Winter 2021 Math 126
Topics in Applied Mathematics
Data-driven Uncertainty Quantification

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 8: Importance Sampling

8.1 Importance Sampling

- Importance sampling is a sampling method with a reduced variance.
- Assume that we are interested in the evaluation of the following integral

$$E_p[f] = \int f(x)p(x)dx$$

  where $p(x)$ is a probability density.
- Let $g(x)$ is another probability density with **a support containing the support of** $f(x)$ and it is easy to draw a sample from $g(x)$.
- Importance sampling use the following idea of a change of variables

$$\int f(x)p(x)dx = \int \frac{f(x)p(x)}{g(x)} g(x)dx$$

8.1 Importance Sampling

If $\{x_i\}$ is IID from $g(x)$, the integral is approximated by

$$E_p[f] \approx \frac{1}{n}\sum_i^n \frac{f(x_i)p(x_i)}{g(x_i)} = \frac{1}{n}\sum_i^n w_i f(x_i)$$

where $w_i = \frac{p(x_i)}{g(x_i)}$ is the weight of x_i. **Note** that there is no $g(x_i)$ in the numerator.

8.1 Importance Sampling

Example: Small tail probabilities: rare events.

- Among many other applications of the importance sampling, a small tail probability is a good example related to rare events.
- Let we are interested in the following probability using a Monte Carlo method

$$\mu(Z > 4.5) = \int_{-\infty}^{4.5} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

  where Z is the standard normal random variable.
- As we know the analytic form of the density that is easy to integrate, the probability we are looking for is 3.39×10^{-6}, a really small probability.

8.1 Importance Sampling

Example: Small tail probabilities: rare events.

- A programming tip: a fast calculation for counting (code example in Matlab/Python)
- What is your expected sample size to calculate the small probability?
- The small probability is of order 10^{-6}. This means that we can expect only a few values larger than 4.5 out of million sample values.
- Try in Matlab/Python

8.1 Importance Sampling

Example: Small tail probabilities: rare events.

- A programming tip: a fast calculation for counting (code example in Matlab/Python)
- What is your expected sample size to calculate the small probability?
- The small probability is of order 10^{-6}. This means that we can expect only a few values larger than 4.5 out of million sample values.
- Try in Matlab/Python
- Importance sampling using a fat tail distribution (we used the Cauchy density) increases the accuracy (or decrease the sample size to estimate the small density).

8.2 Finite Variance

- The variance of the importance sampling estimator is

$$E_g\left[\frac{f^2(x)p^2(x)}{g^2(x)}\right] - E_g\left[\frac{f(x)p(x)}{g(x)}\right]^2.$$

- To have a finite variance, we need to have

$$\int \frac{f^2(x)p^2(x)}{g(x)}dx < \infty.$$

- That is, the ratio $p(x)/g(x)$ must be bounded.
- This implies that the tails of $g(x)$ must be fatter than those of $p(x)$.

8.2 Finite Variance

Theorem. The choice of $g(x)$ that minimizes the variance of the importance sampling estimator is

$$g^*(x) = \frac{|f(x)|p(x)}{\int |f(s)|p(s)ds}$$

Idea of Proof.

- The variance of the importance sampling estimator is given by

  $$E_g\left[\frac{f^2(x)p^2(x)}{g^2(x)}\right] - E_g\left[\frac{f(x)p(x)}{g(x)}\right]^2$$

  where the second term is independent of $g(x)$.
- From the Jensen's inequality, the lower bound of the first term is

  $$E_g\left[\frac{f^2(x)p^2(x)}{g^2(x)}\right] \geq \left(E_g\left[\frac{|f(x)|p(x)}{g(x)}\right]\right)^2$$
- The lower bound is obtained by $g^*(x) = \frac{|f(x)|p(x)}{\int |f(s)|p(s)ds}$.

8.2 Finite Variance

- An alternative importance sampling estimator with increased stability is
$$\frac{\sum_i^n \frac{f(x_i)p(x_i)}{g(x_i)}}{\sum_i^n \frac{p(x_i)}{g(x_i)}}$$
instead of the standard importance sampling estimator
$$\frac{1}{n}\sum_i^n \frac{f(x_i)p(x_i)}{g(x_i)}.$$
- This is motivated by the following convergence
$$\frac{1}{n}\sum_i^n \frac{p(x_i)}{g(x_i)} \to 1 \quad \text{as } n \to \infty$$
- This estimator is **biased** but the bias is small.

8.2 Finite Variance

Also, the following approach is preferred to achieve a stable importance density $g(x)$ with fat tails

$$g(x) = \rho h(x) + (1-\rho) l(x), \quad 0 < \rho < 1$$

where $h(x)$ is close to $p(x)$ and $l(x)$ has fat tails.

8.3 Sequential Importance Sampling

Example: Target tracking (Gordon et al. 1993). We consider a tracking problem where an object (an airplane, a pedestrian or a ship) is observed through some noisy measurement of its angular position Z_t at time t. Of interests are the position (X_t, Y_t) of the object in the plane and its speed $(\dot{X}_t, \dot{Y}_t)$. The model is then discretized as $\dot{X}_t = X_{t+1} - X_t$, $\dot{Y}_t = Y_{t+1} - Y_t$, and

$$\dot{X}_t = \dot{X}_{t-1} + \tau\epsilon_t^x$$

$$\dot{Y}_t = \dot{Y}_{t-1} + \tau\epsilon_t^y$$

$$Z_t = \arctan(Y_t/X_t) + \eta\epsilon_t^z,$$

where $\epsilon_t^x, \epsilon_t^y$, and ϵ_t^z are iid $N(0,1)$ random variables. We are interested in $p_t(\theta_t|z_{1:t})$ where $\theta_t = (\tau, \eta, \dot{X}_t, \dot{Y}_t)$ and $z_{1:t}$ denotes the vector $(z_1, z_2, ..., z_t)$.

8.3 Sequential Importance Sampling

- The importance sampling is useful when a sequence of target distributions $p_t(x)$ (where t is an index for time) is available.
- Let observations v_t is available at times $t = m\Delta t$ where $m \in \mathbb{R}$ and Δt is a time interval.
- Then we have a new posterior density $p_t(x)$ using the new observation v_t

$$p_t(x) = p(x(t)|v_t) \sim p(x(t))p(v_t|x(t))$$

8.3 Sequential Importance Sampling

- Drawing a sample from the posterior at each step is costly.
- Importance sampling is ideally suited for this problem in that the densities p_t and p_{t+1} are defined in the same space.
- That is, use the prior density as your importance density for the posterior density.
- The weight is given by

$$w(t) = w(t-1)f_t(x)/g_t(x)$$

where $f_t(x)$ is the posterior and $g_t(x)$ is the prior.

8.3 Sequential Importance Sampling

Weight Degeneracy.

- $$w(t) = w(t-1)\frac{f_t(x)}{g_t(x)} = w(t-2)\frac{f_t(x)f_{t-1}(x)}{g_t(x)g_{t-1}(x)} \cdots = w(0)\Pi\frac{f_{t-i}(x)}{g_{t-i}(x)}.$$

- Equivalently,

$$w(t) = w(0)\exp\left(\sum \ln(f_{t-i}(x)/g_{t-i}(x))\right).$$

- In the special case where g_{t-i} and f_{t-i} are both independent of time, approximately we have

$$w(t) \sim \exp\left(-tE_g[\ln g(x)/f(x)]\right).$$

- Thus as $t \to \infty$, w degenerates to 0

$$w(t) \to 0.$$

Homework

1. For the normal-Cauchy Bayes estimator

$$\delta(x) = \frac{\int_{-\infty}^{\infty} \frac{\theta}{1+\theta^2} e^{-(x-\theta)^2/2} d\theta}{\int_{-\infty}^{\infty} \frac{1}{1+\theta^2} e^{-(x-\theta)^2/2} d\theta},$$

   which is the posterior mean of θ with a Cauchy prior density and a normal likelihood,

   1.1 Use Monte Carlo integration to calculate the integral.

   1.2 What is the standard error with a sample size $n =$ (a) 100, (b) 1000 and (c) 10000.

2. For a Gaussian random variable X with a mean 0 and a variance σ^2, prove that

$$E[e^{-X^2}] = \frac{1}{\sqrt{2\sigma^2 + 1}}.$$

Homework

3 Make a one-paragraph summary of **particle filtering** (max 5 sentences).

Homework