

Data Science and Statistical Computing

Tutorial 4, Week 8 Solutions

Q1

(a)

If we have that $X \sim \text{Exp}(\lambda = 1)$, then using the hint, we can see that the integral is simply:

$$\begin{aligned}\mu &= \int_0^1 x \lambda e^{-\lambda x} dx = \int_0^1 x f_X(x) dx \\ &= \int_{\mathbb{R}} x \mathbb{1}\{x \in [0, 1]\} f_X(x) dx \\ &= \mathbb{E}[X \mathbb{1}\{X \in [0, 1]\}]\end{aligned}$$

(b)

Using integration by parts,

$$\begin{aligned}u &= x, & dv &= e^{-x} \\ du &= 1, & v &= -e^{-x}\end{aligned}$$

$$\begin{aligned}\int x e^{-x} dx &= -x e^{-x} + \int e^{-x} dx \\ &= -e^{-x}(1 + x) \\ \Rightarrow \int_0^1 x e^{-x} dx &= [-e^{-x}(1 + x)]_0^1 \\ &= -e^{-1}(1 + 1) + e^{-0}(1 + 0) \\ &= 1 - 2e^{-1} \\ &\approx 0.2642\end{aligned}$$

Therefore, $\mathbb{E}[X \mathbb{1}\{X \in [0, 1]\}] = 1 - 2e^{-1} \approx 0.2642$.

(c)

1. Simulate $x_1, \dots, x_n$ from an Exponential($\lambda = 1$) distribution
2. Calculate

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n x_i \mathbb{1}\{x_i \in [0, 1]\}$$

(d)

We know from lectures that

$$\text{Var}(\hat{\mu}_n) = \mathbb{E}[(\hat{\mu}_n - \mu)^2] = \frac{\sigma^2}{n}$$

where here we have

$$\sigma^2 = \text{Var}(X \mathbb{1}\{X \in [0, 1]\})$$

with $X \sim \text{Exp}(\lambda = 1)$.

Based on the hint, we want to find:

$$\text{Var}(X \mathbb{1}\{X \in [0, 1]\}) = \mathbb{E}[(X \mathbb{1}\{X \in [0, 1]\})^2] - \mathbb{E}[X \mathbb{1}\{X \in [0, 1]\}]^2$$

We already know from (b) that $\mathbb{E}[X \mathbb{1}\{X \in [0, 1]\}] = 1 - 2e^{-1}$, so we just need:

$$\begin{aligned}\mathbb{E}[(X \mathbb{1}\{X \in [0, 1]\})^2] &= \int_{\mathbb{R}} (x \mathbb{1}\{x \in [0, 1]\})^2 f_X(x) dx \\ &= \int_0^1 x^2 e^{-x} dx\end{aligned}$$

Again, a simple application of integration by parts gives:

$$\begin{aligned}\int x^2 e^{-x} dx &= -x^2 e^{-x} - \int -2x e^{-x} dx \\ &= -x^2 e^{-x} + 2 \int x e^{-x} dx\end{aligned}$$

We know the remaining integral from (b), so we get:

$$\begin{aligned}\int x^2 e^{-x} dx &= -x^2 e^{-x} + 2(-e^{-x}(1+x)) \\ \Rightarrow \int_0^1 x^2 e^{-x} dx &= [-e^{-x}(2+2x+x^2)]_0^1 \\ &= -e^{-1}(2+2+1) + e^0(2+0+0) \\ &= 2 - 5e^{-1} \\ &\approx 0.1606\end{aligned}$$

So,

$$\begin{aligned}\text{Var}(X \mathbb{1}\{X \in [0, 1]\}) &= \underbrace{(2 - 5e^{-1})}_{\mathbb{E}[(X \mathbb{1}\{X \in [0, 1]\})^2]} - \underbrace{(1 - 2e^{-1})^2}_{\mathbb{E}[X \mathbb{1}\{X \in [0, 1]\}]^2} \\ &= 1 - e^{-1} - 4e^{-2} \\ &\approx 0.09078\end{aligned}$$

Thus,

$$\text{Var}(\hat{\mu}_n) \approx \frac{0.09078}{n}$$

Q2

(a)

Let X have Uniform distribution on $[0, 1]$, so that:

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

Then,

$$\begin{aligned}\mu &= \int_0^1 x e^{-x} dx = \int_0^1 x e^{-x} f_X(x) dx \\ &= \int_{\mathbb{R}} x e^{-x} f_X(x) dx \\ &= \mathbb{E}[X e^{-X}]\end{aligned}$$

(b)

1. Simulate $x_1, \dots, x_n$ from a Uniform(0, 1) distribution
2. Calculate

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n x_i e^{-x_i}$$

(c)

As for Q1(d), we know from lectures that

$$\text{Var}(\hat{\mu}_n) = \mathbb{E}[(\hat{\mu}_n - \mu)^2] = \frac{\sigma^2}{n}$$

However, now

$$\sigma^2 = \text{Var}(X e^{-X})$$

with $X \sim \text{Unif}(0, 1)$.

We will again use the standard formula for the variance:

$$\text{Var}(X e^{-X}) = \mathbb{E}[(X e^{-X})^2] - \mathbb{E}[X e^{-X}]^2$$

We know that Monte Carlo integration is unbiased, so the second term is unchanged in value despite the different problem formulation,

$$\mathbb{E}[X e^{-X}] = 1 - 2e^{-1} \approx 0.2642$$

So we just need:

$$\begin{aligned} \mathbb{E}[(X e^{-X})^2] &= \int x^2 e^{-2x} f_X(x) dx \\ &= \int_0^1 x^2 e^{-2x} dx \end{aligned}$$

Using the hint in the question to save time on integration, we find:

$$\begin{aligned} \int_0^1 x^2 e^{-2x} dx &= \left[-\frac{e^{-2x}}{4} (2x^2 + 2x + 1) \right]_0^1 \\ &= -\frac{5}{4} e^{-2} + \frac{1}{4} \\ &\approx 0.08083 \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}(X e^{-X}) &= \underbrace{\frac{1}{4} - \frac{5}{4} e^{-2}}_{\mathbb{E}[(X e^{-X})^2]} - \underbrace{(1 - 2e^{-1})^2}_{\mathbb{E}[X e^{-X}]^2} \\ &= 4e^{-1} - \frac{21}{4} e^{-2} - \frac{3}{4} \\ &\approx 0.01101 \end{aligned}$$

Thus,

$$\text{Var}(\hat{\mu}_n) \approx \frac{0.01101}{n}$$

(d)

The width of the confidence interval for Monte Carlo integration is

$$z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Letting σ_E^2 and n_E be the variance and simulation size under Exponential sampling, and σ_U^2 and n_U be the variance and simulation size under Uniform sampling, we would need:

$$\begin{aligned} z_{\alpha/2} \frac{\sigma_E}{\sqrt{n_E}} &= z_{\alpha/2} \frac{\sigma_U}{\sqrt{n_U}} \\ \Rightarrow \frac{n_E}{n_U} &= \frac{\sigma_E^2}{\sigma_U^2} \\ \Rightarrow \frac{n_E}{n_U} &= \frac{0.09078}{0.01101} = 8.25 \end{aligned}$$

So we would need to take $8.25 \times$ times as many simulations from an Exponential to achieve the same accuracy as using Uniform simulations!

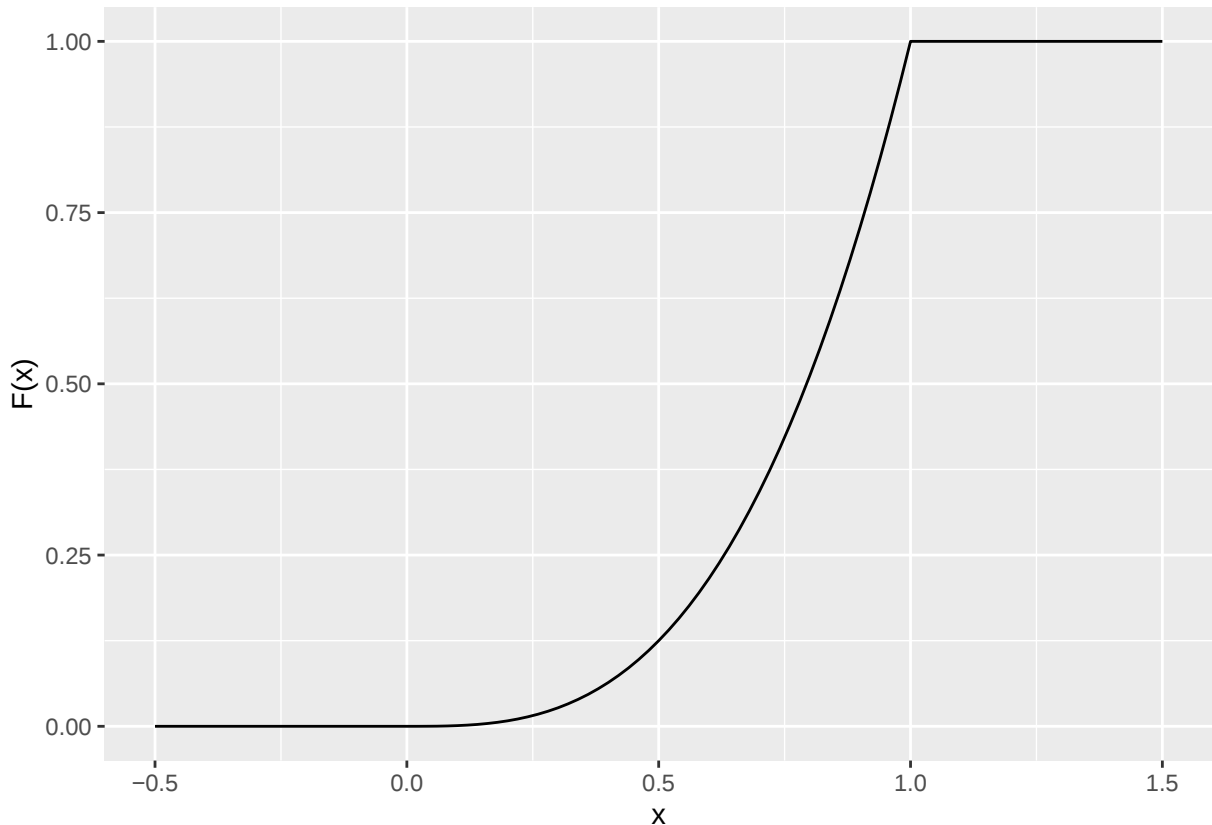
Q3

(a)

Given the probability density function in the question, we need to find the cdf:

$$F_X(x) = \int_{-\infty}^x \alpha z^{\alpha-1} dz = \begin{cases} 0 & \text{if } x < 0 \\ x^\alpha & \text{if } x \in [0, 1] \\ 1 & \text{if } x > 1 \end{cases}$$

For example,



Therefore, $F^{-1}(u) = u^{\frac{1}{\alpha}}$.

Hence, we simulate from this probability density function by simulating $U \sim \text{Unif}(0, 1)$ and then return $X = u^{\frac{1}{\alpha}}$.

(b)

Let X have probability density function $f_X(x)$ as given in the question,

$$\begin{aligned}\mu &= \int_0^1 x e^{-x} dx = \int_0^1 \frac{1}{f_X(x)} x e^{-x} f_X(x) dx \\ &= \int_{\mathbb{R}} \alpha^{-1} x^{-\alpha+1} x e^{-x} f_X(x) dx \\ &= \int_{\mathbb{R}} \alpha^{-1} x^{-\alpha+2} e^{-x} f_X(x) dx \\ &= \mathbb{E} [\alpha^{-1} x^{-\alpha+2} e^{-x}]\end{aligned}$$

(c)

1. Simulate $x_1, \dots, x_n$ from the pdf given in the question, perhaps using inverse sampling as described in (a)
2. Calculate

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \alpha^{-1} x_i^{-\alpha+2} e^{-x_i}$$

(d)

As for Q1(d) & Q2(c), we know from lectures that

$$\text{Var}(\hat{\mu}_n) = \mathbb{E}[(\hat{\mu}_n - \mu)^2] = \frac{\sigma^2}{n}$$

However, now

$$\sigma^2 = \text{Var}(\alpha^{-1} x^{-\alpha+2} e^{-x})$$

with X simulated from $f_X(\cdot)$.

We are told $\alpha = 2$, so we want:

$$\text{Var}(0.5e^{-X}) = \mathbb{E}[(0.5e^{-X})^2] - \mathbb{E}[0.5e^{-X}]^2$$

We know that Monte Carlo integration is unbiased, so the second term is unchanged in value despite the different problem formulation,

$$\mathbb{E}[0.5e^{-X}] = 1 - 2e^{-1} \approx 0.2642$$

So we just need:

$$\begin{aligned}
\mathbb{E} \left[(0.5e^{-X})^2 \right] &= \int 0.25e^{-2x} f_X(x) dx \\
&= \int_0^1 0.25e^{-2x} \times 2x dx \\
&= 0.5 \int_0^1 xe^{-2x} dx \\
&= 0.5 \left[-\frac{e^{-2x}}{4} (2x + 1) \right]_0^1 \quad \text{using hint} \\
&= \frac{1}{8} - \frac{3}{8}e^{-2} \\
&\approx 0.07425
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{Var} (0.5e^{-X}) &= \underbrace{\frac{1}{8} - \frac{3}{8}e^{-2}}_{\mathbb{E}[(0.5e^{-X})^2]} - \underbrace{(1 - 2e^{-1})^2}_{\mathbb{E}[0.5e^{-X}]^2} \\
&= 4e^{-1} - \frac{35}{8}e^{-2} - \frac{7}{8} \\
&\approx 0.004426
\end{aligned}$$

Thus,

$$\text{Var} (\hat{\mu}_n) \approx \frac{0.004426}{n}$$

(e)

The logic is as for Q2(d). So,

$$\frac{0.09078}{0.004426} = 20.51$$

So we would need more than $20\times$ as many simulations from an *Exponential* to achieve the same accuracy.

$$\frac{0.01101}{0.004426} = 2.49$$

So we would need $2.49\times$ as many simulations from a *Uniform* to achieve the same accuracy.