MATH5835M Statistical Computing
Exercise Sheet 1 (answers)

https://www1.maths.leeds.ac.uk/~voss/2023/MATH5835M/
Jochen Voss, J.Voss@leeds.ac.uk
2023/24, semester 1

Answer 1. The standard uniform distribution has density $\varphi(u) = 1_{[0,1]}(u)$ and thus we get

$$\mathbb{E}(X) = \mathbb{E}\bigl(U^4\bigr) = \int_{-\infty}^{\infty} u^4 \, \varphi(u) \, du = \int_0^1 u^4 \, du = 1^5/5 - 0^5/5 = 1/5.$$

To estimate this expectation using R, we can use the following commands:

```
> N <- 1e6
> U <- runif(N)
> X <- U^4
> mean(X)
[1] 0.1999706
```

Answer 2. In this question we want to estimate $p = P\bigl(\sin(X) > 1/2\bigr)$ where $X \sim \mathcal{N}(0,1)$, using Monte Carlo. A basic way to obtain an estimate is as follows:

```
> N <- 1e6
> X <- rnorm(N)
> p <- mean(sin(X) > 1/2)
> p
[1] 0.296202
```

It is not quite clear what "an estimate for p which is correct to n decimal places" means exactly. Here we consider the criterion $\mathrm{RMSE}(Z_N^{\mathrm{MC}}) \leq 10^{-n}$. A more sophisticated approach would be to consider confidence intervals instead.

We know $\mathrm{RMSE}(Z_N^{\mathrm{MC}}) = \sqrt{\mathrm{Var}\bigl(1_{\{\sin(X)>1/2\}}\bigr)/N}$, and thus we have $\mathrm{RMSE}(Z_N^{\mathrm{MC}}) \leq 10^{-n}$, if and only if $N \geq 10^{2n} \,\mathrm{Var}\bigl(1_{\{\sin(X)>1/2\}}$. Estimating the variance numerically, we find:

```
> var(sin(X) > 1/2)
[1] 0.2084666
> N.min <- ceiling(10^(2*(1:6)) * var(sin(X) > 1/2))
> N.min
[1]           21          2085        208467       20846659    2084665837
[6] 208466583663
```

This shows that only 25 samples are required to get the RMSE below 0.1, but nearly 250 billion samples are required to get the RMSE below 10^{-6}.

To find out how long the computation would take on my laptop, I time the case of $N = 10^8$:

```
> library(tictoc)
> tic()
> N <- 1e8
> X <- runif(N)
> p <- mean(sin(X) > 1/2)
> toc()
5.564 sec elapsed
```

Thus, on my laptop, $N = 10^8$ takes around 5 seconds. If we assume that running time is proportional to N, getting the RMSE below 10^{-6} would take $2.5 \cdot 10^{11}/10^8 \cdot 5/60/60 \approx 3.5$ hours. In reality, probably my laptop does not have enough memory to store 250 billion samples and the calculation may crash.

Answer 3. To get the estimate:

```
> N <- 1e6
> X <- rnorm(N)
> Y <- rnorm(N)
> mXY <- pmax(X, Y)
> Z <- mean(mXY)
> Z
[1] 0.5639553
```

The get the error:

```
> RMSE <- sd(mXY) / sqrt(N)
> RMSE
[1] 0.0008246609
```

To check whether the sample size was large enough:

```
> RMSE < 0.01 * Z
[1] TRUE
```