Computational Methods (16:540:540)
Homework 1: Solution

1. The random variable X follows the uniform distribution, $X \sim U(0, 10)$, with the p.d.f.

$$f(x) = \begin{cases} \dfrac{1}{10}, & 0 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

Define $h(x) = 10e^{-2|x-5|}$. Evaluate $E_f[h(x)]$ using

(i) Monte Carlo integration approach.

(ii) Importance sampling approach. Use $N(5, 1)$ as a candidate probability distribution, i.e., $g(x)$

(iii) Compare the answers of (i) and (ii) for different sample sizes.

Solution:

$$\begin{aligned} E_f[h(x)] &= \int_0^{10} h(x) \cdot f(x) dx \\ &= \int_0^{10} 10e^{-2|x-5|} \cdot \frac{1}{10} dx \\ &= \int_0^{10} e^{-2|x-5|} dx \end{aligned}$$

Monte Carlo integration approach and Importance Sampling approach can be used to approximate this integaration.

(i) In Monte Carlo integration approach,
   1, gerneate n random samples $x_1, x_2, ..., x_n$ from $U(0, 10)$.
   2, evaluate $\frac{1}{n} \sum_{i=1}^{n} 10e^{-2|x_i-5|}$.

(ii) By introducing a new variable $g(x) = \frac{1}{\sqrt{2\pi}\times 1} e^{-\frac{(x-5)^2}{2\times 1^2}}$. We have,

$$
\begin{aligned}
E_f[h(x)] &= \int_0^{10} h(x) \cdot f(x) dx \\
&= \int_0^{10} h(x) \cdot \frac{f(x)}{g(x)} \cdot g(x) dx \\
&= \int_0^{10} 10e^{-2|x-5|} \cdot \frac{\frac{1}{10}}{\frac{1}{\sqrt{2\pi}\times 1} e^{-\frac{(x-5)^2}{2\times 1^2}}} \cdot g(x) dx \\
&= \int_0^{10} \frac{\sqrt{2\pi} \times 1 \cdot e^{-2|x-5|}}{e^{-\frac{(x-5)^2}{2\times 1^2}}} \cdot g(x) dx \\
&= E_g[\sqrt{2\pi} \cdot e^{-2|x-5|+\frac{1}{2}(x-5)^2}]
\end{aligned}
$$

Importance Sampling approach,
1, gerneate n random samples $x_1, x_2, ..., x_n$ from $N(5, 1^2)$.
2, evaluate $\frac{1}{n} \sum_{i=1}^{n} \sqrt{2\pi} \cdot e^{-2|x_i-5|+\frac{1}{2}(x_i-5)^2}$.

(iii) The true value of the $E_f[h(x)]$ is

$$
\begin{aligned}
E_f[h(x)] &= \int_0^{10} e^{-2|x-5|} dx \\
&= \int_0^{5} e^{-2(-x+5)} dx + \int_5^{10} e^{-2(x-5)} dx \\
&= \frac{e^{10}-1}{2e^{10}} + \frac{e^{10}-1}{2e^{10}} \\
&= \frac{e^{10}-1}{e^{10}} \approx 0.9999546
\end{aligned}
$$

Sample size n	10	100	200	500	1000	10000	100000	1000000
	Monte Carlo integration approach							
Difference	0.491515	0.183945	0.122688	0.062519	0.046694	0.016578	0.005382	0.001901
	Importance Sampling approach							
Difference	0.151861	0.052294	0.03654	0.022756	0.019084	0.004842	0.001103	0.000623

The table above shows the average percentage differences between the methods and the true value in 25 experiments. According to the table, when the sample size is increasing, the percentage differences of both methods are decreasing. And the estimation accuracy of the Importance sampling approach is about three times better than the Monte Carlo integration approach. From figure 3 and figure, we could notice that the estimated value from the importance sampling approach is more "concentrate" than the result from the Monte Carlo integration approach. The smaller variance indicates the Importance Sampling approach provides a more stable result.

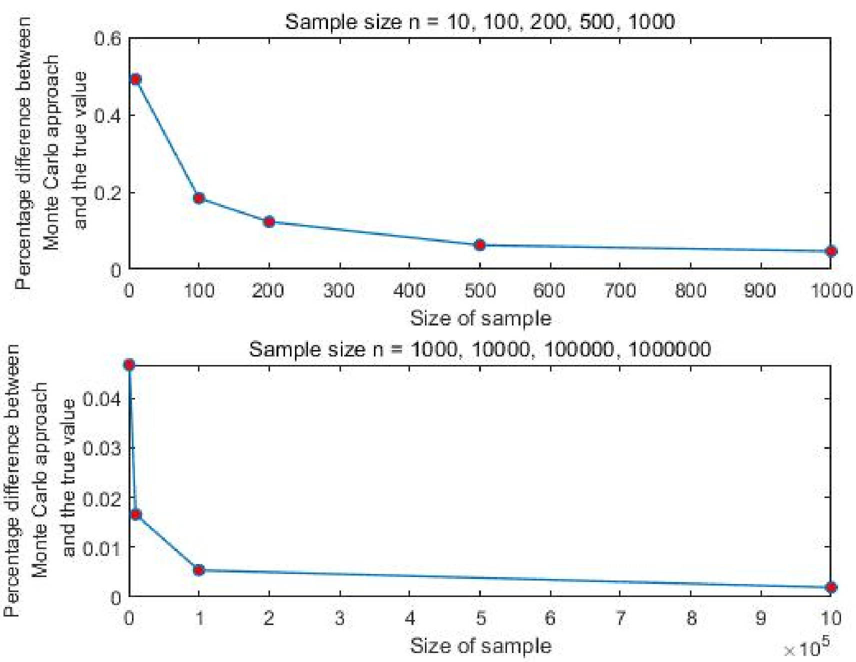

Figure 1: The error rate of the Monte Carlo approach with different sample sizes.

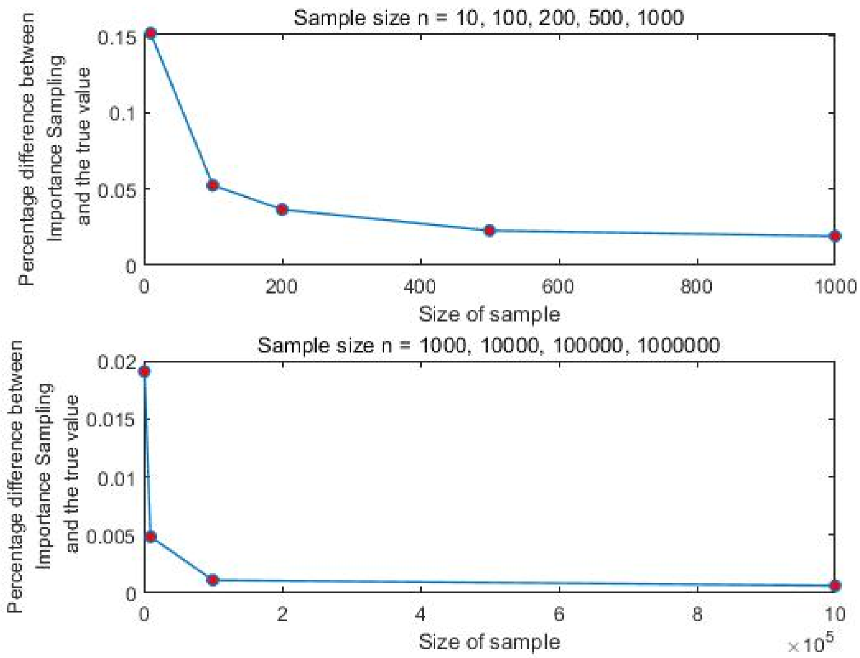

Figure 2: The error rate of the Importance Sampling approach with different sample sizes.

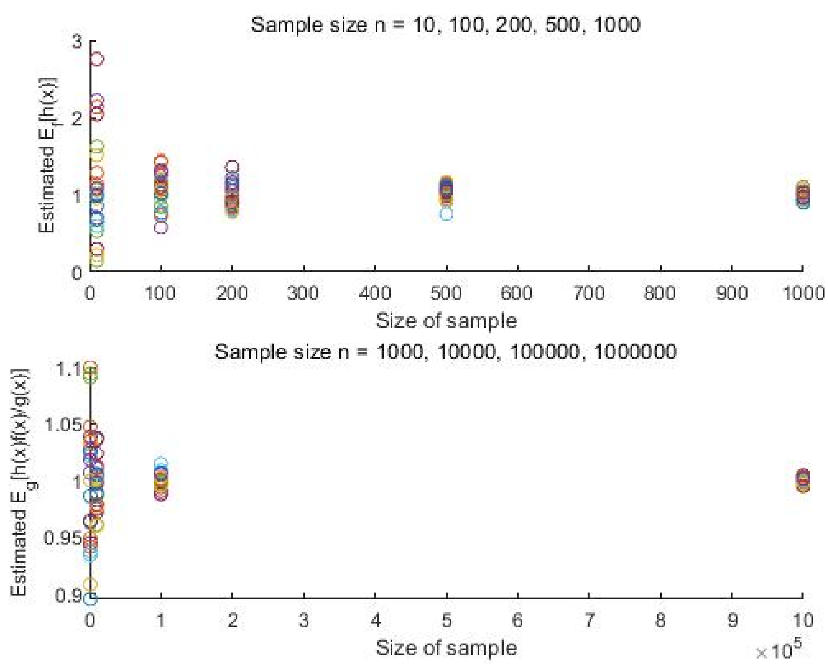

Figure 3: The value of $\hat{E}_f[h(x)]$ from the Monte Carlo approach in 25 experiments with different sample sizes.

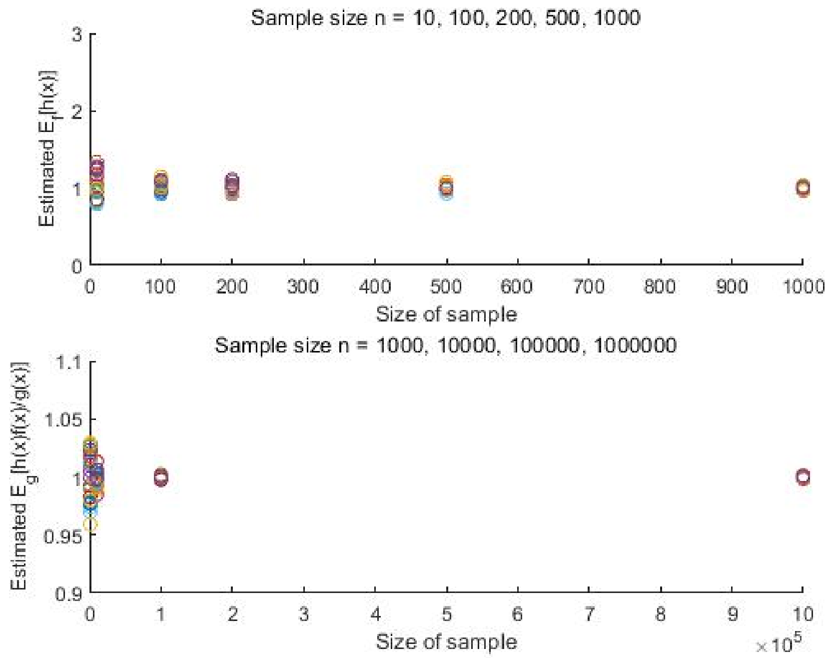

Figure 4: The value of $\hat{E}_g[h(x) \cdot \frac{f(x)}{g(x)}]$ from the Importance Sampling approach in 25 experiments with different sample sizes.

2. Given a p.d.f. of a double exponential distribution for a random variable X,

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < +\infty.$$

Calculate the m_{th} moment of X, where $m = 2, 3, 4, 5$ using the importance sampling approach.
Note that the m_{th} moment for X is defined by

$$E_f[X^m] = \int_{-\infty}^{+\infty} x^m f(x)dx, \quad m = 2, 3, 4, 5.$$

Use $N(0, 2^2)$ as a candidate probability distribution, i.e., $g(x)$. For sample sizes of 100, 1000, and 10,000, report the sample mean and variance of $h(X)$, i.e., $E_f[X^m]$.

Solution: Let $h(x) = X^m$ and $g(x) = \frac{1}{\sqrt{2\pi}\times 2}e^{-\frac{(x-0)^2}{2\times 2^2}}$. We have,

$$\begin{aligned}
E_f[h(x)] &= \int_{-\infty}^{+\infty} h(x) \cdot f(x)dx \\
&= \int_{-\infty}^{+\infty} h(x) \cdot \frac{f(x)}{g(x)} \cdot g(x)dx \\
&= \int_{-\infty}^{+\infty} X^m \cdot \frac{\frac{1}{2}e^{-|x|}}{\frac{1}{\sqrt{2\pi}\times 2}e^{-\frac{(x-0)^2}{2\times 2^2}}} \cdot g(x)dx \\
&= \int_{-\infty}^{+\infty} X^m \cdot \frac{\sqrt{2\pi} \cdot e^{-|x|}}{e^{-\frac{(x-0)^2}{2\times 2^2}}} \cdot g(x)dx \\
&= \sqrt{2\pi} \cdot E_g[X^m \cdot e^{-|x|+\frac{1}{8}x^2}]
\end{aligned}$$

Importance Sampling approach,
1, gerneate n random samples $x_1, x_2, ..., x_n$ from $N(0, 2^2)$.
2, evaluate $\frac{\sqrt{2\pi}}{n} \cdot [x_i^m \cdot e^{-|x_i|+\frac{1}{8}x_i^2}], m = 2, 3, 4, 5.$

When $m = 2$, the true value of the $E_f[X^2]$ is:

$$\begin{aligned}
E_f[X^2] &= \int_{-\infty}^{\infty} x^2 \frac{1}{2} e^{-|x|} dx \\
&= \int_{-\infty}^{0} x^2 \frac{1}{2} e^{-(-x)} dx + \int_{0}^{\infty} x^2 \frac{1}{2} e^{-x} dx \\
&= 1 + \int_{0}^{\infty} x^2 \frac{1}{2} e^{-x} dx \\
&= 1 + 1 = 2.
\end{aligned}$$

When $m = 3$,$E_f[X^3] = 0$, when $m = 4$,$E_f[X^4] = 24$, when $m = 5$,$E_f[X^5] = 0$.
When taking 10000000 sample, the Importance sampling approach provide the value of $E_f[X^m], m = 1, 2, 3, 4, 5$ as followed.

m	1	2	3	4	5
$E_f[X^m]$	0.0043	2.0019	-0.0053	23.6574	-0.2857

The variance can be calculated by

$$V(x) = E(x^2) - E^2(x).$$

n	100	1000	10000	100000	100000
sample mean of $h(X)$	0.009118	-0.01131	-0.0026	-0.00193	-0.00028
sample variance of $h(X)$	1.938731	1.973768	2.03855	2.014994	2.000292