Data Science and Statistical Computing

Tutorial 6, Week 21 Solutions

Q1

We first find $\mathbb{E}_f\left[g(X)\right]$.

$$
\begin{aligned}
\mathbb{E}_f\left[g(X)\right] &= \int_{\mathbb{R}} g(x)f(x)\,dx \\
&= \int_0^1 x^2\,dx \quad \text{(NB limits determined by } f\text{)} \\
&= \left.\frac{x^3}{3}\right|_{x=0}^1 \\
&= \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}_{\tilde{f}}\left[\frac{g(X)f(X)}{\tilde{f}(X)}\right] &= \int_{\mathbb{R}} \frac{g(x)f(x)}{\tilde{f}(x)}\tilde{f}(x)\,dx \\
&= \int_0^{\frac{1}{2}} \frac{x^2}{2}\times 2\,dx \quad \text{(NB limits determined by } \tilde{f}\text{)} \\
&= \int_0^{\frac{1}{2}} x^2\,dx \\
&= \frac{1}{24}
\end{aligned}
$$

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

Q2

(a)

$$
\begin{aligned}
\mathbb{E}_f[X] &= \int_{\Omega} xf(x)\,dx \\
&= \int_0^1 x(2x)\,dx \\
&= \frac{2}{3}
\end{aligned}
$$

(b)

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

$$
\text{Var}(\hat{\mu}_n) = \frac{\text{Var}(g(X))}{n}
$$

Now $g(X) = X$ here, and

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

We know $\mathbb{E}[X]$ from part (a), so just need,

$$\mathbb{E}[X^2] = \int_\Omega x^2 f(x)\, dx$$

$$= \int_0^1 x^2 (2x)\, dx$$

$$= \frac{1}{2}$$

Thus,

$$\mathrm{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18} \approx 0.0556$$

Therefore,

$$\mathrm{Var}(\hat{\mu}_n) = \frac{1}{18n}$$

(c)

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

where

$$w_i = \frac{f(x_i)}{\tilde{f}(x_i)} = \frac{2x_i}{\alpha x_i^{\alpha-1}} = 2\alpha^{-1} x_i^{2-\alpha}$$

Thus, in full,

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n 2\alpha^{-1} x_i^{3-\alpha}$$

Next, for the variance, we know from lectures,

$$\mathrm{Var}(\hat{\mu}_n) = \frac{\sigma_{\tilde{f}}^2}{n} \quad \text{where} \quad \sigma_{\tilde{f}}^2 = \int_{\tilde{\Omega}} \frac{\left(g(x)f(x) - \mu\tilde{f}(x)\right)^2}{\tilde{f}(x)}\, dx,$$

and recall that here $g(x) = x$ and from (a) $\mu = \frac{2}{3}$. Thus, (not necessary to show all these small steps, but providing full detail for solutions ...)

$$\sigma_{\tilde{f}}^2 = \int_{\tilde{\Omega}} \frac{\left(g(x)f(x) - \mu\tilde{f}(x)\right)^2}{\tilde{f}(x)}\, dx$$

$$= \int_0^1 \frac{\left(x(2x) - \frac{2}{3}\alpha x^{\alpha-1}\right)^2}{\alpha x^{\alpha-1}}\, dx$$

$$= \int_0^1 \frac{4}{9}\alpha^{-1} x^{1-\alpha}\left(3x^2 - \alpha x^{\alpha-1}\right)^2\, dx$$

$$= \int_0^1 \frac{4}{9}\alpha^{-1} x^{1-\alpha}\left(9x^4 - 6\alpha x^{\alpha+1} + \alpha^2 x^{2\alpha-2}\right)\, dx$$

$$= \int_0^1 \frac{4\alpha x^{\alpha-1}}{9} + \frac{4x^{5-\alpha}}{\alpha} - \frac{8x^2}{3}\, dx$$

$$= \frac{4x^\alpha}{9} + \frac{4x^{6-\alpha}}{\alpha(6-\alpha)} - \frac{8x^3}{9}\Bigg|_{x=0}^1 \qquad \text{assuming } 0 < \alpha < 6$$

$$= \frac{4}{\alpha(6-\alpha)} - \frac{4}{9}$$

Note: the distribution $\tilde{f}$ only required $\alpha > 0$, but this variance calculation now only holds for $\alpha \in (0,6)$. This arose because the $x = 0$ end of the integral is indeterminate for the term $x^{6-\alpha}$ if $\alpha > 6$ and $\frac{1}{6-\alpha}$ is indeterminate if $\alpha = 6$.

$$\implies \mathrm{Var}(\hat{\mu}_n) = \frac{4}{n\alpha(6-\alpha)} - \frac{4}{9n}, \quad \alpha \in (0,6)$$

(d)

We are interested in values of α for which it is true that,

$$\frac{4}{\alpha(6-\alpha)} - \frac{4}{9} < \frac{1}{18}$$

$$\implies \frac{4}{\alpha(6-\alpha)} < \frac{1}{2}$$

Now, the variance only holds for $\alpha \in (0,6)$, so $\alpha(6-\alpha)$ is always positive. Thus we may rearrange without fear of flipping the inequality. So,

$$\implies 8 < \alpha(6-\alpha) \quad \text{for } \alpha \in (0,6)$$
$$\implies 0 < (2-\alpha)(\alpha - 4) \quad \text{for } \alpha \in (0,6)$$

Now $(2-\alpha)(\alpha-4)$ is quadratic in α with negative coefficient on the square term, so this inequality holds only between the roots.

Thus, the variance of importance sampling is smaller than standard Monte Carlo when using proposal $\tilde{f}(\cdot \mid \alpha)$ for any $\alpha \in (2,4)$.

(e)

$$\tilde{f}_{\text{opt}}(x) = \frac{|g(x)|f(x)}{\int_\Omega |g(x)|f(x)\,dx}$$

Recall, here $g(x) = x$, so

$$\tilde{f}_{\text{opt}}(x) = \frac{|x|2x}{\int_0^1 |x|2x\,dx}$$

But, since $f(x)$ is only non-zero on $[0,1]$, the absolute value can be dropped and we find the optimal proposal is simply:

$$\tilde{f}_{\text{opt}}(x) = 3x^2$$

Yes, this **does** belong to the family $\tilde{f}(\cdot \mid \alpha)$, corresponding to the case $\alpha = 3$.

Using part (c), this tells us the variance when using this estimator is:

$$\text{Var}(\hat{\mu}_n) = \frac{4}{n\alpha(6-\alpha)} - \frac{4}{9n} = \frac{4}{n3(6-3)} - \frac{4}{9n} = 0$$

Zero variance!?! Note, this implies that if I take even one simulation from $\tilde{f}(\cdot \mid \alpha = 3)$ that the estimator $\hat{\mu}_1$ will be exactly correct ... does this make sense? Yes! Because also from part (c), we have,

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n 2\alpha^{-1}x_i^{3-\alpha}$$
$$= \frac{1}{n}\sum_{i=1}^n 2 \times 3^{-1}x_i^{3-3}$$
$$= \frac{1}{n}\sum_{i=1}^n \frac{2}{3}$$
$$= \frac{2}{3} \equiv \mu$$

Note: this is obviously a situation we would never be in when using importance sampling for a real problem which actually required it! Usually, the optimal proposal is inaccessible to us, because to get it we would need to be able to do the original integral anyway ... the point here is to hightlight that if we can get close to approximating the optimal proposal, then clearly we can dramatically reduce the variance of our estimator.

Q3

Here, the weight for a sample x_i from $\tilde{f}$ will be:

$$w_i = \frac{f(x_i)}{\tilde{f}(x_i)} = \frac{\lambda e^{-\lambda x_i}}{\eta e^{-\eta x_i}} = \frac{\lambda}{\eta} e^{-(\lambda - \eta) x_i}$$

Now $x_i \in [0, \infty)$ since the Exponential distribution is on the half real line. Notice the term $e^{-(\lambda - \eta) x_i}$ ... in order for the weights to be bounded for all possible x_i, we need $\lambda - \eta > 0$. Thus, the weights are only bounded for $\eta < \lambda$.