MATH70093– Computational Statistics; Assessed coursework 1—Autumn 2021
Due Monday 1st of November 2021 – deadline specific to your group

Upload your final version only - once the report is uploaded there is no option for re-uploading. Avoid last minute uploads. **Hand-in no more than 8 pages.** Provide your `R` code in the appendix; the appendix does not count towards the page limit. Considerable emphasis will be put on clarity of expression and a clean presentation. Only detailed, well-written answers that clearly explain your reasoning will score highly. The quality and suitability of the chosen methods will be taken into account in the distribution of marks in all questions.

1. *(30% of total marks)* Denote the last 2 digits of your CID as a_1 and a_2, e.g, if your CID is 00268914, then take $a_1 = 1$ and $a_2 = 4$. Let f be a pdf given by

$$f(x) = k\,(1 + cos(x)^3)\,\exp\left(-\frac{a_1+1}{10}x^2\right) \quad \text{for } -1 - a_2/2 \le x \le 1 + a_2/2$$

and $f(x) = 0$ otherwise; where $|x|$ denotes the absolute value of x.

   (a) Use numerical integration to estimate the constant k. Report the estimate.

   (b) Devise an acceptance-rejection algorithm to generate samples from f using a Normal proposal distribution with mean 0 and variance $5/(a_1+1)$. Clearly state how you derived your algorithm. Compare the empirical cdf from 1000 samples generated from this sampler to the true cdf by plotting both curves on the same graph. You can treat a good approximation using numerical integration as the true cdf.

   (c) Let X be a random variable following the distribution with pdf f. Use Monte Carlo integration based on a sample size of 10^4 to compute $E[X^2]$. Compute a 90% confidence interval for your Monte Carlo estimate.

   (d) How many samples from the normal proposal distribution in question (b) would you need to generate in order to obtain a 90% confidence interval with length at most equal to 10^{-4}?

2. *(30% of total marks)* Suppose we are interested in estimating

$$I(\sigma) = \frac{1}{2}\int_{-\infty}^{\infty} x\left[\frac{1}{\pi(1+x^2)} + \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-5)^2}{2\sigma^2}\right)\right]dx \quad \text{for } \sigma \in \mathcal{S} := \{0.01, 0.1, 0.5, 1\}.$$

   (a) Compute the integrals $\{I(\sigma)\}_{\sigma\in\mathcal{S}}$ using the function `integrate`. Comment on the results. Implement a numerical method that gives (roughly) the correct answers.

   (b) Estimate $\{I(\sigma)\}_{\sigma\in\mathcal{S}}$ using importance sampling with a normal distribution with mean 0 and standard deviation 4 as the importance sampling distribution. Report the estimate using a sample size of 10^4.

   (c) Describe how Monte-Carlo integration can be used to estimate $\{I(\sigma)\}_{\sigma\in\mathcal{S}}$. Report the estimate using a sample size of 10^4.

   (d) Compare the performance of the samplers in questions (b) and (c); you might want to compare the sampling errors of the estimates. Discuss your findings.

3. *(30% of total marks)* Consider the two-dimensional density

$$f(x,y) = k\left(1 + \frac{(x-2y)^2}{4}\right)^{-5/2}\left(1 + \frac{(x+2y)^2}{3}\right)^{-2}.$$

   Suppose we are interested in computing $E[(X-Y)^2]$. Describe and implement an importance sampler with proposal distribution being $X, Y \sim \mathcal{N}(0,1)$ independently. Discuss its performance. Suggest and implement at least one possible improvement, discussing its performance. Explain the reasons behind your observation.

4. *(30% of total marks)* Suppose we are interested in estimating $I = E[X\,\mathbb{1}\{X \ge 1\}]$ where $X \sim \mathcal{N}(0,1)$.

   (a) Estimate I using Monte-Carlo integration.

   (b) Estimate I using an importance sampler, with a proposal distribution of your choice. Justify your choice.

   (c) Describe how you could use the technique of *control variate* to estimate I assuming that you know the value of $P(X \ge 1)$. Implement the approach. You can obtain $P(X \ge 1)$ using the function `pnorm`.

   (d) Compare the sampling errors of the estimates obtained in questions (a), (b) and (c).