Expectations & Sampling Methods
Aubrey Yan

The expectation of a random variable or functions of random variables can be difficult to compute analytically when the probability distribution of those variables is not standard well-known distributions. One way to approximate expectations is by drawing samples from the probability distributions. In this quarter, we learned to use Monte Carlo methods to draw samples from "difficult" distributions and use these samples to approximate expectations.

Monte Carlo is the art of approximating an expectation by the sample mean of a function of simulated random variables. Consider a random variable **X** having a pmf/pdf $f_X(x)$, then

$$E(g(X)) = \sum_{x \in X} g(x)f_X(x) \text{ if X is discrete and } E(g(X)) = \int_{x \in X} g(x)f_X(x)dx \text{ if X is continuous.}$$

Now if we were to take an n-sample of X's, and we computed the mean of g(x) over the sample, then we would have the Monte Carlo *estimate* $g_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$ of $E(g(x))$. Also, $g_n(X)$ is unbiased for $E(g(X))$.

Importance Sampling is choosing a good distribution from which to simulate one's random variables. It involves multiplying the integrand by 1 to yield an expectation of a quantity that varies less than the original integrand over the region of integration. For example, let $h(x)$ be a density for the random variable X which takes values only in A so that

$$\int_{x \in A} g(x)dx = E_h\left(\frac{g(X)}{h(X)}\right) \text{ so long as } h(x) \neq 0 \text{ for any } x \in A \text{ for which } g(x) \neq 0, \text{ and where } E_h$$

denotes the expectation with respect to the density h. This gives the Monte Carlot estimator:

$$g^h_n(X) = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i)}{h(X_i)} \text{ where } X_i \sim h(x).$$

Rejection sampling is aiming at generating random numbers from a target probability distribution f(x). As anticipated, we want to draw a random sample having a probability distribution f(x), but we are not able to do that with a deterministic algorithm. Hence we use a "proposal function" g(x), from which it's easier to draw a random variable X. This function g(x), multiplied by a constant M, has to 'envelop' our f(x) so that $M \times g(x) \geq f(x)$. Then, we generate another random variable $U \sim Unif(0, 1)$, we are now provided with a scattering of couples (x, y,) where $y = u \times M \times g(x)$ (so it's our uniform variable scaled by $M \times g(x)$). The idea is that, for all the points (x, y) which lie under the curve f(x), we accept the fact that x belongs to a random sample drawn from f(x) with probability $\frac{f(x)}{M \times g(x)}$. So the decision rule is $u \times M \times g(x) \leq f(x) \Rightarrow u \leq \frac{f(x)}{M \times g(x)}$. Note that the constant M is such that g(x) is as close as possible to f(x), so that the rejection area is the smallest possible. The optimal M is given by $M = sup\left\{\frac{f(x)}{g(x)}\right\}$, and $\frac{1}{M}$ represents the probability of acceptance.