow\sqrt{\mu_2}Y$

UNTIL $U\leq(1+Y^2)(1-F(X))$

RETURN X

12. Assume that f is a monotone density on $[0,\infty)$ with distribution function F. Show that for all $0\leq t<x$,

$$f(x)\leq\frac{1-F(t)}{x-t}.$$

Derive from this the inequality

$$f(x) \leq f(0)(1-F(x-\frac{1}{f(0)})) .$$

Note that these inequalities can be used to derive rejection algorithms from tail inequalities for the distribution function.

4. THE INVERSION-REJECTION METHOD.

4.1. The principle.

Assume that f is a density on R, and that we know a few things about f, but not too much. For example, we may know that f is bounded by M, or that $f \in Lip_1(C)$, or that f is unimodal with mode at 0. We have in addition two black boxes, one for computing f, and one for computing the distribution function F. The rejection method is not applicable because we cannot a priori find an integrable dominating curve as for example in the case of log-concave densities. In many cases, this problem can be overcome by the inversion-rejection method (Devroye, 1984). In its most elementary form, it can be put as follows: consider a countable partition of R into intervals $[x_i, x_{i+1})$ where i can take positive and negative values. This partition is fixed but need not be stored: often we can compute the next point x_i from i and/or the previous point. Generate a uniform [0,1] random variate U, and find the index i for which

$$F(x_i) \leq U < F(x_{i+1}) .$$

Thus, interval $[x_i, x_{i+1})$ is chosen with probability $F(x_{i+1})-F(x_i)$ by inversion. If the x_i's are not stored, then some version of sequential search can be used. After i is selected, return a random variate X with density f restricted to the given interval. What we have gained is the fact that the interval is compact, and that in most cases we can easily find a uniform dominating density and use rejection. For example, if f is known to be bounded by M, then we can use a uniform curve with value M. When $f \in Lip_1(C)$, we can use a triangular dominating curve with value $\min(f(x_i)+C(x-x_i), f(x_{i+1})+C(x_{i+1}-x))$. When f is unimodal, then a dominating curve with value $\max(f(x_i), f(x_{i+1}))$ can always be used.

There are two contributors to the expected time taken by the inversion-rejection algorithm:

(i) $E(N_s)$: the expected number of computations of F in the sequential search.

(ii) $E(N_r)$: the expected number of iterations in the rejection method. It is not difficult to see that this is the area under the dominating curve.

In the example of a density bounded by M but otherwise arbitrary, the area under the dominating curve is ∞. Thus, $E(N_r)=\infty$. Nevertheless $N_r < \infty$ with probability one. This fact does not come as a surprise considering the magnitude of the class of densities involved. For unimodal f, even with an infinite peak at the mode and two big tails, it is always possible to construct a partition such that the area under the dominating piecewise constant function is finite. Thus, in the analysis of the different cases, it will be important to distinguish between the families of densities.

The inversion-rejection method is of the black-box type. Its main disadvantage is that programs for calculating both f and F are needed. On the positive side, the families that can be dealt with can be gigantic. The method is not recommended when speed is the most important issue.

We look at the three families introduced above in separate sub-sections. A little extra time is spent on the important class of unimodal densities. The analysis is in all cases based upon the distributional properties of N_s and N_r.

4.2. Bounded densities.

As our first example, we take the family of densities f on $[0,\infty)$ bounded by M. There is nothing sacred about the positive half of R, the choice is made for convenience only. Assume that $[0,\infty)$ is partitioned by a sequence

$$0=x_0<x_1<x_2< \cdots .$$

Let us write $p_i=F(x_{i+1})-F(x_i)$, $i \geq 0$. In a black box method, the inversion step should preferably be carried out by sequential search, starting from 0. In that case, we have

$$P(N_s \geq j) = \sum_{i=j-1}^{\infty} p_i = \int_{x_{j-1}}^{\infty} f = 1-F(x_{j-1}) \quad (j \geq 1) .$$

Also,

$$E(N_s) = 1+\sum_{i=0}^{\infty} i \ p_i = \sum_{i=0}^{\infty} (1-F(x_i)) .$$

Given that we have chosen the i-th interval, the number of iterations in the rejection step is geometrically distributed with parameter $p_i/(M(x_{i+1}-x_i))$, $i \geq 0$. Thus,

$$P(N_r \geq j) = \sum_{i=0}^{\infty} p_i (1-\frac{p_i}{M(x_{i+1}-x_i)})^j .$$

Also,

$$E(N_r) = \sum_{i=0}^{\infty} p_i \frac{M(x_{i+1}-x_i)}{p_i} = \infty .$$

Example 4.1. Equi-spaced intervals.

When $x_{i+1}-x_i=\delta>0$, we obtain perhaps the simplest algorithm of the inversion-rejection type. We can summarize its performance as follows:

$$E(N_s) = 1+\sum_{i=0}^{\infty} i \int_{\delta i}^{\delta(i+1)} f \le 1+\frac{1}{\delta}\sum_{i=0}^{\infty} \int_{\delta i}^{\delta(i+1)} xf = 1+\frac{E(X)}{\delta} ;$$

$$E(N_s) \ge \frac{E(X)}{\delta} ;$$

$$P(N_r \ge j) = \sum_{i=0}^{\infty} p_i (1-\frac{1}{M\delta}p_i)^j .$$

The sequential search is intimately linked with the size of the tail of the density (as measured by $E(X)$). It seems reasonable to take $\delta=cE(X)$ for some universal constant c. When we take c too large, the probabilities $P(N_r \ge j)$ could be unacceptably high. When c is too small, $E(N_s)$ is too large. What is needed here is a compromise. We cannot choose c so as to minimize $E(N_s+N_r)$ for example, since this is ∞. Another method of design can be followed: fix j, and minimize $P(N_r \ge j)+P(N_s \ge j)$. This is

$$\sum_{i=0}^{\infty} p_i (1-\frac{p_i}{M\delta})^j + \sum_{i=j-1}^{\infty} p_i$$

$$\le \sum_{i=J}^{\infty} p_i + \frac{JM\delta}{j+1}(\frac{j}{j+1})^j + \sum_{i=j-1}^{\infty} p_i$$

where J is a positive integer to be picked later. We have used the following simple inequality:

$$u(1-\frac{u}{a})^j \le \frac{a}{j+1}\left(1-\frac{\frac{a}{j+1}}{a}\right)^j .$$

Since we have difficulty minimizing the original expression and the last upper bound, it seems logical to attempt to minimize yet another bound. This strategy is deliberately suboptimal. What we hope to buy is simplicity and insight. Assume that $\mu=E(X)$ is known. Then the tail sums of p_i's can be bounded from above by Markov's inequality. In particular, using also $(1+\frac{1}{j})^j \ge 2$, $j \ge 1$, the last expression is bounded by

$$\frac{\mu}{\delta J}+\frac{JM\delta}{2(j+1)}+\frac{\mu+2}{\delta(j+1)} .$$

The optimal non-integer J is

$$\sqrt{\frac{2(j+1)\mu}{M\delta^2}}$$

and we will take the ceiling of this. Our upper bound now reads

$$2\sqrt{\frac{M\mu}{2(j+1)}}+\frac{\frac{\mu+2}{\delta}+\frac{M\delta}{2}}{j+1} .$$

The last thing left to do is to minimize this with respect to δ, the interval width. Notice however that this will affect only the second order term in the upper bound (coefficient of $\frac{1}{j+1}$), and not the main asymptotic term. For the choice $\delta=\sqrt{\frac{2\mu+4}{M}}$, the second term is

$$\frac{\sqrt{2M(\mu+2)}}{j+1} .$$

The important observation is that for any choice of δ that is independent of j,

$$P(N_s \geq j)+P(N_r \geq j) \leq 2\sqrt{\frac{M\mu}{2(j+1)}}+O(\frac{1}{j}) .$$

The factor $M\mu$ is scale invariant, and is both a measure of how spread out f is and how difficult f is for the present black box method. For this bound to hold, it is not necessary to know μ. The main term in the upper bound is the contribution from N_r. If we assume the existence of higher moments of the distribution, or the moment-generating function, we can obtain upper bounds which decrease faster than $1/\sqrt{j}$ as $j \to \infty$ (exercise 4.1). ■

There are other obvious choices for interval sizes. For example, we could start with an interval of width δ, and then double the width of consecutive intervals. Because this will be dealt with in greater detail for monotone densities, it will be skipped here. Also, because of the better complexity for monotone densities, it is worthwhile to spend more time there.

4.3. Unimodal and monotone densities.

This entire subsection is an adaptation of Devroye (1984). Let us first reduce the problem to one that is manageable. If we know the position of the mode of a unimodal density, and if we can compute $F(x)$ at all x, which is our standing assumption, then it is obvious that we need only consider monotone densities. These can be conveniently flipped around and/or translated to 0, so that all monotone densities to be considered can be assumed to have a mode at 0 and support on $[0,\infty)$. Unfortunately, compact support cannot be assumed because nonlinear transformations to $[0,1]$ could destroy the monotonicity. One thing we can assume however is that we either have an infinite peak at 0 or an infinite tail but not both. Just use the following splitting device:

Splitting algorithm for monotone densities

[SET-UP]

Choose a number $z>0$. (If f is known to be bounded, set $z \leftarrow 0$, and if f is known to have compact support contained in $[0,c]$, set $z \leftarrow c$.)

$t \leftarrow F(z)$

[GENERATOR]

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

IF $U>t$

THEN generate a random variate X with (bounded monotone) density $f(x)/(1-t)$ on $[z,\infty)$.

ELSE generate a random variate X with (compact support) density $f(x)/t$ on $[0,z]$.

RETURN X

Thus, it suffices to treat compact support and bounded monotone densities separately. We will provide the reader with three general strategies, two for bounded monotone densities, and one for compact support monotone densities. Undoubtedly, there are other strategies that could be preferable for certain densities, so no claims of optimality are made. The emphasis is on the manner in which the problem is attacked, and on the interaction between design and analysis. As we pointed out in the introduction, the whole story is told by the quantities $E(N_s)$ and $E(N_r)$ when they are finite.

4.4. Monotone densities on [0,1].

In this section, we will analyze the following inversion-rejection algorithm:

Inversion-rejection algorithm with intervals shrinking at a geometrical rate

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

$X \leftarrow 1$

REPEAT

  $X \leftarrow \frac{X}{r}$

UNTIL $U \geq F(X)$

REPEAT

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

  $Y \leftarrow X(1+(r-1)V)$ (Y is uniform on $[X, rX)$)

UNTIL $W \leq \frac{f(Y)}{f(X)}$

RETURN Y

The constant $r>1$ is a design constant. For a first quick understanding, one can take $r=2$. In the first REPEAT loop, the inversion loop, the following intervals are considered: $[\frac{1}{r},1),[\frac{1}{r^2},\frac{1}{r}),\ldots$. For the case $r=2$, we have interval halving as we go along. For this algorithm,

$$E(N_s) = \sum_{i=1}^{\infty} i \int_{r^{-i}}^{r^{-(i-1)}} f(x)\, dx ,$$

$$E(N_r) = \sum_{i=1}^{\infty} \frac{r-1}{r^i} f(r^{-i}) .$$

The performance of this algorithm is summarized in Theorem 4.1:

Theorem 4.1.

Let f be a monotone density on [0,1], and define

$$H(f) = \int_0^1 \log(\frac{1}{x}) f(x)\,dx .$$

Then, for the algorithm described above,

$$\frac{H(f)}{\log(r)} \leq E(N_s) \leq 1+\frac{H(f)}{\log(r)}$$

and

$$1 \leq E(N_r) \leq r .$$

The functional $H(f)$ satisfies the following inequalities:

A. $1 \leq H(f)$.

B. $\log\left(\frac{1}{\int_0^\infty x f(x)\,dx}\right) \leq H(f)$ (valid even if f has unbounded support).

C. $H(f) \leq 1+\log(f(0))$.

D. $H(f) \leq \frac{4}{e}+2\int_0^1 \log_+ f(x)\, f(x)\,dx$ (valid even if f is not monotone).

Proof of Theorem 4.1.

For the first part, note that on $[r^{-i}, r^{-(i-1)}]$,

$$\frac{\log(x)}{\log(r)} \leq i \leq 1+\frac{\log(x)}{\log(r)} .$$

Thus, resubstitution in the expression of $E(N_s)$ yields the first inequality. We also see that $E(N_r) \geq 1$. To obtain the upper bound for $E(N_r)$, we use a short geometrical argument:

$$E(N_r) = \sum_{i=1}^{\infty} \frac{r-1}{r^i} f(r^{-i})$$

$$= \sum_{i=1}^{\infty} \int_{r^{-i}}^{r^{-(i-1)}} f(r^{-i})\,dx$$

$$\leq \sum_{i=1}^{\infty} \int_{r^{-(i+1)}}^{r^{-i}} f(x)\,dx \times r$$

$$= r\int_0^{\frac{1}{r}} f(x)\,dx$$

$$\leq r \ .$$

Inequality A uses the fact that $-\log(x)$ and $f(x)$ are both nonincreasing on [0,1], and therefore, by Steffensen's inequality (1925),

$$\int_0^1 -\log(x) f(x)\, dx \geq \int_0^1 -\log(x)\, dx \int_0^1 f(x)\, dx = 1 \ .$$

Inequality B uses the convexity of $-\log(x)$ and Jensen's inequality. If X is a random variable with density f, then

$$H(f) = E(-\log(X)) \geq -\log(E(X)) \ .$$

Inequality C can be obtained as a special case of another inequality of Steffensen's (1918): in its original form, it states that if $0 \leq h \leq 1$, and if g is nonincreasing and integrable on [0,1], then

$$\int_0^1 g(x) h(x)\, dx \leq \int_0^a g(x)\, dx$$

where $a = \int_0^1 h(x)\, dx$. Apply this inequality with $g(x) = -\log(x)$, $h(x) = \frac{f(x)}{f(0)}$. Thus, $a = \frac{1}{f(0)}$. Therefore,

$$\frac{H(f)}{f(0)} \leq \int_0^{\frac{1}{f(0)}} -\log(x)\, dx$$

$$= \int_{\log(f(0))}^{\infty} y e^{-y}\, dy = \frac{1}{f(0)}(1+\log(f(0))) \ .$$

Inequality D is a Young-type inequality which can be found in Hardy, Littlewood and Polya (1952, Theorem 239). ■

In Theorem 4.1, we have shown that $E(N_s) < \infty$ if and only if $H(f) < \infty$. On the other hand, $E(N_r)$ is uniformly bounded over all monotone f on [0,1]. Our main concern is thus with the sequential search. We do at least as well as in the black box method of section 3.2 (Theorem 3.2), where the expected number of iterations in the rejection method was $1+\log(f(0))$. We are guaranteed to have $E(N_s) \leq 1+(1+\log(f(0)))/\log(r)$, and even if $f(0)=\infty$, the inversion-rejection method can have $E(N_s) < \infty$.

Example 4.2. The beta density.

Consider the beta $(1,a+1)$ density $f(x)=(a+1)(1-x)^a$ on $[0,1]$ where $a>0$ is a parameter. We have $f(0)=a+1$, $E(X)=\frac{1}{a+2}$. Thus, by inequalities B and C of Theorem 4.1,

$$\log(a+2) \leq H(f) \leq 1+\log(a+1) .$$

We have $H(f) \sim \log(a)$ as $a \to \infty$: the average time of the given inversion-rejection algorithm grows as $\log(a)$ as $a \to \infty$. ■

In the absence of extra information about the density, it is recommended that r be set equal to 2. This choice also gives small computational advantages. It is important nevertheless to realize that this choice is not optimal in general. For example, assume that we wish to minimize $E(N_s+N_r)$, a criterion in which both contributions are given equal weight because both N_s and N_r count in effect numbers of computations of f and/or F. The minimization problem is rather difficult. But if we work on a good upper bound for $E(N_s+N_r)$, then it is nevertheless possible to obtain:

Theorem 4.2.

For the inversion-rejection algorithm of this section with design constant $r>1$, we have

$$\inf_{r>1} E(N_s+N_r)$$

$$\leq 1+H(f)\left(\frac{1}{\log^2(H(f))}+\frac{1}{\log(H(f))-2\log(\log(H(f)))}\right)$$

$$\sim \frac{H(f)}{\log(H(f))}$$

as $H(f) \to \infty$. The bound is attained for

$$r = \frac{H(f)}{\log^2(H(f))} .$$

Proof of Theorem 4.2.

We start from

$$E(N_s+N_r) \le 1+r+\frac{H(f)}{\log(r)} .$$

Resubstitution of the value of r given in the theorem gives us the inequality. This value was obtained by functional iteration applied to

$$r = \frac{H(f)}{\log^2(r)} ,$$

an equation which must be satisfied for the minimum of the upper bound (set the derivative of the upper bound with respect to r equal to 0). The functional iteration was started at $r=H(f)$. That the value is not bad follows from the fact that for $H(f) \ge e$,

$$1+r+\frac{H(f)}{\log(r)} \ge 1+\frac{H(f)}{\log(H(f))} ,$$

so that at least from an asymptotic point of view no improvement is possible over the given bound. ■

As a curious application of Theorem 4.2, consider the case again of a monotone density on [0,1] with finite $f(0)$. Recalling that $H(f) \le 1+\log(f(0))$, we see that if we take

$$r = \frac{1+f(0)}{\log^2(1+f(0))} ,$$

a choice which is indeed implementable, then

$$E(N_s+N_r)$$
$$\le 1+1+\frac{\log(f(0))}{\log^2(1+\log(f(0)))}+\frac{\log(f(0))}{\log(1+\log(f(0)))-2\log(\log(1+\log(f(0))))}$$
$$\sim \log\frac{(f(0))}{\log(1+\log(f(0)))}$$

as $f(0)\to\infty$. This should be compared with the value of $E(N_r)=1+\log(f(0))$ for the black box rejection algorithm following Theorem 3.1.

For densities that are also known to be convex, a slight improvement in $E(N_r)$ is possible. See exercise 4.5.

4.5. Bounded monotone densities: inversion-rejection based on Newton-Raphson iterations.

In this section, we assume that f is monotone on $[0,\infty)$ and that $f(0)<\infty$. It is possible that f has a large tail. In an attempt to automatically balance $E(N_s)$ against $E(N_r)$, and thus to avoid the eternal problem of having to find a good design constant, we could determine intervals for sequential search based upon Newton-Raphson iterations started at $x_0=0$. Recall the definition of the hazard rate

$$h(x) = \frac{f(x)}{1-F(x)} .$$

If we try to solve $F(x)=1$ for x by Newton-Raphson iterations started at $x_0=0$, we obtain a sequence $x_0 \le x_1 \le x_2 \le \cdots$ where

$$x_{n+1} = x_n + \frac{1-F(x_n)}{f(x_n)} = x_n + \frac{1}{h(x_n)} .$$

The x_n's need not be stored. Obviously, storing them could considerably speed up the algorithm.

Inversion-rejection algorithm for bounded densities based upon Newton-Raphson iterations

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

$X \leftarrow 0$, $R \leftarrow F(X)$, $Z \leftarrow f(X)$

REPEAT

  $X^* \leftarrow X + \frac{1-R}{Z}$, $R^* \leftarrow F(X^*)$, $Z^* \leftarrow f(X^*)$

  IF $U \le R^*$

    THEN Accept $\leftarrow$ True

    ELSE $R \leftarrow R^*$, $Z \leftarrow Z^*$, $X \leftarrow X^*$

UNTIL Accept

REPEAT

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

  $Y \leftarrow X+(X^*-X)V$, $T \leftarrow WZ$ (Y is uniformly distributed on $[X,X^*)$)

  Accept $\leftarrow [T \le Z^*]$ (optional squeeze step)

  IF NOT Accept THEN Accept $\leftarrow [T \le f(Y)]$

UNTIL Accept

RETURN Y

One of the differences with the algorithm of the previous section is that in every iteration of the inversion step, one evaluation of both F and f is required as compared to one evaluation of F. The performance of the algorithm is dealt with in Theorem 4.3.

Theorem 4.3.

Let f be a bounded monotone density on $[0,\infty)$ with mode at 0. For the inversion-rejection algorithm given above,

$$E(N_s) = E(N_r) = \sum_{i=0}^{\infty} (1-F(x_i))$$

where $0=x_0 \le x_1 \le x_2 \le \cdots$ is the sequence of numbers defined by

$$x_{n+1} = x_n + \frac{1-F(x_n)}{f(x_n)} \quad (n \ge 0) .$$

If f is also DHR (has nonincreasing hazard rate), then

$$1 \le E(N_r) = E(N_s) \le 1+E(Xf(0)) .$$

If f is also IHR (has nondecreasing hazard rate), then

$$1 \le E(N_r) = E(N_s) \le \frac{e}{e-1} .$$

Proof of Theorem 4.3.

$$E(N_s) = \sum_{i=1}^{\infty} i\,((1-F(x_{i-1}))-(1-F(x_i))) = \sum_{i=0}^{\infty} (1-F(x_i)) ,$$

$$E(N_r) = \sum_{i=0}^{\infty} f(x_i)(x_{i+1}-x_i) = \sum_{i=0}^{\infty} (1-F(x_i)) .$$

When f is DHR, then

$$E(Xf(0)) = f(0)\int_0^{\infty}(1-F(x))\,dx = \int_0^{\infty} \frac{f(0)}{h(x)} f(x)\,dx \ge 1 .$$

For IHR densities, the inequality should be reversed. Thus, for DHR densities,

$$\sum_{i=0}^{\infty} (1-F(x_i)) \le 1+\sum_{i=1}^{\infty} \frac{\int_{x_{i-1}}^{x_i} (1-F(x))\,dx}{x_i - x_{i-1}}$$

$$= 1+\sum_{i=1}^{\infty} \int_{x_{i-1}}^{x_i} (1-F(x))\,dx \;\; h(x_{i-1})$$

$$\le 1+\int_0^{\infty} f(0)(1-F(x))\,dx = 1+E(Xf(0)) .$$

When f is IHR, then

$$1-F(x_{i+1}) = (1-F(x_i))e^{-\int_{x_i}^{x_{i+1}} h(x)\,dx}$$

$$\leq (1-F(x_i))e^{-h(x_i)(x_{i+1}-x_i)}$$
$$= \frac{1-F(x_i)}{e} .$$

Thus,

$$\sum_{i=0}^{\infty}(1-F(x_i)) \leq \sum_{i=0}^{\infty} e^{-i} = \frac{e}{e-1} . \blacksquare$$

We have thus found an algorithm with a perfect balance between the two parts, since $E(N_s)=E(N_r)$. This does not mean that the algorithm is optimal. However, in many cases, the performance is very good. For example, its expected time is uniformly bounded over all IHR densities. Examples of IHR densities on $[0,\infty)$ are given in the table below.

Name	Density f	Hazard rate h	$E(N_s)=E(N_r)$
Halfnormal	$\sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2}}$		$\leq \frac{e}{e-1}$
Gamma (a), $a \geq 1$	$\frac{x^{a-1}e^{-x}}{\Gamma(a)}$		$\leq \frac{e}{e-1}$
Exponential	e^{-x}	1	$\frac{e}{e-1}$
Weibull (a), $a \geq 1$	$ax^{a-1}e^{-x^a}$	ax^{a-1}	$\leq \frac{e}{e-1}$
Beta $(a,1)$, $a \geq 1$	ax^{a-1} $(0 \leq x \leq 1)$	$\frac{x^{a-1}}{1-x^a}$	$\leq \frac{e}{e-1}$
Beta $(1,a+1)$, $a \geq 0$	$(a+1)(1-x)^a$ $(0 \leq x \leq 1)$	$\frac{a+1}{1-x}$	$\left(1-(1-\frac{1}{a+1})^{a+1}\right)^{-1}$
Truncated extreme value, $a>0$	$\frac{1}{a}e^{x-\frac{e^x-1}{a}}$	$\frac{e^x}{a}$	$\leq \frac{e}{e-1}$

This is not the place to enter into a detailed study of IHR densities. It suffices to state that they are an important family in daily statistics (see e.g. Barlow and Proschan (1965, 1975), and Barlow, Marshall and Proschan (1963)). Some of its salient properties are covered in exercise 4.6. Some entries for $E(N_s)$ in the table given above are explicitly known. They show that the upper bound of Theorem 4.3 is sharp in a strong sense. For example, for the exponential density, we have $x_n=n$, and thus

$$E(N_s) = E(N_r) = \sum_{i=0}^{\infty}(1-F(i)) = \sum_{i=0}^{\infty} e^{-i} = \frac{e}{e-1} .$$

For the beta $(1,a+1)$ density mentioned in the table, we can verify that

$$x_{n+1} = \frac{a}{a+1}x_n + \frac{1}{a+1} ,$$

and thus,

$$x_n = 1-\left(\frac{a}{a+1}\right)^n \quad (n \geq 0) .$$

Thus,

$$E(N_s) = \sum_{i=0}^{\infty}(1-F(x_i)) = \sum_{i=0}^{\infty}(1-x_i)^{a+1}$$

$$= \sum_{i=0}^{\infty}\left(\frac{a}{a+1}\right)^{i(a+1)} = \left(1-\left(1-\frac{1}{a+1}\right)^{a+1}\right)^{-1} .$$

This varies from 1 ($a=0$) to $\frac{e}{e-1}$ ($a \uparrow \infty$) without exceeding $\frac{e}{e-1}$. Thus, once again, the inequality of Theorem 4.3 is tight.

For DHR densities, the upper bound is often very loose, and not as good as the performance bounds obtained for the dynamic thinning method (section VI.2). For example, for the Pareto density $\frac{a}{(1+x)^{a+1}}$ (where $a>0$ is a parameter), we have a hazard rate $h(x)=\frac{a}{1+x}$, and $E(N_s)=\left(1-(1+\frac{1}{a})^{-a}\right)^{-1}$. This can be seen as follows:

$$(x_{n+1}+1) = (x_n+1)(1+\frac{1}{a}) ;$$

$$(x_n+1) = (1+\frac{1}{a})^n \quad (n \geq 0) ;$$

$$E(N_s) = \sum_{i=0}^{\infty}(1+\frac{1}{a})^{-ia} = \left(1-(1+\frac{1}{a})^{-a}\right)^{-1} .$$

The last expression varies from $\frac{e}{e-1}$ ($a \uparrow \infty$) to 2 ($a=1$) and up to ∞ as $a \downarrow 0$.

4.6. Bounded monotone densities: geometrically increasing interval sizes.

For bounded densities, we can use a sequential search from left to right, symmetric to the method used for unbounded but compact support densities. There are two design parameters: $t>0$ and $r>1$, and the consecutive intervals are

$$[0,t),[t,tr),[tr,tr^2),\ldots .$$

A typical choice is $t=1$, $r=2$. General guidelines follow after the performance analysis. Let us begin with the algorithm:

Inversion-rejection method for bounded monotone densities based upon geometrically exploding intervals

Generate a uniform [0,1] random variate U.
$X \leftarrow 0$, $X^* \leftarrow t$
WHILE $U > F(X^*)$ DO
  $X \leftarrow X^*$, $X^* \leftarrow rX^*$
REPEAT
  Generate two iid uniform [0,1] random variates, V, W.
  $Y \leftarrow X + (X^* - X)V$ (Y is uniformly distributed on $[X, X^*)$)
UNTIL $W \leq \frac{f(Y)}{f(X)}$
RETURN Y

Theorem 4.4.

Let f be a bounded monotone density, and let $t > 0$ and $r > 1$ be constants. Define

$$H_t(f) = \int_0^\infty \log_+(\frac{x}{t}) f(x)\, dx \ .$$

Then, for the algorithm given above,

$$1 + \frac{H_t(f)}{\log(r)} \leq E(N_s) \leq 2 + \frac{H_t(f)}{\log(r)} ,$$

and

$$1 \leq tf(0) + \int_t^\infty f(x)\, dx \leq E(N_r) \leq tf(0) + r \ .$$

Proof of Theorem 4.4.

We repeatedly use the fact that $tr^{i-1} \leq x < tr^i$ if and only if $i-1 \leq \log(\frac{x}{t})/\log(r) < i$, $i > 1$. Now,

$$E(N_s) = \int_0^t f(x)\, dx + \sum_{i=1}^\infty (i+1) \int_{tr^{i-1}}^{tr^i} f(x)\, dx = 1 + \sum_{i=1}^\infty i \int_{tr^{i-1}}^{tr^i} f(x)\, dx$$

$$\leq 2 + \int_t^\infty \frac{\log(\frac{x}{t})}{\log(r)} f(x)\, dx = 2 + \frac{H_t(f)}{\log(r)} ,$$

and

$$E(N_s) \geq 1+\int_t^\infty \frac{\log(\frac{x}{t})}{\log(r)} f(x)\,dx = 1+\frac{H_t(f)}{\log(r)} .$$

Also,

$$E(N_r) = tf(0)+\sum_{i=1}^{\infty}(tr^i - tr^{i-1})f(tr^{i-1})$$

$$\leq tf(0)+\sum_{i=1}^{\infty}\frac{tr^i - tr^{i-1}}{tr^{i-1}-tr^{i-2}}\int_{tr^{i-2}}^{tr^{i-1}} f(x)\,dx$$

$$\leq tf(0)+r ,$$

and

$$E(N_r) \geq tf(0)+\sum_{i=1}^{\infty}\frac{tr^i - tr^{i-1}}{tr^i - tr^{i-1}}\int_{tr^{i-1}}^{tr^i} f(x)\,dx$$

$$= tf(0)+\int_t^\infty f(x)\,dx \geq 1 . \blacksquare$$

We would like the algorithm to perform at a scale-invariant speed. This can be achieved for $t = \frac{1}{f(0)}$. In that case, the upper bounds of Theorem 4.4 read:

$$E(N_s) \leq 2+\frac{H^*(f)}{\log(r)} ;$$

$$E(N_r) \leq 1+r ,$$

where

$$H^*(f) = \int_0^\infty \log_+(xf(0))f(x)\,dx$$

is the scale invariant counterpart of the quantity $H(f)$ defined in Theorem 4.1. $H^*(f)$ can be considered as the normalized logarithmic moment for the density f. For the vast majority of distributions, $H^*(f) < \infty$. In fact, one must search hard to find a monotone density for which $H^*(f) = \infty$. The tail of the density must at least of the order of $1/(x \log^2(x))$ as $x \to \infty$, such as is the case for

$$f(x) = \frac{1}{(x+e)\log^2(x+e)} \quad (x > 0) .$$

With little a priori information, we suggest the choice

$$\boxed{\begin{array}{c} r=2 \\ t=\frac{1}{f(0)} \end{array}}$$

It is interesting to derive a good guiding formula for r. We start from the inequality

$$E(N_s)+E(N_r) \leq 3+r+\frac{H^*(f)}{\log(r)},$$

which is minimal for the unique solution $r>1$ for which $r\log^2(r)=H^*(f)$. By functional iteration started at $r=H^*(f)$, we obtain the crude estimate

$$r = \frac{H^*(f)}{\log^2(H^*(f))} .$$

For this choice, we have as $H^*(f)\to\infty$,

$$E(N_s)+E(N_r) \leq (1+o(1))\frac{H^*(f)}{\log(H^*(f))} .$$

Example 4.3. Moment known.

A loose upper bound for $H^*(f)$ is afforded by Jensen's inequality:

$$H^*(f) \leq \int_0^\infty \log(1+xf(0))f(x)\,dx \leq \log(1+E(Xf(0)))$$

where X is a random variable with density f. Thus, the expected time of the algorithm grows at worst as the logarithm of the first moment of the distribution. For example, for the beta $(1,a+1)$ density of Example 4.1, this upper bound is $\log(1+\frac{a+1}{a+2}) \leq \log(2)$ for all $a>0$. This is an example of a family for which the first moment, hence $H^*(f)$, is uniformly bounded. From this,

$$E(N_s) \leq 2+\frac{\log(2)}{\log(r)} ;$$

$$E(N_r) \leq 1+r .$$

The ad hoc choice $r=2$ makes both upper bounds equal to 3. ■

4.7. Lipschitz densities on $[0,\infty)$.

The inversion-rejection method can also be used for Lipschitz densities f on $[0,\infty)$. This class is smaller than the class of bounded densities, but very large compared to the class of monotone densities. The black box method of section 3 for this class required knowledge of a moment of the distribution. In contrast, the method presented here works for all densities $f \in Lip_1(C)$ where only C must be given beforehand. The moments of the distribution need not even exist. If the positive half of the real line is partitioned by

$$0=x_0<x_1<x_2< \cdots ,$$

then, it is easily seen that on $[x_n, x_{n+1}]$,

$$f(x) \leq \min(f(x_n)+C(x-x_n), f(x_{n+1})+C(x_{n+1}-x)) ,$$

and

$$f(x) \leq \sqrt{2C(1-F(x_n))}$$

where the last inequality is based upon Theorem 3.5. The areas under the respective dominating curves are

$$E(N_r) = \sum_{n=0}^{\infty} \frac{1}{2C} \left(c\,\Delta_n (f(x_n)+f(x_{n+1})) - \frac{1}{2}(f(x_n)+f(x_{n+1}))^2 + \frac{C^2\Delta_n^{\ 2}}{2} \right)$$

and

$$E(N_r) = \sum_{n=0}^{\infty} \Delta_n \sqrt{2C(1-F(x_n))} ,$$

where $\Delta_n = x_{n+1}-x_n$. The value of $E(N_s)$ depends only upon the partition, and not upon the inequalities used in the rejection step, and plays no role when the inequalities are compared. Generally speaking, the second inequality is better because it uses more information (the value of F is used). Consider the first inequality. To guarantee that $E(N_r)$ be finite, for the vast majority of Lip_1 densities we need to ask that

$$\sum_{n=0}^{\infty} \Delta_n^{\ 2} < \infty .$$

But, since we require a valid partition of R, we must also have

$$\sum_{n=0}^{\infty} \Delta_n = \infty .$$

In particular, we cannot afford to take $\Delta_n = \delta > 0$ for all n. Consider now Δ_n satisfying the conditions stated above. When $\Delta_n \sim n^{-a}$, then it is necessary that $a \in (\frac{1}{2},1]$. Thus, the intervals shrink rapidly to 0. Consider for example

$$\Delta_n = \frac{c}{n+1} \quad (n \geq 0) .$$

For this choice, the intervals shrink so rapidly that we spend too much time searching unless f has a very small tail. In particular,

$$E(N_s) = \sum_{n=0}^{\infty} P(X \geq \sum_{i=0}^{n} \Delta_i)$$

$$\leq \sum_{n=0}^{\infty} P(X \geq c \log(n+2))$$

$$= \sum_{n=0}^{\infty} P(e^{\frac{X}{c}} \geq n+2)$$

$$\leq E(e^{\frac{X}{c}}) .$$

A similar lower bound for $E(N_s)$ exists, so that we conclude that $E(N_s)<\infty$ if and only if the moment generating function at $\frac{1}{c}$ is finite, i.e.

$$m(\frac{1}{c}) = E(e^{\frac{X}{c}}) < \infty .$$

In other words, f must have a sub-exponential tail for good expected time. Thus, instead of analyzing the first inequality further, we concentrate on the second inequality.

The algorithm based upon the second inequality can be summarized as follows:

Inversion-rejection algorithm for Lipschitz densities

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

$X \leftarrow 0$, $R \leftarrow F(X)$

REPEAT

$X* \leftarrow$ Next (X), $R* \leftarrow F(X*)$ (The function Next computes the next value in the partition.)

IF $U \leq R*$

THEN Accept $\leftarrow$ True

ELSE $R \leftarrow R*$, $X \leftarrow X*$

UNTIL Accept

REPEAT

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

$Y \leftarrow X + V(X*-X)$ (Y is uniformly distributed on $[X, X*)$).

UNTIL $W\sqrt{2C(1-R)} \leq f(Y)$

RETURN Y

There are three partitioning schemes that stand out as being either important or practical. These are defined as follows:

A. $x_n = n\delta$ for some $\delta > 0$ (thus, $x_{n+1} - x_n = \delta$).

B. $x_{n+1} = tr^n$ for some $t > 0, r > 1$, $x_1 = t$ (note that $x_{n+1} = rx_n$ for all $n \geq 1$). The intervals grow exponentially fast.

C. $x_{n+1} = x_n + \sqrt{\frac{1-F(x_n)}{2C}}$ (this choice provides a balance between $E(N_s)$ and $E(N_r)$).

Schemes A and B require additional design constants, whereas scheme C is completely automatic. Which scheme is actually preferable depends upon various factors, foremost among these the size of the tail of the distribution. By imposing conditions on the tail, we can derive upper bounds for $E(N_s)$ and $E(N_r)$. These are collected in Theorem 4.5:

Theorem 4.5.

Let $f \in Lip_1(C)$ be a density on $[0,\infty)$. Let $p>1$ be a constant. When the p-th moment exists, it is denoted by μ_p.

For scheme A,

$$\max(1,\frac{\mu_1}{\delta}) \leq E(N_s) \leq 1+\frac{\mu_1}{\delta} ;$$

$$\delta\sqrt{2C}\,\max(1,\frac{1}{\sqrt{\mu_2}},\frac{\sqrt{\mu_2}}{\delta}) \leq E(N_r) \leq \delta\sqrt{2C}\,(2+\frac{p}{p-1}\frac{(\mu_{2p})^{\frac{1}{2p}}}{\delta}) .$$

In particular, if $\delta=\sqrt{\frac{\mu_1}{\sqrt{8C}}}$, then

$$E(N_s)+E(N_r) \leq 1+(8C)^{\frac{1}{4}}\sqrt{\mu_1}+\sqrt{8C}\,(\mu_4)^{\frac{1}{4}} ,$$

and when $\delta=\frac{1}{\sqrt{8C}}$,

$$E(N_s)+E(N_r) \leq 2+\sqrt{8C}\,((\mu_4)^{\frac{1}{4}}+\mu_1) \leq 2+\sqrt{32C}\,(\mu_4)^{\frac{1}{4}} .$$

For scheme B,

$$E(N_s) \leq 2+E\left(\frac{\log_+(\frac{X}{t})}{\log(r)}\right) ;$$

$$E(N_r) \leq \sqrt{2C}\,(t+\frac{\sqrt{\mu_{2p}}\,r^{p-1}(r-1)}{t^{p-1}(r^{p-1}-1)}) .$$

For scheme C,

$$E(N_s) = E(N_r) \leq \sqrt{8C}\int_0^\infty \sqrt{1-F(x)}\,dx \leq \frac{p}{p-1}\sqrt{8C}\,(\mu_{2p})^{\frac{1}{2p}} .$$

At the same time, even if $\mu_2=\infty$, the following lower bound is valid:

$$\sqrt{2C\mu_2} \leq \frac{1}{2}\sqrt{8C}\int_0^\infty \sqrt{1-F(x)}\,dx \leq E(N_s) = E(N_r) .$$

Proof of Theorem 4.5.

In this proof, X denotes a random variate with density f . Rewrite $E(N_s)$ as follows:

$$E(N_s) = \sum_{n=0}^{\infty} \int_{\delta n}^{\infty} f(x)\,dx = \int_0^{\infty} \left\lfloor \frac{x}{\delta}+1 \right\rfloor dx .$$

This can be obtained by an interchange of the sum and the integral. But then, by Jensen's inequality and trivial bounds,

$$\max(1,\frac{E(X)}{\delta}) \le \int_0^{\infty} \max(1,\frac{x}{\delta}) f(x)\,dx \le E(N_s)$$

$$\le \int_0^{\infty} (\frac{x}{\delta}+1) f(x)\,dx = 1+\frac{E(X)}{\delta} .$$

Next,

$$E(N_r) = \sum_{n=0}^{\infty} \sqrt{2C(1-F(\delta n))}\delta ,$$

so that by Chebyshev's inequality,

$$\frac{E(N_r)}{\delta\sqrt{2C}} \le \sum_{n=0}^{\infty} \min(1,\frac{\sqrt{\mu_{2p}}}{(n\delta)^p})$$

$$\le 1+\frac{1}{\delta}(\mu_{2p})^{\frac{1}{2p}} + \sum_{n=n_0}^{\infty} \frac{\sqrt{\mu_{2p}}}{(n\delta)^p}$$

where $n_0 = \left\lceil \frac{1}{\delta}(\mu_{2p})^{\frac{1}{2p}} \right\rceil$. By a simple argument, we see that

$$\sum_{n=n_0}^{\infty} n^{-p} \le n_0^{-p} + \int_{n_0}^{\infty} x^{-p}\,dx$$

$$= n_0^{-p} + \frac{1}{p-1} n_0^{-(p-1)} .$$

Combining this shows that

$$\frac{E(N_r)}{\delta\sqrt{2C}} \le 1+\frac{1}{\delta}(\mu_{2p})^{\frac{1}{2p}} + 1 + \frac{1}{(p-1)\delta}(\mu_{2p})^{\frac{1}{2p}}$$

$$= 2+\frac{p}{p-1}\frac{(\mu_{2p})^{\frac{1}{2p}}}{\delta} .$$

This brings us to the lower bounds for scheme A. We have, by the Cauchy-Schwarz inequality,

$$\frac{E(N_r)}{\delta\sqrt{2C}} = \sum_{n=0}^{\infty} \sqrt{\int_{\delta n}^{\infty} f}$$

$$\geq \sum_{n=0}^{\infty} \frac{\int_{\delta n}^{\infty} \sqrt{f}\,(x\sqrt{f}\,)}{\sqrt{\int (x\sqrt{f}\,)^2}}$$

$$= \sum_{n=0}^{\infty} \frac{\int_{\delta n}^{\infty} xf}{\sqrt{\mu_2}}$$

$$\geq \frac{1}{\sqrt{\mu_2}} \int xf\,(x)\max(1,\frac{x}{\delta})\,dx$$

$$\geq \frac{1}{\sqrt{\mu_2}} \max(1,\frac{\mu_2}{\delta})$$

$$= \max(\frac{1}{\sqrt{\mu_2}},\frac{\sqrt{\mu_2}}{\delta}) .$$

Also,

$$\frac{E(N_r)}{\delta\sqrt{2C}} = \sum_{n=0}^{\infty} \sqrt{\int_{\delta n}^{\infty} f}$$

$$\geq \sum_{n=0}^{\infty} \int_{\delta n}^{\infty} f$$

$$\geq \max(1,\frac{\mu_1}{\delta}) .$$

For scheme B, we have

$$E(N_s) = 1 + \sum_{n=0}^{\infty} (1-F(tr^n))$$

$$= 1 + \sum_{n=0}^{\infty} \int_{tr^n}^{\infty} f\,(x)\,dx$$

$$\leq 2 + E\left(\frac{\log_+(\frac{X}{t})}{\log(r)} \right) .$$

Also,

$$E(N_r) = \sum_{n=0}^{\infty} \sqrt{2C}\,\sqrt{1-F(tr^n)t(r-1)r^n} + \sqrt{2C}\,t$$

$$\leq \sqrt{2C}\,t + \sqrt{2C} \sum_{n=0}^{\infty} t(r-1)r^n \frac{\sqrt{\mu_{2p}}}{t^p r^{np}}$$

$$= \sqrt{2C}\,(t + \frac{\sqrt{\mu_{2p}}\, r^{p-1}(r-1)}{t^{p-1}(r^{p-1}-1)}) .$$

Finally, we consider scheme C. Consider the graph of $1-\sqrt{1-F(x)}$. Construct for given x_n the triangle with top on the given curve, and base $[x_n, x_{n+1}]$ at height 1. Its area is $\frac{1-F(x_n)}{\sqrt{8C}}$. The triangle lies completely above the given curve because the slope of the hypothenusa is $\sqrt{2C}$, which is at least as steep as the derivative of $1-\sqrt{1-F}$ at any point. To see this, note that the latter derivative at x is

$$\frac{f(x)}{2\sqrt{1-F(x)}} \leq \frac{\sqrt{2C(1-F(x))}}{2\sqrt{1-F(x)}} = \sqrt{\frac{C}{2}} .$$

Thus, the sums of the areas of the triangles is not greater than the integral $\int_0^\infty \sqrt{1-F(x)}\, dx$. But this sum is

$$\sum_{n=0}^{\infty} \frac{1-F(x_n)}{\sqrt{8C}} = \frac{E(N_r)}{\sqrt{8C}} = \frac{E(N_s)}{\sqrt{8C}} .$$

Also, twice the area of the triangles is at least equal to $\int_0^\infty \sqrt{1-F(x)}\, dx$. The bounds in terms of the various moments mentioned are obtained without further trouble. First, by Chebyshev's inequality,

$$\int_0^\infty \sqrt{1-F(x)}\, dx \leq \int_0^\infty \min(1, \frac{\sqrt{\mu_{2p}}}{x^p})\, dx = (\mu_{2p})^{\frac{1}{2p}} + \frac{1}{p-1}(\mu_{2p})^{\frac{1}{2p}} .$$

Also, by the Cauchy-Schwarz inequality,

$$\int_0^\infty \sqrt{\int_x^\infty f} \quad dx \geq (\mu_2)^{-\frac{1}{2}} \int_0^\infty \int_x^\infty y f(y)\, dy \quad dx$$

$$= (\mu_2)^{-\frac{1}{2}} \int_0^\infty \int_0^y dx \quad y f(y)\, dy = \sqrt{\mu_2} . ■$$

We observe that $\sqrt{C}X$ is a scale-invariant quantity. Thus, one upper bound for scheme A (choice $\delta=\frac{1}{\sqrt{8C}}$) and the upper bound for scheme C are scale-invariant: they depend upon the shape of the density only. Scheme C is attractive because no design constants have to be chosen at any time. In scheme A for example, the choice of δ is critical. The geometrically increasing interval sizes of scheme B seem to offer little advantage over the other methods, because $E(N_r)$ is relatively large.

4.8. Exercises.

1. Obtain an upper bound for $P(N_r \geq j)$ in terms of j when equi-spaced intervals are used for bounded densities on $[0,\infty)$ as in Example 4.1. Assume first that the r-th moment μ_r is finite. Assume next that $E(e^{tX})=m(t)<\infty$ for some $t>0$. The interval width δ does not depend upon j. Check that the main term in the upper bound is scale-invariant.
2. Prove inequality D of Theorem 4.1.
3. Give an example of a monotone density on [0,1], unbounded at 0, with $H(f)<\infty$.
4. Inequalities A through C in Theorem 4.1 are best possible: they can be attained for some classes of monotone densities on [0,1]. Describe some classes of densities for which we have equality.
5. When f is a monotone convex density on [0,1], then the inversion-rejection algorithm based on shrinking intervals given in the text can be adapted so that rejection is used with a trapezoidal dominating curve joining $[X, f(X)]$ and $[rX, f(rX)]$ where $r>1$ is the shrinkage parameter used in the original algorithm. Such a change would leave N_s the same. It reduces $E(N_r)$ however. Formally, the algorithm can be written as follows:

Inversion-rejection algorithm with intervals shrinking at a geometrical rate

Generate a uniform [0,1] random variate U.
$X \leftarrow 1$
REPEAT
    $X \leftarrow \frac{X}{r}$
UNTIL $U \geq F(X)$
$Z \leftarrow f(X), Z* \leftarrow f(rX)$
REPEAT
    Generate three independent uniform [0,1] random variates, U, V, W.
    $R \leftarrow \min(U, V\frac{Z+Z*}{Z-Z*})$
    $Y \leftarrow X(1+(r-1)R)$ (Y has the given trapezoidal density)
    $T \leftarrow W(Z+(Z*-Z)R)$
    Accept $\leftarrow [T \leq Z*]$ (optional squeeze step)
    IF NOT Accept THEN Accept $\leftarrow [W \leq f(Y)]$
UNTIL Accept
RETURN Y

Prove that $E(N_r) \leq \frac{1}{2}(1+r)$. In other words, for large values of r, this corresponds to an improvement of the order of 50%.

6. **IHR densities.** Prove the following statements:

   A. If X has an IHR density on $[0,\infty)$, then $Xf(0)$ is stochastically smaller than an exponential variate, i.e. for all $x>0$, $P(Xf(0)>x) \leq e^{-x}$. Conclude that for $r>0$, $E(X^r) \leq \frac{\Gamma(r+1)}{f(0)^r}$.

   B. For $r>0$, $E(X^r) \leq \Gamma(r+1) E^r(X)$ (Barlow, Marshall and Proschan, 1963).

   C. The convolution of two IHR densities is again IHR.

   D. Let Y,Z be independent IHR random variables with hazard rates h_Y and h_Z. Then, if h_{Y+Z} is the hazard rate of their sum, $h_{Y+Z} \leq \min(h_Y, h_Z)$.

   E. Construct an IHR density which is continuous, unbounded, and has infinitely many peaks.

7. Show how to choose r and t in the inversion-rejection algorithm with geometrically exploding intervals so as to obtain performance that is sub-logarithmic in the first moment of the distribution in the following sense:

$$E(N_r)+E(N_s) \leq C \frac{\log(1+\mu f(0))}{\log(\log(e+\mu f(0)))} ,$$

where $\mu=E(X)$, C is some universal constant, and X is a random variable with density f.

8. **Bounded convex monotone densities.** Give an algorithm analogous to that studied in Theorem 4.4 for this class of densities: its sole difference is that the rejection step uses a trapezoidal dominating curve. For this algorithm, in the notation of Theorem 4.4, prove the inequality

$$E(N_r) \leq \frac{1}{2}(tf(0)+r+1) .$$

9. Prove that if $\Delta_n = \frac{c}{n+1}$ in the algorithm for Lipschitz densities, then $E(N_s)<\infty$ if and only if $E(e^{\frac{X}{c}})<\infty$.

10. Suggest good choices for t and r in scheme B of Theorem 4.5. These choices should preferably minimize $E(N_s)+E(N_r)$, or the upper bound for this sum given in the theorem. The resulting upper bound should be scale-invariant.

11. Consider a density f on $[0,\infty)$ which is in $Lip_\alpha(C)$ for some $\alpha \in (0,1]$. Using the inequality of Theorem 3.5 for such densities, give an algorithm generalizing scheme C of Theorem 4.5 for Lip_1 densities. Make sure that $E(N_s)=E(N_r)$ and give an upper bound for $E(N_s)$ which generalizes the upper bound of Theorem 4.5.

12. The lower bound for scheme C in Theorem 4.5 shows that when $\mu_2=\infty$, then $E(N_s)=\infty$. This is a nearly optimal result, in that for most densities with finite second moment, $E(N_s)<\infty$. For example, if $\mu_{2+\epsilon}<\infty$ for some

$\epsilon>0$, then $E(N_s)<\infty$. Find densities for which $\mu_2<\infty$, yet $E(N_s)=\infty$.