IMPORTANCE SAMPLING.

Sometimes, it is necessary to compute $\mu = E(f(x))$, is nearly zero outside a region A for which $P(X \in A)$ is small. The set A may have small volume, or it may be in the tail of the X distribution. A plain Monte Carlo sample from the distn. of X could fail to have even one point inside the region A; which we will call our important region. We do this by sampling from a distn. that over weights the important region, hence the name importance sampling.

Suppose that, we require to find $\mu = E(f(X))$

$$= \int_D f(x)\, p(x)\, dx$$

, where $p(\cdot)$ is a pdf on $D \subseteq \mathbb{R}$ & f is the integrand. We take $p(x) = 0, \forall\, x \notin D$. If q is a (+)ve pdf on $\mathbb{R}$, then

$$\mu = \int_D f(x)\, p(x)\, dx = \int_D \frac{f(x)\, p(x)}{q(x)}\, q(x)\, dx = E_q\left[\frac{f(X)\, p(X)}{q(X)}\right]$$

where $E_q(\cdot)$ is Exp for $X \sim q$. Our target is to find $E_p(f(X))$. By making a multiplicative adjustment to f we compensate for sampling from q instead of p. The adjustment factor $\frac{p(x)}{q(x)}$ is called likelihood ratio; q is the importance distribution & p is the nominal distribution; $q(x) > 0$ whenever $f(x)p(x) \neq 0$. So, for $Q = \{x \mid q(x) > 0\}$ we have $x \in Q$ whenever $f(x)p(x) \neq 0$. So, if $x \in D \cap Q^c$, we know $f(x) = 0$, while if $x \in Q \cap D^c$ we have $p(x) = 0$.

Scanned by CamScanner

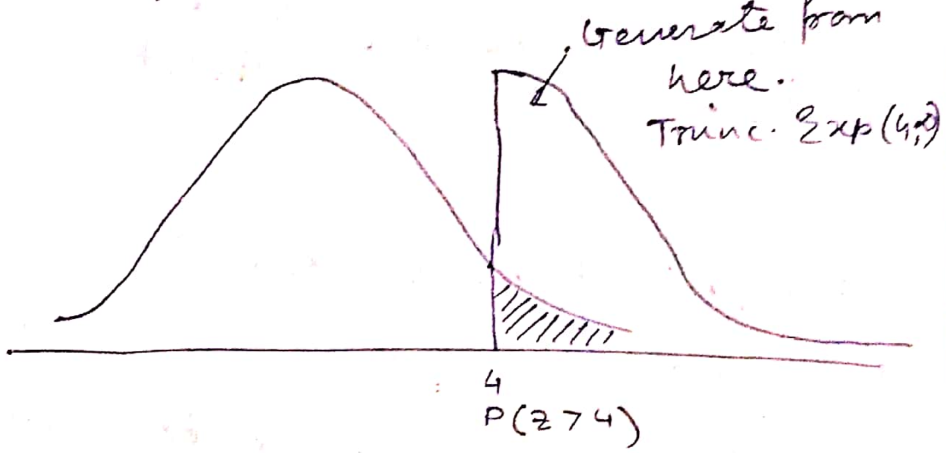
Now, $E_q\left[\frac{f(x)p(x)}{q(x)}\right] = \int_Q \frac{f(x)p(x)}{q(x)} q(x)\,dx = \int_Q f(x)p(x)\,dx$

$$= \int_D f(x)p(x)\,dx + \int_{Q\cap D^c} f(x)\,p(x)\,dx - \int_{D\cap Q^c} f(x)\,p(x)\,dx$$

$$= \int_D f(x)\,p(x)\,dx = \mu.$$

The importance sampling estimate of $\mu = E_p(f(X))$ is

$$\hat{\mu}_q = \frac{1}{n}\sum_{i=1}^{n} \frac{f(X_i)\,p(X_i)}{q(X_i)}, \quad X_i \sim q.$$



Eg- Suppose $X_1, \cdots, X_n \sim N(0,\theta)$, $\theta \sim Ga(3,\frac{1}{2})$. We plan to use the importance density $g \sim Ga(\alpha = cs^2, \beta = c)$, where $c>0$ is a const & s^2 is the sample variance. The mean of the Importance density is then s^2. We will determine c that would optimize the estimation precision.

Scanned by CamScanner

The joint density of the data, given θ is

$$p(x_1, \ldots, x_n | \theta) = \left(\frac{1}{2\pi\theta}\right)^{n/2} exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right)$$

$$\propto \theta^{-n/2} exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right).$$

Then the posterior is proportional to:

$$p(\theta | x_1, \ldots, x_n) = p(x_1, \ldots, x_n | \theta)\, p(\theta).$$

$$\propto \theta^{-n/2} exp\left(-\frac{1}{2\theta}\sum_i x_i^2\right) \cdot 2^{-3} \frac{\theta^2 e^{-\theta/2}}{\Gamma(3)}.$$

$$\propto \theta^{(4-n)/2} \cdot exp\left\{-\frac{1}{2\theta}\left(\theta^2 + \sum_i x_i^2\right)\right\}.$$

& hence the log-posterior is

$$C + \frac{4-n}{2}\log(\theta) - \frac{1}{2\theta}\left(\theta^2 + \sum_i x_i^2\right).$$

R-code to demonstrate importance sampling:

```
IS = function(X, c, res).
{   n = length(X).
    log.posterior = function(t) ((4-n)/2) * log(t) - (1/(2*t
                          * (t^2 + sum(X^2))).
    a = c * var(X);
    b = c;
    log.q = function(t) dgamma(t, a, b, log=TRUE).
    log.w = function(t) log.posterior(t) - log.q(t).
    U = rgamma(res, a, b) }
# Importance sampling estimate.
    ~~I = mean(exp~~ LP = log.w(U).
    W = max(LP)
    LP = LP - W.
    I = mean(exp(LP) * U)/mean(exp(LP)).
# X = data, c = tuning parameter for q, res = no.
                        monte [illegible] samples.
```