Mandatory2

Tannaz Rafiei

2024-10-21

Problem 1:

part b

In this part, we are tasked with plotting the function $f(x)$ for values of $x \in [-5, 5]$ and explaining how we can use **antithetic random variables** to estimate the integral.

Function $f(x)$: The function $f(x)$ is derived from the integrand in the given expression for the integral I. The integrand is:

$$f(x) = \frac{e^{-x}}{(1 + x^2)^2}$$

This function combines an exponential decay term e^{-x} with a rational function $\frac{1}{(1+x^2)^2}$. The term e^{-x} ensures that as x increases, the function decreases exponentially. The denominator ensures that the function's magnitude is reduced as x grows larger, further emphasizing the decay of the function for large x.

Why Antithetic Random Variables? The goal of **Monte Carlo integration** is to approximate an integral by taking random samples from a probability distribution and computing the average of the function values at these sampled points. Typically, independent random variables are used for sampling. However, we can reduce the **variance** of our Monte Carlo estimator and improve the accuracy of the estimate by using **antithetic random variables**.

Antithetic Random Variables: Antithetic random variables are pairs of random variables that are negatively correlated. When using them for Monte Carlo integration, the idea is to generate a sequence of random variables $X_1, X_2, \ldots, X_n$, and for each X_i, generate its antithetic counterpart $1 - X_i$. The key advantage is that by averaging over both X_i and its antithetic counterpart $1 - X_i$, we can reduce the variance in our estimate of the integral, leading to a more accurate result.

In essence, this approach works because the errors associated with one random variable can often cancel out the errors from its antithetic counterpart, thereby improving the overall accuracy of the estimate.

Why is this approach reasonable? This approach is reasonable because reducing the variance of the Monte Carlo estimator improves the accuracy of the integral estimate without requiring more samples. By using antithetic random variables, we are leveraging negative correlation to stabilize our estimate, leading to a more efficient and reliable approximation of the integral. This method is particularly useful in cases where the function $f(x)$ being integrated behaves symmetrically or exhibits patterns that can benefit from the use of antithetic sampling.

d) For which c is g(x) a probability density function (pdf)? Now, we try to improve our estimate of I by importance sampling. Let $f(x) = \frac{e^{-x}}{(1+x^2)x}$ and consider the function $g(x) = \frac{c}{1+x^2}$ for $-2-\sqrt{3} \leq x \leq -1/\sqrt{3}$, and $g(x) = 0$ otherwise. For which c is $g(x)$ a probability density function (pdf)?

We do integration by substitution with $x = \tan(u) = \frac{\sin(u)}{\cos(u)}$ and

$$\frac{dx}{du} = \frac{\cos(u)\cos(u) - \sin(u)(-\sin(u))}{\cos^2(u)} = \frac{\cos^2(u) + \sin^2(u)}{\cos^2(u)} = \frac{1}{\cos^2(u)}.$$

And thus,

$$\int \frac{1}{1+x^2} dx = \int \frac{1}{1 + \sin^2(u)/\cos^2(u)} \frac{1}{\cos^2(u)} du$$

$$= \int \frac{1}{\frac{\cos^2(u)+\sin^2(u)}{\cos^2(u)}} \frac{1}{\cos^2(u)} du$$

$$= \int 1 du = u = \tan^{-1}(x).$$

The function $g(x)$ is a probability density function if $\int g(x)dx = \int_{-1/\sqrt{3}}^{2-\sqrt{3}} g(x)dx = 1$. We have:

$$\int_{-1/\sqrt{3}}^{2-\sqrt{3}} \frac{c}{1+x^2} dx = c\big[\tan^{-1}(x)\big]_{-1/\sqrt{3}}^{2-\sqrt{3}}$$

$$= c\big(\tan^{-1}(-1/\sqrt{3}) - \tan^{-1}(2-\sqrt{3})\big)$$

$$= c\big(-\frac{\pi}{6} - \big(-\frac{5\pi}{12}\big)\big) = c\big(\frac{3\pi}{12}\big) = \frac{c\pi}{4} = 1.$$

And thus $c = 4$.

---

e) Plot $f(x)$ and $g(x)$ for values $x \in [-5, 5]$ in one plot with different colors. Do you think $g(x)$ is a good choice for the importance function? Why? $g(x)$ is a good choice for the importance sampling function if it closely matches the shape of $f(x)$ over the domain of interest. The goal is for $g(x)$ to resemble $f(x)$ where $f(x)$ has higher values, to reduce the variance in the estimation.

We will examine the plots in the next section using R code to see how well $g(x)$ approximates $f(x)$.

---

f) Find the inverse cumulative distribution function $G^{-1}(x)$. To find the inverse cumulative distribution function $G^{-1}(x)$, we first integrate the probability density function $g(x)$ to get the cumulative distribution function (CDF).

Given:

$$g(x) = \frac{6}{\pi(1 + x^2)}$$

We integrate this over the domain $x \in \left[-\frac{2}{\sqrt{3}}, -\frac{1}{\sqrt{3}}\right]$:

$$G(x) = \int_{-\frac{2}{\sqrt{3}}}^{x} g(t)dt = \frac{6}{\pi}\left(\arctan(x) - \arctan\left(-\frac{2}{\sqrt{3}}\right)\right)$$

To invert this, set $G(x) = u$ and solve for x:

$$u = \frac{6}{\pi}\left(\arctan(x) - \arctan\left(-\frac{2}{\sqrt{3}}\right)\right)$$

Solve for x:

$$x = \tan\left(\frac{\pi}{6}u + \arctan\left(-\frac{2}{\sqrt{3}}\right)\right)$$

This gives us the inverse CDF $G^{-1}(x)$.

Problem_2

part_b) For a Poisson process, the number of events $N(t)$ in time t follows a **Poisson distribution** with parameter λt, where λ is the rate (intensity) of the process.

Expected Value:

$$\mathbb{E}[N(t)] = \lambda \cdot t$$

Standard Deviation:

$$\mathrm{Var}[N(t)] = \lambda \cdot t$$

Therefore, the **standard deviation** is:

$$\sigma[N(t)] = \sqrt{\lambda \cdot t}$$

Given $\lambda = 3$, we can calculate the expected value and standard deviation for $N(5)$ and $N(20)$.

1. For $N(5)$:

- Expected value:

$$\mathbb{E}[N(5)] = 3 \times 5 = 15$$

- Standard deviation:

$$\sigma[N(5)] = \sqrt{3 \times 5} = \sqrt{15} \approx 3.87$$

2. For $N(20)$:

- Expected value:

$$\mathbb{E}[N(20)] = 3 \times 20 = 60$$

- Standard deviation:

$$\sigma[N(20)] = \sqrt{3 \times 20} = \sqrt{60} \approx 7.75$$

Problem_3

part_a) Find the integral by Monte Carlo integration using $n = 5000$ simulations based on independent uniformly and exponentially distributed random numbers.

Note that we can write this integral as the product of two integrals, i.e.,

$$I = \int_4^1 \int_2^{-2} \int_1^0 \frac{1}{1+x^2+y^2+z^2} e^{-1/4w} dw\, dz\, dy\, dx = \int_4^1 \int_2^{-2} \int_1^0 \frac{1}{1+x^2+y^2+z^2} dz\, dy\, dx \int_0^{10} e^{-1/4w} dw = I_1 \cdot I_2.$$

First Integral I_1:

$$I_1 = \int_4^1 \int_2^{-2} \int_1^0 \frac{1}{1+x^2+y^2+z^2} dz\, dy\, dx.$$

The first integral I_1 can be estimated by using uniform distributed variables, namely:

$$X_i \sim U[1,4],\ Y_i \sim U[-2,2],\ Z_i \sim U[0,1].$$

The Monte Carlo estimate of I_1 is given by:

$$\hat{I}_1 = \frac{(4-1)(2-(-2))(1-0)}{n} \sum_{i=1}^{n} \frac{1}{1+X_i^2+Y_i^2+Z_i^2}.$$

Second Integral I_2:

$$I_2 = \int_0^{10} e^{-1/4w} dw = \int_0^{\infty} I(w<10) 4 e^{-1/4w} dw = \int_0^{\infty} I(w<10) h(w) dw,$$

where $h(w)$ is the density function of an exponential distribution with $\beta = 4$.

We can rewrite I_2 as:

$$\hat{I}_2 = \frac{1}{n} \sum_{i=1}^{n} 4 \cdot I(W_i < 10),$$

where $W_i \sim \mathrm{Exp}(\beta = 4)$.

Thus, the estimate of I is:

$$\hat{I} = \hat{I}_1 \cdot \hat{I}_2.$$

part_b)

Generating χ^2-Distributed Random Variables with 6 Degrees of Freedom

Given standard uniformly distributed random variables, we can generate χ^2-distributed random variables with 6 degrees of freedom as follows:

Key Relationships

1. If $U \sim U[0,1]$, then:
$$-\beta \ln(U) \sim \mathrm{Exp}(\beta)$$
2. If $X \sim \mathrm{Exp}(\beta)$, then:
$$2\beta X \sim \chi_2^2$$

Procedure

1. Let $U_i \sim U[0,1]$ for $i = 1, 2, 3$.
2. Define:
$$Y_i = 2(-\beta \ln(U_i))$$
Substituting $\beta = 1$, this simplifies to:
$$Y_i = -2 \ln(U_i) \sim \chi_2^2$$
3. Sum the Y_i values to generate a χ^2-distributed random variable with 6 degrees of freedom:
$$Y_1 + Y_2 + Y_3 = -2(\ln(U_1) + \ln(U_2) + \ln(U_3)) = -2 \ln(U_1 U_2 U_3) \sim \chi_6^2$$

Thus, summing the transformed uniform random variables gives a χ^2-distributed random variable with 6 degrees of freedom.