Importance Sampling: Unnormalized Densities

(Chapter 7)

Michael Tsiang

Stats 102C: Introduction to Monte Carlo Methods

Acknowledgements: Qing Zhou

Outline

1. Unnormalized Densities
   - Example 1: $q(x) = 3e^{-x^2/2}$ for $x \geq 0$
   - Example 2: $q(x) = e^{-x^2/2}$ for $x \in \mathbb{R}$
   - Example 3: $q(x) = e^{-5x}$ for $x \geq 0$
   - Example 4: $q(x) = x^3(1-x)^2$ for $x \in [0, 1]$
   - Example 5: $q(x) = x^3 e^{-x/2}$ for $x \geq 0$
2. Self-Normalized Importance Sampling
   - Example 6: Folded Normal Distribution

Unnormalized Densities

$f(x) = 2 \cdot \text{normal}$

- The PDF of the folded normal distribution is given by

$$f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, \quad \text{for } x \geq 0.$$

$=1$

- The PDF $f(x)$ is called a **normalized density**, since

$$\int_0^\infty f(x)\,\mathrm{d}x = 1.$$

- Note that

$$\int_0^\infty \sqrt{\frac{2}{\pi}} e^{-x^2/2}\,\mathrm{d}x = 1, \quad \text{so} \quad \int_0^\infty e^{-x^2/2}\,\mathrm{d}x = \sqrt{\frac{\pi}{2}}.$$

Unnormalized Densities

- Let $q(x)$ be defined as

$$q(x) = e^{-x^2/2}, \quad \text{for } x \geq 0.$$

$\neq 1$

- Then $q(x)$ is called an **unnormalized density**, since
  - $q(x) > 0$, for $x \geq 0$
  - $\displaystyle\int_0^\infty q(x)\,\mathrm{d}x = \sqrt{\frac{\pi}{2}} \neq 1.$

unnormalized density

$\int f(x) = A \underbrace{\int q(x)\,dx}$

Z_q

- The quantity

$$Z_q := \int_0^\infty q(x)\,\mathrm{d}x = \sqrt{\frac{\pi}{2}}$$

is the **normalizing constant**.

$\int q(x)dx$

$Z_q = \int_0^\infty q(x)dx \qquad Z_q \cdot \frac{1}{Z_q} = \int f(x) = 1$

Unnormalized Densities

Definition

Let $q(x)$ be a function defined on a region D. Suppose that

- $q(x) > 0$, for $x \in D$ $\int_D q(x)\,dx \neq 1$
- $\displaystyle\int_D q(x)\,\mathrm{d}x = Z_q < \infty$.

Then $q(x)$ is an **unnormalized density** on D. The corresponding normalized density is

$$f(x) = \frac{q(x)}{Z_q}.$$

$A \underbrace{\int q(x)\,dx}_{Z_q} = \int f(x) = 1$

Unnormalized Densities

If $q(x)$ is an unnormalized density, then the density

$$f(x) = \frac{q(x)}{Z_q}$$

is normalized, since

$$\int_D f(x)\,\mathrm{d}x = \int_D \frac{q(x)}{Z_q}\,\mathrm{d}x = \frac{1}{Z_q}\int_D q(x)\,\mathrm{d}x = \frac{1}{Z_q}Z_q = 1.$$

Note that:

$q(x) \rightarrow \int q(x)$

- For any normalized density, there are many unnormalized densities.
- For any unnormalized density, there is only one normalized density.

Example 1: $q(x) = 3e^{-x^2/2}$ for $x \geq 0$

$\int_0^\infty 3e^{-x^2/2} dx \times ? = 1$

Let $q(x)$ be defined as

$$q(x) = 3e^{-x^2/2}, \quad \text{for } x \geq 0.$$

We want to find the Z_q such that $\dfrac{q(x)}{Z_q}$ is a probability density.

Example 1: $q(x) = 3e^{-x^2/2}$ for $x \geq 0$

$\int_0^\infty 3 \cdot e^{-x^2/2}\, dx$

- We recognize that $e^{-x^2/2}$ for $x \geq 0$ is the "core" of the folded normal distribution.
- We previously found $\int_0^\infty e^{-x^2/2}\,\mathrm{d}x = \sqrt{\frac{\pi}{2}}$.
- Then Z_q is the normalizing constant

$$
\begin{aligned}
Z_q &= \int_0^\infty q(x)\,\mathrm{d}x = \int_0^\infty 3e^{-x^2/2}\,\mathrm{d}x \\
&= 3\int_0^\infty e^{-x^2/2}\,\mathrm{d}x \\
&= 3\sqrt{\frac{\pi}{2}}.
\end{aligned}
$$

$q(x) = e^{-x^2/2}$ is unnormalized pdf

$\therefore$ $Z_q = 3\sqrt{\frac{\pi}{2}}$ = normalizing constant

Example 2: $q(x) = e^{-x^2/2}$ for $x \in \mathbb{R}$

Let $q(x)$ be defined as

$\sqrt{\frac{\pi}{2}}$

$$q(x) = e^{-x^2/2}, \quad \text{for } x \in \mathbb{R}.$$

We want to find the Z_q such that $\dfrac{q(x)}{Z_q}$ is a probability density.

Example 2: $q(x) = e^{-x^2/2}$ for $x \in \mathbb{R}$

core

- We recognize that $e^{-x^2/2}$ for $x \in \mathbb{R}$ is the "core" of the standard normal distribution $\mathcal{N}(0,1)$:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \text{for } x \in \mathbb{R}.$$

core

- Since $f(x)$ is a PDF, then

core

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \underline{e^{-x^2/2}}\,\mathrm{d}x = 1.$$

- Then the normalizing constant for $q(x)$ is

$$= \frac{1}{\frac{1}{\sqrt{2\pi}}}$$

$$Z_q = \int_{-\infty}^{\infty} q(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi}.$$

Example 3: $q(x) = e^{-5x}$ for $x \geq 0$

Let $q(x)$ be defined as

$$q(x) = e^{-5x}, \quad \text{for } x \geq 0.$$

We want to find the Z_q such that $\dfrac{q(x)}{Z_q}$ is a probability density.

$Z_q = \int_0^{\infty} e^{-5x}\,dx$ core of exp(5)

Example 3: $q(x) = e^{-5x}$ for $x \geq 0$

- We recognize that e^{-5x} for $x \geq 0$ is the "core" of the exponential distribution $\mathrm{Exp}(\lambda)$:

$$\boxed{f(x) = \lambda e^{-\lambda x},} \quad \text{for } x \geq 0,$$

with $\lambda = 5$.

- Since $f(x)$ is a PDF, then

$$\int_0^\infty f(x)\,\mathrm{d}x = \int_0^\infty 5e^{-5x}\,\mathrm{d}x = 1.$$

- Then the normalizing constant for $q(x)$ is

$$Z_q = \int_0^\infty q(x)\,\mathrm{d}x = \int_0^\infty e^{-5x}\,\mathrm{d}x = \frac{1}{5}. = \frac{1}{5}$$

Gamma and Beta Distributions

Recall:

The PDF of a $\mathbf{Gamma}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ distribution has the form

$$\boxed{f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},} \quad \text{for } x \geq 0,$$

and the PDF of a $\mathbf{Beta}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ distribution has the form

$$\boxed{f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}} \quad \text{for } 0 \leq x \leq 1,$$

where $\boxed{\Gamma(z) = \int_0^\infty x^{z-1} e^{-z} \, \mathrm{d}x.}$

If n is a positive integer, $\Gamma(n) = (n-1)!$.

$\alpha = 4$ $\beta = 3$

Example 4: $q(x) = x^3(1-x)^2$ for $x \in [0,1]$

Beta $(4,3)$

$$\frac{\Gamma(7)}{\Gamma(4)\Gamma(3)} = \frac{6\cdot 5\cdot \cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}{\cancel{3}\cdot\cancel{2}\cdot\cancel{1}\cdot\cancel{2}\cdot 1} = 60$$

$\therefore qz = \frac{1}{60}$

Let $q(x)$ be defined as

$$q(x) = x^3(1-x)^2, \quad \text{for } x \in [0,1].$$

We want to find the Z_q such that $\dfrac{q(x)}{Z_q}$ is a probability density.

Example 4: $q(x) = x^3(1-x)^2$ for $x \in [0,1]$

- We recognize that $x^3(1-x)^2$ for $x \in [0,1]$ is the "core" of the Beta distribution $\mathrm{Beta}(\alpha, \beta)$:

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } 0 \le x \le 1,$$

with

$$\alpha - 1 = 3 \quad \text{and} \quad \beta - 1 = 2,$$

so $\alpha = 4$ and $\beta = 3$.

$\int_0^1 \frac{1}{60} q(x) = 1$

- Since $f(x)$ is a PDF, then

$$\int_0^1 f(x)\,\mathrm{d}x = \int_0^1 \frac{\Gamma(7)}{\Gamma(4)\Gamma(3)} x^3(1-x)^2\,\mathrm{d}x = 1.$$

- Then the normalizing constant for $q(x)$ is

$\frac{1}{\frac{\Gamma(7)}{\Gamma(4)\Gamma(3)}} = \frac{1}{60}$

$$Z_q = \int_0^1 x^3(1-x)^2\,\mathrm{d}x = \frac{\Gamma(4)\Gamma(3)}{\Gamma(7)} = \frac{3!2!}{6!} = \frac{1}{60}.$$

Example 5: $q(x) = x^3 e^{-x/2}$ for $x \geq 0$

$$\boxed{f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},}$$ for $x \geq 0$,

$x^{?} e^{-?} \longrightarrow$ gamma

$q(x) = x^{4-1} e^{-\frac{1}{2}x} \cdot \frac{(\frac{1}{2})^4}{\Gamma(4)} \rightarrow \alpha = 4 \quad \beta = \frac{1}{2}$

$\rightarrow \frac{\frac{1}{16}}{3 \cdot 2 \cdot 1} = \frac{1}{96}$

Let $q(x)$ be defined as

$$q(x) = x^3 e^{-x/2}, \quad \text{for } x \geq 0.$$

We want to find the Z_q such that $\dfrac{q(x)}{Z_q}$ is a probability density.

Example 5: $q(x) = x^3 e^{-x/2}$ for $x \geq 0$

- We recognize that $x^3 e^{-x/2}$ for $x \geq 0$ is the "core" of the gamma distribution $\mathrm{Gamma}(\alpha, \beta)$:

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad \text{for } x \geq 0,$$

with $\alpha = 4$ and $\beta = \dfrac{1}{2}$.

$$\frac{\Gamma(4)}{(\frac{1}{2})^4}$$

- Since $f(x)$ is a PDF, then

$$\int_0^\infty f(x)\,\mathrm{d}x = \int_0^\infty \frac{(\frac{1}{2})^4}{\Gamma(4)} \boxed{x^3 e^{-x/2}\,\mathrm{d}x} = 1.$$

- Then the normalizing constant for $q(x)$ is

$$Z_q = \int_0^\infty x^3 e^{-x/2}\,\mathrm{d}x = \frac{\Gamma(4)}{(\frac{1}{2})^4} = 3! \cdot 2^4 = 96.$$

Outline

1. Unnormalized Densities
   - Example 1: $q(x) = 3e^{-x^2/2}$ for $x \geq 0$
   - Example 2: $q(x) = e^{-x^2/2}$ for $x \in \mathbb{R}$
   - Example 3: $q(x) = e^{-5x}$ for $x \geq 0$
   - Example 4: $q(x) = x^3(1-x)^2$ for $x \in [0, 1]$
   - Example 5: $q(x) = x^3 e^{-x/2}$ for $x \geq 0$

2. Self-Normalized Importance Sampling
   - Example 6: Folded Normal Distribution

Self-Normalized Importance Sampling

- Let $f(x)$ be a normalized density, for $x \in D$, where D is the support of X.

- Let $q(x)$ be an unnormalized density for $f(x)$ with normalizing constant $Z_q = \int_D q(x)\,\mathrm{d}x$, i.e., $f(x) = \dfrac{q(x)}{Z_q}$.

- Suppose we want to estimate

$$E_f[h(X)] = \int_D h(x) f(x)\,\mathrm{d}x = \int_D h(x) \frac{q(x)}{Z_q}\,\mathrm{d}x,$$

  but Z_q is unknown and we are not able to sample from $f(x)$ directly.

- How can we estimate $E_f[h(X)]$ when we only know the unnormalized density $q(x)$?

Self-Normalized Importance Sampling

trail dis $\longrightarrow g(x) \longrightarrow$ find q_g

- We can use a modified version of importance sampling!

$\longrightarrow q(x)$

- Let $g(x) = \dfrac{r(x)}{Z_r}$ be a trial distribution, where $r(x)$ is an unnormalized density for $g(x)$, with

$$Z_r = \int r(x)\,\mathrm{d}x.$$

- The normalizing constant Z_r may be unknown.

trail dis

Self-Normalized Importance Sampling

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

   weight

   unnormalized

   $$w(X^{(i)}) = \frac{q(X^{(i)})}{r(X^{(i)})}, \text{ for } i = 1, 2, \dots, n.$$

2. Estimate $E_f[h(X)]$ by the **self-normalized importance sampling estimator**

   $$\widehat{E_f[h(X)]} = \frac{\sum_{i=1}^{n} w(X^{(i)}) h(X^{(i)})}{\sum_{i=1}^{n} w(X^{(i)})}.$$

Self-Normalized Importance Sampling

Proof (Self-Normalized Importance Sampling, Part 1).

We will show that $\widehat{E_f[h(X)]} \xrightarrow{\text{a.s.}} E_f[h(X)]$.

$g(x) = \frac{r(x)}{Z(r)} \rightarrow \frac{g(x)}{r(x)} = \frac{1}{Z(r)}$

By the Strong Law of Large Numbers:

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} w(X^{(i)}) \quad &\xrightarrow{\text{a.s.}} \quad E_g\left[\frac{q(X)}{r(X)}\right] = \int \frac{q(x)}{r(x)} g(x)\,\mathrm{d}x \\
&= \frac{Z_q}{Z_r}
\end{aligned}
$$

$Z_q \rightarrow \int q(x)\,dx$

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} w(X^{(i)})h(X^{(i)}) \quad &\xrightarrow{\text{a.s.}} \quad E_g\left[\frac{q(X)}{r(X)} h(X)\right] = \int \frac{q(x)}{r(x)} h(x) g(x)\,\mathrm{d}x \\
&= \frac{1}{Z_r}\int q(x)h(x)\,\mathrm{d}x
\end{aligned}
$$

$\frac{g(x)}{r(x)} = \frac{1}{Z_r}$

Self-Normalized Importance Sampling

Proof (Self-Normalized Importance Sampling, Part 2).

Thus

$$\frac{\sum_{i=1}^{n} w(X^{(i)})h(X^{(i)})}{\sum_{i=1}^{n} w(X^{(i)})} \xrightarrow{\text{a.s.}} \frac{1}{Z_q}\int q(x)h(x)\,\mathrm{d}x = \int \frac{q(x)}{Z_q}h(x)\,\mathrm{d}x$$

$$= \int h(x)f(x)\,\mathrm{d}x$$

$$= E_f[h(X)],$$

$$\frac{1}{Zr} \div \frac{Zq}{Zr} = \frac{1}{Zq}$$

as desired. $\square$

$$\frac{E_g\left(\frac{q(x)}{r(x)}h(x)\right)}{E_g\left(\frac{q(x)}{r(x)}\right)} = E_f(h(X))$$

from $g \to f$

Example 6: Folded Normal Distribution

$Z_q = \frac{1}{\sqrt{\frac{\pi}{2}}} = \sqrt{\frac{2}{\pi}}$

- Let $q(x)$ be an unnormalized density for $f(x)$, given by

$$q(x) = e^{-x^2/2}, \quad \text{for } x \geq 0.$$

- We want to use self-normalized importance sampling to estimate

$$E_f(X) = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0^\infty x \frac{q(x)}{Z_q}\,\mathrm{d}x.$$

- We previously found the theoretical $E_f(X) = \sqrt{\frac{2}{\pi}}$.

$E_f(X) = \int \frac{q(x)}{Z_q} h(x)\,dx = \int_0^\infty \frac{e^{-x^2/2}}{Z_q} x\,dx$

$h(x) = x \quad = \int_0^\infty x \cdot \sqrt{\frac{\pi}{2}}\, e^{-x^2/2}\,dx$

Example 6: Folded Normal Distribution

- We previously found $\int_0^\infty e^{-x^2/2}\,\mathrm{d}x = \sqrt{\frac{\pi}{2}}$.

- Consider the unnormalized trial density from $g(x)$

$$r(x) = e^{-2x}, \text{ for } x \geq 0.$$

- We recognize that $r(x)$ is an unnormalized density for $g(x) \sim \mathrm{Exp}(\lambda = 2)$, but we do not need to know the normalizing constant Z_r.

Example 6: Folded Normal Distribution

Self-normalized importance sampling to estimate $E_f(X)$:

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

$$\begin{aligned} w(X^{(i)}) &= \frac{q(X^{(i)})}{r(X^{(i)})} \\ &= \frac{e^{-(X^{(i)})^2/2}}{e^{-2X^{(i)}}} \\ &= \exp\left[-\frac{(X^{(i)})^2}{2} + 2X^{(i)}\right]. \end{aligned}$$

2. Estimate $E_f(X)$ by

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

Example 6: Folded Normal Distribution

R Code to estimate $E_f(X)$ (self-normalized importance sampling):

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

> n <- 10000 # Specify the number of points to generate

> # Generate n points from Exp(lambda = 2)
> X <- rexp(n, rate = 2)   trail

> # Compute importance weights
> W <- exp(-X^2 / 2 + 2 * X)   w(x^i)

> # Compute sum(w(X) * X) / sum(w(X))
> sum(W * X) / sum(W)
[1] 0.800945

> # Theoretical value
> sqrt(2 / pi)
[1] 0.7978846
```