Abstract

The Rejection and Acceptance Complement Methods.

Exact Sampling Methods

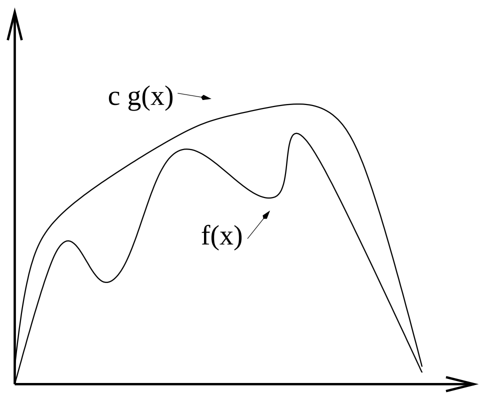
The Metropolis algorithm and its variants are great, for they can be used to generate ergodic Markov chains with, in principle, any pre specified stationary distribution. Just choose an arbitrary starting point and eventually the chain will begin sampling from its stationary distribution. The problem is, of course, that (usually) we can't be 100% sure about when this will happen. Exact sampling methods, when available, are therefore more reliable and, often more efficient than MCMC methods. For this reason MCMC sampling is almost always used in combination with exact methods. This is specially true for the Gibbs sampler where we need to generate from the full conditionals. What exact method to use depends on the particular problem. There are literally hundreds of methods available. The current bible is Devroye's

(*http://omega.stat.psu.edu:8008/summer99/lecture4/http://cgm.cs.mcgill.ca/ luc/*) book which is unfortunately not available online. Besides the only truly universal exact generator (the inverse cdf method introduced in lecture I

(*http://omega.stat.psu.edu:8008/summer99/lecture4/../lecture1/l1.html*)) the most flexible exact algorithm is the rejection method.

The Rejection Method

The idea, I think first implemented by Von Neumann, is very simple. To sample from $f(x)$ find another simpler density $g(x)$ from which you know how to sample and such that $cg(x) \geq f(x)$ for some $c > 1$ (we say that cg envelopes f). Like in the following picture:



Then generate uniformly under the graph of the envelope and accept only samples that fall under the graph of f. In other words reject a sample if it falls in between the envelope and the function f.

Copying from the bible:

Theorem 1 *If f and g are densities on R^p with, $f(x) \leq cg(x)$ for all $x \in R^p$, some $c \geq 1$. Then the following algorithm*

The Rejection Method

```
{
REPEAT

        {
        x ← sample from g
        u ← unif(0,1)
        y ← c g(x)/f(x)
        }

   UNTIL uy ≤ 1
RETURN x }
```

will produce a sample $X = x$ with density $f(x)$.

Proof: Consider the following two lemmas

Lemma 1 *Let X be a random p-vector with density $f(x)$. Let U be a uniform(0,1) r.v. independent of X, and let $c > 0$. Then,*

I. *$(X, cUf(X))$ is uniform on, $A = \{(x,u) : x \in R^p, 0 \leq u \leq cf(x)\}$*

II. *If (X,U) is uniform on A, then X has density $f(x)$.*

Proof:

I. *Let $B \subset A$, then*

$$\begin{aligned} P[(X, cUf(X)) \in B] &= \int P[(x, cUf(x)) \in B | X = x] f(x) dx \\ &= \int \left\{ \int_{B_x = \{u:(x,u) \in B\}} \frac{1}{cf(x)} du \right\} f(x) dx \\ &= \frac{1}{c} \int_B du dx \end{aligned}$$

where the second equality follows from Tonelli's theorem, the independence of X and U and the fact that $cf(x)U$ is uniform$(0, cf(x))$. But,

$$|A| = \int_A du dx = \int_{R^d} \left(\int_0^{cf(x)} du \right) dx = c$$

thus, for any $B \subset A$ measurable we have,

$$P\left((X, cUf(X)) \in B\right) = \frac{|B|}{|A|}$$

which means that $(X, cUf(X))$ is uniform on A.

II. *We just need to show that, $\forall B$ measurable, $P[X \in B] = \int_B f(x)dx$. But,*

$$\begin{aligned}
P[X \in B] &= P[X \in B, 0 \leq U \leq cf(X)] \\
&= P[(X,U) \in B_1 = \{(x,u) : x \in B, 0 \leq u \leq cf(x)\}] \\
&= \frac{\int\int_{B_1} dudx}{\int\int_A dudx} \\
&= \frac{1}{c}\int_B cf(x)dx = \int_B f(x)dx.
\end{aligned}$$

Lemma 2 *Let $X_1, X_2, \ldots$ be iid vectors in R^p. Let $A \subset R^p$ s.t. $P[X \in A] = \alpha > 0$. Let Y be the first $X_i \in A$. Then, $\forall B \subset R^p$ measurable,*

$$P[Y \in B] = \frac{P[X_1 \in A \bigcap B]}{\alpha}.$$

Moreover, if X_1 is uniform on $A_0 \supset A$ then Y is uniform on A.

Proof:

$$\begin{aligned}
P[Y \in B] &= \sum_{i=1}^{\infty} P[X_1 \notin A, \ldots, X_{i-1} \notin A, X_i \in B \bigcap A] \\
&= \sum_{i=1}^{\infty} (i-\alpha)^{i-1} P[X_1 \in B \bigcap A] \\
&= \frac{P[X_1 \in B \bigcap A)}{1-(1-\alpha)}
\end{aligned}$$

Also, if X_1 is uniform on $A_0 \supset A$ then for all measurable B,

$$\begin{aligned}
P[Y \in B] &= \frac{P[X_1 \in ABA_0)}{\alpha} = \frac{|ABA_0|/|A_0|}{|A_0A|/|A_0|} \\
&= \frac{|AB|}{|A|}
\end{aligned}$$

therefore, Y is uniform on A.

We are now ready for the Theorem.

Proof: (that the Rejection Method is valid). In the first line the algorithm generates (by Lemma 1, I) $(X, cUg(X))$ uniform on $C = \{(x,u) : x \in R^p, 0 \leq u \leq cg(x)\}$. However at exit time (by Lemma 2), $(X, cUg(X))$ is uniform on $C = \{(x,u) : x \in R^p, 0 \leq u \leq f(x)\}$. Notice that, by the assumption that $f(x) \leq cg(x)$ (i.e. that g envelopes f), $C \supset A$. Thus, (by Lemma 1, II) X has density $f(x)$.

Best g is f itself

Let N be the number of pairs (x, u) generated by the rejection algorithm to exit. We have,

$$\begin{aligned} P[N \geq i] &= P[\text{ reject at least } (i-1) \text{ pairs}] \\ &= P[UY \geq 1]^{i-1} = (1-\alpha)^{i-1} \end{aligned}$$

where,

$$\begin{aligned} \alpha &= P[Uc\frac{g(X)}{f(X)} \leq 1] = \int P[Ucg(x) \leq f(x)]g(x)]dx \\ &= \int \frac{f(x)}{cg(x)} g(x)dx = \frac{1}{c}. \end{aligned}$$

The expected number of pairs generated by the algorithm is, $< N >$ where,

$$< N >= \sum_{i=1}^{\infty} P[N \geq i] = \sum_{i=1}^{\infty} (1-\alpha)^{i-1} = \frac{1}{1-(1-\alpha)} = \frac{1}{\alpha} = c$$

Thus, c is the expected number of rejections and therefore it should be kept as small as possible, i.e. as close as possible to its minimum value of 1. Since, $f(x) \leq cg(x)$, in order for c to be small g must be close to f.

Example: To sample from the standard Gaussian we can use the rejection method with g as the Laplace distribution. Notice that,

$$f(x) = (2\pi)^{-1/2} \exp(-x^2/2)$$

and in order to get an upper bound for f we need a lower bound for the energy, i.e. for $x^2/2$. But that follows easily from,

$$\frac{1}{2}(|x|-1)^2 = \frac{x^2}{2} + \frac{1}{2} - |x| \geq 0$$

from where we get,

$$\exp(-x^2/2) \leq \exp(\frac{1}{2} - |x|) = (2\pi)^{1/2} \left(\frac{2e}{\pi}\right)^{1/2} \left(\frac{1}{2}\exp(-|x|)\right)$$

the last term in parenthesis above, is the density of the Laplace distribution which the above inequality shows to envelope the $N(0,1)$ with,

$$c = \left(\frac{2e}{\pi}\right)^{1/2} \approx 1.3155$$

Javascript Demo Implementation of the above in Javascript. Solution by Ke Yang

(*http://omega.stat.psu.edu:8008/summer99/lecture4/kyang-1.html*)

There are several variations of the rejection method in the bible. One fairly general method that never rejects any samples is:

The Acceptance Complement Method

Suppose that we don't know how to envelope f but,

$$f(x) = f_1(x) + f_2(x)$$

with, $f_1(x) \geq 0, f_2(x) \geq 0$ and $f_1(x) \geq g(x)$ where g is a density. Furthermore suppose that we know how to sample from g and from f_2 (properly normalized), then,

The Acceptance Complement Method

```
{
x ← sample from g
u ← unif(0,1)
IF u > f1(x)/g(x) THEN x ← sample from f2/∫f2
RETURN x
}
```

Theorem 2 *$X = x$ is a sample from f.*

Proof: Let $\alpha = \int f_2$ and suppose that Y has density g. We have,

$$\begin{aligned} P[X \in B] &= P[Y \in B, U \leq \frac{f_1(Y)}{g(Y)}] + P[U > \frac{f_1(Y)}{g(Y)}] \int_B f_2(x)dx/\alpha \\ &= \int_B \frac{f_1(x)}{g(x)} g(x)dx + \left(1 - \int \frac{f_1(y)}{g(y)} g(y)dy\right) \int_B f_2/\alpha \\ &= \int_B f_1 + \int_B f_2 = \int_B f \end{aligned}$$

Example: $f_1(x) = \min\{f(x), g(x)\}$ and $f_2(x) = (f(x) - g(x))_+$. Then, clearly $f_1(x) \leq g(x)$ and $f(x) = f_1(x) + f_2(x)$. So if we know how to sample from g and f_2, we are done.

Another useful general split is available for *almost flat densities* on $[-1, 1]$, i.e. densities $f(x)$ such that,

$$\sup f(x) - \inf f(x) \leq \frac{1}{2}$$

then we may take,

$$\begin{aligned} g(x) &= 1/2 \text{ for } |x| \leq 1 \\ f_1(x) &= f(x) - (M - \frac{1}{2}), \text{ with } M = \sup f(x) \\ f_2(x) &= M - \frac{1}{2} \text{ for } |x| \leq 1 \end{aligned}$$

which works since g and f_2 are proportional to densities uniform on $[-1, 1]$ (so easy to sample from). That f_1 is bounded above by g just follows from the almost flat condition and that $f_1 > 0$ follows from the fact that,

$$0 \leq \inf f(x) \leq \frac{1}{2} \leq \sup f(x) \leq 1$$

Hence, the method can be used to generate from many densities symmetric about with a single mode at 0, for which $M = f(0)$ and $\inf\ f = f(1)$. For example we can generate from the truncated Cauchy for $|x| \leq 1$. We have,

$$f(x) = \frac{2}{\pi(1+x^2)} \quad \text{for } |x| \leq 1$$

which is almost flat since $2/\pi - 1/\pi = 1/\pi \approx 0.318 < 0.5$. Using the property that $1/X$ is Cauchy when X is Cauchy we can generate a complete Cauchy by just using the Acceptance Complement Method to generate from the truncated Cauchy and then with probability $1/2$ return $1/X$ instead of X.

Example: Javascript implementation of the Cauchy with the above method.

A Trivial Perfect MCMC Method

If there is available a good envelope for f then the rejection method is preferable to asymptotic methods based on Markov chains. The problem is that often good envelopes are not easy to find. A bad envelope can be easier to find and still be useful when combined with a Markov chain method. For example it may be possible to show that $f(x) \leq 10000 g(x)$ which means that we can use the rejection method but expect to generate, on the average, $10000 (X, U)$ pairs before accepting one vector X. In this case we can still use the rejection method to generate a single observation of X which is warranted to have the correct density f. Then, use the observation generated by the rejection method as the initial point for a Markov chain method with statitionary distribution f and harvest the complete path of the chain that is now sampling from its asymptotic distribution.