Random Number Generation

Inverse CDF

Let X have a continuous, strictly increasing cdf $F(x)$ for $a < x < b$, where we allow $a = -\infty$ and or $b = +\infty$. If $U \sim$ uniform$(0,1)$, then the cdf of $Y := F^{-1}(U)$ is $F(x)$. To see this, write

$$F_Y(y) = P(Y \le y) = P(F^{-1}(U) \le y) = P(U \le F(y))$$
$$= \int_0^{F(y)} 1\, du = F(y).$$

Example. Let $X \sim \exp(\lambda = 1)$ so that $F(x) = 1 - e^{-x}$, $x \ge 0$. To find $F^{-1}(u)$, solve $u = F(x)$ for x, i.e., $u = 1 - e^{-x}$ or $e^{-x} = 1 - u$ or $-x = \ln(1-u)$ or $x = -\ln(1-u)$. So $Y = -\ln(1-U) \sim \exp(\lambda = 1)$. In addition, since $V := 1 - U$ is also uniform$(0,1)$, it suffices to generate $V \sim$ uniform$(0,1)$ and use $Y = -\ln V$ instead. This how the MATLAB function exprnd works.

Example. Let $X \sim$ Cauchy$(\lambda = 1)$ with cdf $F(x) = \frac{1}{\pi}\tan^{-1}x + \frac{1}{2}$. From $u = \frac{1}{\pi}\tan^{-1}x + \frac{1}{2}$, $x = \tan[\pi(u - 1/2)]$.

Example. Let $X \sim N(0,1)$ with cdf $F(x) = \int_{-\infty}^{x} e^{-t^2/2}\,dt/\sqrt{2\pi}$. We cannot solve $u = F(x)$ in closed form, although we could do it numerically, which would be slow. Note, however, that although MATLAB provides F^{-1} (see norminv), MATLAB does not use this method.

Mixtures

Suppose X has a density f of the form

$$f(x) = \sum_{k=1}^{n} p_k f_k(x),$$

where the p_k are positive and $\sum_{k=1}^{n} p_k = 1$, and f_k is the pdf of a RV Y_k that can be simulated easily. Generate a discrete RV N with $P(N=k) = p_k$. If $N=k$, generate Y_k with density f_k and put $X = Y_N = Y_k$. Then

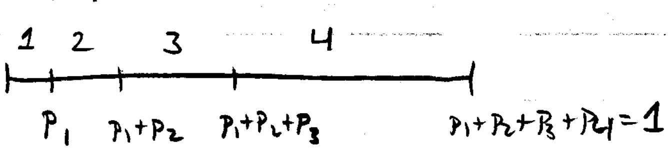
$$P(X \le x) = P(Y_N \le x) = \sum_{k=1}^{n} P(Y_N \le x \mid N=k)\, p_k$$

$$= \sum_{k=1}^{n} P(Y_k \le x \mid N=k)\, p_k = \sum_{k=1}^{n}\left(\int_{-\infty}^{x} f_k(y)\, dy\right) p_k$$

& so $\displaystyle \frac{d}{dx} P(X \le x) = \sum_{k=1}^{n} f_k(x)\, p_k = f(x)$ as required.

Discrete RVs

To generate the discrete RV N above, let $U \sim \text{unif}(0,1)$ and put $N := \min\{1 \le \ell \le n : \sum_{k=1}^{\ell} p_k \ge U\}$. Then since $N = \ell \iff \sum_{k=1}^{\ell} p_k \ge U > \sum_{k=1}^{\ell-1} p_k$ & so $P(N=\ell) = p_\ell$ as required.



In MATLAB, if $P = [p_1, \ldots, p_n]$, and $S = \text{cumsum}(P)$, and $U = \text{rand}(1)$; $i = \text{find}(U \le S)$, then $N = i(1)$ has $P(N=k) = p_k$.

Special Tricks

1) If U and V are independent uniform$(0,1)$ and
$X := \sqrt{-2 \ln U}\, \cos(2\pi V)$, $Y := \sqrt{-2\ln U}\, \sin(2\pi V)$,
then it can be shown that X and Y are independent $N(0,1)$ RVs.

2) If $X_1, X_2, \ldots, X_n$ are iid $\exp(\lambda)$, then $Y := X_1 + \cdots + X_n$ is Erlang(n, λ).

3) If $X_1, X_2, \ldots, X_n$ are iid $N(0,1)$, then $Y = X_1^2 + \cdots + X_n^2$ is chi-squared with n degrees of freedom.

4) If $U_1, U_2, \ldots, U_n$ are iid uniform$(0,1)$, then by the central limit theorem $Y := U_1 + \cdots + U_n$ is approximately $N(n/2, n/12)$. Hence, $X := (Y - n/2)/\sqrt{n/12}$ is approximately $N(0,1)$.

Rejection (or Acceptance/Rejection)

Let X and Y be RV's with respective pdfs p and q. Let $f = \alpha p$ and $g = \beta q$, where α and β are arbitrary positive constants. Put

$$K := \sup_x \frac{f(x)}{g(x)} < \infty$$

so that $f(x) \le K g(x)$. Thus, $f(x) > 0 \Rightarrow g(x) > 0$.
Fix any $K' \ge K$. For $n = 1, 2, \ldots$, let U_n be iid uniform$(0,1)$ and Y_n be iid with density q.
Let $N = \min\{n \ge 1 : U_n \le \dfrac{f(Y_n)}{K' g(Y_n)}\}$. Put $X := Y_N$.

If we put $Z_n := I_{\{U_n > \frac{f(Y_n)}{K' g(Y_n)}\}}$, then

Z_n is a Bernoulli sequence. The first time $Z_n = 0$, N, is a geometric$_1$ RV with parameter

$$\theta := P\left(U_n > \frac{f(Y_n)}{K' g(Y_n)}\right),$$

and so

$$E[N] = \frac{1}{1-\theta} = \frac{1}{P\left(U_n \le \frac{f(Y_n)}{K' g(Y_n)}\right)}$$

called the acceptance prob.

Now, the acceptance probability can be found using the law of total probability:

$$P\left(U_n \le \frac{f(Y_n)}{K' g(Y_n)}\right) = \int P\left(U_n \le \frac{f(Y_n)}{K' g(Y_n)} \,\middle|\, Y_n = y\right) g(y)\,dy$$

$$= \int P\left(U_n \le \frac{f(y)}{K' g(y)}\right) g(y)\,dy$$

$$= \int \frac{f(y)}{K' g(y)} g(y)\,dy$$

$$= \frac{1}{K'\beta} \int f(y)\,dy = \frac{\alpha}{K'\beta}$$

To minimize $E[N]$, maximize $\alpha/(K'\beta)$.

Next,

$$P(X \le x) = P(Y_N \le x)$$

$$= \sum_{n=1}^{\infty} P(Y_N \le x \mid N=n)\, P(N=n)$$

Now,

$$P(Y_N \le x \mid N=n) = P(Y_n \le x \mid N=n)$$

$$= P(Y_n \le x \mid Z_1=1, \ldots, Z_{n-1}=1, Z_n=0)$$
$$= P(Y_n \le x \mid Z_n = 0).$$

We then write

$$P(Y_n \le x, Z_n=0) = P\left(Y_n \le x,\ U_n \le \frac{f(Y_n)}{K'g(Y_n)}\right)$$

$$= \int P\left(Y_n \le x, U_n \le \frac{f(Y_n)}{K'g(Y_n)} \,\Big|\, Y_n = y\right) q(y)\,dy$$

$$= \int P\left(y \le x,\ U_n \le \frac{f(y)}{K'g(y)}\right) q(y)\,dy$$

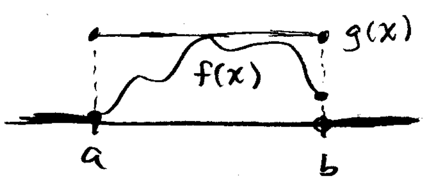
$$= \int_{-\infty}^{x} P\left(U_n \le \frac{f(y)}{K'g(y)}\right) q(y)\,dy$$

$$= \int_{-\infty}^{x} \frac{f(y)}{K'g(y)}\, q(y)\,dy$$

$$= \int_{-\infty}^{x} p(y)\,dy \cdot \underbrace{\frac{\alpha}{K'\beta}}_{=P\left(U_n \le \frac{f(Y_n)}{K'g(Y_n)}\right) = P(Z_n=0)}$$

$\therefore\ P(Y_n \le x \mid Z_n=0) = F_X(x)$
$\therefore\ P(Y_N \le x / N=n) = F_X(x)$
$\therefore\ P(Y_N \le x) = \sum_{n=1}^{\infty} F_X(x) P(N=n) = F_X(x)$

<u>Example</u>



$\bar{f} := \max_{a \le x \le b} f(x)$

$g(x) = \bar{f}$ for $a \le x \le b$

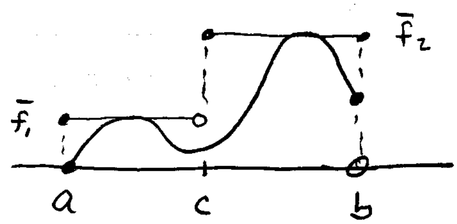
Take $K' = K = \sup_{a \le x \le b} \frac{f(x)}{g(x)} = 1$

Then $U_n \sim \text{unif}(0,1)$, $Y_n \sim \text{uniform}[a,b]$, The acceptance condition is

$$U_n \le \frac{f(Y_n)}{1\cdot \bar f} \qquad \text{or} \qquad \bar f\, U_n \le f(Y_n)$$

This is equivalent to having $V_n = \bar f\, U_n \sim \text{unif}(0,\bar f)$ independent of $Y_n \sim \text{unif}[a,b]$. Thus, (Y_n, V_n) is unif $\sim [a,b]\times[0,\bar f]$ and the acceptance condition is that (Y_n, V_n) lie on or below the curve $f(x)$.

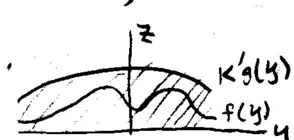
<u>Example</u>



In this case $q(y) = \dfrac{\bar f_1}{\bar f_1+\bar f_2}\cdot f_1(y) + \dfrac{\bar f_2}{\bar f_1+\bar f_2} f_2(y)$,

where $f_1 \sim \text{unif}[a,c]$ and $f_2 \sim \text{unif}[c,b]$.

<u>Prop.</u> In general, $(Y_n, K'g(Y_n)U_n)$ is uniformly distributed on $A := \{(y,z) : 0 \le z \le K'g(y)\}$.



<u>Proof.</u> Let B be a bounded subset of A. We must show that $P((Y_n, K'g(Y_n)U_n)\in B) = \text{area}(B)/\text{area}(A)$, where $\text{area}(A) = \int K'g(y)\,dy = K'\beta\int q(y)\,dy = K'\beta$.
Now

$$P((Y_n,K'g(Y_n)U_n)\in B) = \int P((y,K'g(y)U_n)\in B)\,q(y)\,dy$$
$$= \int P(K'g(y)U_n \in B_y)\,q(y)\,dy, \qquad B_y := \{z : (y,z)\in B\}$$

Note that $z\in B_y \Leftrightarrow (y,z)\in B \subset A \Rightarrow 0\le z\le K'g(y)$. Also,

$K'g(y)U_n \sim \text{unif}[0, K'g(y)]$. Thus,

$$P((Y_n, K'g(Y_n)U_n) \in B) = \int \left(\int I_{B_y}(z) \frac{1}{K'g(y)} dz \right) g(y)\, dy,$$

$$= \int \left(\int I_B(y,z)\, dz \right) dy \Big/ K'\beta \qquad \text{since } B_y \subset [0, K'g(y)]$$

$$= \frac{\text{area}(B)}{K'\beta}. \qquad \blacksquare$$

Simulating Random Vectors

Suppose $f_{XYZ}(x,y,z)$ is given. Write it as

$$f_{XYZ}(x,y,z) = f_{Z|YX}(z|y,x) f_{Y|X}(y|x) f_X(x)$$

Then simulate $X \sim f_X(x)$. If $X=x$, simulate Y according to the one-dimensional pdf $f_{Y|X}(\cdot|x)$ to get $Y=y$. Then simulate Z according to the one-dimensional density $f_{Z|YX}(\cdot|y,x)$.

Of course, this idea generalizes to simulate $X_1, \ldots, X_n$. Note, however, that to generate X_n, we must store the previously generated values of $X_1 = x_1, \ldots, X_{n-1} = x_{n-1}$. For large n, the storage requirements become prohibitive. Fortunately, in the Markov case, since

$$f_{X_k|X_{k-1},\ldots,X_1}(x_k|x_{k-1},\ldots,x_1) = f_k(x_k|x_{k-1}),$$

we only need to store the most recently generated value.

The Gaussian Case

To simulate $X \sim N(0, C)$, let $Y_1, \ldots, Y_n$ be iid $N(0,1)$ and put $Y := [Y_1, \ldots, Y_n]'$. Then $C^{-1/2}Y \sim N(0,C)$ since $E[(C^{-1/2}Y)(C^{-1/2}Y)'] = C^{-1/2} C C^{-1/2} = I$. If $\Lambda = P'CP$, $C = P\Lambda P'$ or $C = P\Lambda^{1/2}P'P\Lambda^{1/2}P' \Rightarrow C^{1/2} = P\Lambda^{1/2}P'$ or $C^{-1/2} = P\Lambda^{-1/2}P'$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $P'P = PP' = I$.