April 29, 2022

EE-698R: Advanced Topics in Machine Learning

Rejection and Importance Sampling

Deepankar Vishwakarma (180224) and Jaskaran Kalra (180322)

1. Rejection Sampling

Goal: We wish to sample from a distribution p(x) and that sampling directly from p(x) is difficult.

Imagine a distribution $x \sim p(x)$ (it a single or multi-dimensional). If we take a point $(x, p(x))$ from this distribution, we can claim that in that region of distribution these points will be uniformly distributed i.e.

$$(x, p(x)) \sim U(.) \tag{1}$$

1.1. Method for Rejection Sampling:-

(a) Take a simpler distribution $q(x)$ which is easy to sample from and scale it by some constant k such that $k.q(x) >= p(x)$ as shown in the diagram below

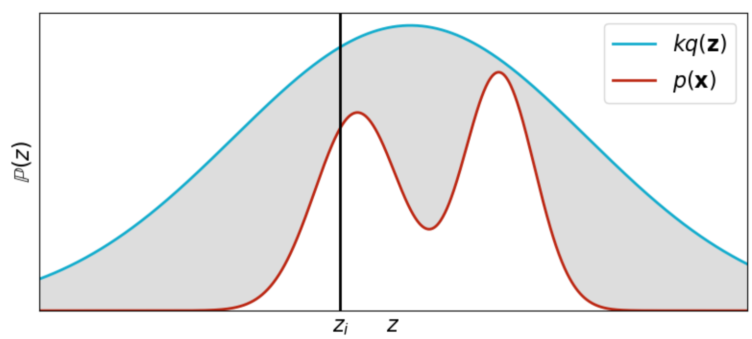

Fig-1

Now from the claim we made earlier $(x, k.q(x))$ is also uniformly distributed i.e.

$$(x, k.q(x)) \sim U(.) \tag{2}$$

Let us call $k.q(x)$ as $u(x)$. Now if $(x, u(x))$ lies within the red marked boundary region in the Fig-1 then accept, else reject the point.

1.2. Sampling from this new distribution $(x, u(x))$ within $p(x)$ region:-

Method-1.

The steps include-:
- Sample $x_0 \sim q(x)$
- Compute $q(x_0)$
- Compute $p(x_0)$
- Then sample $u \sim U(0, k(x_0))$
- Now if $u > p(x_0)$ then reject else accept.

Method-2.

The steps include-:
- Sample $x_0 \sim q(x)$
- Then sample $u_0 \sim U(0, 1)$
- Compute $q(x_0)$
- Compute $p(x_0)$
- Now if $u_0 > p(x_0)/k.q(x_0)$ then reject x_0 else accept.

1.3. Some important points and observations-:

1. Now for fraction of samples wasted we have:
   Area under $p(x) = 1$
   Area under $u(x) = \int k.q(x)dx = k$
   Hence, fraction of samples wasted $= k - 1/k$, where k can be chosen to be $max(p(x)/q(x))$
2. For probability of acceptance of the sample:

$$P(accepting\ a\ sample\ x_0) = \int P(accept \mid x_0).P(x_0)d(x_0) \tag{3}$$

We know that-

$$P(accept \mid x_0) = p(x_0)/k.q(x_0) \tag{4}$$

Also,

$$P(x_0) = q(x_0) \tag{5}$$

hence,

$$\int (p(x_0)/k.q(x_0)).q(x_0)d(x_0) = 1/k \tag{6}$$

Therefore, we can clearly say that if k is large then probability of acceptance is very small.

2. Importance Sampling

Goal: To find $E_{x \sim p(x)}[f(x)]$, where $p(x)$ is some distribution which is difficult to sample from but is easy to compute. A natural solution to this is by using Rejection Sampling to sample from the distribution and to calculate the expectation using -

$$E_{x \sim p(x)}[f(x)] \approx \frac{1}{S} \sum_{s=1}^{S} f(x^{(s)}) \tag{7}$$

where $x^{(s)} \sim p(x)$ are samples obtained from rejection sampling.

Importance Sampling : We assume a distribution $q(x)$ which is easy to sample from, and

$$E_{x\sim p(x)}[f(x)] = \int f(x)p(x)dx \approx \frac{1}{S}\sum_{s=1}^{S} f(x^{(s)}) \qquad ; x^{(s)} \sim p(x) \tag{8}$$

$$E_{x\sim p(x)}[f(x)] = \int f(x)\frac{p(x)}{q(x)}q(x)dx \approx \frac{1}{S}\sum_{s=1}^{S} f(x^{(s)})\frac{p(x^{(s)})}{q(x^{(s)})} \qquad ; x^{(s)} \sim q(x) \tag{9}$$

In eqn. 9 we, divided and multiplied with $q(x)$, a known distribution which we can sample from.

Limitation:

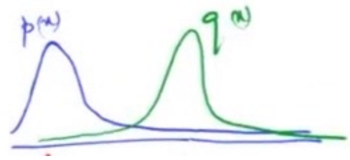

Fig-2

Let's say we have $p(x)$ and $q(x)$ as shown in Fig. 2. When we do Monte Carlo approximation to compute the expectation value, most of the samples from $q(x)$ come from the region where $q(x)$ has a comparatively larger value whereas in this region $p(x)$ has a very small value. On the other hand, in the region where $p(x)$ has a larger value $q(x)$ is very small and thus only a very samples while computing the expectation will be drawn from this region. This is problematic as we won't be representing the original distribution properly. Therefore, for this method to work efficiently, $q(x)$ should be nearly equal or at least close to $p(x)$. $q(x)$ should be significant where $p(x)$ is significant.

Now let us assume a case where the normalization constant of the distribution $p(x) = \frac{\tilde{p}(x)}{Z_p}$ is not given. We choose a $q(x) = \frac{\tilde{q}(x)}{Z_q}$. Now,

$$E[f] = \int f(x)\frac{\tilde{p}(x)}{Z_p}\frac{q(x)}{\frac{\tilde{q}(x)}{Z_q}}dx \tag{10}$$

$$= \frac{1}{S}\frac{Z_q}{Z_p}\sum_{s=1}^{S} f(x^{(s)})\frac{\tilde{p}(x^{(s)})}{\tilde{q}(x^{(s)})} \tag{11}$$

where $\frac{\tilde{p}(x^{(s)})}{\tilde{q}(x^{(s)})}$ are called the importance weights. In eqn (11), Z_q/Z_p is difficult to compute. Here's what we do-

$$\frac{Z_p}{Z_q} = \frac{\int \tilde{p}(x)dx}{\frac{\tilde{q}(x)}{q(x)}} \tag{12}$$

$$= \int \frac{\tilde{p}(x)}{\tilde{q}(x)}q(x)dx \tag{13}$$

$$= \frac{1}{S}\sum_{s=1}^{S}\frac{\tilde{p}(x^{(s)})}{\tilde{q}(x^{(s)})} \qquad ; x^{(s)} \sim q(x) \tag{14}$$

Combining eqns (11) and (14), we get:

$$E[f] = \frac{\sum_{s=1}^{S}\tilde{r}^{(s)}f(x^{(s)})}{\sum_{s=1}^{S}\tilde{r}^{(s)}} \qquad ; x^{(s)} \sim q(x) \quad \text{and} \quad \tilde{r}^{(s)} = \frac{\tilde{p}(x^{(s)})}{\tilde{q}(x^{(s)})} \tag{15}$$

Therefore, when you don't have the normalization constants, you can still use Importance sampling technique.