MTH 511a - 2020: Lecture 13

Instructor: Dootika Vats

1 Importance Sampling

1.1 Weighted Importance Sampling

Often for many distributions, we do not know the target distribution fully, but only know it up to a normalizing constant. That is, the target density is

$$\pi(x) = a\tilde{\pi}(x)$$

for some unknown a and the proposal density is

$$g(x) = b\tilde{g}(x)$$

for some known or unknown b. For simplicity, we will assume that b is unknown. Suppose for some function h, the following integral is of interest:

$$\theta := \int_{\mathcal{X}} h(x)\pi(x)dx\,.$$

We still want to use g as the importance distribution, so that

$$\theta = \int_{\mathcal{X}} \frac{h(x)\pi(x)}{g(x)} g(x)\,dx\,.$$

Since a and b are unknown, we can't evaluate $\pi(x)$ and $g(x)$. So our original estimator does not work anymore! We can evaluate $\tilde{\pi}(x)$ and $\tilde{g}(x)$. So if we can estimate a and b as well, that will allow us estimate θ. Instead, we will estimate b/a, which also works!

Consider $Z_1, \dots, Z_t \sim G$. The *weighted importance sampling* estimator of θ is

$$\hat{\theta}_w := \frac{\displaystyle\sum_{t=1}^{N} \frac{h(Z_t)\tilde{\pi}(Z_t)}{\tilde{g}(Z_t)}}{\displaystyle\sum_{t=1}^{N} \frac{\tilde{\pi}(Z_t)}{\tilde{g}(Z_t)}}\,.$$

Theorem 1. *As $N \to \infty$, $\hat{\theta}_w \xrightarrow{p} \theta$.*

Proof. As a first step, note that at $N \to \infty$

$$\frac{1}{N}\sum_{t=1}^{N} \frac{h(Z_t)\tilde{\pi}(Z_t)}{\tilde{g}(Z_t)} \xrightarrow{p} \mathrm{E}_g\left[\frac{h(Z)\tilde{\pi}(Z)}{\tilde{g}(Z)}\right] \quad \text{and} \quad \frac{1}{N}\sum_{t=1}^{N} \frac{\tilde{\pi}(Z_t)}{\tilde{g}(Z_t)} \xrightarrow{p} \mathrm{E}_g\left[\frac{\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]$$

By an application of the continuous mapping theorem, as $N \to \infty$

$$\hat{\theta}_w \xrightarrow{p} \frac{\mathrm{E}_g\left[\dfrac{h(Z)\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]}{\mathrm{E}_g\left[\dfrac{\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]}$$

We need to show that

$$\theta = \frac{\mathrm{E}_g\left[\dfrac{h(Z)\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]}{\mathrm{E}_g\left[\dfrac{\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]}$$

So we will first find both expectations. First

$$\begin{aligned}
\mathrm{E}_g\left[\frac{h(Z)\tilde{\pi}(Z)}{\tilde{g}(Z)}\right] &= \int_{\mathcal{X}} \frac{h(z)\tilde{\pi}(z)}{\tilde{g}(z)}\, g(z)dz \\
&= \frac{b}{a}\int_{\mathcal{X}} \frac{h(z)\pi(z)}{g(z)}\, g(z)dz \\
&= \frac{b}{a}\theta\,.
\end{aligned}$$

Second,

$$\begin{aligned}
\mathrm{E}_g\left[\frac{\tilde{\pi}(Z)}{\tilde{g}(Z)}\right] &= \int_{\mathcal{X}} \frac{\tilde{\pi}(z)}{\tilde{g}(z)} g(z)dz \\
&= \frac{b}{a}\int_{\mathcal{X}} \pi(z)dz = \frac{b}{a}\,.
\end{aligned}$$

So,

$$\frac{\mathrm{E}_g\left[\dfrac{h(Z)\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]}{\mathrm{E}_g\left[\dfrac{\tilde{\pi}(Z)}{\tilde{g}(Z)}\right]} = \frac{\frac{b}{a}\theta}{\frac{b}{a}} = \theta\,.$$

$\square$

We will denote

$$w(Z) = \frac{\tilde{\pi}(Z)}{\tilde{g}(Z)}\,.$$

Then $w(Z)$ is called the importance sampling weight.

However, not knowing b and a comes at a cost. Unlike the simple importance sampling estimator, the weighted importance sampling estimator $\hat{\theta}_w$ is not unbiased.

To see this heuristically, let's assume that we obtained two different samples from G for the numerator and the denominator. That is, assume that $Z_1, \dots, Z_N \sim G$ and $T_1, \dots, T_N \sim G$. This allows us to break the expectation as follows:

$$\begin{aligned}
\mathrm{E}_g\left[\frac{\sum_{t=1}^{N} h(Z_t)w(Z_t)}{\sum_{t=1}^{N} w(T_t)}\right] &= \mathrm{E}_g\left[\sum_{t=1}^{N} h(Z_t)w(Z_t)\right] \mathrm{E}_g\left[\frac{1}{\sum_{t=1}^{N} w(T_t)}\right] \\
&\leq \mathrm{E}_g\left[\sum_{t=1}^{N} h(Z_t)w(Z_t)\right] \frac{1}{\mathrm{E}_g\left[\sum_{t=1}^{N} w(T_t)\right]} \qquad \text{By Jensen's inequality} \\
&= \theta\,,
\end{aligned}$$

where the equality only holds for a constant function only. Thus, in general, $\hat{\theta}_w$ underestimates the true value of θ.

An exact expression for the variance is difficult to obtain. But an approximate variance expression is

$$\begin{aligned}
\mathrm{Var}_g\left[\hat{\theta}_w\right] &\approx \frac{\mathrm{Var}_g\left(h(Z)w(Z)\right)}{\left[\mathrm{E}_g(w(Z))\right]^2}\left(1 + \frac{\mathrm{Var}_g(w(Z))}{\left[\mathrm{E}_g(w(Z))\right]^2}\right) \\
&= \frac{\mathrm{Var}_g\left(h(Z)w(Z)\right)}{b^2/a^2}\left(1 + \frac{\mathrm{Var}_g(w(Z))}{\left[\mathrm{E}_g(w(Z))\right]^2}\right) \\
&= \mathrm{Var}_g\left(ah(Z)w(Z)/b\right)\left(1 + \frac{\mathrm{Var}_g(w(Z))}{\left[\mathrm{E}_g(w(Z))\right]^2}\right) \\
&= \mathrm{Var}_g\left(h(Z)\frac{\pi(Z)}{g(Z)}\right)\left(1 + \frac{\mathrm{Var}_g(w(Z))}{\left[\mathrm{E}_g(w(Z))\right]^2}\right) \\
&= \sigma_g^2\left(1 + \frac{\mathrm{Var}_g(w(Z))}{\left[\mathrm{E}_g(w(Z))\right]^2}\right) \\
&\geq \sigma_g^2\,,
\end{aligned}$$

where recall that σ_g^2 is the variance coming from the simple importance sampling estimator. Thus, if we can do simple importance sampling, we should do it. We should weighted importance sampling only when necessary.

Example 1. Consider the target distribution

$$\pi(x) \propto e^{(\sin(xy))} \quad 0 \leq x, y \leq \pi$$

The target distribution is not a product of two marginals, so we have to implement multivariate importance sampling. Also, we do not know the normalizing constants. So for some unknown $a > 0$

$$\pi(x) = a\, e^{(\sin(xy))} \quad 0 \leq x, y \leq \pi$$

Suppose $h(x, y) = xy$. We will use a weighted importance sampling distribution. Consider the importance distribution $U[0, \pi] \times U[0, \pi]$. So that

$$g(z, t) = \frac{1}{\pi^2} I(0 < z < \pi) I(0 < t < \pi) := b \text{ (assume this is unknown)}$$

Sample $(Z_1, T_1), \ldots, (Z_N, T_N) \sim U[0, \pi] \times U[0, \pi]$. The weights are

$$w(Z_t, T_t) = \frac{\tilde{\pi}(Z_t, T_t)}{\tilde{g}(Z_t, T_t)} = e^{\sin(Z_t T_t)} .$$

The final estimator is

$$\hat{\theta}_w = \frac{\sum_{t=1}^{N} Z_t T_t w(Z_t, T_t)}{\sum_{t=1}^{N} w(Z_t, T_t)} .$$

1.2 Questions to think about

- How would you choose a good proposal for weighted importance sampling? Would finding a proposal that yields a small variance suffice?
- Do you have intuition as to why the variance of the weight importance estimator is larger than the variance of the simple importance sampler for the same importance proposal?
- Implement the above example to convince yourself that weighted importance sampling underestimates the truth and converges to the truth from below.