Data Science and Statistical Modelling

Tutorial 3, Week 6

Q1

At the end of the last tutorial, we were interested in evaluating the integral

$$\int_0^1 xe^{-x}\,dx$$

We did so by writing it as an expectation with respect to an Exponential distribution involving an indicator function.

An alternative way to tackle the above Monte Carlo integral would be to simulate from a finitely supported distribution, so that you don't need to use the indicator function.

a) Express the integral above as an expectation with respect to a Uniform distribution.

b) Write down a Monte Carlo integration algorithm that estimates the integral using simulations from a Uniform distribution.

c) Compute the variance of this new Monte Carlo estimator based on Uniform simulations.

Hint: To save some time with integrals, you may use the fact that:

$$\int x^2 e^{-2x}\,dx = -\frac{e^{-2x}}{4}(2x^2 + 2x + 1)$$

(Note: The variance of the estimator based on Exponential simulations in the previous tutorial was approximately $0.09078/n$)

d) How many more/fewer simulations would I need with Exponential simulation to achieve the same size confidence interval as under Uniform simulation?

Q2

The following is a valid probability density function for all $\alpha > 0$,

$$f_X(x) = \begin{cases} \alpha x^{\alpha-1} & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

a) Describe how to use inverse transform sampling to generate simulations of a random variable that follow this density.

b) Express the integral from Q1 as an expectation with respect to a random variable with the above pdf.

c) Write down a Monte Carlo integration algorithm that estimates the integral using simulations of the above random variable.

d) Let $\alpha = 2$. Compute the variance of this new Monte Carlo estimator.

Hint: To save some time with integrals, you may use the fact that:

$$\int xe^{-2x}\,dx = -\frac{e^{-2x}}{4}(2x + 1)$$

e) How many more/fewer simulations would I need with Exponential simulation or with Uniform simulation to achieve the same size confidence interval as when simulating from f_X?

Q3

In an example in lectures we showed how to inverse transform sample from a standard Cauchy distribution, and used these samples to rejection sample from a standard Normal distribution. We will try another approach to sample standard Normals here.

a) The standard Laplace distribution (sometimes called the double exponential distribution) has pdf:

$$g(x) = \frac{1}{2}\exp(-|x|) \quad \forall\, x \in \mathbb{R}$$

Derive an inverse transform sampling scheme to generate simulations from this distribution.

b) If $f(x)$ is the pdf of a standard Normal distribution, show $\exists\, c < \infty$ such that

$$f(x) \le cg(x) \,\forall\, x \in \mathbb{R}$$

and find the smallest such c.

c) Write down the steps to produce a rejection sampled simulation from a standard Normal distribution via an inverse transform sampled standard Laplace simulation.

d) In lectures, the standard Cauchy distribution had $c \approx 1.521$ when used to rejection sample standard Normals. In light of your analysis, would you favour simulation via the standard Cauchy or standard Laplace and why? Compare the expected number of iterations required to produce one sample for both methods.

Q4

Let X be a random variable on $\mathbb{R}$ with cdf $F(\cdot)$. The so-called *truncated* version of a random variable arises when we restrict the support of X to some range $X \in [a, b] \subset \mathbb{R}$. In other words, it is the random variable $(X \,|\, a \le X \le b)$

a) Prove that the cdf of the truncated random variable is:

$$\mathbb{P}(X \le x \,|\, a \le X \le b) = \begin{cases} 0 & \text{if } x < a \\ \dfrac{F(x) - F(a)}{F(b) - F(a)} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

b) Hence or otherwise, find an inverse transform sampler for any truncated distribution in terms of $F(\cdot)$.

c) Write down an inverse transform sampler for an Exponential distribution truncated to $[1, \infty)$

d) Does the result in (c) confirm any properties of the Exponential distribution?

Q5

Let $f(x)$ be the Uniform$(0, 1)$ distribution, $g(x)$ the Uniform$(0, \frac{1}{2})$ distribution, and consider the functional $h(x) = x^2$. Show that,

$$\mathbb{E}_g\left[\frac{h(X)f(X)}{g(X)}\right] \ne \mathbb{E}_f\left[h(X)\right]$$

Note: *This is to illustrate the importance of the requirement that $g(\cdot)$ be a pdf such that $g(x) > 0$ whenever $h(x)f(x) \ne 0$ in order for importance sampling to be valid! This condition is clearly violated for the choices of $f(\cdot), g(\cdot)$ and $h(\cdot)$ above.*

Q6

We consider a very simple problem to keep the algebra easy(ish), but this is still a lengthy question. Please remember, you would not really need importance sampling in such a simple problem, we do it here to highlight some interesting points!

Let,

$$f(x) = \begin{cases} 2x & \text{for } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

a) Compute $\mu = \mathbb{E}_f[X]$ exactly by solving the relevant integral.

b) Assume you have n Monte Carlo simulations $\{x_1, \ldots, x_n\}$ from $f(\cdot)$. Write down the equation for the Monte Carlo estimator of μ, $\hat{\mu}_n$, as well as the variance of $\hat{\mu}_n$ as a function of n.

You decide to use a proposal distribution from the family of distributions having probability density function of the form:

$$g(x) = \begin{cases} \alpha x^{\alpha-1} & \text{for } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha > 0$ is a parameter (we know this distribution is easy to inverse transform sample from Question 2 above).

c) Assume you have n simulations $\{x_1, \ldots, x_n\}$ from $g(\cdot)$. Write down the equation for the importance sampling estimator of μ, $\hat{\mu}_n$, and derive the variance of $\hat{\mu}_n$ as a function of α and n (*Hint:* Theorem 5.4).

d) For what range of choices of $\alpha > 0$ does the importance sampling estimator have lower variance than the standard Monte Carlo estimator?

e) By using the formula for $g_{\text{opt}}(x)$ in lectures, write down the *optimal* proposal distribution for this problem.

   Does it belong to the family $g(\cdot \mid \alpha)$ above? If so, determine the value for α and use part (c) to write down the estimator variance. Does this variance make sense?!

   Hint: take any single random draw $x_1 \in [0, 1]$ and compute $\hat{\mu}_1$ by importance sampling with the optimal proposal.

Q7

Let $f(x) = \lambda e^{-\lambda x}, x \geq 0$ for some fixed parameter value $\lambda > 0$. If you use proposal density $g(x) = \eta e^{-\eta x}, x \geq 0$, what are the weights for an importance sample, x_i, drawn from g? For what values of η are you guaranteed the weights are bounded?

> **Note:** *Think about what can happen if the weights are not bounded. It is often the case that if the weights are unbounded then the variance of the estimator is infinite, but requiring the weights to be bounded restricts us to using importance sampling in the same situations as rejection sampling (there are approaches to manage this, but it is beyond the scope of this second year course).*