Lecture - 4

December 21, 2011

General Sampling Methods

We assume availability of uniform random number. A typical simulation uses methods for transforming samples from uniform distribution to samples from other distributions. The most widely used general techniques are

- Inverse transform method
- Acceptance Rejection Method

Inverse transform method

Suppose we want to generate a random variable X with the property that $P(X \leq x) = F(x) \quad \forall x$. The inverse transform method sets : $X = F^{-1}(U)$, $U \sim \mathcal{U}(0,1)$, where $\mathcal{U}(0,1)$ is a uniform distribution on $[0,1]$.

Theorem 0.1. *Suppose, $U \sim \mathcal{U}(0,1)$ and F is a continuous increasing function. Then $F^{-1}(u)$ is a sample from F.*

Proof: Let P denotes the underlying probability.

$U \sim \mathcal{U}(0,1)$ means $P(U \leq \xi) = \xi \;\; 0 \leq \xi \leq 1$. Therefore,

$$P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

Examples:

1. **Exponential Distribution** : The exponential distribution with mean θ has distribution $F(x) = (1 - e^{-\theta x})$. This is the distribution of the time between jumps of a poisson process with rate $\frac{1}{\theta}$. Inverting yields

$$X = -\frac{1}{\theta}\ln(1-U)$$

2. **Arc Sin Law** : The time at which a standard brownian motion attains its maximum over the time interval [0,1] and has distribution

$$F(x) = \frac{2}{\pi}\arcsin\sqrt{x}; \quad 0 \leq x \leq 1$$

The inverse transform method for sampling from this distribution is :

$$\begin{aligned} X &= \sin^2(\frac{U\pi}{2}) \\ &= \frac{1}{2} - \frac{1}{2}\cos(U\pi); \quad U \sim \mathcal{U}[0,1] \end{aligned}$$

3. **Rayleigh Distribution:** Distribution function of Rayleigh Distribution can be given by :

$$F(x) = 1 - e^{-2x(x-b)}; \quad x \geq b.$$

Solving the equation $F(x) = u$, $u \in (0,1)$ results in a quadratic with roots :

$$x = \frac{b}{2} \pm \frac{\sqrt{b^2 - 2\ln(1-u)}}{2}$$

The inverse is given by larger of the two roots. In particular, we need $x \geq b$, since the maximum of the Brownian path must be as large as the terminal value. Alternatively, replacing u with $(1-u)$ we get,

$$x = \frac{b}{2} \pm \frac{\sqrt{b^2 - 2\ln u}}{2}$$

Note that even if the inverse of F is not known explicitly, the inverse transform method is still applicable through numerical evaluation of F^{-1} computing $F^{-1}(u)$ is equivalent to finding a root x of the equation $F(x) - u = 0$. For a distribution F with density f, Newton's method for finding roots produce a sequence of iterates.

4. **Discrete Distribution :** Consider the following example for a discrete random variable whose possible values are $c_1 < c_2 < \cdots < c_n$. Let p_i be the probability attached to c_i, $i = 1, 2, \cdots, n$ and set $q_0 = 0$. Also, $q_i = \sum_{j=1}^{n} p_j, i = 1, 2, \cdots, n$.

   There are cumulative probability associated with c_i, i.e. $q_i = F(c_i)$; $i = 1, \cdots, n$. To sample from this distribution :

   (a) Generate a uniform $U \sim \mathcal{U}[0,1]$.

   (b) Find $k \in \{1, 2, \cdots, n\}$ such that $q_{k-1} < U < q_k$.

   (c) Set $X = c_k$.

5. **Conditional distribution :** Suppose X has a distribution F and consider the problem of sampling X unconditionally. If $U \sim \mathcal{U}[0,1]$, then the random variable defined by $V = F(a) + (F(b) - F(a))U$ is uniformly distributed between $F(a)$ and $F(b)$ and $F^{-1}(V)$ has the desired conditional distribution. To see this observe that

$$\begin{aligned} P(F^{-1}(V) \leq x) &= P(F(a) + (F(b) - F(a))U \leq F(x)) \\ &= P(U \leq \frac{F(x) - F(a)}{F(b) - F(a)}) \\ &= \frac{F(x) - F(a)}{F(b) - F(a)} \end{aligned}$$

   This is precisely the distribution of X given $a < X < b$.

Acceptance Rejection Method

Introduce by Von-Neumann, this method is the most widely applicable mechanism for generality random samples. This method generates samples from a target distribution by generating condidates from a more convenient distribution and then rejecting a random subset of generated condidates. The rejection mechanism is designed so that to accepted samples are indeed distributed according to the target distribution. this technique is by no means restricted to univariate distribution.

Suppose we wish to generate samples from a density f defined on same set χ. this could be a subset of the real line or R^d or more generalset.

Let g be a density on χ from which we know how to generate samples and with the property $f(x) \leq cg(x)$, $\forall x \in \chi$.

In acceptance-rejection method we generate a sample χ from g and accept the sample with probability $\frac{f(x)}{cg(x)}$. This can be implemented by sampling U uniformly over $(0, 1)$. If x is rejected, a new candidate is sampled from g and the acceptance test applied again. the process repeats until the acceptance test is passed and accepted value is returned as asample from f.

Algorithm :

(a) Generate X from distribution g.

(b) Generate v from $u[0, 1]$.

(c) If $U \leq \frac{f(x)}{cg(x)}$,return X, otherwise go to step 1.

To verify the validity of the acceptance rejection method, let Y be the sample returned by the algorithm and observe that Y has the distribution of X conditional on $U \leq \frac{f(x)}{cg(x)}$.

Thus for any $A \subseteq X$,

$$
\begin{aligned}
P[Y \in A] &= P\left[X \in A | U \leq \frac{f(x)}{cg(x)}\right] \\
&= \frac{P\left[X \in A, U \leq \frac{f(x)}{cg(x)}\right]}{P\left[U \leq \frac{f(x)}{cg(x)}\right]} \\
&= \frac{\int_A P[U \leq \frac{f(x)}{cg(x)} | X = x] g(x) dx}{P\left[U \leq \frac{f(x)}{cg(x)}\right]} \\
&= \int_A f(x) dx
\end{aligned}
$$