

# MA40189 - Solution Sheet Seven

Simon Shaw, [s.shaw@bath.ac.uk](mailto:s.shaw@bath.ac.uk)  
<https://people.bath.ac.uk/masss/ma40189.html>

2021/22 Semester II

1. Consider the integral

$$I = \int_0^{10} \exp(-2|x-5|) dx.$$

(a) Suppose that  $X \sim U(0, 10)$  with probability density function

$$f(x) = \begin{cases} \frac{1}{10} & 0 \leq x \leq 10, \\ 0 & \text{otherwise.} \end{cases}$$

Show that  $I = E_f(10 \exp(-2|X-5|))$ , the expectation of  $10 \exp(-2|X-5|)$  calculated with respect to the probability density function  $f(x)$ . Hence, explain how to use the method of Monte Carlo integration to estimate  $I$ .

Note that

$$\begin{aligned} I &= \int_0^{10} \exp(-2|x-5|) dx = \int_0^{10} 10 \exp(-2|x-5|) \frac{1}{10} dx \\ &= E_f(10 \exp(-2|X-5|)). \end{aligned}$$

We can then estimate  $I$  using Monte Carlo integration as follows:

- (i) sample  $n$  values  $x_1, \dots, x_n$  independently from  $U(0, 10)$
- (ii) compute  $10 \exp(-2|x_i-5|)$  for each  $x_i$
- (iii) estimate  $I$  by  $\hat{I}$ , the sample mean of the  $10 \exp(-2|x_i-5|)$ s, so that

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n 10 \exp(-2|x_i-5|). \quad (1)$$

(b) Explain how to estimate  $I$  using the method of importance sampling where we sample from the  $N(5, 1)$  distribution.

For importance sampling, we want to write  $I$  as an expectation calculated with respect to the  $N(5, 1)$  distribution which has support on  $(-\infty, \infty)$ . If we let  $\mathbb{I}_{\{0 \leq x \leq 10\}}$  be the indicator function that  $0 \leq x \leq 10$  then

$$I = \int_0^{10} \exp(-2|x-5|) dx = \int_{-\infty}^{\infty} \exp(-2|x-5|) \mathbb{I}_{\{0 \leq x \leq 10\}} dx$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{10 \exp(-2|x-5|) \frac{1}{10} \mathbb{I}_{\{0 \leq x \leq 10\}}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-5)^2\right\}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-5)^2\right\} dx \\
&= \int_{-\infty}^{\infty} \frac{g(x)f(x)}{q(x)} q(x) dx = E_q\left(\frac{g(X)f(X)}{q(X)}\right)
\end{aligned}$$

where  $g(x) = 10 \exp(-2|x-5|)$ ,  $f(x) = \frac{1}{10} \mathbb{I}_{\{0 \leq x \leq 10\}}$  (the density of the  $U(0, 10)$  distribution) and  $q(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-5)^2\right\}$  (the density of the  $N(5, 1)$  distribution). We can then estimate  $I$  using importance sampling as follows:

- (i) sample  $n$  values  $x_1, \dots, x_n$  independently from  $N(5, 1)$
- (ii) compute

$$\frac{g(x_i)f(x_i)}{q(x_i)} = \frac{10 \exp(-2|x_i-5|) \frac{1}{10} \mathbb{I}_{\{0 \leq x_i \leq 10\}}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i-5)^2\right\}}$$

for each  $x_i$

- (iii) estimate  $I$  by  $\tilde{I}$ , the sample mean of the  $\frac{g(x_i)f(x_i)}{q(x_i)}$ s, so that

$$\tilde{I} = \frac{1}{n} \sum_{i=1}^n \frac{10 \exp(-2|x_i-5|) \frac{1}{10} \mathbb{I}_{\{0 \leq x_i \leq 10\}}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_i-5)^2\right\}}. \quad (2)$$

- (c) Use **R** to explore the methods used in parts (a) and (b). Which method do you prefer?

We can write a function `monte0` in R to estimate  $I$  by  $\hat{I}$ , as given by equation (1).

```

# Calculates m replications of the Monte Carlo
# integration in part (a) with n = n1
monte0 <- function(n1,m){
  vect <- c(1:m)
  for (i in 1:m)
  {
    x <- runif(n1,0,10)
    vect[i] <- sum(10*exp(-2*abs(x-5)))/n1
  }
  return(vect)
}

```

Similarly, the function `import0` estimates  $I$  by  $\tilde{I}$ , as given by equation (2). Notice that `dunif(x, 0, 10) =  $\frac{1}{10} \mathbb{I}_{\{0 \leq x \leq 10\}}$` .

```

# Calculates m replications of the importance
# sampling in part (b) with n = n1
import0 <- function(n1,m){
  vect <- c(1:m)
  for (i in 1:m)
  {
    x <- rnorm(n1,5,1)
    weight <- dunif(x,0,10)/dnorm(x,5,1)
    vect[i] <- sum(10*exp(-2*abs(x-5))*weight)/n1
  }
}

```

```

return(vect)
}

```

We explore the two methods for  $n = 100, 500, 1000$ , and  $5000$ , in each case taking 1000 replications. The results are shown in the following table, note that the true value of  $I$  is  $1 - \exp(-10) = 0.9999546$ .

| Method | $n$  | Mean of 1000 samples | Variance of 1000 samples |
|--------|------|----------------------|--------------------------|
| Monte  | 100  | 1.013746             | 0.04300273               |
| Import | 100  | 0.9997384            | 0.003624815              |
| Monte  | 500  | 0.999168             | 0.008210746              |
| Import | 500  | 1.00074              | 0.0007107777             |
| Monte  | 1000 | 1.00279              | 0.004010403              |
| Import | 1000 | 1.000122             | 0.0003694998             |
| Monte  | 5000 | 1.000884             | 0.000773445              |
| Import | 5000 | 0.9997536            | 0.00007040271            |

We recall from lectures, that both methods give unbiased estimates of  $I$  so we compare their performance by considering the respective variances. From the table, we observe that the sample variance for the Monte Carlo integration is about ten times that of the importance sampling method.

- for Monte Carlo integration as  $Var_f(\hat{I}) = \frac{1}{n}Var_f(g(X))$ , where  $g(x) = 10 \exp(-2|x - 5|)$ , then  $Var_f(g(X)) \approx 4$
- for importance sampling we have that  $Var_q(\tilde{I}) = \frac{1}{n}Var_q\left(\frac{g(X)f(X)}{q(X)}\right)$  so that  $Var_q\left(\frac{g(X)f(X)}{q(X)}\right) \approx 0.4$

We thus see that the importance sampling method is preferable to the Monte Carlo integration. Figure 1 demonstrates why this is the case. The function  $g(x) = 10 \exp(-2|x - 5|)$  has a peak at  $x = 5$  and decays quickly elsewhere in the interval  $[0, 10]$  so that when we sample from  $U(0, 10)$  many of the sampled values will be where  $g(x)$  is very close to zero. We are likely to increase the precision of our estimate if we sample from a distribution which attaches more weight, or *importance*, to the region where  $g(x)$  is not close to zero. As Figure 1 shows, the  $N(5, 1)$  does this, leading to the reduced variance in the importance sampling scheme.

2. Suppose that  $f(\theta | x) = cg(\theta)$  where  $g(\theta) \propto f(x | \theta)f(\theta)$  and  $c$  is the normalising constant. For a function  $h(\theta)$  we wish to find  $E(h(\theta) | X) = E_f(h(\theta))$  where

$$E_f(h(\theta)) = \int_{-\infty}^{\infty} h(\theta)f(\theta | x) d\theta.$$

Suppose that we sample  $N$  values  $\theta_1, \dots, \theta_N$  independently from a distribution with probability density function  $q(\theta)$  (you may assume that the support of  $q(\theta)$  is the same as  $f(\theta | x)$ ).

- (a) Show that  $E_f(h(\theta)) = cE_q\left(\frac{h(\theta)g(\theta)}{q(\theta)}\right)$ , where  $E_q(\cdot)$  denotes expectation calculated with respect to the probability density function  $q(\theta)$ . If  $c$  is known, explain how to use  $\theta_1, \dots, \theta_N$  to estimate  $E_f(h(\theta))$ .

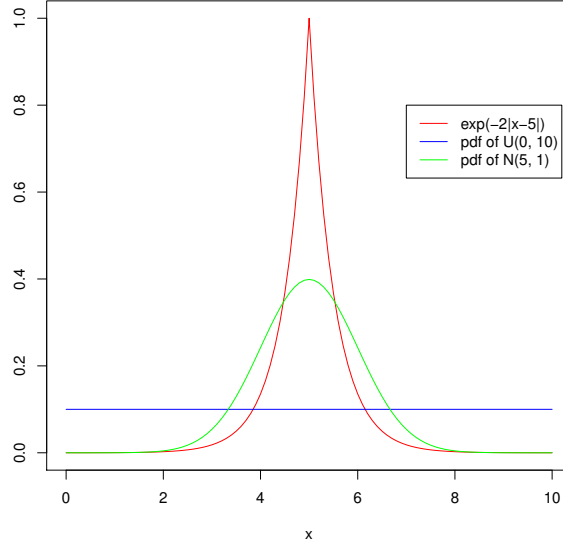


Figure 1: A plot of  $\frac{g(x)}{10} = \exp(-2|x-5|)$  with the probability density function of the  $U(0, 10)$  used for the Monte Carlo integration in part (a) and the probability density function of the  $N(5, 1)$  used for the importance sampling in part (b).

Note that, as  $f(\theta | x) = cg(\theta)$ , then

$$\begin{aligned} E_f(h(\theta)) &= c \int_{-\infty}^{\infty} h(\theta)g(\theta) d\theta \\ &= c \int_{-\infty}^{\infty} \frac{h(\theta)g(\theta)}{q(\theta)} q(\theta) d\theta = cE_q \left( \frac{h(\theta)g(\theta)}{q(\theta)} \right). \end{aligned}$$

We compute  $\frac{h(\theta_i)g(\theta_i)}{q(\theta_i)}$  for each  $i = 1, \dots, N$  and estimate  $E_q \left( \frac{h(\theta)g(\theta)}{q(\theta)} \right)$  by

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \frac{h(\theta_i)g(\theta_i)}{q(\theta_i)} \quad (3)$$

and, since  $c$  is known, estimate  $E_f(h(\theta))$  by  $c\hat{I}$ .

(b) **Suppose that  $c$  is unknown. By noting that**

$$c^{-1} = \int_{-\infty}^{\infty} g(\theta) d\theta,$$

**explain how to use  $\theta_1, \dots, \theta_N$  to estimate  $c^{-1}$ .**

We have that

$$1 = \int_{-\infty}^{\infty} f(\theta | x) d\theta = \int_{-\infty}^{\infty} cg(\theta) d\theta$$

so that

$$\begin{aligned} c^{-1} &= \int_{-\infty}^{\infty} g(\theta) d\theta \\ &= \int_{-\infty}^{\infty} \frac{g(\theta)}{q(\theta)} q(\theta) d\theta = E_q \left( \frac{g(\theta)}{q(\theta)} \right). \end{aligned}$$

Thus, we compute  $\frac{g(\theta_i)}{q(\theta_i)}$  for each  $i = 1, \dots, N$  and estimate  $c^{-1} = E_q \left( \frac{g(\theta)}{q(\theta)} \right)$  by

$$\tilde{I} = \frac{1}{N} \sum_{i=1}^N \frac{g(\theta_i)}{q(\theta_i)}. \quad (4)$$

- (c) **In the case when  $c$  is unknown, suggest a sensible way of using  $\theta_1, \dots, \theta_N$  to estimate  $E_f(h(\theta))$ .**

By the Law of Large Numbers,  $\hat{I}$ , as given by equation (3), almost surely converges to  $E_q \left( \frac{h(\theta)g(\theta)}{q(\theta)} \right)$  whilst  $\tilde{I}$ , as given by equation (4), almost surely converges to  $c^{-1}$  so that a sensible estimate of  $E_f(h(\theta))$  is

$$\check{I} = \frac{\hat{I}}{\tilde{I}} = \frac{\sum_{i=1}^N h(\theta_i) w_i}{\sum_{i=1}^N w_i}$$

where  $w_i = g(\theta_i)/q(\theta_i)$ . Estimating  $E_f(h(\theta))$  by  $\check{I}$  is often called **self-normalised importance sampling** and provides a method of estimating posterior expectations without needing to know the normalising constant  $c$ . It can be shown that (provided that  $q(\theta) > 0$  whenever  $g(\theta) > 0$ )  $\check{I}$  converges to  $E_f(h(\theta))$ .

### 3. Let $X \sim N(0, 1)$ .

- (a) **Explain how to use Monte Carlo integration to find  $P(X > a)$  for  $a \in \mathbb{R}$ .**

Note that

$$\begin{aligned} P(X > a) &= \int_a^{\infty} f(x) dx \\ &= \int_{-\infty}^{\infty} \mathbb{I}_{\{x > a\}} f(x) dx \\ &= E_f(\mathbb{I}_{\{X > a\}}) \end{aligned}$$

where  $\mathbb{I}_{\{X > a\}}$  is the indicator function of the event  $\{X > a\}$  and  $E_f(\cdot)$  denotes expectation with respect to the distribution  $f$ . We thus may use Monte Carlo integration to estimate  $P(X > a)$  as follows:

- (i) sample  $n$  values  $x_1, \dots, x_n$  independently from  $f(x)$ , the  $N(0, 1)$
  - (ii) compute  $\mathbb{I}_{\{x_i > a\}}$  for each  $x_i$ , so  $\mathbb{I}_{\{x_i > a\}} = 1$  if  $x_i > a$  and is 0 otherwise
  - (iii) estimate  $P(X > a) = E_f(\mathbb{I}_{\{X > a\}})$  by the sample mean of the  $\mathbb{I}_{\{x_i > a\}}$ ,  $\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{x_i > a\}}$ .
- (b) **What are the difficulties with the method for large values of  $a$ ?**

We can write a function `monte` in R to estimate  $P(X > a)$ .

```

# Calculates m replications of the Monte Carlo
# integration in part (a) with a = y and n = n1
monte <- function(y,n1,m){
  vect<-c(1:m)
  for (i in 1:m)
  {
    x <- rnorm(n1)
    total <- length(x[x > y])
    vect[i] <- total/n1
  }
  return(vect)
}

```

We consider sampling  $n = 1000$  observations to estimate  $P(X > a)$  for  $a = 0, 1.96, 3$  and  $4.5$ . We'll repeat each experiment 1000 times to explore the sampling properties of the estimate and also calculate the number of samples which contained observations greater than  $a$ . For example, for  $a = 4.5$  we compute

```

> sample<-monte(4.5,1000,1000)
> mean(sample)
[1] 5e-06
> var(sample)
[1] 4.97998e-09
> length(sample[sample > 0])
[1] 5

```

We summarise our results.

| $a$  | $P(X > a)$                | Mean of<br>1000 samples | Variance of<br>1000 samples | No of non-<br>zero samples |
|------|---------------------------|-------------------------|-----------------------------|----------------------------|
| 0    | 0.5                       | 0.499914                | 0.000245                    | 1000                       |
| 1.96 | 0.0249979                 | 0.025027                | $2.437865 \times 10^{-5}$   | 1000                       |
| 3    | 0.001349898               | 0.001395                | $1.394369 \times 10^{-6}$   | 754                        |
| 4.5  | $3.397673 \times 10^{-6}$ | $5 \times 10^{-6}$      | $4.97998 \times 10^{-9}$    | 5                          |

As  $a$  increases then  $P(X > a)$  decreases and we are interested in estimating the probability of an increasingly rare event. If we let  $t = \sum_{i=1}^n \mathbb{I}_{\{x_i > a\}}$  then  $t$  denotes the number of the  $n$  observations that were greater than  $a$  and our estimate of  $P(X > a)$  is  $t/n$ . For large  $a$ , we will need to take  $n$  to be extremely large to even observe a non-zero  $t$  let alone obtain a stable estimate. For example, in our 1000 samples with  $n = 1000$  and  $a = 4.5$ , only five of the samples contained an observation greater than 4.5. An alternate approach to estimating  $P(X > a)$  which doesn't require  $n$  to be extremely large may be to use importance sampling where we sample from a distribution that puts more mass on the event  $\{X > a\}$ . Two possible approaches are considered in parts (c) and (d).

- (c) **Suppose instead we consider finding  $P(X > a)$  using the method of importance sampling where we sample from  $N(\mu, 1)$ . By considering the cases when  $a = 3$  and  $\mu = 4$  and that when  $a = 4.5$  and  $\mu = 4.5$  comment on the advantages of this approach over direct Monte Carlo integration.**

Let  $q(x)$  denote the probability density function of the  $N(\mu, 1)$ . Then, as  $q(x) > 0$

for all  $x \in (-\infty, \infty)$ , we have

$$\begin{aligned} E_f(\mathbb{I}_{\{X>a\}}) &= \int_{-\infty}^{\infty} \mathbb{I}_{\{x>a\}} f(x) dx = \int_{-\infty}^{\infty} \left( \frac{\mathbb{I}_{\{x>a\}} f(x)}{q(x)} \right) q(x) dx \\ &= E_q \left( \frac{\mathbb{I}_{\{X>a\}} f(X)}{q(X)} \right) \end{aligned}$$

where

$$\frac{f(X)}{q(X)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}X^2\right\}}{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(X-\mu)^2\right\}} = \exp\left\{\frac{\mu^2}{2} - \mu X\right\}.$$

Thus,

$$E_f(\mathbb{I}_{\{X>a\}}) = E_q \left( \mathbb{I}_{\{X>a\}} \exp\left\{\frac{\mu^2}{2} - \mu X\right\} \right).$$

We thus use importance sampling to estimate  $P(X > a) = E_f(\mathbb{I}_{\{X>a\}})$  as follows:

- (i) sample  $n$  values  $x_1, \dots, x_n$  independently from  $q(x)$ , the  $N(\mu, 1)$
- (ii) compute  $\mathbb{I}_{\{x_i>a\}} \exp\left\{\frac{\mu^2}{2} - \mu x_i\right\}$  for each  $x_i$
- (iii) estimate  $P(X > a) = E_f(\mathbb{I}_{\{X>a\}})$  by

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{x_i>a\}} \exp\left\{\frac{\mu^2}{2} - \mu x_i\right\}.$$

We write a function `import` in R to implement this algorithm.

```
# Calculates m replications of the importance sampling
# in part (c) with a = y1, mu = y2 and n = n1
import <- function(y1,y2,n1,m)
{
  vect <- c(1:m)
  for (i in 1:m)
  {
    x <- rnorm(n1,y2)
    weight <- dnorm(x)/dnorm(x, y2)
    vect[i] <- sum(weight[x > y1])/n1
  }
  return(vect)
}
```

Suppose we consider  $a = 3$  and  $\mu = 4$  so that  $P(X > 3) = 0.001349898$ . We compare the importance sampling method with direct Monte Carlo integration for various choices of  $n$ .

| Method | $n$  | Mean of<br>1000 samples | Variance of<br>1000 samples | No of non-<br>zero samples |
|--------|------|-------------------------|-----------------------------|----------------------------|
| Monte  | 100  | 0.00119                 | $1.16956 \times 10^{-5}$    | 113                        |
| Import | 100  | 0.001349189             | $9.391652 \times 10^{-8}$   | 1000                       |
| Monte  | 500  | 0.001232                | $2.308484 \times 10^{-6}$   | 472                        |
| Import | 500  | 0.001349657             | $1.949415 \times 10^{-8}$   | 1000                       |
| Monte  | 1000 | 0.001352                | $1.373469 \times 10^{-6}$   | 733                        |
| Import | 1000 | 0.001348711             | $9.348038 \times 10^{-9}$   | 1000                       |

We observe that, even for  $n = 100$  observations, the method of importance sampling is doing well and that the variance of the estimate is much reduced when compared to direct Monte Carlo integration. We repeat the process with  $a = 4.5$  and  $\mu = 4.5$  so that  $P(X > 4.5) = 3.397673 \times 10^{-6}$ .

| Method | $n$  | Mean of 1000 samples      | Variance of 1000 samples   | No of non-zero samples |
|--------|------|---------------------------|----------------------------|------------------------|
| Monte  | 100  |                           |                            | 0                      |
| Import | 100  | $3.42554 \times 10^{-6}$  | $6.096552 \times 10^{-13}$ | 1000                   |
| Monte  | 500  |                           |                            | 0                      |
| Import | 500  | $3.395555 \times 10^{-6}$ | $1.202796 \times 10^{-13}$ | 1000                   |
| Monte  | 1000 | $2 \times 10^{-6}$        | $1.997998 \times 10^{-9}$  | 2                      |
| Import | 1000 | $3.397125 \times 10^{-6}$ | $5.644111 \times 10^{-14}$ | 1000                   |

Notice that even for sizes of  $n$  when Monte Carlo fails (as we haven't observed any non-zero values), the importance sampling method is working well. When we can use Monte Carlo, for the same  $n$ , the variance of the estimate using importance sampling is vastly reduced.

- (d) **Explain how to find  $P(X > 4.5)$  using the method of importance sampling where we sample from the Exp(1) distribution truncated at 4.5 which has probability density function**

$$q(x) = \begin{cases} \exp\{-(x - 4.5)\} & x > 4.5, \\ 0 & \text{otherwise.} \end{cases}$$

In this part, the idea is to remove the need of the indicator function when estimating  $P(X > a)$  by sampling from a distribution,  $q(x)$ , whose support is  $(a, \infty)$  so that

$$P(X > a) = \int_a^\infty f(x) dx = \int_a^\infty \frac{f(x)}{q(x)} q(x) dx = E_q \left( \frac{f(X)}{q(X)} \right).$$

For  $a = 4.5$  we have that, for  $X > 4.5$ ,

$$\frac{f(X)}{q(X)} = \frac{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}X^2\right\}}{\exp\{-(X - 4.5)\}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}X^2 + X - 4.5\right\}$$

so that our importance sampling method is as follows:

- (i) sample  $n$  values  $x_1, \dots, x_n$  independently from  $q(x)$ , the Exp(1) distribution truncated at 4.5
- (ii) compute  $\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_i^2 + x_i - 4.5\right\}$  for each  $x_i$
- (iii) estimate  $P(X > a) = E_f(\mathbb{I}_{\{X > a\}})$  by

$$\hat{I} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x_i^2 + x_i - 4.5\right\}.$$

We write a function `import1` in R to implement this algorithm.<sup>1</sup> To demonstrate the convergence of the estimate in Figure 2 we plot the cumulative sum of  $\hat{I}$  for each  $n$ .

<sup>1</sup>This example is inspired by Example 3.5 of Robert, C. P. and Casella G. (2010). *Introducing Monte Carlo Methods with R*, Springer, New York.



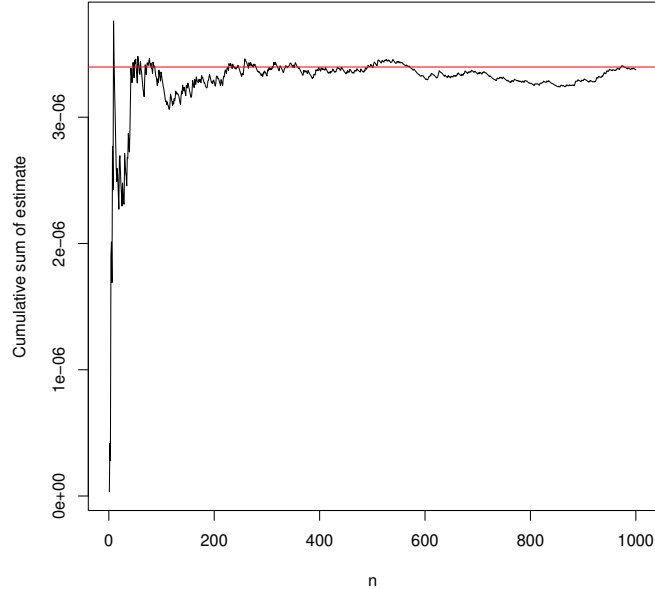


Figure 2: Convergence, as  $n$  increases, of the importance sampling estimate  $\hat{I}$  for  $P(X > 4.5)$  obtained by sampling from the  $\text{Exp}(1)$  distribution truncated at 4.5.

```
# Calculates the importance sampling in part (d) for
# P(X > y) and q(x) the Exp(1) truncated at y
# Plots the cumulative estimate of P(X > y) with n = n1
import1 <- function(y,n1){
  x<-rexp(n1)+y
  weight=dnorm(x)/dexp(x-y)
  plot(cumsum(weight)/1:n1,type="l",ylab="Cumulative sum of estimate",xlab="n")
  abline(a=pnorm(-y),b=0,col="red")
}
```

Figure 2 shows that the estimate is converging to the true value and that this happens quickly, certainly when compared with the direct Monte Carlo integration approach of part (a) whose limitations were noted in part (b).

4. Let  $E_f(g(X))$  denote the expectation of  $g(X)$  with respect to the probability density function  $f(x)$  so that

$$E_f(g(X)) = \int_X g(x)f(x) dx.$$

Suppose that we sample  $N$  independent observations from a distribution  $q(x)$  and use the method of importance sampling to estimate  $E_f(g(X))$ . Our estimator is thus

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N \frac{g(X_i)f(X_i)}{q(X_i)}.$$

**Prove that the choice of  $q(x)$  which minimises the variance of  $\hat{I}$  (with respect to  $q(x)$ ) is**

$$q^*(x) = \frac{|g(x)|f(x)}{\int_X |g(z)|f(z) dz}.$$

Let  $Var_q(\cdot)$  denote variance calculated with respect to  $q(x)$ . Then, as  $X_1, \dots, X_N$  are independent observations from  $q(x)$ ,

$$Var_q(\hat{I}) = \frac{1}{N^2} \sum_{i=1}^N Var_q \left( \frac{g(X_i)f(X_i)}{q(X_i)} \right) = \frac{1}{N} Var_q \left( \frac{g(X)f(X)}{q(X)} \right).$$

Hence, the choice of  $q(x)$  which minimises the variance of  $\hat{I}$  is equivalent to that which minimises

$$\begin{aligned} Var_q \left( \frac{g(X)f(X)}{q(X)} \right) &= E_q \left( \frac{g^2(X)f^2(X)}{q^2(X)} \right) - E_q^2 \left( \frac{g(X)f(X)}{q(X)} \right) \\ &= E_q \left( \frac{g^2(X)f^2(X)}{q^2(X)} \right) - E_f^2(g(X)). \end{aligned}$$

Thus, as  $E_f^2(g(X))$  does not depend upon the choice of  $q(x)$ , then we need only consider finding a  $q(x)$  minimising  $E_q \left( \frac{g^2(X)f^2(X)}{q^2(X)} \right)$ . Now, for any  $q(x)$ , as  $Var_q \left( \frac{|g(X)|f(X)}{q(X)} \right) \geq 0$ ,

$$E_q \left( \frac{g^2(X)f^2(X)}{q^2(X)} \right) \geq E_q^2 \left( \frac{|g(X)|f(X)}{q(X)} \right) = \left( \int_X |g(x)|f(x) dx \right)^2. \quad (5)$$

Notice that the lower bound does not depend upon the choice of  $q(x)$ . Thus, if there exists a  $q(x)$  which achieves this minimum bound then it must minimise the variance of  $\hat{I}$ .

If  $q^*(x) = \frac{|g(x)|f(x)}{\int_X |g(z)|f(z) dz}$  then

$$\frac{|g(X)|f(X)}{q^*(X)} = \int_X |g(z)|f(z) dz$$

so that, as  $\int_X |g(z)|f(z) dz$  is a constant,

$$E_q \left( \frac{g^2(X)f^2(X)}{(q^*(X))^2} \right) = E_q \left( \left( \int_X |g(z)|f(z) dz \right)^2 \right) = \left( \int_X |g(z)|f(z) dz \right)^2$$

and the lower bound in (5) is achieved showing that  $q^*(x)$  minimises the variance of  $\hat{I}$ .<sup>2</sup>

---

<sup>2</sup>Notice that this is really a theoretical rather than a practical result. For example, if  $g(x) > 0$  then  $q^*(x) = (g(x)f(x))/E_f(g(X))$  and so requires  $E_f(g(X))$  which is the integral we're trying to evaluate.