

# 1 Importance Sampling

In many applications we want to compute  $\mu = E(f(\mathbf{X}))$  where  $f(\mathbf{x})$  is **nearly zero outside a region**  $A$  for which  $P(\mathbf{X} \in A)$  is **small**. The set  $A$  may have **small volume**, or it may be in the tail of the distribution of  $\mathbf{X}$ . A **plain** Monte Carlo sample from the distribution of  $\mathbf{X}$  could **fail to have even one point** inside the region  $A$ . It is clear intuitively that we must get **some samples** from the **interesting or important** region. We do this by sampling from a distribution that **over-weights the important region**, hence the name importance sampling. Having oversampled the important region, we **have to adjust our estimate** somehow to account for having sampled from this other distribution.

## 1.1 Basic Importance Sampling

Suppose that our problem is to find  $\mu = E(f(\mathbf{X})) = \int_{\mathcal{D}} f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$  where  $p$  is a probability density function on  $\mathcal{D} \subset \mathbb{R}^d$ . We take  $p(\mathbf{x}) = 0$  for all  $\mathbf{x} \notin \mathcal{D}$ . If  $q$  is a positive probability density function on  $\mathbb{R}^d$ , then

$$\mu = \int_{\mathcal{D}} f(\mathbf{x})p(\mathbf{x})d\mathbf{x} = \int_{\mathcal{D}} \frac{f(\mathbf{x})p(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})d\mathbf{x} = E_q\left(\frac{f(\mathbf{X})p(\mathbf{X})}{q(\mathbf{X})}\right),$$

where  $E_q(\cdot)$  denotes **expectation** for  $\mathbf{X} \sim q$ . We also write  $E_q(\cdot)$  and  $Var_q(\cdot)$  for expectation and variance, respectively, when  $\mathbf{X} \sim q$ . Our **original goal** then is to find  $E_p(f(\mathbf{X}))$ . By making a **multiplicative adjustment** to  $f$  we **compensate** for sampling from  $q$  **instead** of  $p$ . The **adjustment factor**  $p(\mathbf{x})/q(\mathbf{x})$  is called the **likelihood ratio**. The distribution  $q$  and  $p$  are called the **importance distribution** and the **nominal distribution**, respectively. The importance sampling estimate of  $\mu = E_p(f(\mathbf{X}))$  is

$$\hat{\mu}_{\text{imp}} = \frac{1}{n} \sum_{i=1}^n \frac{f(\mathbf{X}_i)p(\mathbf{X}_i)}{q(\mathbf{X}_i)} = \frac{1}{n} \sum_{i=1}^n h(\mathbf{X}_i),$$

where  $h(\mathbf{x}) = \frac{f(\mathbf{x})p(\mathbf{x})}{q(\mathbf{x})}$  and  $\mathbf{X}_i \sim q$ .

It is easy to see that  $\hat{\mu}_{\text{imp}}$  is **unbiased** for  $\mu$ , as

$$E(\hat{\mu}_{\text{imp}}) = E_q(h(\mathbf{X})) = \mu.$$

The **variance** of  $\hat{\mu}_{\text{imp}}$  can be expressed as  $\sigma_q^2/n$ , where

$$\sigma_q^2 = Var(h(\mathbf{X})) = \int_{\mathcal{D}} \frac{f^2(\mathbf{x})p^2(\mathbf{x})}{q(\mathbf{x})}d\mathbf{x} - \mu^2 = \int_{\mathcal{D}} \frac{(f(\mathbf{x})p(\mathbf{x}) - \mu q(\mathbf{x}))^2}{q(\mathbf{x})}d\mathbf{x}.$$

To construct a **confidence interval** for  $\mu$ , we need to estimate  $\sigma_q^2$ . The natural variance estimator

is

$$\hat{\sigma}_q^2 = \frac{1}{n} \sum_{i=1}^n \left( \frac{f(\mathbf{X}_i)p(\mathbf{X}_i)}{q(\mathbf{X}_i)} - \hat{\mu}_{\text{imp}} \right)^2.$$

Therefore, an asymptotic 99% confidence interval for  $\mu$  is  $\hat{\mu}_{\text{imp}} \mp 2.58\hat{\sigma}_q^2/\sqrt{n}$ .

**Remark 1.** The importance distribution  $q$  **does not have to be positive** everywhere. It is **enough** to have  $q(\mathbf{x}) > 0$  **whenever**  $f(\mathbf{x})p(\mathbf{x}) \neq 0$ . †

**Remark 2.** The expression for the variance of  $\hat{\mu}_{\text{imp}}$  guides us in selecting a good importance sampling rule. The first expression of  $\sigma_q^2$  suggests that a **better**  $q$  is **one that gives a smaller value** of  $\int_{\mathcal{D}} (fp)^2/q d\mathbf{x}$ .

The second integral expression of  $\sigma_q^2$  illustrates **how importance sampling can succeed or fail**. The numerator in the **integrand is small** when  $f(\mathbf{x})p(\mathbf{x}) - \mu q(\mathbf{x})$  is **close to zero**, that is, when  $q(\mathbf{x})$  is **nearly proportional** to  $f(\mathbf{x})p(\mathbf{x})$ . From the denominator, we see that regions with **small values** of  $q(\mathbf{x})$  **greatly magnify** whatever **lack of proportionality** appears in the numerator. †

**Example 1.** (Gaussian  $p$  and  $q$ : A word of caution) The effect of **light-tailed**  $q$  can be illustrated by this example. Suppose that  $f(x) = x$ , and  $p(x) = \exp(-x^2/2)/\sqrt{2\pi}$ . If  $q(x) = \exp(-x^2/(2\sigma^2))/(\sigma\sqrt{2\pi})$  with  $\sigma > 0$  then

$$\begin{aligned} \sigma_q^2 &= \int_{-\infty}^{\infty} x^2 \frac{(\exp(-x^2/2)/\sqrt{2\pi})^2}{\exp(-x^2/(2\sigma^2))/(\sigma\sqrt{2\pi})} dx \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp(-x^2(2 - \sigma^{-2})/2) dx \\ &= \begin{cases} \frac{\sigma}{(2 - \sigma^{-2})^{3/2}} & \text{if } \sigma^2 > \frac{1}{2} \\ \infty & \text{otherwise.} \end{cases} \quad || \end{aligned}$$

## 1.2 Self-normalized Importance Sampling

Sometimes we can **only compute an unnormalized version** of  $p$ ,  $p_u(\mathbf{x}) = cp(\mathbf{x})$ , where  $c > 0$  is **unknown**. Also suppose that we can compute  $q_u(\mathbf{x}) = bq(\mathbf{x})$ , where  $b > 0$  might be **unknown**. If we are fortunate or clever enough to have  $b = c$ , then  $p(\mathbf{x})/q(\mathbf{x}) = p_u(\mathbf{x})/q_u(\mathbf{x})$  and we can still use  $\hat{\mu}_{\text{imp}}$ . Otherwise we may compute the **ratio**  $w_u(\mathbf{x}) = p_u(\mathbf{x})/q_u(\mathbf{x}) = (c/b)p(\mathbf{x})/q(\mathbf{x})$  and consider the **self-normalized importance sampling estimator**

$$\tilde{\mu}_{\text{imp}} = \frac{\sum_{i=1}^n f(\mathbf{X}_i)w_u(\mathbf{X}_i)}{\sum_{i=1}^n w_u(\mathbf{X}_i)} = \frac{\sum_{i=1}^n f(\mathbf{X}_i)w(\mathbf{X}_i)}{\sum_{i=1}^n w(\mathbf{X}_i)}.$$

In general  $\tilde{\mu}_{\text{imp}}$  is a **biased** estimator of  $\mu$ .

**Theorem 1.** *Let  $p$  be a probability density function on  $\mathbb{R}^d$  and let  $f(\mathbf{x})$  be a function such that  $\mu = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$  exists. Suppose that  $q(\mathbf{x})$  is a probability density function on  $\mathbb{R}^d$  with  $q(\mathbf{x}) > 0$  whenever  $p(\mathbf{x}) > 0$ . Let  $X_1, \dots, X_n \sim q$  be independent and let  $\tilde{\mu}_{\text{imp}}$  be the self-normalized importance sampling estimator. Then*

$$P\left(\lim_{n \rightarrow \infty} \tilde{\mu}_{\text{imp}} = \mu\right) = 1.$$

*Proof.* The proof is simple using strong law of large numbers.

**Remark 3.** The **self-normalized** importance sampler  $\tilde{\mu}_{\text{imp}}$  **requires a stronger condition** on  $q$  than the unbiased importance sampler  $\hat{\mu}_{\text{imp}}$  does. We now need  $q(\mathbf{x}) > 0$  whenever  $p(\mathbf{x}) > 0$  even if  $f(\mathbf{x})$  is zero. †

### 1.3 Importance Sampling Diagnostic

Importance sampling uses **unequally weighted** observations. The **weights** are  $w_i = p(\mathbf{x}_i)/q(\mathbf{x}_i) \geq 0$  for  $i = 1, \dots, n$ . In extreme settings, one of the  $w_i$  may be **vastly larger** than all the others and then we have **effectively only got one observation**. We would like to have a **diagnostic** to tell when the weights are problematic. It is even possible that  $w_1 = w_2 = \dots = w_n = 0$ . In that case, importance sampling has clearly failed and we do not need a diagnostic to tell us so. Hence, we may **assume** that  $\sum_{i=1}^n w_i > 0$ .

Consider a hypothetical linear combination

$$S_w = \frac{\sum_{i=1}^n w_i Z_i}{\sum_{i=1}^n w_i},$$

where  $Z_i$  are independent random variables with common mean and common variance  $\sigma^2 > 0$  and  $w_i > 0$  are weights. The **unweighted average** of  $n_e$  independent random variables  $Z_i$  has variance  $\sigma^2/n_e$ . Setting  $\text{Var}(S_w) = \sigma^2/n_e$  and solving for  $n_e$  yields the **effective sample size**

$$n_e = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n w_i^2} = \frac{n\bar{w}^2}{w^2}.$$

If the **weights are too imbalanced** then the result is similar to averaging only  $n_e \ll n$  observations and might therefore be **unreliable**. The point at which  $n_e$  becomes alarmingly small is hard to specify, because it is application specific.