Importance Sampling: Estimating Volume and Normalizing Constants (Chapter 8)

Michael Tsiang

Stats 102C: Introduction to Monte Carlo Methods

Acknowledgements: Qing Zhou

Outline

Estimating Volume

- Let D denote a region in $\mathbb{R}^n$, and define:

$$h(\boldsymbol{x}) = I(\boldsymbol{x} \in D) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in D \\ 0 & \text{otherwise} \end{cases}$$

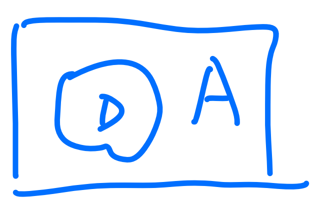
- Then the volume of D is

$$V_D = \int_D \mathrm{d}\boldsymbol{x} = \int I(\boldsymbol{x} \in D)\, \mathrm{d}\boldsymbol{x}.$$

- For example, for $D \subset \mathbb{R}^2$, the volume of D is

$$V_D = \iint_D \mathrm{d}x_1\, \mathrm{d}x_2 = \iint I[(x_1, x_2) \in D]\, \mathrm{d}x_1\, \mathrm{d}x_2.$$

region A



Estimating Volume

Key Idea: Express V_D as the expectation of a random variable!

Find a region A such that:

1. D is contained in A: $D \subset A$.
2. The volume of A, denoted by V_A, is easy to calculate.

Then:

$$
\begin{aligned}
V_D &= \int I(\boldsymbol{x} \in D)\,\mathrm{d}\boldsymbol{x} \\
&= \int I(\boldsymbol{x} \in D)\frac{V_A}{V_A}\,\mathrm{d}\boldsymbol{x} \\
&= V_A \int \overbrace{I(\boldsymbol{x} \in D)\frac{1}{V_A}}\,\mathrm{d}\boldsymbol{x} \\
&= V_A\, E[I(X \in D)],
\end{aligned}
$$

1

density of uniform (A)

where $X \sim \mathrm{Unif}(A)$.

$V_A \; E[I(X \in D)]$

$X \sim$ unifor (A)

Estimating Volume

Estimating Volume by Importance Sampling

1. Generate $X^{(1)}, X^{(2)}, \ldots, X^{(n)} \sim \text{Unif}(A)$, and compute the importance weights

$$w^{(i)} = V_A \cdot I(X^{(i)} \in D) = \begin{cases} V_A & \text{if } X^{(i)} \in D \\ 0 & \text{otherwise,} \end{cases}$$

   for $i = 1, 2, \ldots, n$.

2. Estimate V_D by

$$\widehat{V}_D = \frac{1}{n} \sum_{i=1}^{n} w^{(i)}.$$

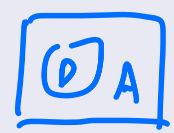
Estimating Volume

Proof (Estimating Volume by Importance Sampling).

It suffices to show that $E(\widehat{V}_D) = V_D$ and $\mathrm{Var}(\widehat{V}_D) \xrightarrow{n \to \infty} 0$.

We compute

$$
\begin{aligned}
E(\widehat{V}_D) &= E(w) \\
&= E[V_A \cdot I(X \in D)] \\
&= V_A \cdot E[I(X \in D)] \quad \text{Bernoulli} \\
&= V_A \cdot [1 \cdot P(X \in D) + \cancel{0 \cdot P(X \notin D)}] \\
&= V_A \cdot P(X \in D) \\
(X \sim \mathrm{Unif}(A)) \quad &= V_A \cdot \frac{V_D}{V_A} \quad \text{unbiased} \\
&= V_D,
\end{aligned}
$$



so $E(\widehat{V}_D) = V_D$ (i.e., $\widehat{V}_D$ is unbiased).

Estimating Volume

Proof (Estimating Volume by Importance Sampling).

We also compute

$$\begin{aligned}
\mathrm{Var}(\widehat{V}_D) &= \frac{1}{n}\mathrm{Var}(w) \\
&= \frac{1}{n}\mathrm{Var}[V_A \cdot I(X \in D)] \\
&= \frac{1}{n}V_A^2 \cdot \mathrm{Var}[I(X \in D)] \quad n(1-p)p \\
&= \frac{1}{n}V_A^2 \cdot \left\{E[I(X \in D)^2] - E[I(X \in D)]^2\right\} \\
(I(X \in D)^2 = I(X \in D)) \quad &= \frac{1}{n}V_A^2 \cdot \left\{E[I(X \in D)] - E[I(X \in D)]^2\right\} \\
(X \sim \mathrm{Unif}(A)) \quad &= \frac{1}{n}V_A^2 \cdot \left[\frac{V_D}{V_A} - \left(\frac{V_D}{V_A}\right)^2\right] \\
&= \boxed{\frac{1}{n}V_D(V_A - V_D).} \quad E(I(X \in D)) = \frac{V_D}{V_A}
\end{aligned}$$

so $\mathrm{Var}(\widehat{V}_D) = \frac{1}{n}V_D(V_A - V_D)$, and $\mathrm{Var}(\widehat{V}_D) \overset{n\to\infty}{\longrightarrow} 0$. □

as $D \to A$: $V_A - V_D = 0$

Estimating Volume $\therefore \mathrm{Var}(\hat{V}_D) = 0$

We calculated $\mathrm{Var}(\widehat{V}_D) = \frac{1}{n} V_D(V_A - V_D)$, which shows:

- As $n \to \infty$, $\mathrm{Var}(\widehat{V}_D) \longrightarrow 0$, so we can choose a sufficiently large n to make our estimator as precise as we want.
- If $V_A - V_D$ is small, then $\mathrm{Var}(\widehat{V}_D)$ is small, so we can reduce the variance of $\widehat{V}_D$ by choosing the region A to be close to D.

given a D : choose $\boxed{A} \longrightarrow D$

to increase sufficienty

Estimating Volume

Since V_D is unknown, we can approximate $\mathrm{Var}(\widehat{V}_D)$ in two ways:

- The sample variance:

$$\mathrm{Var}(\widehat{V}_D) = \frac{1}{n}\mathrm{Var}(w) \approx \frac{1}{n}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(w^{(i)} - \frac{1}{n}\sum_{j=1}^{n} w^{(j)}\right)^2\right]$$

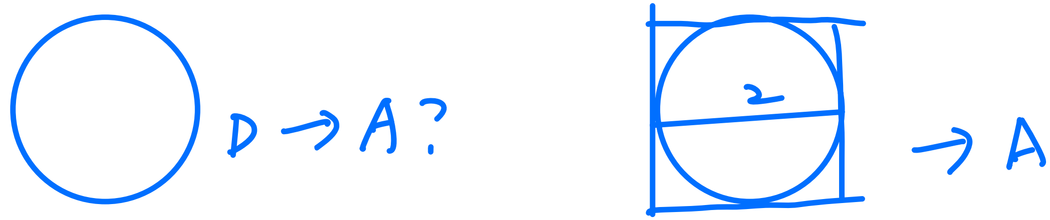
- The plug-in estimator:

$$\mathrm{Var}(\widehat{V}_D) = \frac{1}{n}V_D(V_A - V_D) \approx \frac{1}{n}\widehat{V}_D(V_A - \widehat{V}_D).$$

$$\mathrm{Var}(\hat{V}_D) = \frac{1}{n} V_D(V_A - V_D)$$

plug $\hat{V}_D \to V_D$ $\hat{V}_A \to V_A$

Example 1: Estimating the Area of a Circle



- Let D denote the unit disc $D = \{(x, y) : x^2 + y^2 \leq 1\}$.
- We want to estimate the area of D using importance sampling.
- The theoretical area is $V_D = \pi r^2 = \pi$. (area of D
- Consider the bounding rectangle (square) $A : [-1, 1] \times [-1, 1]$.
- Then $V_A = 2 \cdot 2 = 4$. A (area)

find E(VD)

Example 1: Estimating the Area of a Circle

Estimating V_D using importance sampling:

1. Generate

$$\begin{aligned} X^{(1)}, X^{(2)}, \ldots, X^{(n)} &\sim \mathrm{Unif}[-1, 1] \\ Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)} &\sim \mathrm{Unif}[-1, 1] \end{aligned}$$

   so $(X^{(i)}, Y^{(i)}) \sim \mathrm{Unif}(A)$, and compute the importance weights, for $i = 1, 2, \ldots, n$

$$\begin{aligned} w^{(i)} &= V_A \cdot I[(X^{(i)}, Y^{(i)}) \in D] \\ &= 4 \cdot I[(X^{(i)})^2 + (Y^{(i)})^2 \leq 1] \\ &= \begin{cases} 4 & \text{if } (X^{(i)})^2 + (Y^{(i)})^2 \leq 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

2. Estimate V_D by

$$w^{(i)} = V_A \cdot I(X^{(i)} \in D) = \begin{cases} V_A & \text{if } X^{(i)} \in D \\ 0 & \text{otherwise,} \end{cases}$$

$$\widehat{V}_D = \frac{1}{n} \sum_{i=1}^{n} w^{(i)}.$$

Example 6: Folded Normal Distribution

R Code to estimate V_D:

```
> set.seed(9999) # for reproduceability

> n <- 1000 # Specify the number of points to generate

> # Generate n points from A: [-1,1]x[-1,1]
> X <- runif(n, -1, 1)
> Y <- runif(n, -1, 1)

> # Compute the area of A
> V_A <- 4

> # Compute importance weights
> w <- 4 * (X^2 + Y^2 <= 1)

> # Compute mean of weights (estimate of V_D)
> Vhat_D <- mean(w)
> Vhat_D
[1] 3.164
```

A region

$Var(\hat{V}_D) = \frac{1}{n}\pi(4-\pi)$

$mean(\hat{V}_D) = \frac{1}{n}\sum_{i=1}^{n} 4\cdot(x^2+y^2\leq 1)$

- The theoretical variance of $\widehat{V}_D$ is

$\pi r^2 \quad (4-\pi r^2) = \frac{1}{n}\pi(4\cdot\pi)$

$$\mathrm{Var}(\widehat{V}_D) = \frac{1}{n}V_D(V_A - V_D) = \frac{1}{n}\pi(4-\pi).$$

- What sample size n do we need such that $\mathrm{Var}(\widehat{V}_D) < 0.01$?

- Can choose the sample size to be

if know $Var(\hat{V}_D) = 0.01$

$$n > \frac{\pi(4-\pi)}{0.01} \approx 269.68,$$

given $Var(\hat{V}_D)$

so $n \geq 270$.

$= \frac{1}{n} V_D\,(V_A - V_D) = 0.01$

$= \frac{1}{n}(\pi(4-\pi)) = 0.01$

solve for n

Example 6: Folded Normal Distribution

$mean(V_D) = mean(W) = \frac{1}{n}\sum w_i$

$Var(V_D) = \frac{1}{n}(\hat{V}_A - \hat{V}_D) V_D$

R Code to estimate $\mathrm{Var}(\widehat{V}_D)$:

$= Var(w)/n$

```
> var(w) / n # sample variance
[1] 0.002647752
```

mean(V_D)

```
> Vhat_D * (V_A - Vhat_D) / n # plug-in estimate
[1] 0.002645104
```

plug-in estimate

```
> pi * (4 - pi) / n # Theoretical variance
[1] 0.002696766
```

Outline

Estimating Normalizing Constants

$\int f(x) = 1$

- Let $q(x)$ be an unnormalized density on a region D, and let

$$Z_q = \int_D q(x)\,\mathrm{d}x$$

denote its normalizing constant.

- We want to estimate the normalizing constant Z_q.

① find trial → ② weight

Estimating Normalizing Constants

To use importance sampling, we require a **normalized** trial $g(x)$ distribution $g(x)$ to generate samples.

Estimating Normalizing Constants by Importance Sampling

1. Generate $X^{(1)}, X^{(2)}, \dots, X^{(n)} \sim g(x)$, and compute the importance weights

$$w^{(i)} = \frac{q(X^{(i)})}{g(X^{(i)})}, \rightarrow \text{find trail}$$

for $i = 1, 2, \dots, n$.

2. Estimate Z_q by

$$E(\hat{Z}_q) = \widehat{Z}_q = \frac{1}{n}\sum_{i=1}^{n} w^{(i)}. \quad \text{mean}(w^i)$$

The variance of the estimator is $\mathrm{Var}(\widehat{Z}_q) = \frac{1}{n}\mathrm{Var}(w)$.

Estimating Normalizing Constants

Proof (Estimating Normalizing Constants by Importance Sampling).

By the Strong Law of Large Numbers,

$$\widehat{Z}_q \xrightarrow{\text{a.s.}} E_g(w), = \text{mean}(w^i) = \frac{1}{n}\sum_{1}^{n} w^i$$

so it suffices to show that $E_g(w) = Z_q$. $= E(\hat{Z}_q)$

Indeed,

$$E_g(w) = E_g\left[\frac{q(X)}{g(X)}\right] = \int \frac{q(x)}{\cancel{g(x)}} \cancel{g(x)}\,\mathrm{d}x = \int q(x)\,\mathrm{d}x = Z_q.$$

Thus $\widehat{Z}_q \xrightarrow{\text{a.s.}} Z_q$, as desired. □

Estimating Normalizing Constants

What if we only find $r(x)$, an unnormalized density for $g(x)$?

- Let $\int r(x)\,\mathrm{d}x = Z_r \neq 1$ and $g(x) = \dfrac{r(x)}{Z_r}$, with Z_r unknown.
- Suppose we generate from $g(x)$ and compute weights

$$w^{(i)} = \frac{q(X^{(i)})}{r(X^{(i)})}.$$

  for $i = 1, 2, \ldots, n$.

- Is $\widehat{Z}_q = \dfrac{1}{n}\sum_{i=1}^{n} w^{(i)}$ still a good estimate of Z_q?

Estimating Normalizing Constants

If we use unnormalized $r(x)$:

- By the Strong Law of Large Numbers,

$$\begin{aligned}\widehat{Z}_q \quad \xrightarrow{\text{a.s.}} \quad E_g(w) &= E_g\left[\frac{q(X)}{r(X)}\right] \\ &= \int \frac{q(x)}{r(x)} g(x)\,\mathrm{d}x \\ &= \int \frac{q(x)}{r(x)} \cdot \frac{r(x)}{Z_r}\,\mathrm{d}x \\ &= \frac{Z_q}{Z_r}.\end{aligned}$$

- So $\widehat{Z}_q \xrightarrow{\text{a.s.}} \dfrac{Z_q}{Z_r} (\neq Z_q)$. Using $r(x)$ produces a bad estimate!