STAT/Q SCI 403: Introduction to Resampling Methods **Spring 2018**

Lecture 4: Importance Sampling and Rejection Sampling

Instructor: Yen-Chi Chen

4.1 Importance Sampling

In Lecture 2, we have learned the Monte Carlo Simulation approach to evaluate an integration. We briefly mentioned the *importance sampling* in that lecture and here we will study more about this approach.

Let X be a random variable with PDF p. Consider evaluating the following quantity:

$$I = \mathbb{E}(f(X)) = \int f(x)p(x)dx,$$

where f is a known function. In the example of Lecture 2, we are interested in evaluating

$$\int_0^1 e^{-x^3}dx = \mathbb{E}(f(X)),$$

where $f(x) = e^{-x^3}$ and X is a uniform random variable over $[0,1]$.

Here is how the importance sampling works. We first pick a proposal density (also called sampling density) q and generate random numbers $Y_1, \cdots, Y_N$ IID from q. Then the importance sampling estimator is

$$\widehat{I}_N = \frac{1}{N}\sum_{i=1}^N f(Y_i)\cdot\frac{p(Y_i)}{q(Y_i)}.$$

When $p = q$, this reduces to the simple estimator that uses sample means of $f(Y_i)$ to estimate its expectation.

Does this estimator a good estimator? Let's study its bias and variance. For the bias,

$$\begin{aligned}
\mathbb{E}(\widehat{I}_N) - I &= \mathbb{E}\left(f(Y_i)\cdot\frac{p(Y_i)}{q(Y_i)}\right) - I\\
&= \int f(y)\frac{p(y)}{q(y)}q(y)dy - I\\
&= \int f(y)p(y)dy - I = 0.
\end{aligned}$$

Thus, it is an unbiased estimator!

How about the variance?

$$\begin{aligned}
\mathsf{Var}(\widehat{I}_N) &= \frac{1}{N}\mathsf{Var}\left(f(Y_i)\cdot\frac{p(Y_i)}{q(Y_i)}\right)\\
&= \frac{1}{N}\left\{\mathbb{E}\left(f^2(Y_i)\cdot\frac{p^2(Y_i)}{q^2(Y_i)}\right) - \underbrace{\mathbb{E}^2\left(f(Y_i)\cdot\frac{p(Y_i)}{q(Y_i)}\right)}_{I^2}\right\}\\
&= \frac{1}{N}\left(\int\frac{f^2(y)p^2(y)}{q(y)}dy - I^2\right).
\end{aligned}$$

So only the first quantity depends on the choice of proposal density q. Thus, if we have multiple proposal density, say q_1, q_2, q_3, the best proposal will be the one that minimizes the integration $\int \frac{f^2(y)p^2(y)}{q(y)} dy$.

You may be curious about the optimal proposal density (the q that minimizes the variance). And here is a striking result about this optimal proposal density. First, we recall the Cauchy-Scharwz inequality–for any two functions $A(y)$ and $B(y)$,

$$\int A^2(y)dy \int B^2(y)dy \geq \left(\int A(y)B(y)dy\right)^2$$

and the $=$ holds whenever $A(y) \propto \cdot B(y)$ for some constant. One way to think about this is to view them as vectors–for any two vectors u, v, $\|u\|^2\|v\|^2 \geq \|u \cdot v\|^2$ and the equality holds whenever u and v are parallel to each other. Identifying $A^2(y) = \frac{f^2(y)p^2(y)}{q(y)}$ and $B^2(y) = q(y)$, we have

$$\int \frac{f^2(y)p^2(y)}{q(y)} dy \underbrace{\int q(y)dy}_{=1} \geq \left(\int \frac{f^2(y)p^2(y)}{q(y)} q(y)dy\right)^2 = I^2.$$

Namely, this tells us that the optimal choice $q_{\mathsf{opt}}(y)$ leads to

$$\mathsf{Var}(\widehat{I}_{N,\mathsf{opt}}) = \frac{1}{N}\left(I^2 - I^2\right) = 0,$$

a zero-variance estimator! Moreover, the optimal q satisfies

$$\sqrt{\frac{f^2(y)p^2(y)}{q_{\mathsf{opt}}(y)}} = A(y) \propto B(y) = \sqrt{q_{\mathsf{opt}}(y)},$$

implying

$$q_{\mathsf{opt}}(y) \propto f(y)p(y) \Longrightarrow p_{\mathsf{opt}}(y) = \frac{f(y)p(y)}{\int f(y)p(y)dy}. \tag{4.1}$$

This gives us a good news–the optimal proposal density has 0 variance and it is unbiased. Thus, we only need to sample it once and we can obtain the actual value of I. However, even if we know the closed form of $q_{\mathsf{opt}}(y)$, how to sample from this density is still unclear. In the next section, we will talk about a method called *Rejection Sampling*, which is an approach that can tackle this problem.

4.2 Rejection Sampling

Given a density function $f(x)$, the rejection sampling is a method that can generate data points from this density function f.

Here is how one can generate a random variable from f.

1. We first choose a number $M \geq \sup_x \frac{f(x)}{p(x)}$ and a proposal density p where we know how to draw sample from (q can be the density of a standard normal distribution).
2. Generate a random number Y from p and another random number U from $\mathsf{Uni}[0,1]$.
3. If $U < \frac{f(Y)}{M \cdot p(Y)}$, we set $X = Y$. Otherwise go back to the previous step to draw another new pair of Y and U.

The above procedure is called *rejection sampling* (or rejection-acceptance sampling). If we want to generate $X_1, \cdots, X_n$ from f, we can apply the above procedure multiple times until we accept n points.

Does this approach work? Now we consider the CDF of X.

$$\begin{aligned} P(X \leq x) &= P(Y \leq x | \mathsf{accept} Y) \\ &= P\left(Y \leq x | U < \frac{f(Y)}{M \cdot p(Y)}\right) \\ &= \frac{P\left(Y \leq x, U < \frac{f(Y)}{M \cdot p(Y)}\right)}{P\left(U < \frac{f(Y)}{M \cdot p(Y)}\right)}. \end{aligned} \tag{4.2}$$

Note that in the last equality, we used the definition of conditional probability.

For the numerator, using the feature of conditional probability,

$$\begin{aligned} P\left(Y \leq x, U < \frac{f(Y)}{M \cdot p(Y)}\right) &= \int P\left(Y \leq x, U < \frac{f(Y)}{M \cdot p(Y)} | Y = y\right) p(y)dy \\ &= \int P\left(y \leq x, U < \frac{f(y)}{M \cdot p(y)}\right) p(y)dy \\ &= \int I(y \leq x) P\left(U < \frac{f(y)}{M \cdot p(y)}\right) p(y)dy \\ &= \int_{-\infty}^{x} \frac{f(y)}{M \cdot p(y)} p(y)dy \\ &= \frac{1}{M} \int_{-\infty}^{x} f(y)dy \end{aligned}$$

Note that in the fourth equality, we use the fact that the choice of $M : M \geq \sup_x \frac{f(x)}{p(x)}$ ensures

$$\frac{f(y)}{M \cdot p(y)} \leq 1 \quad \forall y.$$

For the denominator, using the similar trick,

$$\begin{aligned} P\left(U < \frac{f(Y)}{M \cdot p(Y)}\right) &= \int P\left(U < \frac{f(Y)}{M \cdot p(Y)} | Y = y\right) p(y)dy \\ &= \int P\left(U < \frac{f(y)}{M \cdot p(y)}\right) p(y)dy \\ &= \int \frac{f(y)}{M \cdot p(y)} p(y)dy \\ &= \frac{1}{M} \int f(y)dy = \frac{1}{M}. \end{aligned}$$

Thus, putting altogether into equation (4.2), we obtain

$$P(X \leq x) = \frac{P\left(Y \leq x, U < \frac{f(Y)}{M \cdot p(Y)}\right)}{P\left(U < \frac{f(Y)}{M \cdot p(Y)}\right)} = \frac{\frac{1}{M}\int_{-\infty}^{x} f(y)dy}{\frac{1}{M}} = \int_{-\infty}^{x} f(y)dy,$$

which means that the random variable X does have the density f.

Here are some features about the rejection sampling:

- Using the rejection sampling, we can generate sample from any density f *as long as we know the closed form of* f.
- If we do not choose M well, we may reject many realizations of Y, U to obtain a single realization of X.
- There is an upper on M at the first step: $M \geq \sup_x \frac{f(x)}{p(x)}$.
- In practice, we want to choose M as small as possible because a small M leads to a higher chance of accepting Y. To see this, note that the denominator $P\left(U < \frac{f(Y)}{M \cdot p(Y)}\right) = P(\mathsf{Accept} Y) = \frac{1}{M}$. Thus, a small M leads to a large accepting probability.
- If you want to learn more about rejection sampling, I would recommend http://www.columbia.edu/~ks20/4703-Sigman/4703-07-Notes-ARM.pdf.