MAT 653: Statistical Simulation

Instructor: Dr. Wei Li

2023-09-22

Back **Updated:** 2023-09-22 Statistical Simulation, Wei Li

Classical Monte Carlo Integration

The general problem is to evaluate

$$E_f[h(X)] = \int_{\mathcal{X}} h(x)f(x)\,dx,$$

where h is some known function, X is r.v. $\sim f$ and $\mathcal{X} = supp(f)$. In the following discussion, the underlying $\mathcal{X}$ is implicitly assumed.

The principle of Monte Carlo method for evaluating $E_f[h(X)]$ is to generate an i.i.d. random sample sequence $(X_1, \cdots, X_n)$ from f and propose to estimate it by empirical average

$$\overline{h}_n = \frac{1}{n}\sum_{j=1}^{n} h(X_j)$$

this is called the **Monte Carlo estimator** for $E_f[h(X)]$. This works because by Law of Large numbers, $\overline{h}_n \to E_f[h(X)]$.

Properties of estimator $\overline{h}_n$

1. $E_f(\overline{h}_n) = \mathbb{E}_f[h(X)]$

Proof: $\mathbb{E}_f(\overline{h}_n) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_f[h(X_j)] = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}_f[h(X)] = E_f[h(X)]$

2. $\mathrm{Var}_f(\overline{h}_n) = \dfrac{1}{n}\mathrm{Var}_f[h(X)]$

Proof:

$$\begin{aligned}\mathbb{V}ar_f(\overline{h}_n) &= \mathbb{V}ar_f\Big[\frac{1}{n}\sum_{j=1}^{n} h(X_j)\Big] = \frac{1}{n^2}\mathbb{V}ar_f\Big[\sum_{j=1}^{n} h(X_j)\Big] \\ &\overset{(*)}{=} \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{V}ar_f[h(X_j)] = \frac{1}{n^2}\sum_{j=1}^{n}\mathbb{V}ar_f[h(X)] = \frac{1}{n}\mathbb{V}ar_f[h(X)]\end{aligned}$$

where $(*)$ holds since $X_j \overset{iid}{\sim} f$, $Cov(X_i, X_j) = 0 \quad (i \neq j)$

And $\mathrm{Var}_f(\overline{h}_n)$ can also be estimated from $(X_1, \cdots, X_n)$ by $\displaystyle\frac{1}{n^2}\sum_{j=1}^{n}\big[h(X_j) - \frac{1}{n}\sum_{i=1}^{n} h(X_i)\big]^2$.

Proof:

$$\begin{aligned}
\mathbb{V}ar_f(\overline{h}_n) &= \frac{1}{n}\mathbb{V}ar_f[h(X)] \\
&= \frac{1}{n}\mathbb{E}_f\{[h(X)-\mathbb{E}_f(h(X))]^2\} \quad (\textit{where } \mathbb{E}_f(h(X)) \textit{can be estimated by } \overline{h}_n) \\
&\leftarrow \frac{1}{n}\mathbb{E}_f\{[h(X)-\overline{h}_n]^2\} \qquad (\textit{now estimate } \mathbb{E}_f[h(X)-\overline{h}_n]^2) \\
&\leftarrow \frac{1}{n^2}\sum_{j=1}^{n}\left[h(X_j)-\overline{h}_n\right]^2 \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\left[h(X_j)-\frac{1}{n}\sum_{i=1}^{n}h(X_i)\right]^2
\end{aligned}$$

Example: For $\mathcal{N}(0,1)$, want to estimate c.d.f $\phi(t)=\displaystyle\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}e^{-x}\,dx$.

Say $f(x)=\frac{1}{\sqrt{2\pi}}e^{-x}$, we can rewrite $\phi(t)=\displaystyle\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}e^{-x}\,dx=\int_{-\infty}^{\infty}\mathbf{1}_{(-\infty,t)}(x)\,f(x)\,dx$.

Therefore, we can generate a sample of size n, $\{X_i\}_{i=1}^n \overset{iid}{\sim} f$, then use $\overline{h}_n=\displaystyle\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{(-\infty,t)}(X_i)$ to estimate $\phi(t)$.

Example (Drawback of classical M.C. integration): If $Z\sim\mathcal{N}(0,1)$, and we are asked to evaluate $P(Z>4.5)$.

$P(Z>4.5)=\mathbb{E}_f[\mathbf{1}_{(4.5,\infty)}(Z)]$

Even though we can use classical M.C. integration: generate a sample of size n, $\{Z_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$, then use $\overline{h}_n=\displaystyle\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{(4.5,\infty)}(Z_i)$ to estimate $P(Z>4.5)$. But actually, simulating $\{Z_i\}_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$ probably only produce a hit $Z_i\in(4.5,\infty)$ once in large amount of iterations. However, using importance sampling can solve this issue.

Importance Sampling

Given another p.d.f g s.t. $supp(g)\supseteq supp(f)$.

$$\mathbb{E}_f[h(X)]=\int h(x)f(x)\,dx=\int h(x)\frac{f(x)}{g(x)}g(x)\,dx=\mathbb{E}_g\left[h(X)\frac{f(X)}{g(X)}\right]$$

Now, we can generate a sample of size n, $\{X_i\}_{i=1}^n \overset{iid}{\sim} g$, then use $\overline{h}_n=\displaystyle\frac{1}{n}\sum_{i=1}^{n}h(X_i)\frac{f(X_i)}{g(X_i)}$ to estimate $\mathbb{E}_f[h(X)]=\int h(x)f(x)\,dx$.

Where g is called **importance function**, $\displaystyle\frac{f(X_i)}{g(X_i)}$ is called **importance weight** for X_i and $\left(X_i,\displaystyle\frac{f(X_i)}{g(X_i)}\right)$ is called **importance sample**. $\displaystyle\frac{1}{n}\sum_{i=1}^{n}h(X_i)\frac{f(X_i)}{g(X_i)}$ is called **importance sampling estimator** for $\mathbb{E}_f[h(X)]$.

Example (continue the previous example): recall that $Z\sim\mathcal{N}(0,1)$, and we are asked to evaluate $P(Z>4.5)$.

Now we can take g to be density funCtion of $\mathbf{exp}(1)$ (right) truncated at 4.5:

$$\begin{aligned} g(y) = P\left(exp(1) = y | exp(1) > 4.5\right) &= \frac{P\left(exp(1) = y, exp(1) > 4.5\right)}{P\left(exp(1) > 4.5\right)} = \frac{P\left(exp(1) = y\right)}{P\left(exp(1) > 4.5\right)} \\ &= \frac{e^{-y}}{1 - \int_0^{4.5} e^{-x}\,dx} = \frac{e^{-y}}{1 - (1 - e^{-4.5})} = e^{-(y-4.5)} \quad (y > 4.5) \end{aligned}$$

Now we can generate a sample of size n, $\{Y_i\}_{i=1}^n \overset{iid}{\sim} g$, recall $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x}$, then the importance sampling estimator for $P(Z > 4.5) = \mathbb{E}_f[\mathbf{1}_{(4.5,\infty)}(Z)]$ becomes

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(4.5,\infty)}(Y_i)\frac{f(Y_i)}{g(Y_i)} = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{(4.5,\infty)}(Y_i)\frac{e^{-\frac{Y_i^2}{2}+Y_i-4.5}}{\sqrt{2\pi}} = \frac{1}{n}\sum_{i=1}^n \frac{e^{-\frac{Y_i^2}{2}+Y_i-4.5}}{\sqrt{2\pi}}$$

where here $\mathbf{1}_{(4.5,\infty)}(Y_i) = 1$ since $\{Y_i\}_{i=1}^n \overset{iid}{\sim} g$.

Remark: we need $supp(g) \supseteq supp(f)$ to make sure the value of importance weight $\frac{f(X_i)}{g(X_i)}$ is meaningful. Actually here having a weaker condition $supp(g) \supseteq supp(h \times f)$ suffices since we have the fraction $\frac{h(X_i)f(X_i)}{g(X_i)}$.

Self-normalized version of importance sampling

Again, we want to evaluate $\mathbb{E}_f[h(X)]$. But now $f \propto \tilde{f}$, $g \propto \tilde{g}$. Say $f = c_o f_0$ where f_0 is the unnormalized p.d.f, c_0 is the normalizing constant. And $g = c_1 g_1$, where g_0 is the unnormalized p.d.f, c_1 is the normalizing constant.

Now, suppose we know f_0 and g_0, but possibly not c_0 and c_1. And suppose we know how to generate random sample from g.

$$\begin{aligned} \mathbb{E}_f[h(X)] &= \int h(x)f(x)\,dx \\ &= \frac{\int \frac{h(x)f(x)}{g(x)} g(x)\,dx}{\int \frac{f(x)}{g(x)} g(x)\,dx} \\ &= \frac{\int \frac{f_0(x)}{g_0(x)} h(x)g(x)\,dx}{\int \frac{f_0(x)}{g_0(x)} g(x)\,dx} \quad (now\ \ w(x) := \frac{f_0(x)}{g_0(x)}) \\ &= \frac{\int w(x)h(x)g(x)\,dx}{\int w(x)g(x)\,dx} = \frac{\mathbb{E}_g[w(x)h(x)]}{\mathbb{E}_g[w(x)]} \end{aligned}$$

Then we can generate a sample of size n, $\{X_i\}_{i=1}^n \overset{iid}{\sim} g$, and use $\hat{\mu} = \frac{\frac{1}{n}\sum_{i=1}^n w(X_i)h(X_i)}{\frac{1}{n}\sum_{i=1}^n w(X_i)}$ to estimate $\mu = \mathbb{E}_f[h(X)]$. Here $\hat{\mu}$ is called **Self-normalized importance sampling estimator** for μ.

The reason why $\hat{\mu} \to \mu$.

$$\hat{\mu} = \frac{\frac{1}{n}\sum_{i=1}^n \frac{f_0(X_i)}{g_0(X_i)} h(X_i)}{\frac{1}{n}\sum_{i=1}^n \frac{f_0(X_i)}{g_0(X_i)}} = \frac{\frac{1}{n}\sum_{i=1}^n \frac{f(X_i)}{g(X_i)} h(X_i)}{\frac{1}{n}\sum_{i=1}^n \frac{f(X_i)}{g(X_i)}} \overset{LLN}{\to} \frac{\mathbb{E}_g\left[\frac{f(X)}{g(X)} h(X)\right]}{\mathbb{E}_g\left[\frac{f(X)}{g(X)}\right]}$$

The last term equals to $\int \frac{h(x)f(x)}{g(x)} g(x)\,dx / \int \frac{f(x)}{g(x)} g(x)\,dx = \int h(x)f(x)\,dx = \mathbb{E}_f[h(X)] = \mu$.

Self-normalize Importance Sampling

Target: $E_f(h(x))$

$f \propto f_0$, g is importance pdf, and $supp(f) \subset supp(g)$.

Self-normalized importance sampling $\Longrightarrow$ $\left(x_i, \frac{f_0(x_i)}{g_0(x_i)}\right)$ importance sample. It turns out this can be recycled by multinomial resampling into a sample that is from f.

Step 1: Sample $x_i \sim g$, obtain $\left(x_i, \frac{f_0(x_i)}{g_0(x_i)}\right)_{i=1}^n$

Step 2: importance weights $w_i = \frac{f_0(x_i)}{g_0(x_i)}$. Let $\hat{w}_i = \frac{\frac{1}{n} w_i}{\frac{1}{n}\sum_j^n w_j}$. Notice that $\hat{w}_i \in (0,1), \sum \hat{w}_i = 1$

Step 3: draw a random sample of size m with replacement from $x_1, \cdots, x_n$ with weighted probabilities by $\hat{w}_1, \cdots, \hat{w}_n$: that is for $k = 1, 2, \ldots, m$,

$$X_k^* = \begin{cases} x_1 & \text{with prob} = \hat{w}_1 \\ x_2 & \text{with prob} = \hat{w}_2 \\ \vdots & \\ x_n & \text{with prob} = \hat{w}_n \end{cases}$$

It turns out that $(X_1^*, X_2^*, \cdots, X_m^*)$ is a random sample from f.

Remark: for this sampling to work, the target sample size m should be no more than 10% of the original sample size n.

Back **Updated:** 2023-09-22 Statistical Simulation, Wei Li