CHAPTER 3

Algorithms for Distributions

In this chapter we discuss calculating the probability density and mass functions and the cdf and quantile functions for a wide variety of distributions as well as how to generate observations from these distributions. The distributions considered are all listed in a table at the end of the chapter.

The chapter also makes extensive use of the gamma and beta functions and the incomplete gamma and beta functions defined at the beginning of the table of distributions. We will use f, F, and Q to denote the pdf (or pmf), cdf, and quantile functions, respectively. We will put the name of the random variable as a subscript and in parentheses we'll put the argument of the function followed by a semicolon and then the parameters of the distribution. Thus for example, $f_{\chi^2}(x;\nu)$ denotes the pdf of the χ^2 distribution with ν degrees of freedom.

Calculating Mass and Density Functions

Calculating probability mass functions and density functions must be done with care primarily because many of them are expressed as the ratio of quantities that can be very large. The most common example of this is in evaluating factorials or the gamma function (which is the same thing for integer arguments since $\Gamma(n+1) = n!$. For example, in single precision arithmetic, 34! is the largest factorial that can be computed before overflow occurs. Thus most mass or density functions are calculated by exponentiating the logarithm of the function. For example, the pdf of the χ^2 distribution can be written

$$\begin{aligned} f_{\chi^2}(x;\nu) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{-x/2} x^{(\nu/2)-1}, \quad x > 0, \\ &= \exp((\nu/2) - 1)\log x - x/2 - (\nu/2)\log 2 - \log\Gamma(\nu/2)) \end{aligned}$$

This last equation illustrates that we'll often need to be able to calculate the log of the gamma function.

Calculating the Log of the Gamma Function

One way to do this is ("Numerical Recipes", eq. 6.1.5)

$$\log\Gamma(z+1) = (z+\frac{1}{2})\log(z+\gamma+\frac{1}{2}) - (z+\gamma+\frac{1}{2}) + \log(\sqrt{2\pi}) + \log\left[c_0 + \sum_{j=1}^{n}\frac{c_j}{z+j}\right],$$

for certain values of γ, n, and $c_0, c_1, \ldots, c_n$.

Recursive Calculation of Probability Mass Functions

One interesting simplification for many probability mass functions is that $f(x) = \Pr(X = x)$ can be easily calculated recursively in x. As an example, for the binomial distribution with parameters n and p, we have

$$\frac{\Pr(X = x+1)}{\Pr(X = x)} = \frac{n-x}{x+1}\frac{p}{1-p}, \qquad 0 \le x < n,$$

which, if we define $\phi = p/(1-p)$ and $l_x = \log \Pr(X = x)$ gives the recursion

$$l_{x+1} = l_x + \log\phi + \log(n-x) - \log(x+1).$$

This recursion is particularly useful if the entire pmf is desired. Note also that the recursion can start from either end, that is for increasing x starting at $x = 0$ or for decreasing x starting at $x = n$.

Calculating the cdfs

The cdf of many distributions, including the Z, t, χ^2, F, binomial, and Poisson are related to the incomplete gamma and beta functions defined at the beginning of the table of distributions at the end of this chapter. It is not difficult to show that

$$F_Z(x) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{2} IG\left(\dfrac{x^2}{2}; \dfrac{1}{2}\right), & x \ge 0 \\ 1 - F_Z(-x), & x < 0, \end{cases}$$

$$F_t(x;\nu) = \begin{cases} 1 - \dfrac{IB\left(\dfrac{\nu}{\nu + x^2}; \dfrac{\nu}{2}, \dfrac{1}{2}\right)}{2}, & x \ge 0 \\ 1 - F_t(-x;\nu), & x < 0, \end{cases}$$

$$F_{\chi^2}(x;\nu) = IG\left(\frac{x}{2}; \frac{\nu}{2}\right),$$

$$F_F(x;\nu_1,\nu_2) = 1 - IB\left(\frac{\nu_2}{\nu_2 + \nu_1 x}; \frac{\nu_2}{2}, \frac{\nu_1}{2}\right),$$

$$F_{Bin}(k;n,p) = 1 - IB(p; k+1, n-k),$$

$$F_{Poiss}(k;\lambda) = 1 - F_{\chi^2}(2\lambda; 2(k+1)) = 1 - IG(\lambda; k+1).$$

Thus to evaluate these six cdfs we need only be able to calculate IG and IB.

Calculating the Incomplete Gamma and Beta Functions

There is a vast literature on calculating IG and IB. Popular methods include ("Numerical Recipes", eqs. 6.2.5, 6.2.6, and 6.3.5)

$$\mathrm{IG}(x;a) = \begin{cases} e^{-x}x^a \sum_{n=0}^{\infty} \dfrac{1}{\Gamma(a+1+n)} x^n, & x \le a+1 \\ 1 - \dfrac{1}{\Gamma(a)} e^{-x} x^a \left(\dfrac{1}{x+}\,\dfrac{1-a}{1+}\,\dfrac{1}{x+}\,\dfrac{2-a}{1+}\,\dfrac{2}{x+}\cdots\right), & x > a+1, \end{cases}$$

$$\mathrm{IB}(x;a,b) = \begin{cases} \dfrac{x^a(1-x)^b}{a\beta(a,b)} \left[\dfrac{1}{1+}\,\dfrac{d_1}{1+}\,\dfrac{d_2}{1+}\cdots\right], & x < \dfrac{a+1}{a+b+1}, \\ 1 - \mathrm{IB}(1-x; b, a), & x \ge \dfrac{a+1}{a+b+1}, \end{cases}$$

where for $m = 0, 1, 2, \ldots$

$$d_{2m+1} = -\frac{(a+m)(a+b+m)x}{(a+2m)(a+2m+1)},$$

$$d_{2m} = \frac{m(b-m)x}{(a+2m-1)(a+2m)},$$

and we have used the continued fraction notation

$$a_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\cdots = a_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}}.$$

If we let f_n be the result of evaluating such a continued fraction through the terms a_n and b_n, then it can be shown by induction on n that $f_n{=}A_n/B_n$, where the A_n's and B_n's satisfy the recursion

$$A_j = b_j A_{j-1} + a_j A_{j-2} \qquad B_j = b_j B_{j-1} + a_j B_{j-2}, \quad j = 1, \ldots, n,$$

where $A_{-1} = 1$, $B_{-1} = 0$, $A_0 = b_0$, and $B_0 = 1$. Often this recursion is rescaled at each step, i.e. consecutive pairs of A's and B's are divided by the latest B (unless it is zero). This does not effect the value of f_n and in fact makes $A_n = f_n$, which can be tested for convergence of the continued fraction.

Calculating the Quantile Functions

Continuous Distributions

The quantile function $Q(u)$ of a random variable X having strictly increasing cdf F is the value q of X having $F(q) = u$, i.e. $Q(u) = F^{-1}(x)$. The usual method for finding the quantile function is to use one of the standard root-finding methods to solve $F(q) - u = 0$ for a specified u. For example, Newton's method finds q as the limit of the sequence

$$q_i = q_{i-1} - \frac{F(q_{i-1}) - u}{f(q_{i-1})}, \quad i = 1, 2, \ldots,$$

where the starting value q_0 is obtained by some approximation to q. Notice that since the cdf's of Z, t, χ^2, and F are related to IG and IB, then so are their inverses. Thus we could just write procedures to find these latter two. However, since a great deal of research has been devoted to finding approximations to the quantile functions of various distributions, we have listed them for a variety of distributions in the table at the end of the chapter.

Discrete Distributions

For a discrete random variable X having possible values $x_1, x_2, \ldots$ and corresponding probabilities $f(x_1), f(x_2), \ldots$ and cumulative probabilities $F(x_1), F(x_2), \ldots$, the quantile function $Q(u)$ of X is the smallest x such that $F(x) \geq u$. Thus the straightforward method of finding $Q(u)$ is to successively compare u to $F(x_1)$, $F(x_2)$, and so on until the first x is found for which $F(x) \geq u$.

Calculating Power and Sample Sizes

These operations require the use of the noncentral distributions listed in the table at the end of the chapter.

Random Number Generators

Methods for generating random numbers fall into two classes; 1) special purpose methods that are tailored to a specific distribution, and 2) general methods that can be applied to a wide variety of distributions.

Generating numbers from the uniform distribution plays a particularly important role because of the following theorem.

Theorem 1. PROBABILITY INTEGRAL TRANSFORM

Let X be a random variable having cdf F and quantile function Q, that is $Q(u) = \inf\{x : F(x) \geq u\}$ for $u \in [0,1]$. If U is a random variable having the U[0,1] distribution, then $Q(U)$ has the cdf F.

Thus in theory we have a method for generating random numbers $X_1, \ldots, X_n$ from any distribution, namely to first generate $U[0,1]$'s $U_1, \ldots, U_n$ and then to get $X_i = Q(U_i)$. This is easy to do for several of the distributions in the table at the end of this chapter including the exponential, weibull, and others. Unfortunately, there are many distributions for which it is too difficult to calculate the quantile function (the normal distribution for example which requires the evaluation of the incomplete gamma function) and special generators must be developed on a case by case basis.

Generating Normal Data

For many distributions it is possible to find a simple transformation of other easy to generate variables that has the desired distribution. For the normal distribution, there are two famous examples of this. The first is called the Box-Muller generator which relies on the fact that if U_1 and U_2 are independent $U(0,1)$ random variables, then

$$X_1 = \sqrt{-2\log U_1}\cos(2\pi U_2), \qquad X_2 = \sqrt{-2\log U_1}\sin(2\pi U_2),$$

are independent $N(0,1)$ random variables. Thus to generate a pair of normals, we need two uniforms, a log, a cosine, and a sine.

Another transformation can be found which also generates normals in pairs but reduces the amount of work required. Called Marsaglia's power method, it proceeds as follows. First, it generates a point (V_1, V_2) that is uniformly distributed in the unit circle (the circle of radius one centered at the origin). By uniform over a geometric region, we mean that the probability that a point falls into a subset of the region is equal to the "area" of the subset divided by the "area" of the entire region (for three dimensional regions, area would actually be volume). If U_1 and U_2 are independent $U(0,1)$ variables, then the point (V_1, V_2) where $V_i = 2(U_i - .5)$ is uniformly distributed in the unit square (square having vertices (1,0), (0,1), $(-1,0)$, and $(-1,-1)$ and if $W = V_1^2 + V_2^2 < 1$ then the point is also uniform on the unit circle. Thus we generate pairs of uniforms until we find one in the circle. Note that the chance that a point is in the circle is equal to the ratio of the area of the circle (which is π) to the area of the square (which is four) and this ratio is approximately 0.785.

Once we have a point in the circle, we can get independent $N(0,1)$'s X_1 and X_2 by

$$X_1 = V_1\sqrt{\frac{-2\log W}{W}}, \qquad X_2 = V_2\sqrt{\frac{-2\log W}{W}}.$$

This method avoids the calculation of the cosine and sine in the Box-Muller method at the expense of having to generate a few more uniforms (which we'll see is very simple).

Uniform Random Number Generators

Most authors attribute the first calculation-based (as opposed to mechanical devices such as dice) random number generator to von Neumann (1951) which consisted of generating a sequence of 'pseudo-random' numbers by starting with a four digit integer (called a 'seed') and squaring it and then using the four middle digits of the result as the next number in the sequence. To get a random number in the interval from zero to one, one can divide the elements of the sequence by 10,000. The numbers are called 'pseudo-random' since they are in fact generated by a deterministic method.

After 40 years of experimentation with such recursive methods, most uniform random number generators in common use are now based on an integer recursion of the form

$$X_{i+1} = (aX_i + b)\mathrm{mod}M, \qquad i \geq 0,$$

which is called a mixed linear congruential generator, and a vast literature exists on how the integers a, b, and M should be chosen (see Knuth, etc). The starting value X_0 is called a seed and again needs to be supplied by the user. Values between zero and one are then found by dividing the X's by M.

The basic considerations used in choosing a, b, and M include:

1. Because of the modular arithmetic involved, there are only M values that the X's can assume. Thus the value of M needs to be very large and typically are of the order of 2^{32}.

2. As soon as the recursion encounters an X that has occurred previously, then the sequence will repeat. Thus we want to choose a and b so that we are guaranteed of a long cycle length.

3. We must choose a, b, and M so that in any section of the sequence of uniforms, the numbers will appear to be "uniform" (each number in (0,1) seems to appear equally often) and "independent" (there is no discernible pattern in the numbers in the sequence).

Sampling from Discrete Distributions

Let X be a discrete random variable having possible values $x_1, x_2, \ldots$ and corresponding probabilities $f(x_1), f(x_2), \ldots$ and cumulative probabilities $F(x_1), F(x_2), \ldots$. Then the quantile function $Q(u)$ of X is given by the smallest x such that $F(x) \geq u$. Thus a straightforward method for simulating X is to first generate $U \sim U(0,1)$ and then successively compare U to $F(x_1)$, $F(x_2)$, and so on, until finding the first x for which $F(x) \geq U$.

Generating Multinomial Data

A wide variety of random number generation problems can be phrased in terms of randomly selecting (with replacement) one element from a set of K elements. For example a random variable X having the $B(n,p)$ distribution can be written as $X = \sum_{i=1}^{n} Y_i$ where each of the Y's is either one (with probability p) or zero (with probability $1-p$). A multinomial random vector $(X_1, \ldots, X_k)$ with n trials and probability vector $(p_1, \ldots, p_K)$ is similar except that X_i is the number of outcomes of type i (out of K possible outcomes).

In this section we consider methods for randomly selecting an outcome from a set of K outcomes when the probabilities of the outcomes are $p_1, \ldots, p_K$.

Equally Likely Outcomes

If the K outcomes are equally likely, then all we need to do is generate a random integer i from one to K and increment X_i. A random integer from 1 to K is easily obtained from $i = [UK]+1$ where U is $U[0,1]$. Simple extensions of equally likely outcomes are also simple to implement as well. For example, if we have three outcomes and the probabilities are 1/4, 1/2, and 1/4, then we can generate a random integer M from one to four and select outcome one if $M = 1$, outcome two if $M = 2$ or $M = 3$, and outcome three if $M = 4$.

Table Look-Up Methods

The same idea can be used anytime the probabilities $p_1, \ldots, p_K$ have only a small number of decimal places. For example, if there are three decimal places, that is, $p_i = n_i/1000$, for $i = 1, \ldots, K$, then we can form a vector l of length 1,000 whose first n_1 elements are all one, the next n_2 are all two, and so on. Then we can generate a random integer M from one to 1,000 and obtain an outcome equal to $l(M)$.

Generating Random Permutations

A method for producing permutations $i_1, \ldots, i_n$ of the integers one through n is said to produce random permutations if all $n!$ such permutations are equally likely to occur. An obvious way to generate random permutations is to generate a sequence of integers where at the ith step, the selected integer is chosen randomly from the set of integers that have not been selected so far. One simple method for doing this that doesn't require keeping track of what has been selected so far at any step is to start by constructing a vector x of length n where $x_i = i$, and then for $i = 1, \ldots, n-1$ randomly selecting an integer from i to n to be the index of the x to swap with x_i.

Sampling Without Replacement

There are many problems where we need to sample without replacement from some finite population. One example is generating a hypergeometric variable, that is, the number of 'defectives' in a random sample (without replacement) of n objects from a population of N objects wherein there are D defectives and $N-D$ nondefectives. A simple way to do this is to modify the method for generating random permutations. We start with a vector x of length N whose first D elements are ones and last $N-D$ are zeros. Then we can get the indices of the x's to include in our sample by running the random permutation procedure except running the loop from one to n rather than the full one to N.

The Poisson Distribution

The sampling methods given above can be used to generate observations from all of the common discrete distributions except for the Poisson which because of its infinite support and not being expressible in terms of simple sampling methods, requires special methods. Most methods use the fact that a Poisson variable X having intensity λ is the number of events in the interval [0,1] where the times between events are iid Exponential with parameter λ. Thus we can get a Poisson variable X as one less than the number of exponentials needed to get a sum greater than one. Since an exponential with parameter λ is just $-\log U/\lambda$ where U is uniform on [0,1], this means we need to generate uniforms until $-\sum_i \log U_i/\lambda > 1$ or equivalently until $\prod_i U_i < e^{-1/\lambda}$.

Some General Generation Methods

Rejection Methods

If it is hard to generate realizations from a random variable X having pdf f and we can find another pdf g defined over the same range as f for which it is easy to generate realizations and we can find a constant $c > 1$ such that

$$h(x) = cg(x) \geq f(x), \qquad \text{for all } x,$$

then we can generate a realization X from f via the following algorithm, called a rejection or acceptance-rejection method:

1. Generate Y from g.
2. Let $Z = Uh(Y)$ where U is $U(0,1)$ independent of Y.
3. Let $X = Y$ if $Z < f(Y)$, otherwise go back to 1.

This process generates a point (Y,Z) somewhere under the function h, and the basic idea is to use the abscissa of the point as our random number if the point also falls under f at that abscissa. The variable Y has density g while the conditional distribution of Z given Y is $U(0,h(Y))$ and thus the joint pdf of Y and Z is

$$f_{Y,Z}(y,z) = \frac{1}{h(y)}g(y) = \frac{1}{c}, \qquad y \in \mathcal{R},\ z \in [0,h(y)],$$

that is, the point is uniformly distributed above the region in the plane where $y \in \mathcal{R}$ and $z \in [0,h(y)]$. This means that $c\Pr(X \leq x)$ is the area under f for $Y \leq x$, which is $cF_X(x)$, and thus X has cdf F_X as desired.

The probability that a point will be rejected is the area between h and f divided by the area under h, that is,

$$\Pr(\text{rejection}) = \frac{\int_{-\infty}^{\infty}(h(x)-f(x))dx}{\int_{-\infty}^{\infty}h(x)dx} = 1 - \frac{1}{c},$$

and thus we should try to find a c that is as small as possible.

If we know that we can find an "enveloping function" g, then we can find the "best" value of c by finding

$$c = \max_x \left(\frac{f(x)}{g(x)}\right),$$

since we want $cg(x) \geq f(x)$.

To illustrate the rejection method, consider generating $X \sim \Gamma(\nu,1)$, that is from the pdf

$$f(x) = \frac{x^{\nu-1}e^{-x}}{\Gamma(\nu)}, \qquad x > 0,\ \nu > 1.$$

Note that if $X \sim \Gamma(\nu,1)$ then $X/\lambda \sim \Gamma(\nu,\lambda)$ so we need only consider the $\Gamma(\nu,1)$ case to take care of the two parameter case having $\nu > 1$. It is easy to show that for any ν we can find a c so that c times an exponential pdf having parameter $1/\nu$ is uniformly greater than f. Thus we find c as the maximum of the ratio of $f(x)$ to $g(x) = e^{-x/\nu}/\nu$. To find where this maximum occurs, let $y = f(x)/g(x)$ and note that

$$\log y = (\nu-1)\log x - x + \frac{x}{\nu} + \log(\nu/\Gamma(\nu)),$$

$$\frac{\partial \log y}{\partial x} = \frac{\nu-1}{x} - 1 + \frac{1}{\nu}, \quad \frac{\partial^2 \log y}{\partial x^2} = \frac{1-\nu}{x^2},$$

and thus the maximum of the ratio occurs at $x = \nu$ and has value $c = \nu^\nu e^{1-\nu}/\Gamma(\nu)$. This gives the algorithm:

1. Generate $Y \sim \text{Exp}(1/\nu)$, that is, $Y = -\nu\log(U_1)$, where $U_1 \sim U(0,1)$.
2. Let $Z = ce^{-Y/\nu}U_2/\nu)$, where $U_2 \sim U(0,1)$ independent of U_1.
3. Let $X = Y$ if and only if $Z < f(Y)$.

The Decomposition Method

Many times it is possible to decompose a difficult-to-simulate-from pdf f into a mixture of k pdf's, that is, we can write

$$f(x) = \sum_{i=1}^k \alpha_i f_i(x), \qquad \alpha_i \in [0,1],\ \sum_{i=1}^k \alpha_i = 1.$$

Then if we first pick pdf f_i with probability α_i and then generate a realization X from f_i, it is easy to see that in fact X is a realization from f, since

$$\begin{aligned}\Pr(X \leq x) &= \sum_{i=1}^k \Pr(X \leq x | X \text{ from } f_i)\Pr(X \text{ from } f_i) \\ &= \sum_{i=1}^k \alpha_i \int_{-\infty}^x f_i(z)dz = \int_{-\infty}^x \sum_{i=1}^k \alpha_i f_i(z)dz = \int_{-\infty}^x f(z)dz = F_X(x),\end{aligned}$$

as desired. Note that the method works for discrete distributions as well and that the method is very effective if the most likely pdf's are easy to simulate from.

To illustrate the method, consider simulating from the standard exponential pdf $f(x) = e^{-x}$ for $x > 0$. Since $Q(u) = -\log(1-U)$, it is easy to generate data from f directly using the quantile method. However, by using a very simple decomposition method, it is possible to avoid the log an appreciable percent of the time. If we decompose the region under the exponential pdf into 1) the rectangle having lower left corner (0,0) and upper right corner $(1, 1/e)$, 2) the "wedge" above the rectangle, and 3) the tail to the right of $x = 1$, we can decompose $f(x)$ into $\alpha_1 f_1(x) + \alpha_2 f_2(x) + \alpha_3 f_3(x)$ where $\alpha_1 = \alpha_3 = 1/e \doteq 0.368$, $\alpha_2 = 1 - 2/e \doteq 0.264$, and

$$f_1(x) = 1, \qquad x \in [0,1],$$

$$f_2(x) = \frac{e^{-x} - e^{-1}}{1 - 2/e}, \qquad x \in [0,1],$$

$$f_3(x) = e^{-(x-1)}, \qquad x \geq 1.$$

Thus, approximately 37% of the time we need only generate a $U(0,1)$, while approximately 26% of the time we need a realization from f_2, and 37% of the time from f_3. Now the quantile function of f_3 is $1 - \log(1-u)$, and thus we can generate a standard exponential via the quantile method. Finally, it remains to see how to simulate from f_2 the approximately 26% of the time we need to and be sure that this is simple enough that the decomposition method does in fact result in a savings over the quantile method.

It is easy to show that f_2 is the pdf of the minimum of N $U(0,1)$'s where N has the truncated Poisson distribution with parameter one, that is,

$$\Pr(N = n) = \frac{1}{n!(e-2)}, \qquad n = 2, 3, \ldots.$$

To three decimal places, the cumulative probabilities for $n = 2$ through $n = 6$ are .696, .928, .986, .998, and 1.000. Thus we can easily generate the Poisson using the direct quantile method for discrete variables, that is, generate a $U(0,1)$ and find the smallest value of n so that $\Pr(N \leq n) \geq U$. Almost always this will give a small value of N. In fact,

$$\mathrm{E}(N) = \sum_{n=2}^{\infty} n \frac{1}{n!(e-2)} = \frac{1}{e-2} \sum_{n=2}^{\infty} \frac{1}{(n-1)!} = \frac{1}{e-2} \sum_{n=1}^{\infty} \frac{1}{n!} = \frac{e-1}{e-2} \doteq 2.392.$$

This means that on the average we will need to generate approximately 3.392 $U(0,1)$'s to get one standard exponential (one to get N and 2.392 to get X), as well as do the comparisons in getting N and the minimum of N uniforms.

While this does not lead to great savings, it does illustrate the idea of the decomposition method.

Combining Rejection and Decomposition

Using the decomposition method leads to the need to generate data from several distributions and it is natural to in turn use rejection and/or decomposition on these distributions.

Generating Multivariate Data

Multivariate Normal

To generate a p-dimensional normal random vector X with given mean vector μ and covariance matrix Σ, we can first generate a p-dimensional normal random vector e having mean vector zero and identity covariance matrix, that is, e is a vector of p iid $N(0,1)$'s, and then calculate

$$X = \Sigma^{1/2} e + \mu,$$

where $\Sigma^{1/2}$ is the positive definite square root of Σ, that is Σ can be written as the product of $\Sigma^{1/2}$ times the transpose of $\Sigma^{1/2}$. This method relies on the facts that if Z is an r dimensional random vector and A is an $(n \times r)$ matrix of constants, then

$$\mathrm{Var}(AZ) = A\mathrm{Var}(Z)A^T,$$

and that a matrix times a normal random vector is also normally distributed.

An important special case is a bivariate normal vector $(X, Y)^T$ having means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation coefficient ρ. Thus

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}, \qquad \Sigma^{1/2} = \begin{bmatrix} \sigma_X & 0 \\ \rho\sigma_Y & \sigma_Y\sqrt{1-\rho^2} \end{bmatrix},$$

which gives

$$X = \sigma_X e_1 + \mu_X, \qquad Y = \rho\sigma_Y e_1 + \sigma_Y\sqrt{1-\rho^2}e_2 + \mu_Y,$$

where e_1 and e_2 are independent N(0,1)'s.

The Gibbs Sampler

Simulating data of possibly very high dimension (such as simulating visual images) has been made possible in recent years by the creation of methods collectively called Monte Carlo Markov Chain (MCMC) methods, the most famous of which is called Gibbs sampling, the basic idea of which is the following. Suppose we want to generate n observations from a random vector $X = (X_1, \ldots, X_p)^T$ of dimension p. In situations other than the multivariate normal case, this is usually a very difficult problem. Further, there are many situations where we want to get observations when we don't even know the complete joint distribution of X. If we do know the conditional distribution of each X_i given the other $p-1$ X's, then we can do a series of "Gibbs steps." We start with some initial values of all the X's and then one first generates an X_1 from its conditional distribution given the others (using the inital values for the others), then generate an X_2 from its conditional distribution given the just obtained value of X_1 and the initial values for the other X's, then an X_3 from X_1, X_2, and the inital values of the others, and so on. The second Gibbs step is similar except that one uses the X's from the first step the same way the initial values were used during the first step. Thus, each time one is generating an individual X_i, it is from a univariate distribution. Further, there are many situations where the conditional distributions are known when the complete p-dimensional distribution is unknown.

The remarkable fact about this Gibbs sampling procedure is that in a great many important situations, after a reasonable number of Gibbs steps, the distribution of the generated vectors is indistinguishable from the desired joint distribution. See Casella and George (1992, *American Statistician*, pg. 167–174) for details.

Table of Distributions

The table makes extensive use of the beta and gamma functions:

$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad \Gamma(a) = \int_0^\infty z^{a-1}e^{-z}\,dz, \quad a, b > 0,$$

and the incomplete gamma and beta functions:

$$\mathrm{IG}(x; a) = \frac{1}{\Gamma(a)}\int_0^x e^{-t}t^{a-1}\,dt, \quad \mathrm{IB}(x; a, b) = \frac{1}{\beta(a, b)}\int_0^x t^{a-1}(1-t)^{b-1}\,dt,$$

for $a, b > 0,\ 0 < x < 1$.

Four Discrete Distributions

Binomial with Parameters n and p

X is 1) the number of 1's in a sample (with replacement) from a 0–1 population having proportion p of 1's, or 2) the number of "successes" in n independent "trials" where each trial can be a success (with probability p) or a failure (with probability $1-p$).

$$\mathrm{E}(X) = np, \qquad \mathrm{Var}(X) = np(1-p).$$

$$f_{Bin}(x; n, p) = \binom{n}{p} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Hypergeometric with Parameters N, n, and M

X is the number of 1's in a sample (without replacement) of size n from a 0–1 population of size N, where the number of 1's in the population is M.

$$E(X) = np, \quad p = M/N, \qquad \mathrm{Var}(X) = \frac{N-n}{N-1}np(1-p).$$

$$f_{Hyp}(x; N, n, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad \max(0, M-(N-n)) \le x \le \min(n, M).$$

Negative Binomal with parameters n and p

X is the number of the binomial trial where the nth success occurs when on each trial the probability of a success is p.

$$E(X) = n/p, \qquad \mathrm{Var}(X) = \frac{n(1-p)}{p^2}.$$

$$f_{Nbin}(x; n, p) = \binom{x-1}{n-1} p^n (1-p)^{x-n}, \qquad x = n, n+1, n+2, \ldots.$$

Poisson with parameter λ

X arises in two ways: (1) If we have a series of events in which the times between events are independent and have the exponential distribution with mean μ, then the number of events in a time interval of length T has the Poisson distribution with $\lambda = T/\mu$. (2) If a binomial parameter n is large and p is small, then the binomial distribution can be approximated by the Poisson distribution with $\lambda = np$.

$$E(X) = \lambda, \qquad \mathrm{Var}(X) = \lambda.$$

$$f_{Poiss}(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$

Normal and Those Derived from Normal

Chi-Square with ν Degrees of Freedom

$\chi^2_\nu = \sum_{i=1}^{\nu} X_i^2$, where $X_1, \ldots, X_\nu$ are iid $N(0,1)$.

$$E(\chi^2_\nu) = \nu, \qquad \mathrm{Var}(\chi^2_\nu) = 2\nu.$$

$$f_{\chi^2}(x; \nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} e^{-x/2} x^{(\nu/2)-1}, \quad F_{\chi^2}(x; \nu) = \mathrm{IG}\left(\frac{x}{2}; \frac{\nu}{2}\right), \quad x > 0.$$

Define the Wilson-Hilferty approximation

$$x_0 = \nu \left[z \left(\frac{2}{9\nu}\right)^{1/2} + 1 - \frac{2}{9\nu} \right]^3, \qquad z = Q_Z(u).$$

Then (Kennedy and Gentle, pg. 118):

$$q_0(u;\chi^2_\nu) = \begin{cases} \left(u\nu\Gamma\left(\frac{\nu}{2}\right)2^{(\nu-2)/2}\right)^{2/\nu}, & \nu < -1.24\log(u), \\ x_0, & \nu \geq -1.24\log(u) \text{ and } x_0 \leq 2.2\nu+6 \\ -2\log\left(\frac{(1-u)\Gamma(\nu/2)}{(x_0/2)^{(\nu/2)-1}}\right), & \nu \geq -1.24\log(u) \text{ and } x_0 > 2.2\nu+6. \end{cases}$$

Noncentral Chi-Square with ν Degrees of Freedom and Noncentrality Parameter λ

$\chi^2_{\nu,\lambda} = \sum_{i=1}^{\nu} X_i^2$, $X_1,\ldots,X_\nu$ independent, $X_i \sim N(\mu_i, 1)$, $\lambda = \sum_{i=1}^{\nu}\mu_i^2$

$$\mathrm{E}(\chi^2_{\nu,\lambda}) = \nu+\lambda, \quad \mathrm{Var}(\chi^2_{\nu,\lambda}) = 2(\nu+2\lambda).$$

$$f_{\chi^2}(x;\nu,\lambda) = \sum_{j=0}^{\infty} f_{\mathrm{Poiss}}(j;\lambda/2) f_{\chi^2}(x;\nu+2j)$$

$$F_{\chi^2}(x;\nu,\lambda) = \sum_{j=0}^{\infty} f_{\mathrm{Poiss}}(j;\lambda/2) F_{\chi^2}(x;\nu+2j)$$

$$q_0(u;\chi^2_{\nu,\lambda}) = \frac{1+b}{2}\left[Q_Z(u) + \sqrt{\frac{2a}{1+b}-1}\right]^2, \quad \text{(A\&S, 26.4.31)}$$

where $a = \nu+\lambda$, $b = \lambda/(\nu+\lambda)$.

***F* with ν_1 and ν_2 Degrees of Freedom**

$$F_{\nu_1,\nu_2,\lambda} = \frac{Z_1/\nu_1}{Z_2/\nu_2}, \qquad Z_1 \sim \chi^2_{\nu_1},\ Z_2 \sim \chi^2_{\nu_2},\ Z_1, Z_2 \text{ independent.}$$

$$\mathrm{E}(F_{\nu_1,\nu_2}) = \frac{\nu_2}{\nu_2-2}, \quad \nu_2 > 2, \quad \mathrm{Var}(F_{\nu_1,\nu_2}) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}, \quad \nu_2 > 4.$$

$$f_F(x;\nu_1,\nu_2) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} x^{(\nu_1/2)-1}}{\beta\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)\left(1+\frac{\nu_1 x}{\nu_2}\right)^{(\nu_1+\nu_2)/2}}, \qquad x > 0$$

$$F_F(x;\nu_1,\nu_2) = 1 - \mathrm{IB}\left(\frac{\nu_2}{\nu_2+\nu_1 x}; \frac{\nu_2}{2}, \frac{\nu_1}{2}\right)$$

For $\nu_1 > 1$ and $\nu_2 > 1$, with $z = Q_Z(u)$, $a = \nu_2/2$, and $b = \nu_1/2$, we have (A & S, 26.5.22, 26.6.16):

$$q_0(u;F_{\nu_1,\nu_2}) = e^{2w},$$

where

$$w = \frac{z(h+\lambda)^{1/2}}{h} - \left(\frac{1}{2b-1} - \frac{1}{2a-1}\right)\left(\lambda + \frac{5}{6} - \frac{2}{3h}\right)$$

$$h = 2\left(\frac{1}{2a-1} + \frac{1}{2b-1}\right)^{-1}, \quad \lambda = \frac{z^2-3}{6}.$$

If $\nu_1 = 1$ or $\nu_2 = 1$, we have

$$Q_F(u; 1, \nu) = \left[Q_t\left(\frac{1+u}{2}; \nu\right)\right]^2, \qquad Q_F(u; \nu, 1) = \left[\frac{1}{Q_t\left(1 - \frac{u}{2}; \nu\right)}\right]^2.$$

Noncentral F with ν_1 and ν_2 Degrees of Freedom and Noncentrality Parameter λ

$$F_{\nu_1,\nu_2,\lambda} = \frac{Z_1/\nu_1}{Z_2/\nu_2}, \qquad Z_1 \sim \chi^2_{\nu_1,\lambda},\ Z_2 \sim \chi^2_{\nu_2},\ Z_1, Z_2 \text{ independent}$$

$$\mathrm{E}(F_{\nu_1,\nu_2,\lambda}) = \frac{\nu_2(\nu_1 + \lambda)}{\nu_1(\nu_2 - 2)}, \quad \nu_2 > 2,$$

$$\mathrm{Var}(F_{\nu_1,\nu_2,\lambda}) = 2\left(\frac{\nu_2}{\nu_1}\right)^2 \frac{(\nu_1 + \lambda)^2 + (\nu_1 + 2\lambda)(\nu_2 - 2)}{(\nu_2 - 2)^2(\nu_2 - 4)}, \quad \nu_2 > 4.$$

$$f_F(x; \nu_1, \nu_2, \lambda) = \sum_{j=0}^{\infty} \frac{(\lambda/2)^j e^{-\lambda/2} \left(\frac{\nu_1}{\nu_2}\right)^{(\nu_1+2j)/2} x^{(\nu_1+2j-2)/2}}{j!\beta\left(\frac{\nu_2}{2}, \frac{\nu_1 + 2j}{2}\right)\left(1 + \frac{\nu_1 x}{\nu_2}\right)^{(\nu_1+\nu_2+2j)/2}}, \ x > 0$$

$$F_F(x; \nu_1, \nu_2, \lambda) = \sum_{j=0}^{\infty} f_{\mathrm{Poiss}}(j; \lambda/2)\mathrm{IB}\left(\frac{\nu_1 x}{\nu_1 x + \nu_2}; \frac{\nu_1}{2} + j, \frac{\nu_2}{2}\right)$$

Normal Distribution with Mean μ and Variance σ^2

$$f_N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \qquad x \in \mathcal{R},\ \mu \in \mathcal{R},\ \sigma^2 > 0$$

$$F_Z(x) = \begin{cases} \frac{1}{2} + \frac{1}{2}\mathrm{IG}\left(\frac{x^2}{2}; \frac{1}{2}\right), & x \geq 0 \\ 1 - F_Z(-x), & x < 0, \end{cases}$$

Defining $t = \sqrt{-2\log u}$, we have (A & S, eq. 26.2.22):

$$q_0(u; Z) = \begin{cases} \frac{2.30753 + .27061t}{1 + .99229t + .04481t^2} - t, & 0 < u \leq 0.5, \\ -q_0(1 - u; Z), & .5 < u < 1. \end{cases}$$

Student t with ν Degrees of Freedom

$$t_{\nu,\lambda} = \frac{X}{\sqrt{Y/\nu}}, \qquad X \sim N(0, 1),\ Y \sim \chi^2_\nu,\ X, Y \text{ independent}$$

$$\text{Odd moments are zero}, \mathrm{Var}(t_\nu) = \frac{v}{\nu - 2}, \quad \nu > 2.$$

$$f_t(x;\nu) = \frac{1}{\sqrt{\nu}\beta(1/2,\nu/2)}(1+(x^2/\nu))^{-(\nu+1)/2}, \quad -\infty < x < \infty$$

$$F_t(x;\nu) = \begin{cases} 1 - \mathrm{IB}\left(\nu/(\nu+x^2);\nu/2,1/2\right)/2, & x \geq 0 \\ 1 - F_t(-x;\nu), & x < 0, \end{cases}$$

Defining $z = q_0(u;Z)$, one approximation is given by (A & S, eq. 26.7.5)

$$q_0(u;t_\nu) = \begin{cases} z + \sum_{j=1}^{4} \frac{g_j(z)}{\nu^j}, & .5 \leq u < 1, \\ -q_0(1-u;t_\nu), & 0 \leq .5, \end{cases}$$

$$g_1(z) = \frac{1}{4}(z^3+z), \qquad g_2(z) = \frac{1}{96}(5z^5+16z^3+3z),$$

$$g_3(z) = \frac{1}{384}(3z^7+19z^5+17z^3-15z),$$

$$g_4(z) = \frac{1}{92160}(79z^9+776z^7+1482z^5-1920z^3-945z).$$

Noncentral t with ν Degrees of Freedom and Noncentrality Parameter λ

$$t_{\nu,\lambda} = \frac{X}{\sqrt{Y/\nu}}, \qquad X \sim N(\lambda,1),\ Y \sim \chi^2_\nu,\ X,Y \text{ independent}$$

$$\mathrm{E}(t_{\nu,\lambda}) = \left(\frac{\nu}{2}\right)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\lambda, \quad \mathrm{Var}(t_{\nu,\lambda}) = \frac{\nu}{\nu-2}(1+\lambda^2) - \mathrm{E}^2(t_{\nu,\lambda}).$$

$$f_t(x;\nu,\lambda) = \frac{\nu^{\nu/2}}{\sqrt{\pi}\Gamma(\nu/2)} \frac{e^{-\lambda^2/2}}{(\nu+t^2)^{(\nu+1)/2}} \sum_{s=0}^{\infty} \Gamma\left(\frac{\nu+s+1}{2}\right)\left(\frac{\lambda^s}{s!}\right)\left(\frac{2t^2}{\nu+t^2}\right)^{s/2}$$

$$F_t(x;\nu,\lambda) = 1 - \sum_{j=0}^{\infty} e^{-\lambda^2/2} \frac{\left(\lambda^2/2\right)^j}{2j!} \mathrm{IB}\left(\frac{\nu}{\nu+x^2};\frac{\nu}{2},j+\frac{1}{2}\right) \doteq F_Z(a/b)$$

where $a = x\left(1 - \frac{1}{4\nu}\right) - \lambda$ and $b = \left(1 + \frac{x^2}{2\nu}\right)^{1/2}$.

Other Continuous Distributions

Beta(α,β)

$$f_\beta(x;\alpha,\beta) = \frac{1}{\beta(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \qquad x \in [0,1],\ \alpha > 0,\ \beta > 0$$

$$\mu_X = \frac{\alpha}{\alpha+\beta}, \qquad \sigma^2_X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Cauchy(θ,σ)

$$f_{Cauch}(x;\theta,\sigma) = \frac{1}{\pi\sigma} \frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}, \qquad x \in \mathcal{R},\ \theta \in \mathcal{R},\ \sigma > 0$$

Gumbel(α, β)

$Y = \alpha - \beta X$, where X is exponential with parameter 1.

$$f(x; \alpha, \beta) = \frac{1}{\beta} e^{-e^{-(x-\alpha)/\beta}} e^{-(x-\alpha)/\beta}, \qquad x \in \mathcal{R},\ \alpha \in \mathcal{R},\ \beta > 0$$

$$F(x; \alpha, \beta) = e^{-e^{-(x-\alpha)/\beta}}$$

$$Q(u) = \alpha - \beta \log(-\log u), \qquad 0 \le u \le 1$$

$$\mu_X = \alpha + \beta\gamma,\ \gamma \approx .577216,\ \sigma_X^2 = \frac{\pi^2\beta^2}{6}$$

Laplace(μ, σ)

$$f(x; \mu, \sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, \qquad x \in \mathcal{R},\ \mu \in \mathcal{R},\ \sigma > 0$$

$$F(x; \mu, \sigma) = \begin{cases} \frac{1}{2} e^{(x-\mu)/\sigma}, & \text{if } x < \mu \\ 1 - \frac{1}{2} e^{-(x-\mu)/\sigma}, & \text{if } x \ge \mu \end{cases}$$

$$Q(u) = \begin{cases} \sigma \log(2u) + \mu, & \text{if } 0 < u < \frac{1}{2} \\ -\sigma \log 2(1-u) + \mu, & \text{if } \frac{1}{2} \le u < 1 \end{cases}$$

Logistic(μ, β)

For $x \in \mathcal{R}$, $\mu \in \mathcal{R}$, and $\beta > 0$,

$$f(x; \mu, \beta) = \frac{1}{\beta} \frac{e^{-(x-\mu)/\beta}}{[1 + e^{-(x-\mu)/\beta}]^2}, \quad F(x; \mu, \beta) = \frac{1}{1 + e^{-(x-\mu)/\beta}}, \quad Q(u) = \mu - \beta \log \frac{1-u}{u}.$$

Lognormal(μ, σ^2)

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma x} e^{(-\log x - \mu)^2/(2\sigma^2)}, \qquad x > 0,\ \mu \in \mathcal{R},\ \sigma > 0$$

$$\mu_X = e^{\mu + (\sigma^2/2)},\ \sigma_X^2 = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$$

Weibull(α, β)

For $x > 0$, $\alpha > 0$, and $\beta > 0$,

$$f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}, \quad F(x) = 1 - e^{-\alpha x^\beta}, \quad Q(u) = \left[\frac{-\log(1-u)}{\alpha}\right]^{1/\beta}.$$