Monte Carlo Methods for Uncertainty Quantification: Practical 1

1. Suppose that a and b are two scalar random variables with zero mean, i.e. $\mathbb{E}[a] = \mathbb{E}[b] = 0$.

   Note that
   $$\mathbb{E}[\,(a - \lambda\,b)^2\,] \;\geq\; 0$$
   for any constant λ.

   By considering
   $$\lambda = \pm\sqrt{\mathbb{E}[a^2]/\mathbb{E}[b^2]}$$

   prove that
   $$|\mathbb{E}[ab]| \leq \sqrt{\mathbb{E}[a^2]\;\mathbb{E}[b^2]}.$$

   Hence, prove that for any two scalar random variables x and y, the correlation
   $$\rho \equiv \frac{\mathbb{E}[\,(x - \mathbb{E}[x])\,(y - \mathbb{E}[y])\,]}{\sqrt{\mathbb{V}[x]\,\mathbb{V}[y]}}$$
   has magnitude $|\rho| \leq 1$.

2. Suppose that a and b are again two scalar random variables with zero mean.

   Using the result from the previous question, prove that
   $$\sqrt{\mathbb{E}[\,(a+b)^2\,]} \;\leq\; \sqrt{\mathbb{E}[a^2]} + \sqrt{\mathbb{E}[b^2]}.$$

   Hence, prove that for any two scalar random variables x and y,
   $$\sqrt{\mathbb{V}[x+y]} \;\leq\; \sqrt{\mathbb{V}[x]} + \sqrt{\mathbb{V}[y]}.$$

3. Suppose that $a_1, a_2, \ldots, a_N$ are independent samples from the same distribution with zero mean and variance σ^2, i.e. $\mathbb{E}[a] = 0$, and $\mathbb{V}[a] = \mathbb{E}[a^2] = \sigma^2$.

   If we define
   $$\bar{a}_N \equiv N^{-1}\sum_{n=1}^{N} a_n, \quad \widetilde{\sigma}^2 \equiv N^{-1}\sum_{n=1}^{N}(a_n - \bar{a}_N)^2$$

   then prove that
   $$\widetilde{\sigma}^2 = \left(N^{-1}\sum_{n=1}^{N} a_n^2\right) - \bar{a}_N^2,$$

   and
   $$\mathbb{E}[\widetilde{\sigma}^2] = \frac{N-1}{N}\,\sigma^2.$$

4. The probability density function for the Normal distribution with zero mean and unit variance is

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{1}{2}x^2\right)$$

Note that the integral of the p.d.f. is equal to 1, the mean is zero, and the variance is 1, so we have

$$
\begin{aligned}
\mathbb{E}[1] &= \int_{-\infty}^{\infty} \phi(x)\, \mathrm{d}x = 1, \\
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x\,\phi(x)\, \mathrm{d}x = 0, \\
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2\,\phi(x)\, \mathrm{d}x = 1.
\end{aligned}
$$

Use integration by parts to prove that for $n > 2$,

$$\mathbb{E}[X^n] = (n-1)\,\mathbb{E}[X^{n-2}],$$

and hence determine $\mathbb{V}[X^2]$.

5. (a) In Matlab, generate 10^6 independent unit Normal random variables using **randn**, and then use these to estimate the value of

$$\mathbb{E}[X^2 + X + 1]$$

where X is a unit Normal random variable.

(b) Determine the error in the estimate.

(c) How does this compare to the error bound given by the Central Limit Theorem?

(Note that you can use the result from the previous question to determine $\mathbb{V}[X^2 + X + 1]$ which will help you to compute the variance of the Monte Carlo estimate.)

2

6. Let U be uniformly distributed on $[0, 1]$. You are to use Monte Carlo simulation to estimate the value of

$$\overline{f} = \mathbb{E}[f(U)] = \int_0^1 f(U) \, dU$$

where

$$f(x) = x \, \cos \pi x.$$

(a) Calculate analytically the exact value for $\overline{f}$ and

$$\sigma^2 = \mathbb{E}[(f(U) - \overline{f})^2] = \int_0^1 (f(U) - \overline{f})^2 \, dU$$

(b) Using the Matlab `rand` function, write a Matlab program which computes

$$Y_m = N^{-1} \sum_{n=1}^N f(U^{(m,n)})$$

for 1000 different sets of 1000 independent random variables $U^{(m,n)}$.

(c) Sort the Y_m into ascending order, and then plot $C_m = (m - 1/2)/1000$ versus Y_m – this is the numerical cumulative distribution function.

Superimpose on the same plot the cumulative distribution function you would expect from the Central Limit Theorem (using the Matlab `normcdf` or `norminv` functions) and comment on your results.

You may like to experiment by trying larger or smaller sets of points to improve your understanding of the asymptotic behaviour described by the CLT.

(d) Modify your code to use a single set of 10^6 random numbers, and plot

$$Y_N = N^{-1} \sum_{n=1}^N f(U^{(n)})$$

versus N for $N = 10^3 - 10^6$. This should demonstrate the convergence to the true value predicted by the Strong Law of Large Numbers.

Following the Matlab code in lecture 1, for each N, also compute an unbiased estimate for the variance σ^2 and hence add to the plot upper and lower confidence bounds based on 3 standard deviations of the variation in the mean.

Add a line corresponding to the true value. Does this lie inside the bounds?