Lecture Notes III – Importance sampling and rejection sampling

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

April 2025

Importance sampling

Rejection sampling

Estimating $\mathbb{E}_p[f(X)]$ revisited

Given a function $f(x)$ and a distribution F known (and its density $p(x) = F'(x)$).
Want

$$\mathbb{E}_p[f(X)] \;\equiv\; \mu_f \;\equiv\; \int_{-\infty}^{\infty} f(x)p(x)dx.$$

(Note: the interval can be any interval $[a, b]$)

Idea 1 Estimate μ_f by sample average $\hat{\mu}_N$

- Algorithm 1
  1. Sample $X_{1:N} \sim p$
  2. $\hat{\mu}_N \;=\; \frac{1}{N}\sum_{i=1}^{N} f(X_i)$

Idea 2 Let $q(x)$ be any other density, with $\boxed{q(x) > 0 \text{ whenever } p(x) > 0}$.
Then,

$$\mu_f \;=\; \int_{-\infty}^{\infty} \underbrace{\frac{f(x)p(x)}{q(x)}}_{\tilde{f}(x)} q(x)dx \;=\; \mathbb{E}_q[\tilde{f}(X)]$$

- Algorithm 2
  1. Sample $X_{1:N} \sim q$
  2. $\hat{\mu}_{N,q} \;=\; \frac{1}{N}\sum_{i=1}^{N} \frac{f(X_i)p(X_i)}{q(X_i)}$

- When could Algorithm 2 be better than Algorithm 1?

The variance of $\hat{\mu}_{N,q}$

$\text{Var}\,\hat{\mu}_{f,q} \;=\; \frac{1}{N}\left(\int \tilde{f}^2(x)q(x)dx - \mu_f^2\right)$

- Only $M \;=\; \int \tilde{f}^2(x)q(x)dx$ depends on q

$$M \;=\; \int \frac{f^2(x)p^2(x)}{q^2(x)}q(x)dx \;=\; \int \frac{f^2(x)p^2(x)}{q(x)}dx$$

- Want q that makes M as small as possible

The variance of $\hat{\mu}_{N,q}$

$\text{Var}\,\hat{\mu}_{f,q} \;=\; \frac{1}{N}\left(\int \tilde{f}^2(x)q(x)dx - \mu_f^2\right)$

- Only $M \;=\; \int \tilde{f}^2(x)q(x)dx$ depends on q

$$M \;=\; \int \frac{f^2(x)p^2(x)}{q^2(x)}q(x)dx \;=\; \int \frac{f^2(x)p^2(x)}{q(x)}dx$$

- Want q that makes M as small as possible
- $\boxed{q^*(x) \propto f(x)p(x)}$
- Let's see why. First we need to normalize q

$$q^*(x) \;=\; \frac{f(x)p(x)}{Z} \quad \text{with } Z \;=\; \int f(x)p(x)dx \;=\; \mu_f \tag{1}$$

- Now calculate M for q^*

$$M^* \;=\; \int \frac{f^2(x)p^2(x)}{\frac{f(x)p(x)}{\mu_f}}dx \;=\; \mu_f \underbrace{\int f(x)p(x)dx}_{\mu_f} \;=\; \mu_f^2 \tag{2}$$

- Finally, the variance $\text{Var}\,\hat{\mu}_{f,q} \;=\; \frac{1}{N}(M^* - \mu_f^2) \;=\; 0$!

- Theoretically, $N = 1$ sample is enough!
- But, q^* depends on the true μ_f that we are trying to estimate!

Importance sampling in practice

Theory $\boxed{q^*(x) \propto f(x)p(x)}$

- In practice, we want a distribution q so that
  - $q \approx q^*$
  - q is easy to sample from
  - $q(x)$ is easy to calculate for any x
- Rule of thumb 1: q should have modes where $f(x)p(x)$ is large
- Rule of thumb 2: avoid $q(x) \ll p(x)f(x)$ (tails of q should not decrease too fast!)
  When $f(x)p(x)$ is far from uniform, even a very rough approximation can reduce variance by orders of magnitude.

Links to some examples

Monte Carlo course notes, Ch. 4, by Tom Kennedy, U of Arizona Monte Carlo course notes, Ch. 6, by Tom Kennedy, U of Arizona
Monte Carlo course notes by Eric Anderson, U. C. Berkeley

Examples

Rejection sampling

- Sampling from $F(x)$ when we only know $p(x) = F'(x)$
- (or, when X is multidimensional)

Given $p(x)$ a density
Want An algorithm to get samples from $p(x)$

Rejection sampling

- Sampling from $F(x)$ when we only know $p(x) = F'(x)$
- (or, when X is multidimensional)

Given $p(x)$ a density

Want An algorithm to get samples from $p(x)$

- Will use $q(x)$ another distribution
- Let $\boxed{M = \sup_x \frac{p(x)}{q(x)}}$

  Note that $p(x) > 0$ implies $q(x) > 0$

- Rejection Sampling Algorithm
  1. sample $Y \sim q$
  2. sample $U \sim \mathrm{unif}[0, 1]$
  3. if $U < \frac{1}{M} \frac{p(Y)}{q(Y)}$ then output $X = Y$

     else go to step 1.

Why does this work?

- Denote F_{RS} the CDF of the samples from the Rejection Sampling algorithm.

$$F_{RS}(x) \;=\; Pr[X \leq x \,|\, Y \text{ accepted }] \;=\; \frac{Pr\left[X \leq x,\; U < \frac{1}{M}\frac{p(Y)}{q(Y)}\right]}{Pr\left[U < \frac{1}{M}\frac{p(Y)}{q(Y)}\right]} \tag{3}$$

Pr is the probability under the joint distribution of U and Y

- the denominator

$$Pr\left[U < \frac{1}{M}\frac{p(Y)}{q(Y)}\right] \;=\; \int \left(Pr_{U\sim \mathrm{unif}[0,1]}\left[U < \frac{1}{M}\frac{p(y)}{q(y)}\right]\right) q(y)dy \tag{4}$$

$$= \int \frac{1}{M}\frac{p(y)}{q(y)} q(y)dy \;=\; \frac{1}{M}\int p(y)dy \;=\; \frac{1}{M} \tag{5}$$

- now the numerator

$$Pr\left[X \leq x,\; U < \frac{1}{M}\frac{p(Y)}{q(Y)}\right] \;=\; \int \left(Pr_{U\sim \mathrm{unif}[0,1]}\left[U < \frac{1}{M}\frac{p(y)}{q(y)}\right]\right) I(y \leq x)q(y)dy \tag{6}$$

$$= \int \frac{1}{M}\frac{p(y)}{q(y)} I(y \leq x)q(y)dy \tag{7}$$

$$= \frac{1}{M}\int p(y)I(y \leq x)dy \;=\; \frac{1}{M}\int_{-\infty}^{x} p(y)dy \;=\; \frac{1}{M}F(x)$$

- Therefore $F_{RS}(x) = F(x)$ for all x QED

Practical Rejection Sampling. What is a good q?

- Intuitively, q should be as close to p as possible
- Acceptance probability $Pr\left[U < \frac{1}{M}\frac{p(Y)}{q(Y)}\right] = \frac{1}{M}$
- Therefore, a good q will have $M = \sup_x \frac{p(x)}{q(x)}$ small
  (Note that $M \geq 1$)

Examples