Generating Random Deviates

Often there are no direct ways of sampling from a desired distribution (e.g. inverse cdf or relationship with other distributions).

So we need other approaches to generation for other distributions.

Acceptance-Rejection (von Neumann, 1951)

Want to simulate from a distribution with density $f(x)$.

Need to find a "dominating" or majorizing distribution $g(x)$ where g is easy to sample from and

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

for all x and some constant $c > 1$.

Sampling scheme

1) Sample x from $g(x)$ and compute the ratio

$$r(x) = \frac{f(x)}{cg(x)} = \frac{f(x)}{h(x)} \leq 1$$

2) Sample $u \sim U(0,1)$

If $u \leq r(x)$ accept and return x

If $u > r(x)$ reject and go back to 1)

Note that step 2) is equivalent to flipping a biased coin with success probability r.

The resultant sample is a draw from $f(x)$.

Proof:

Let I be the indicator of whether a sample x is accepted. Then

$$P[I=1] = \int P\left[I = 1 \middle| X = x\right] g(x)\,dx$$
$$= \int \frac{f(x)}{cg(x)} g(x)\,dx = \frac{1}{c}$$

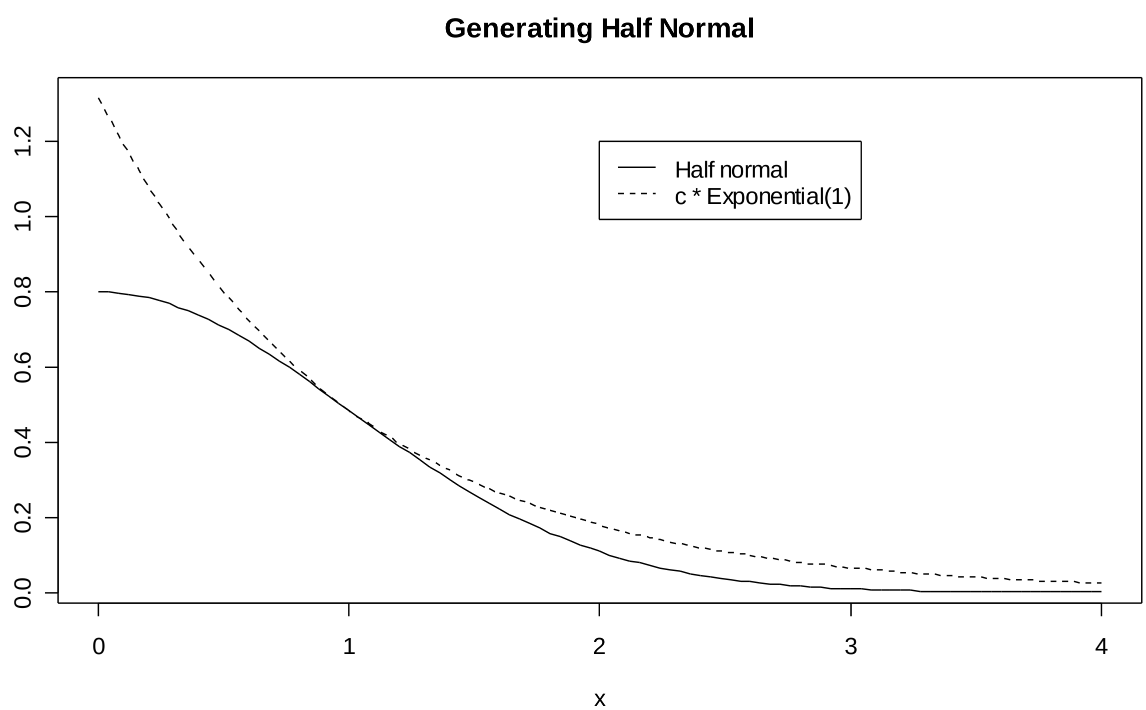
Next,

$$p\left(x \middle| I=1\right) = \frac{f(x)}{cg(x)} g(x) \Big/ P[I=1]$$

$$= \frac{cf(x)}{c} = f(x)$$

See Flury (1990) for a more geometrical proof.

Its based on the idea of drawing uniform points (x,y) under the curve $h(x)$ and only accepting the points that also lie under the curve $f(x)$.



The number of draws needed until an acceptance occurs is Geometric$(1/c)$ and thus the expected number of draws until a sample is accepted is c.

The acceptance probability satisfies

$$\frac{1}{c} = \frac{\int f(x)dx}{\int cg(x)dx} = \frac{\text{Area under } f(x)}{\text{Area under } h(x)}$$

One consequence of this is that c should be made as small as possible to minimize the number rejections.

The optimal c is given by

$$c = \sup \frac{f(x)}{g(x)}$$

Note that the best c need not be determined, just one that satisfies

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

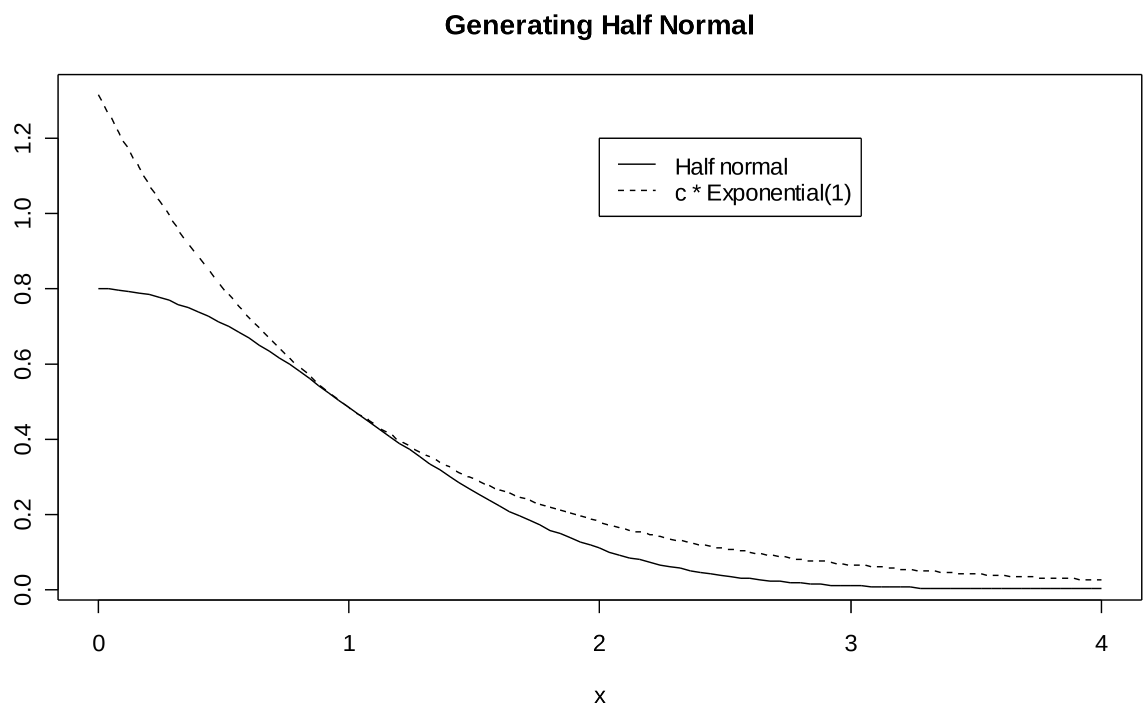
for all x.

Example: Generating from the half normal distribution.

$$f(x) = 2\phi(x) I(x \geq 0)$$
$$= \sqrt{\frac{2}{\pi}} \exp\left(-\frac{x^2}{2}\right) I(x \geq 0)$$

Lets use an Exp(1) as the dominating density

$$g(x) = e^{-x} I(x \geq 0)$$



The optimal c is

$$c = \sqrt{\frac{2}{\pi}} \exp(1/2) \approx 1.315489$$

Thus the acceptance –rejection scheme is

1) Draw $x \sim \text{Exp}(1)$

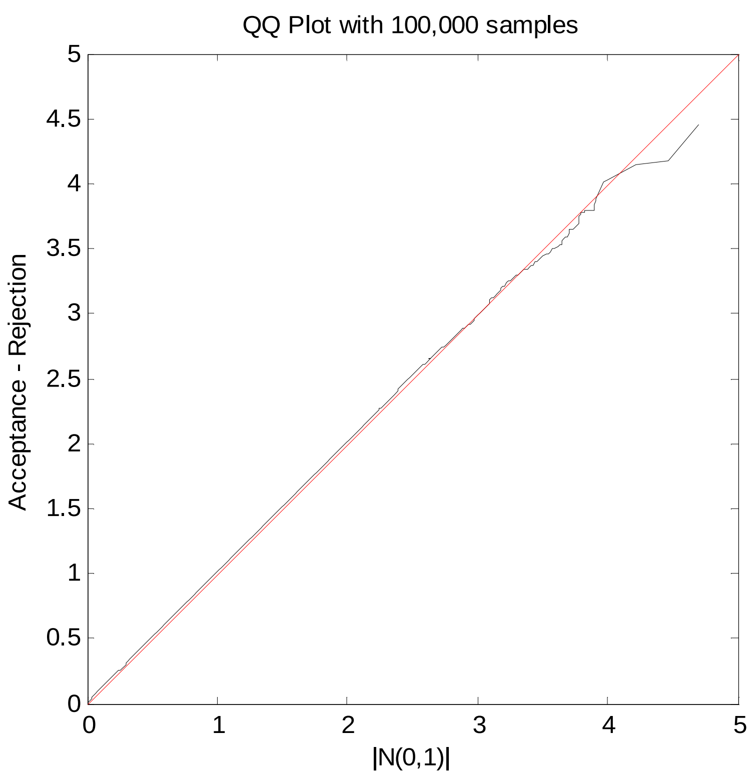
$$r(x) = \exp\left(-0.5(x-1)^2\right)$$

2) Draw $u \sim U(0,1)$

If $u \le r(x)$ accept and return x

If $u > r(x)$ reject and go back to 1)

Note that this scheme isn't needed for this example as the half normal distribution is the distribution of the absolute value of $N(0,1)$



In the above is was assumed that $f(x)$ was a density function. In fact $f(x)$ only needs to be known up to a multiplicative constant

$$l(x) = bf(x)$$

where b may be unknown.

One place where this is useful is with posterior distributions as

$$p(x|y) \propto \pi(x) f(y|x)$$

The normalizing constant may be difficult to determine exactly.

However it is not necessary to do so. Modify the procedure as follows.

Find c such that

$$l(x) \leq cg(x) = h(x)$$

for all x and some constant $c > 1$.

Sampling scheme

1) Sample x from $g(x)$ and compute the ratio

$$r(x) = \frac{l(x)}{cg(x)} = \frac{l(x)}{h(x)} \leq 1$$

2) Sample $u \sim U(0,1)$

If $u \le r(x)$ accept and return x

If $u > r(x)$ reject and go back to 1)

Do everything the same except use $l(x)$ instead of $f(x)$

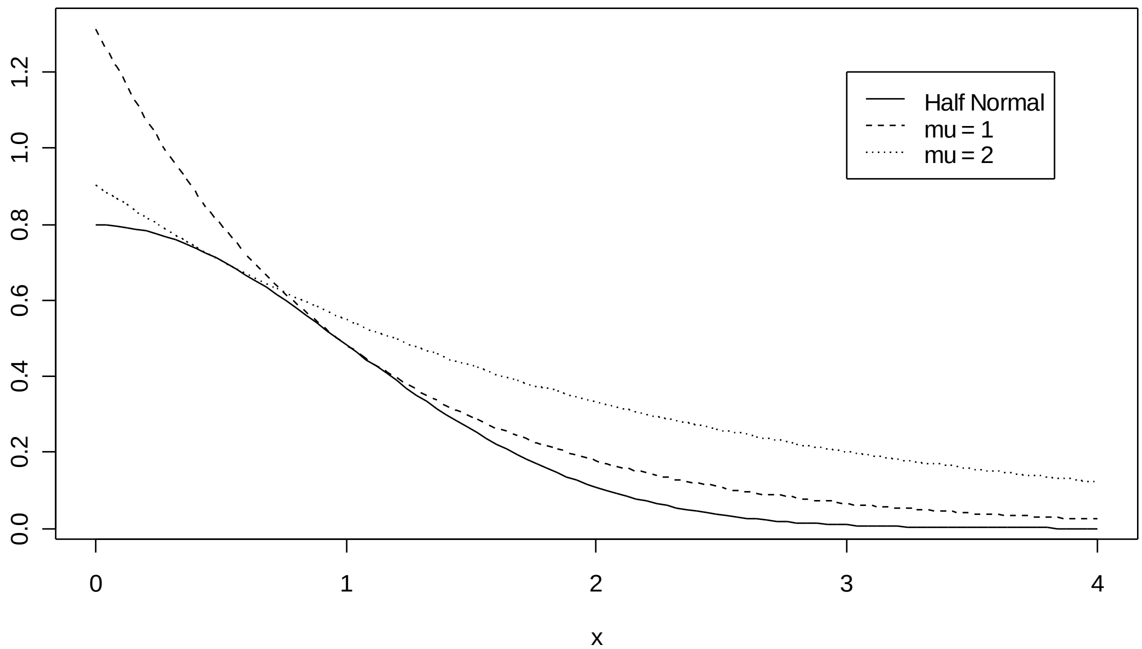
The acceptance probability for this scheme is b/c.

In addition to the constant c chosen, the distribution $g(x)$ will also affect the acceptance rate. (c is chosen conditional on $g(x)$)

A good choice $g(x)$ will normally be "close to" $f(x)$. You want to minimize the separation between the two densities.

Often will look for much member of a parametric family will minimize c.

For example, for the Half normal problem, which $\text{Exp}(\mu)$, will minimize $c(\mu)$.



In fact $\mu = 1$ will minimize $c(\mu)$ for this problem.

Note that so far I'm been appeared to have been focusing on continuous random variables.

In fact acceptance-rejection works fine with discrete random variables and with variables over more than one dimension.

The proof goes through in this more general place by replacing integration over a density to integration over a more general measure.

For discrete random variables, you get a sum over the probability mass function.

With higher dimensional problems, the majorization constants tend to be higher, implying the procedure is less efficient.

Log-concave densities

There is a class of densities where it is easy to set up an acceptance-rejection scheme.

It the case when the log of the density is concave on the support of the distributions

If $f(x)$ is log concave, any tangent line to $\log f(x)$ will lie above $\log f(x)$ (call it $l(x) = a + bx$).

Thus $h(x) = e^{l(x)} = e^{a} e^{bx}$ lies above $f(x)$.

$h(x)$ looks like a scaled exponential density.

This suggests that exponential distributions can be used as the majorizing distributions.

A strictly log concave density is unimodal.

The mode may occur at either endpoint or in the middle.

If the mode occurs at an endpoint, a single exponential can be used (as with the half normal example)

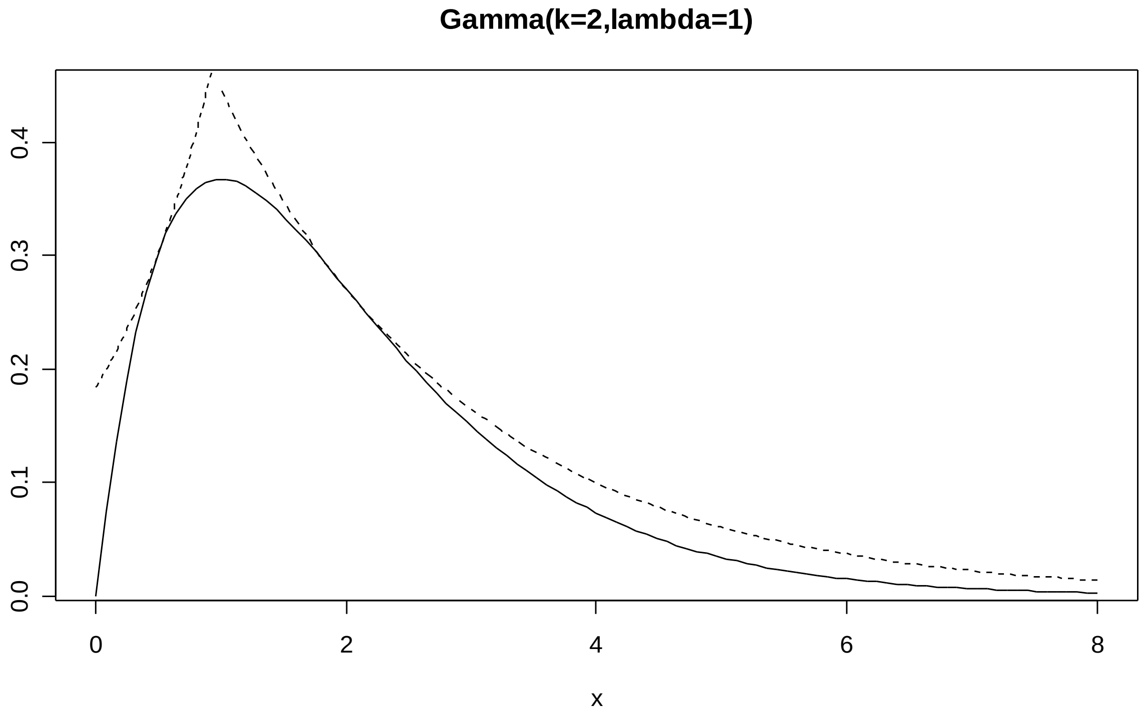
If the mode occurs in the middle of the range, two exponential envelopes are needed (one for left of the mode, the other for the right of the mode)

Example: $\text{Gamma}(k = 2, \lambda = 1)$

The mode for a Gamma is $(k-1)\lambda$ (so 1 for this example)

Left side: $g_l(x) = \dfrac{1}{\mu_l} \exp\left((x-1)/\mu_l\right) I(x < 1)$

Right side: $g_r(x) = \dfrac{1}{\mu_r} \exp\left(-(x-1)/\mu_r\right) I(x \geq 1)$



The choice of μ_l and μ_r depend on where you want the majorized densities $g_l(x)$ and $g_r(x)$ to be tangent to $f(x)$

In the above the tangent points are $x_l = 0.5$ and $x_r = 2$.

The total area under $h(x) = c_l g_l(x) + c_r g_r(x)$ is $c = c_l + c_r$

For the example, $c_l = 0.5$ and $c_r = 0.8925206$, so the rejection rate for this sampler is just under 30%.

To determine which exponentials to use, involves solving the systems (for given x_l and x_r)

$$f(x_l) = c_l g_l(x_l) \qquad f(x_r) = c_r g_r(x_r)$$
$$f'(x_l) = c_l g_l'(x_l) \qquad f'(x_r) = c_r g_r'(x_r)$$

Solving gives

$$\lambda_l = \frac{f'(x_l)}{f(x_l)} \qquad \lambda_r = -\frac{f'(x_r)}{f(x_r)}$$

$$c_l = \frac{f(x_l)^2}{f'(x_l)} e^{-\lambda_l (x_l - m)} \qquad c_r = -\frac{f(x_r)^2}{f'(x_r)} e^{\lambda_r (x_r - m)}$$

where $\lambda_i = 1/\mu_i$

The optimal choices for x_l and x_r can be found by minimizing c_l and c_r separately.

Discrete log concave distributions

Random variable defined on non-negative integers

Log concave defined as

$$\log f(x) \geq \frac{1}{2}\left[\log f(x-1) + \log f(x+1)\right]$$

which is equivalent to

$$f(x)^2 \geq f(x-1)\, f(x+1)$$

for all integers x.

A possible majorizing distribution in the discrete case is the geometric distribution

$$P[X = x] = p(1-p)^x\,; \quad x = 0, 1, 2, \ldots$$

See Lange for choices of p_l, p_r, x_l, x_r

Ratio Method

This is another method that is useful when the distribution you are interested in $f(x)$, is only known up to an unknown constant, $h(x) = cf(x)$

Define

$$S_h = \left\{(u,v) : 0 < u \leq \sqrt{h(v/u)}\right\}$$

If this set is bounded, we can draw uniform points from this set to generate X.

Proposition 20.5.1

Suppose

$$k_u = \sup_x \sqrt{h(x)}$$

and

$$k_v = \sup_x |x| \sqrt{h(x)}$$

are both finite. Then the rectangle $[0, k_u] \times [-k_v, k_v]$ encloses S_h.

If $h(x) = 0$ for $x < 0$, then the rectangle $[0, k_u] \times [0, k_v]$ encloses S_h.

If the point (U,V) sampled uniformly from the enclosing set falls with in S_h, then the ratio $X = V/U$ is distributed according to $f(x)$.

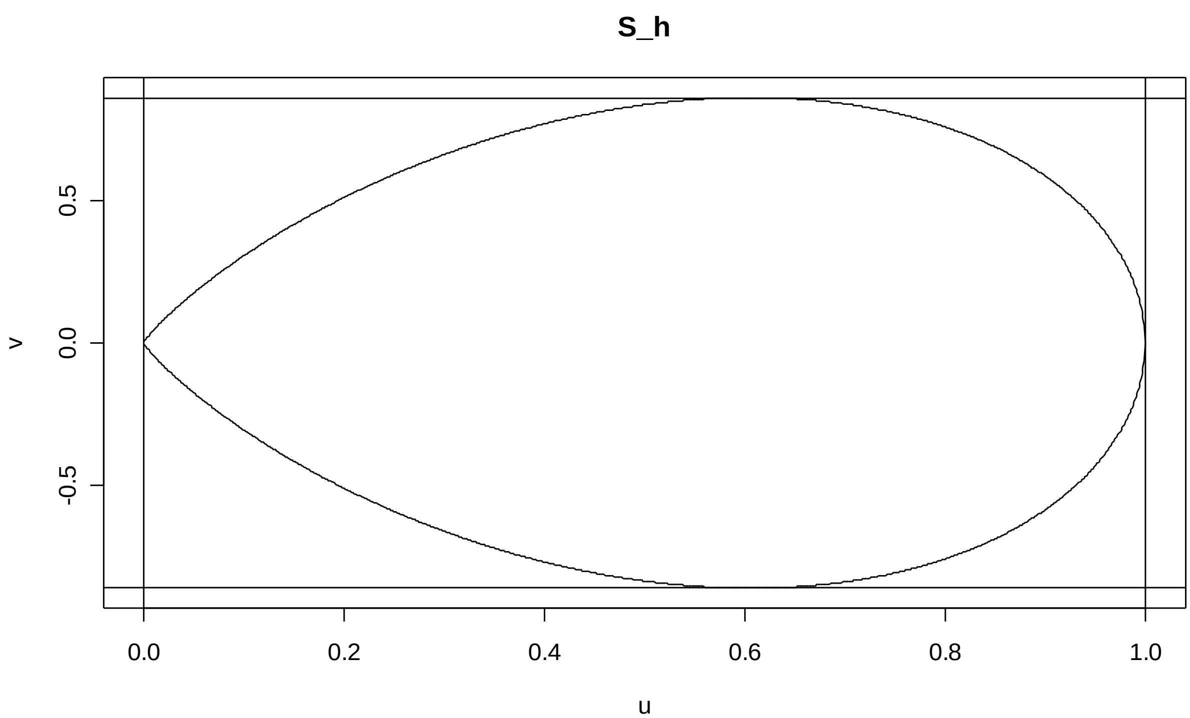
Example: Standard normal

$$h(x) = \exp\left(-\frac{1}{2}x^2\right)$$

$$\sqrt{h(x)} = \exp\left(-\frac{1}{4}x^2\right)$$

$$k_u = \sqrt{h(0)} = 1$$

$$k_v = \sqrt{2}\exp(-0.5)$$



Generate

$$U \sim U(0,1)$$

$$V \sim U\left(-\sqrt{2}\exp(-0.5), \sqrt{2}\exp(-0.5)\right)$$

Accept $X = V/U$ if

$$U \le \exp\left(-0.25V^2/U^2\right)$$

The acceptance rate for this procedure is about 62.6%