then identify a marginal or conditional distribution that corresponds to the desired distribution.

The transformations must be done with care and must respect the nonrandomness in the underlying uniform generator.

For example, for a double exponential distribution, with density

$$p(x) = \frac{1}{2}e^{-|x|},$$

consider the simple method:
Generate U_1 and U_2; set $X = \log(U_1)$; then if $U_2 > 0.5$, set $X = -X$.

Often such mathematically correct transformations must be performed with special care on the computer. In this example, if the uniform stream is from a linear congruential generator with a relatively small multiplier, the method will yield a stream of double exponentials in which all extreme values are positive. (Because if U_1 is very small, U_2 will be also.)

Inverse CDF Method

If X is a scalar random variable with a continuous cumulative distribution function (CDF) P_X, then the random variable

$$U = P_X(X)$$

has a U(0, 1) distribution.

This fact provides a very simple relationship with a uniform random variable U and a random variable X with CDF P_X, namely,

$$X = P_X^{-1}(U), \tag{7.3}$$

where the inverse of the CDF exists. Use of this straightforward transformation is called the *inverse CDF* technique. The log transformation mentioned above that yields an exponential random variable uses the inverse CDF.

For a discrete random variable, although the inverse of the CDF does not exist, the inverse CDF method can still be used. The value of the discrete random variable is chosen as the smallest value within its countable range such that the CDF is no less than the value of the uniform variate.

For a multivariate random variable, the inverse CDF method yields a level curve in the range of the random variable; hence, the method is not directly useful for multivariate random variables. Multivariate random variates can be generated using the inverse CDF method first on a univariate marginal and then on a sequence of univariate conditionals.

Acceptance/Rejection Methods

Acceptance/rejection methods for generating realizations of a random variable X make use of realizations of another random variable Y whose PDF g_Y is similar to the PDF of X, p_X. The random variable Y is chosen so that we can easily generate realizations of it and so that its density g_Y can be scaled to majorize p_X using some constant c; that is, so that $cg_Y(x) \geq p_X(x)$ for all x. The density g_Y is called the *majorizing* density, and cg_Y is called the majorizing function. The majorizing density is also called the "proposal density". The density of interest, p_X, is called the "target density". The support of the target density must be contained in the support of the majorizing density; for densities with infinite support, the majorizing density must likewise have infinite support. In the case of infinite support, it is critical that the majorizing density not approach zero faster than the target density.

Acceptance/rejection methods can also be used for discrete random variables. We use the term "probability density" to include a probability mass function, and all of the discussion in this section applies equally to probability functions and probability densities.

Unlike the inverse CDF method, acceptance/rejection methods apply immediately to multivariate random variables.

Algorithm 7.1 The Acceptance/Rejection Method to Convert Uniform Random Numbers

1. Generate y from the distribution with density function g_Y.
2. Generate u from a uniform (0,1) distribution.
3. If $u \leq p_X(y)/cg_Y(y)$, then
   3.a. take y as the desired realization;
   otherwise,
   3.b. return to step 1. ∎

It is easy to see that Algorithm 7.1 produces a random variable with the density p_X. Let Z be the random variable delivered. For any x, because Y (from the density g) and U are independent, we have

$$\begin{aligned}
\Pr(Z \leq x) &= \Pr\left(Y \leq x \mid U \leq \frac{p_X(Y)}{cg_Y(Y)}\right) \\
&= \frac{\int_{-\infty}^{x} \int_{0}^{p_X(t)/cg_Y(t)} g_Y(t)\,\mathrm{d}s\,\mathrm{d}t}{\int_{-\infty}^{\infty} \int_{0}^{p_X(t)/cg_Y(t)} g_Y(t)\,\mathrm{d}s\,\mathrm{d}t} \\
&= \int_{-\infty}^{x} p_X(t)\,\mathrm{d}t,
\end{aligned}$$

which is the CDF corresponding to p_X. (Differentiating this quantity with respect to x yields $p_X(x)$.) Therefore, Z has the desired distribution.

It is easy to see that the random variable corresponding to the number of passes through the steps of Algorithm 7.1 until the desired variate is delivered has a geometric distribution. This random variable is a measure of the inefficiency of the algorithm. Because both cg_Y and p_X are densities, it is easy to see that the expected value of this random variable is c. (See Exercise 7.3.)

TABLE 6.1 Some methods for generating a random variable X from familiar distributions.

Distribution	Method
Uniform	See [195, 227, 383, 538, 539, 557]. For $X \sim \text{Unif}(a, b)$; draw $U \sim \text{Unif}(0, 1)$; then let $X = a + (b - a)U$.
Normal(μ, σ^2) and Lognormal(μ, σ^2)	Draw $U_1, U_2 \sim$ i.i.d. Unif(0, 1); then $X_1 = \mu + \sigma\sqrt{-2\log U_1}\cos\{2\pi U_2\}$ and $X_2 = \mu + \sigma\sqrt{-2\log U_1}\sin\{2\pi U_2\}$ are independent $N(\mu, \sigma^2)$. If $X \sim N(\mu, \sigma^2)$ then $\exp\{X\} \sim \text{Lognormal}(\mu, \sigma^2)$.
Multivariate $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Generate standard multivariate normal vector, $\mathbf{Y}$, coordinatewise; then $\mathbf{X} = \boldsymbol{\Sigma}^{-1/2}\mathbf{Y} + \boldsymbol{\mu}$.
Cauchy(α, β)	Draw $U \sim \text{Unif}(0, 1)$; then $X = \alpha + \beta\tan\{\pi(U - \frac{1}{2})\}$.
Exponential(λ)	Draw $U \sim \text{Unif}(0, 1)$; then $X = -(\log U)/\lambda$.
Poisson(λ)	Draw $U_1, U_2, \ldots \sim$ i.i.d. Unif(0, 1); then $X = j - 1$, where j is the lowest index for which $\prod_{i=1}^{j} U_i < e^{-\lambda}$.
Gamma(r, λ)	See Example 6.1, references, or for integer r, $X = -(1/\lambda)\sum_{i=1}^{r}\log U_i$ for $U_1, \ldots, U_r \sim$ i.i.d. Unif(0, 1).
Chi-square (df $= k$)	Draw $Y_1, \ldots, Y_k \sim$ i.i.d. $N(0, 1)$, then $X = \sum_{i=1}^{k} Y_i^2$; or draw $X \sim \text{Gamma}(k/2, \frac{1}{2})$.
Student's t (df $= k$) and $F_{k,m}$ distribution	Draw $Y \sim N(0, 1)$, $Z \sim \chi_k^2$, $W \sim \chi_m^2$ independently, then $X = Y/\sqrt{Z/k}$ has the t distribution and $F = (Z/k)/(W/m)$ has the F distribution.
Beta(a, b)	Draw $Y \sim \text{Gamma}(a, 1)$ and $Z \sim \text{Gamma}(b, 1)$ independently; then $X = Y/(Y + Z)$.
Bernoulli(p) and Binomial(n, p)	Draw $U \sim \text{Unif}(0, 1)$; then $X = 1_{\{U<p\}}$ is Bernoulli(p). The sum of n independent Bernoulli(p) draws has a Binomial(n, p) distribution.
Negative Binomial(r, p)	Draw $U_1, \ldots, U_r \sim$ i.i.d. Unif(0, 1); then $X = \sum_{i=1}^{r}\lfloor(\log U_i)/\log\{1-p\}\rfloor$, and $\lfloor\cdot\rfloor$ means greatest integer.
Multinomial(1, $(p_1, \ldots, p_k)$)	Partition [0, 1] into k segments so the ith segment has length p_i. Draw $U \sim \text{Unif}(0, 1)$; then let X equal the index of the segment into which U falls. Tally such draws for Multinomial(n, $(p_1, \ldots, p_k)$).
Dirichlet($\alpha_1, \ldots, \alpha_k$)	Draw independent $Y_i \sim \text{Gamma}(\alpha_i, 1)$ for $i = 1, \ldots, k$; then $\mathbf{X}^{\mathrm{T}} = \left(Y_1/\sum_{i=1}^{k} Y_i, \ldots, Y_k/\sum_{i=1}^{k} Y_i\right)$.

Although this approach is not exact, we include it in this section because the degree of approximation is deterministic and can be reduced to any desired level by increasing m sufficiently. Compared to the alternatives, this simulation method is not appealing because it requires a complete approximation to F regardless of the desired sample size, it does not generalize to multiple dimensions, and it is less efficient than other approaches.

6.2.3 Rejection Sampling

If $f(x)$ can be calculated, at least up to a proportionality constant, then we can use *rejection sampling* to obtain a random draw from exactly the target distribution. This strategy relies on sampling candidates from an easier distribution and then correcting the sampling probability through random rejection of some candidates.

Let g denote another density from which we know how to sample and for which we can easily calculate $g(x)$. Let $e(\cdot)$ denote an *envelope*, having the property $e(x) = g(x)/\alpha \geq f(x)$ for all x for which $f(x) > 0$ for a given constant $\alpha \leq 1$. Rejection sampling proceeds as follows:

1. Sample $Y \sim g$.
2. Sample $U \sim \text{Unif}(0, 1)$.
3. Reject Y if $U > f(Y)/e(Y)$. In this case, do not record the value of Y as an element in the target random sample. Instead, return to step 1.
4. Otherwise, keep the value of Y. Set $X = Y$, and consider X to be an element of the target random sample. Return to step 1 until you have accumulated a sample of the desired size.

The draws kept using this algorithm constitute an i.i.d. sample from the target density f; there is no approximation involved. To see this, note that the probability that a kept draw falls at or below a value y is

$$\begin{aligned} P[X \leq y] &= P\left[Y \leq y \,\middle|\, U \leq \frac{f(Y)}{e(Y)}\right] \\ &= P\left[Y \leq y \text{ and } U \leq \frac{f(Y)}{e(Y)}\right] \Big/ P\left[U \leq \frac{f(Y)}{e(Y)}\right] \\ &= \int_{-\infty}^{y}\int_{0}^{f(z)/e(z)} du\; g(z)\,dz \Big/ \int_{-\infty}^{\infty}\int_{0}^{f(z)/e(z)} du\; g(z)\,dz \qquad (6.4) \\ &= \int_{-\infty}^{y} f(z)\,dz, \qquad (6.5) \end{aligned}$$

which is the desired probability. Thus, the sampling distribution is exact, and α can be interpreted as the expected proportion of candidates that are accepted. Hence α is a measure of the efficiency of the algorithm. We may continue the rejection sampling procedure until it yields exactly the desired number of sampled points, but this requires a random total number of iterations that will depend on the proportion of rejections.

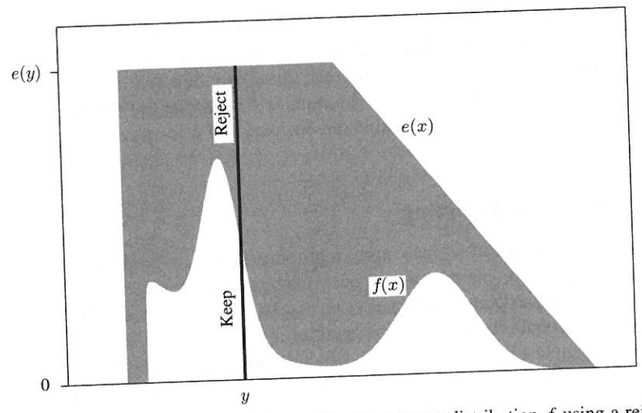

FIGURE 6.1 Illustration of rejection sampling for a target distribution f using a rejection sampling envelope e.

Recall the rejection rule in step 3 for determining the fate of a candidate draw, $Y = y$. Sampling $U \sim \text{Unif}(0, 1)$ and obeying this rule is equivalent to sampling $U|y \sim \text{Unif}(0, e(y))$ and keeping the value y if $U < f(y)$. Consider Figure 6.1. Suppose the value y falls at the point indicated by the vertical bar. Then imagine sampling $U|Y = y$ uniformly along the vertical bar. The rejection rule eliminates this Y draw with probability proportional to the length of the bar above $f(y)$ relative to the overall bar length. Therefore, one can view rejection sampling as sampling uniformly from the two-dimensional region under the curve e and then throwing away any draws falling above f and below e. Since sampling from f is equivalent to sampling uniformly from the two-dimensional region under the curve labeled $f(x)$ and then ignoring the vertical coordinate, rejection sampling provides draws exactly from f.

The shaded region in Figure 6.1 above f and below e indicates the waste. The draw $Y = y$ is very likely to be rejected when $e(y)$ is far larger than $f(y)$. Envelopes that exceed f everywhere by at most a slim margin produce fewer wasted (i.e., rejected) draws and correspond to α values near 1.

Suppose now that the target distribution f is only known up to a proportionality constant c. That is, suppose we are only able to compute easily $q(x) = f(x)/c$, where c is unknown. Such densities arise, for example, in Bayesian inference when f is a posterior distribution known to equal the product of the prior and the likelihood scaled by some normalizing constant. Fortunately, rejection sampling can be applied in such cases. We find an envelope e such that $e(x) \geq q(x)$ for all x for which $q(x) > 0$. A draw $Y = y$ is rejected when $U > q(y)/e(y)$. The sampling probability remains correct because the unknown constant c cancels out in the numerator and denominator of (6.4) when f is replaced by q. The proportion of kept draws is α/c.

Multivariate targets can also be sampled using rejection sampling, provided that a suitable multivariate envelope can be constructed. The rejection sampling algorithm is conceptually unchanged.

To produce an envelope we must know enough about the target to bound it. This may require optimization or a clever approximation to f or q in order to ensure that e can be constructed to exceed the target everywhere. Note that when the target is continuous and log-concave, it is unimodal. If we select x_1 and x_2 on opposite sides of that mode, then the function obtained by connecting the line segments that are tangent to $\log f$ or $\log q$ at x_1 and x_2 yields a piecewise exponential envelope with exponential tails. Deriving this envelope does not require knowing the maximum of the target density; it merely requires checking that x_1 and x_2 lie on opposite sides of it. The adaptive rejection sampling method described in Section 6.2.3.2 exploits this idea to generate good envelopes.

To summarize, good rejection sampling envelopes have three properties: They are easily constructed or confirmed to exceed the target everywhere, they are easy to sample, and they generate few rejected draws.

Example 6.1 (Gamma Deviates) Consider the problem of generating a Gamma(r, 1) random variable when $r \geq 1$. When Y is generated according to the density

$$f(y) = \frac{t(y)^{r-1} t'(y) \exp\{-t(y)\}}{\Gamma(r)} \tag{6.6}$$

for $t(y) = a(1 + by)^3$ for $-1/b < y < \infty, a = r - \frac{1}{3}$, and $b = 1/\sqrt{9a}$, then $X = t(Y)$ will have a Gamma(r, 1) distribution [443]. Marsaglia and Tsang describe how to use this fact in a rejection sampling framework [444]. Adopt (6.6) as the target distribution because transforming draws from f gives the desired gamma draws.

Simplifying f and ignoring the normalizing constant, we wish to generate from the density that is proportional to $q(y) = \exp\{a \log\{t(y)/a\} - t(y) + a\}$. Conveniently, q fits snugly under the function $e(y) = \exp\{-y^2/2\}$, which is the unscaled standard normal density. Therefore, rejection sampling amounts to sampling a standard normal random variable, Z, and a standard uniform random variable, U, then setting $X = t(Z)$ if

$$U \leq \frac{q(Z)}{e(Z)} = \exp\left\{\frac{Z^2}{2} + a \log\left\{\frac{t(Z)}{a}\right\} - t(Z) + a\right\} \tag{6.7}$$

and $t(Z) > 0$. Otherwise, the draw is rejected and the process begun anew. An accepted draw has density Gamma(r, 1). Draws from Gamma(r, 1) can be rescaled to obtain draws from Gamma(r, λ).

In a simulation when $r = 4$, over 99% of candidate draws are accepted and a plot of $e(y)$ and $q(y)$ against y shows that the two curves are nearly superimposed. Even in the worst case ($r = 1$), the envelope is excellent, with less than 5% waste.□

Example 6.2 (Sampling a Bayesian Posterior) Suppose 10 independent observations (8, 3, 4, 3, 1, 7, 2, 6, 2, 7) are observed from the model $X_i|\lambda \sim \text{Poisson}(\lambda)$. A lognormal prior distribution for λ is assumed: $\log \lambda \sim N(\log 4, 0.5^2)$. Denote the likelihood as $L(\lambda|\mathbf{x})$ and the prior as $f(\lambda)$. We know that $\hat{\lambda} = \bar{x} = 4.3$ maximizes $L(\lambda|\mathbf{x})$ with respect to λ; therefore the unnormalized posterior, $q(\lambda|\mathbf{x}) = f(\lambda)L(\lambda|\mathbf{x})$

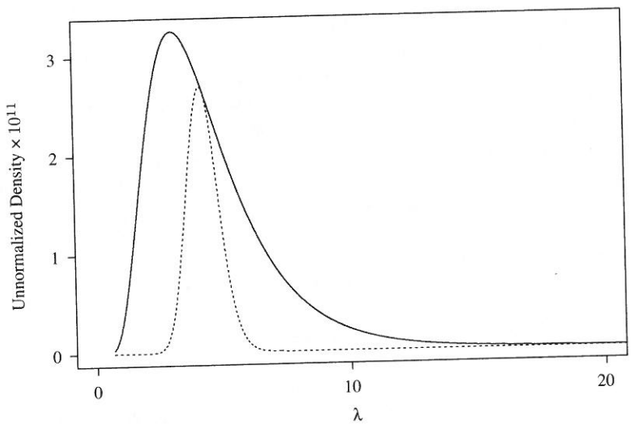

FIGURE 6.2 Unnormalized target (dotted) and envelope (solid) for rejection sampling in Example 6.2.

is bounded above by $e(\lambda) = f(\lambda)L(4.3|\mathbf{x})$. Figure 6.2 shows q and e. Note that the prior is proportional to e. Thus, rejection sampling begins by sampling λ_i from the lognormal prior and U_i from a standard uniform distribution. Then λ_i is kept if $U_i < q(\lambda_i|\mathbf{x})/e(\lambda_i) = L(\lambda_i|\mathbf{x})/L(4.3|\mathbf{x})$. Otherwise, λ_i is rejected and the process is begun anew. Any kept λ_i is a draw from the posterior. Although not efficient—only about 30% of candidate draws are kept—this approach is easy and exact. □

6.2.3.1 *Squeezed Rejection Sampling* Ordinary rejection sampling requires one evaluation of f for every candidate draw Y. In cases where evaluating f is computationally expensive but rejection sampling is otherwise appealing, improved simulation speed is achieved by *squeezed rejection sampling* [383, 441, 442].

This strategy preempts the evaluation of f in some instances by employing a nonnegative *squeezing function*, s. For s to be a suitable squeezing function, $s(x)$ must not exceed $f(x)$ anywhere on the support of f. An envelope, e, is also used; as with ordinary rejection sampling, $e(x) = g(x)/\alpha \geq f(x)$ on the support of f.

The algorithm proceeds as follows:

1. Sample $Y \sim g$.
2. Sample $U \sim \text{Unif}(0, 1)$.
3. If $U \leq s(Y)/e(Y)$, keep the value of Y. Set $X = Y$ and consider X to be an element in the target random sample. Then go to step 6.
4. Otherwise, determine whether $U \leq f(Y)/e(Y)$. If this inequality holds, keep the value of Y, setting $X = Y$. Consider X to be an element in the target random sample; then go to step 6.
5. If Y has not yet been kept, reject it as an element in the target random sample.
6. Return to step 1 until you have accumulated a sample of the desired size.

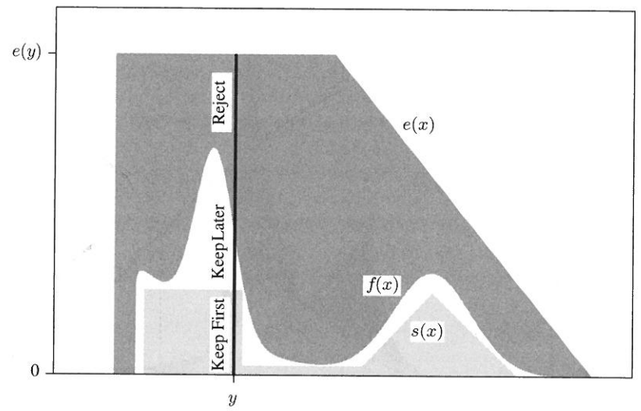

FIGURE 6.3 Illustration of squeezed rejection sampling for a target distribution, f, using envelope e and squeezing function s. Keep First and Keep Later correspond to steps 3 and 4 of the algorithm, respectively.

Note that when $Y = y$, this candidate draw is kept with overall probability $f(y)/e(y)$, and rejected with probability $[e(y) - f(y)]/e(y)$. These are the same probabilities as with simple rejection sampling. Step 3 allows a decision to keep Y to be made on the basis of an evaluation of s, rather than of f. When s nestles up just underneath f everywhere, we achieve the largest decrease in the number of evaluations of f.

Figure 6.3 illustrates the procedure. When a candidate $Y = y$ is sampled, the algorithm proceeds in a manner equivalent to sampling a Unif(0, $e(y)$) random variable. If this uniform variate falls below $s(y)$, the candidate is kept immediately. The lighter shaded region indicates where candidates are immediately kept. If the candidate is not immediately kept, then a second test must be employed to determine whether the uniform variate falls under $f(y)$ or not. Finally, the darker shaded region indicates where candidates are ultimately rejected.

As with rejection sampling, the proportion of candidate draws kept is α. The proportion of iterations in which evaluation of f is avoided is $\int s(x)\,dx / \int e(x)\,dx$.

Squeezed rejection sampling can also be carried out when the target is known only up to a proportionality constant. In this case, the envelope and squeezing function sandwich the unnormalized target. The method is still exact, and the same efficiency considerations apply.

Generalizations for sampling multivariate targets are straightforward.

6.2.3.2 *Adaptive Rejection Sampling* Clearly the most challenging aspect of the rejection sampling strategy is the construction of a suitable envelope. Gilks and Wild proposed an automatic envelope generation strategy for squeezed rejection sampling for a continuous, differentiable, log-concave density on a connected region of support [244].