Importance Sampling

David S. Rosenberg

NYU: CDS

January 9, 2021

Problem statement

- Consider the function $f : \mathcal{X} \to \mathbb{R}$.
- We'd like to estimate $\mathbb{E}f(X)$ for $X \sim P$
- With a sample $X_1, \ldots, X_n \sim P$, the Monte Carlo estimate of $\mathbb{E}f(X)$ is

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

- Unfortunately we don't have such a sample.
- Is there anything we can do with a sample from a different distribution $Y_1, \ldots, Y_n \sim Q$?

Change of measure

Theorem (Change of measure)

Let $p(x)$ and $q(x)$ be PMFs on $\mathcal{X}$, for some finite $\mathcal{X}$. Assume that $p(x) > 0 \implies q(x) > 0$ for all $x \in \mathcal{X}$. Then for any $f : \mathcal{X} \to \mathbb{R}$,

$$\mathbb{E}_{X \sim p(x)} f(X) = \mathbb{E}_{X \sim q(x)} \left[f(X) \frac{p(X)}{q(X)} \right]$$

Change of measure

Theorem (Change of measure)

Let $p(x)$ and $q(x)$ be PMFs on $\mathcal{X}$, for some finite $\mathcal{X}$. Assume that $p(x) > 0 \implies q(x) > 0$ for all $x \in \mathcal{X}$. Then for any $f : \mathcal{X} \to \mathbb{R}$,

$$\mathbb{E}_{X\sim p(x)} f(X) = \mathbb{E}_{X\sim q(x)} \left[f(X) \frac{p(X)}{q(X)} \right]$$

This is a special case of the Radon-Nikodym theorem, which holds in a much more general setting. Here's the development from [Sch95, Sec A.6] for the case of probability measures: Let μ_1 and μ_2 be measures on $(S, \mathcal{A})$. We say that μ_2 is **absolutely continuous** with respect to μ_1, denoted $\mu_2 \ll \mu_1$ if $\forall A \in \mathcal{A}$, $\mu_1(A) = 0$ implies $\mu_2(A) = 0$.

Theorem (Radon-Nikodym theorem for probabilities)

[Sch95, Thm A.74]Let μ_1 and μ_2 be probability measures on $(S, \mathcal{A})$ such that $\mu_2 \ll \mu_1$. Then there exists measurable function $f : S \to [0, \infty)$ such that for any $g : S \to \mathbb{R}$ with $\int |g(x)|\, d\mu_2(x) < \infty$ we have

$$\int g(x)\, d\mu_2(x) = \int g(x) f(x)\, d\mu_1(x).$$

The function f is called the ***Radon-Nikodym derivative*** *of μ_2 with respect to μ_1 and is unique μ_1-a.s. We often denote f by $\frac{d\mu_2}{d\mu_1}$.*

Proof of change of measure

Proof.

Verifying that summand is 0 whenever $q(x) = 0$ or $p(x) = 0$, we have

$$
\begin{aligned}
\mathbb{E}_{X \sim p(x)} f(X) &= \sum_{x \in \mathcal{X}} p(x) f(x) \\
&= \sum_{x \in \mathcal{X}} q(x) \frac{p(x)}{q(x)} f(x) \\
&= \mathbb{E}_{X \sim q(x)} \left[\frac{p(x)}{q(x)} f(x) \right]
\end{aligned}
$$

$\square$

Proof of change of measure

Proof.
Verifying that summand is 0 whenever $q(x)=0$ or $p(x)=0$, we have

$$\begin{aligned}
\mathbb{E}_{X\sim p(x)} f(X) &= \sum_{x\in\mathcal{X}} p(x) f(x) \\
&= \sum_{x\in\mathcal{X}} q(x)\frac{p(x)}{q(x)} f(x) \\
&= \mathbb{E}_{X\sim q(x)}\left[\frac{p(x)}{q(x)} f(x)\right]
\end{aligned}$$

□

With a bit more detail, we have:

$$\begin{aligned}
\mathbb{E}_{X\sim p(x)} f(X) &= \sum_{x:p(x)>0} p(x) f(x) && \text{no contribution when} p(x)=0 \\
&= \sum_{x:p(x)>0,q(x)>0} p(x) f(x) && \text{since } p(x)>0 \implies q(x)>0 \\
&= \sum_{x:p(x)>0,q(x)>0} q(x)\frac{p(x)}{q(x)} f(x) \\
&= \sum_{q(x)>0} q(x)\frac{p(x)}{q(x)} f(x) && \text{summand 0 when } p(x)=0 \\
&= \mathbb{E}_{X\sim q(x)}\left[\frac{p(x)}{q(x)} f(x)\right] && \text{no contribution when } q(x)=0.
\end{aligned}$$

Importance sampling

- Suppose that $p(x) > 0 \implies q(x) > 0$ for all $x \in \mathcal{X}$. Then for any $f : \mathcal{X} \to \mathbb{R}$,

$$\mathbb{E}_{X \sim p(x)} f(X) = \mathbb{E}_{X \sim q(x)} \left[f(X) \frac{p(X)}{q(X)} \right]$$

- If we have a sample $X_1, \ldots, X_n \sim q(x)$, then a Monte Carlo estimate of the RHS

$$\hat{\mu}_{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} f(X_i) \frac{p(X_i)}{q(X_i)}$$

  is called an **importance sampling** estimator for $\mathbb{E}_{X \sim p(x)} f(X)$.
- The ratios $f(X_i)/q(X_i)$ are called **importance weights**.
- Importance sampling is a weighted average of $f(X_i)$'s.

Importance sampling estimator is unbiased and consistent

- $\hat{\mu}_{\text{IS}}$is **unbiased** by the change of measure theorem:

$$\mathbb{E}\hat{\mu}_{\text{IS}} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{X_i \sim q(x)}\left[f(X_i)\frac{p(X_i)}{q(X_i)}\right] = \mathbb{E}_{X \sim p(x)} f(X),$$

- $\hat{\mu}_{\text{IS}}$ is **consistent** by the Law of Large Numbers: $\hat{\mu}_{\text{IS}} \xrightarrow{P} \mathbb{E}\hat{\mu}_{\text{IS}}$ as $n \to \infty$.

Variance of importance sampling

- Since summands are independent, the variance is given by

$$
\begin{aligned}
\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i)\frac{p(X_i)}{q(X_i)}\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}\left(f(X)\frac{p(X)}{q(X)}\right) \\
&= \frac{1}{n}\operatorname{Var}\left(f(X)\frac{p(X)}{q(X)}\right)
\end{aligned}
$$

- Here we see a potential issue with importance sampling: if $q(x)$ is much smaller than $p(x)$, the importance weight can get very large, which can make the variance very large

References

Resources

References I

[Sch95] Mark J. Schervish, *Theory of statistics*, Springer Series in Statistics, Springer New York, 1995.