Sampling methods

by Agner Fog

This document is published at www.agner.org/random, January 2004, as part of a software package.

Introduction

A C++ class library of non-uniform random number generators is available from www.agner.org/random. The distributions include normal, poisson, binomial, hypergeometric, Fisher's noncentral hypergeometric and Wallenius' noncentral hypergeometric distributions as well as multivariate versions of these distributions. This document describes the various methods used for generating variates with some of these distributions. The reader is referred to the file `distrib.pdf` for a definition of these distributions and to the file `nchyp.pdf` (Fog 2004) for theory of Fisher's and Wallenius' noncentral hypergeometric distributions.

Two characteristics are important when selecting the optimal variate generation method, the set-up time and the execution time (Stadlober 1989). The set-up time is the computer time used by the non-random part of the calculations, which only depends on the parameters. The execution time is the time spent on the rest of the calculations, which involves random factors. The set-up time is unimportant if the generator is called repeatedly with the same parameters because the results of these calculations can be saved and re-used. The set-up time is important, however, if the parameters often change. Some of the generators are provided in two different versions, one with a short set-up time, which is optimal when parameters often change, and one with a long set-up time but a short execution time, which is optimal when the generator is called repeatedly with the same parameters. The former generators are implemented in the C++ classes `StochasticLib`, while the latter are implemented in `StochasticLib2`. Generators for Fisher's and Wallenius' noncentral hypergeometric distributions are implemented in `StochasticLib3`.

Some of the functions in this library are adapted from the Win-rand package by Ernst Stadlober. See www.cis.tugraz.at/stat/stadl/random.html. The Win-rand package also contains many other distributions than the ones supplied here, and a graphic interface to show the distribution curves.

Inversion method

Assume that we want to generate a variate with the frequency function $\mathrm{f}(x)$. The distribution function is

$$\mathrm{F}(x) = \int_0^x \mathrm{f}(t)dt \tag{1}$$

for a continuous distribution, or

$$\mathrm{F}(x) = \sum_{j=0}^{\lfloor x \rfloor} \mathrm{f}(j) \tag{2}$$

for a discrete distribution. If u is a variate with the uniform distribution $u \sim \mathrm{uniform}(0,1)$, then a variate with the desired distribution can be generated by applying the inverse distribution function to u:

$$x = \mathrm{F}^{-1}(u) \tag{3}$$

This method requires that F^{-1} is easy to calculate.

Inversion by chop down search from 0

When $f(x)$ is a discrete function (x integer) then (3) can be calculated by successively adding $f(0) + f(1) + ... + f(x)$ until the sum exceeds u.

Each $f(x)$ is computed from the previous one according to a recursion formula:

Poisson:

$$f(x) = f(x-1) \cdot \frac{L}{x} \tag{4}$$

Binomial:

$$f(x) = f(x-1) \cdot \frac{n-x+1}{x} \cdot \frac{p}{1-p} \tag{5}$$

Hypergeometric:

$$f(x) = f(x-1) \cdot \frac{(m-x+1)(n-x+1)}{x(L+x)}, \quad L = N - m - n \tag{6}$$

Fisher's noncentral hypergeometric:

$$f(x) = f(x-1) \cdot \frac{(m-x+1)(n-x+1)}{x(L+x)} \omega, \quad L = N - m - n \tag{7}$$

Since division is slow, this method can be speeded up by multiplying u with the denominator in the recursion formula instead of dividing $f(x)$ with the denominator. We then have to check for overflow unless the parameters are limited to safe intervals.

The computation time is proportional to the mean of x. Therefore this method is used only when the mean is low.

This method is used for the poisson and binomial distributions in `StochasticLib` when the mean is below a certain threshold.

Chop down search from the mode

This is very similar to the method above. $f(x)$ values are added consecutively until the sum exceeds u. The difference is that the x values are taken in a different order. Rather than starting at $x = 0$ we start at the mode M, and take $f(M) + f(M\text{-}1) + f(M\text{+}1) + f(M\text{-}2) + f(M\text{+}2) + ...$ until the sum exceeds u.

The advantage of this method is that the search is likely to end earlier when we take the most probable x-values first. The disadvantage is that the set-up time is higher.

This method is used when the mean is low for hypergeometric in `StochasticLib` and also for poisson and binomial in `StochasticLib2`.

Rejection method

The principle of the rejection method is that the frequency function $f(x)$ is approximated by another distribution function $h(x)$ which is easier to calculate, and then a correction is made by randomly accepting x values with a probability $p(x) = \frac{f(x)}{k\,h(x)}$, and rejecting x values with probability 1-$p(x)$. The constant k is chosen so that $k\,h(x) \geq f(x)$ for all values of x. Whenever an x value is rejected, the procedure starts over again with a new x. The accepted x values have the distribution $f(x)$ because the acceptance probability $p(x)$ is proportional to $f(x) / h(x)$. x values with distribution $h(x)$ are generated by simple inversion from a random number u with uniform distribution: $x = H^{-1}(u)$, where H is the integral of h and H^{-1} is the inverse of H. Acceptance/rejection is done on the basis of a second random number v with uniform distribution. x is accepted if $v < p(x)$ and rejected if $v \geq p(x)$.

The rejection method can be improved by quick acceptance and quick rejection schemes. Quick acceptance is based on a minorizing function $f_a(x) \leq f(x)$, and quick rejection is based on a majorizing function $f_r(x) \geq f(x)$. These can be simple linear functions or other functions that are easy to calculate. The acceptance condition can be written as $vkh(x) < f(x)$. The time-consuming calculation of $f(x)$ is avoided by quick acceptance when $vkh(x) < f_a(x)$ and quick rejection when $vk\,h(x) \geq f_r(x)$. Only when $vkh(x)$ is between these two values do we need to calculate $f(x)$ in the final acceptance/rejection decision.

The advantage of the rejection method is that the calculation time does not grow with x. It is therefore used when the variance is so large that the chop-down search would be more time consuming.

The disadvantage of the rejection method is that it is difficult to find a good hat function $h(x)$ which is easy to calculate and at the same time approximates $f(x)$ so good that the rejection rate will be low. A bad hat function will lead to a high rejection rate and hence a long execution time.

Various improvements of the rejection method are applied, as explained below.

Patchwork rejection method

(Stadlober & Zechner 1998). This variation of the rejection method uses a simple hat function, which has a large uniform section in the center and exponential tails. The area under the curve of the frequency function $f(x)$ is cut into pieces which are rotated and rearranged geometrically like a jigsaw puzzle to fill as much of the area under the hat function as possible. A point under the curve of the hat function is generated on the basis of the two uniform random numbers u and v. It is then determined whether this point falls on a piece of our jigsaw puzzle or in the rejection area between the pieces. If the point falls on a rearranged piece then the transformations of the rearrangement are undone in order to find the corresponding x. Quick acceptance and quick rejection areas are used.

The advantage of this method is that the rejection rate is low. The disadvantage is that the set-up time is quite high because it takes a lot of calculations to determine where the borders of each piece are. Hence, this method is only advantageous if the function is called many times with the same parameters.

The patchwork rejection method is used for the poisson, binomial and hypergeometric functions in `StochasticLib2` when the mean is high.

Ratio-of-uniforms rejection method

(Stadlober 1989, 1990). Let u and v be two independent random numbers with uniform distribution in the intervals $0 < u \le 1$, and $-1 \le v \le 1$. Do the transformation $x = sv/u + a$, $y = u^2$. The rectangle in the (u,v) plane is transformed into a hat function $y = \mathrm{h}(x)$ in the (x,y) plane. All (x,y) points will fall under the hat curve $y = \mathrm{h}(x)$ which is uniform in the center and falls like x^{-2} in the tails. $\mathrm{h}(x)$ is a useful hat function for the rejection method. The acceptance condition is $v < \mathrm{f}(x)/k$. s and a are chosen so that $\mathrm{f}(x) \le k\,\mathrm{h}(x)$ for all x, where

$$\mathrm{h}(x) = \begin{cases} 1 & \text{for } a - s \le x \le a + s \\ \dfrac{s^2}{(x-a)^2} & \text{elsewhere} \end{cases} \tag{8}$$

The advantage of this method is that the calculations are simple and fast, and the rejection rate is reasonable. Quick acceptance and quick rejection areas can be applied. For discrete distributions, $\mathrm{f}(x)$ is replaced by $\mathrm{f}(\lfloor x \rfloor)$.

The following values are used for the hat parameters for the poisson, binomial and hypergeometric distributions

$$a = \mu + \tfrac{1}{2}, \text{ and} \tag{9}$$

$$s = \sqrt{\frac{2}{e}\left(\sigma^2 + \tfrac{1}{2}\right)} + s_1, \quad s_1 = \frac{3}{2} - \sqrt{\frac{3}{e}} \tag{10}$$

where μ is the mean and σ^2 is the variance of the distribution (Ahrens & Dieter 1989). These values are reasonably close to the optimal values. It is possible to calculate the optimal values for a and s (Stadlober 1989, 1990), but this adds to the set-up time with only a marginal improvement in execution time. The optimal value of k is of course the maximum of $\mathrm{f}(x)$: $k = \mathrm{f}(M)$, where M is the mode.

This method is used for the poisson, binomial and hypergeometric distributions in StochasticLib.

Box-Muller transformation

A special transformation method is used for the normal distribution (Devroye 1986).

1. Generate two independent uniform variates u and v in the interval [-1,1]
2. Set $z = u^2+v^2$.
3. If $z \ge 1$ then reject and go to step 1
4. Set $w = \sqrt{\dfrac{-2\log z}{z}}$
5. $x_1 = uw$ and $x_2 = vw$ are independent variates with distribution $\mathrm{normal}(0,1)$

Conditional method for multivariate distributions

The conditional method for sampling from a multivariate distribution is the method where x_1 is sampled first from the marginal distribution, then x_2 is sampled according to the conditional distribution of x_2 given x_1, and so forth (Johnson 1987).

This method is useful for sampling from the multivariate binomial (multinomial) and the multivariate hypergeometric distributions, because the marginal and conditional distributions are all univariate binomial, respectively hypergeometric, distributions.

Fisher's and Wallenius' noncentral hypergeometric distributions

Methods for the univariate and multivariate Fisher's and Wallenius' noncentral hypergeometric distributions are described in detail in the file nchyp.pdf (Fog 2004). These methods are implemented in the C++ class `StochasticLib3`.

References

Ahrens, J H & Dieter, U (1989): A convenient sampling method with bounded computation times for Poisson distributions. *American Journal of Mathematical and Management Sciences*, pp. 1-13.

Devroye, Luc (1986): *Non-uniform random variate generation*. New York: Springer.

Fog, Agner (2004): *The noncentral hypergeometric probability distributions.* www.agner.org/random/theory.

Johnson, Mark E (1987): *Multivariate Statistical Simulation.* New York: Wiley & Sons.

Levin, Bruce (1984): Simple improvements on Cornfield's approximation to the mean of a noncentral hypergeometric random variable. *Biometrika*, vol. 71, no. 3. pp. 630-632.

Liao, Jason G; Rosen, Ori (2002): Fast and Stable Algorithms for Computing and Sampling from the Noncentral Hypergeometric Distribution. *The American Statistician*, vol. 55, no. 4, pp. 366-369.

Stadlober, Ernst (1989): Sampling from poisson, binomial and hypergeometric distributions: Ratio of uniforms as a simple and fast alternative. *Berichte der mathematisch-statistischen Sektion in der Forschungsgesellschaft Joanneum*, no. 303, Graz.

Stadlober, Ernst (1990): The ratio of uniforms approach for generating discrete random variates. *Journal of Computational and Applied Mathematics*, vol. 31, no. 1, 1990, pp. 181-189.

Stadlober, Ernst & Zechner, Heinz (1999): The Patchwork Rejection Technique for Sampling from Unimodal Distributions. *ACM Transactions on Modeling and Computer Simulation*, Vol. 9, No. 1, pp. 59-80.