Statistical Data Analysis
Problem sheet #3
Due Monday, 31 October 2022

Exercise 1 [5 marks]: Suppose the random variable x is uniformly distributed in the interval $[\alpha, \beta]$, with $\alpha, \beta > 0$. Find the expectation value of $1/x$, and compare the answer to $1/E[x]$ using $\alpha = 1$, $\beta = 2$.

Exercise 2 [8 marks]: Consider a discrete vector random variable $\mathbf{n} = (n_1, \ldots, n_M)$ that follows a multinomial distribution with N trials and probabilities for the individual outcomes $\mathbf{p} = (p_1, \ldots, p_M)$ with N trials:

$$P(\mathbf{n}; \mathbf{p}, N) = \frac{N!}{n_1! \cdots n_M!} p_1^{n_1} \cdots p_M^{n_M} \ .$$

Recall the expectation value and covariance for the multinomial distribution are

$$\begin{aligned} E[n_i] &= N p_i \ , \\ \mathrm{cov}[n_i, n_j] &= N p_i (\delta_{ij} - p_j) \ . \end{aligned}$$

Let the random variable u be defined as the first K of the M possible outcomes, with $K < M$:

$$u = \sum_{k=1}^{K} n_k \ .$$

Using error propagation, find the variance of u. Show that this is equal to the variance of a binomially distributed variable with $p_K = \sum_{i=1}^{K} p_i$ and N trials.

Exercise 3 [7 marks]: Consider a continuous random variable x that follows the pdf $f(x)$ with cumulative distribution $F(x)$, and suppose r follows a uniform distribution on $[0, 1]$. Prove (as was claimed in the lectures) that if we set $F(x) = r$ and solve for x, that $x(r)$ follows the pdf $f(x)$. To do this, use the method discussed in the lectures for finding the pdf of a function, and use the inverse function theorem, which says that

$$\frac{d}{dr} F^{-1}(r) = \frac{1}{\frac{dF}{dx}(x(r))} \ .$$

Exercise 4 (computing warm up [0 marks]): There is nothing to turn in for this exercise – it is just a warm-up exercise to ensure that you have your computing environment set up.

Starting with `simpleMC.py`, `simpleMC.ipynb` or `simpleMC.cc` from the course website, generate 10000 random values uniformly distributed between 0 and 1 and display the result as a histogram with 100 bins. (This is what simpleMC already does; you just need to ensure that you can run it.)