

# Assignment 1: Algorithms Galore!

Computational Statistics  
Instructor: Luiz Max de Carvalho

November 5, 2021

**Hand-in date: 17/11/2021.**

## General guidance

- State and prove all non-trivial mathematical results necessary to substantiate your arguments;
- Do not forget to add appropriate scholarly references *at the end* of the document;
- Mathematical expressions also receive punctuation;
- Please hand in a single PDF file as your final main document.  
Code appendices are welcome, *in addition* to the main PDF document.

## Background

By now we have seen quite a few methods for computing integrals *via* Monte Carlo. Each method has its own advantages and drawbacks. It is important that we understand these properties in order to apply the methods effectively. In this assignment we will continue studying the problem of computing the average distance between two points on a disk, this time from the perspective of method comparison. That is to say that in this assignment you will experience the microcosm of comparing several methods for solving a problem for which we happen to know the right answer in closed-form.

Recall that we want to compute

$$\begin{aligned} I &= \frac{1}{\pi^2 R^4} \int_0^R \int_0^R \int_0^{2\pi} \int_0^{2\pi} \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \phi(\theta_1, \theta_2)} r_1 r_2 d\theta_1 d\theta_2 dr_1 dr_2, \\ &= \frac{2^7}{45\pi} R, \end{aligned}$$

where  $\phi(\theta_1, \theta_2)$  is the central angle between  $r_1$  and  $r_2$ .

Here we will do something a bit risky: we will compare a few methods to compute  $I$  using a bunch of different methods without knowing in advance which will work best or if there are going to be any differences at all. Welcome to Science!

## Methods

Here we will list a selection of methods that will be randomly assigned to each student, along with some questions that need to be answered for that particular method.

- **Rejection sampling**

- Justify your choice of proposal distribution and show that it conforms to the necessary conditions for the algorithm to work; in particular, try to find a proposal that gives the highest acceptance probability.

- **Importance sampling**

- Justify your choice of proposal based on the variance of the resulting estimator.

- **Gibbs sampling**

- Write your full conditionals out and show that they adhere to the Hammersley-Clifford condition.

- **Metropolis-Hastings**

- Justify your choice of proposal; test different ones if you need to.

- **Static Hamiltonian Monte Carlo**

- Comment on the choice of step size ( $\varepsilon$ ) and integration time ( $\tau$ ).

## Questions

1. You have been (randomly) assigned a method from the previous section. Represent  $I$  as  $\int_{\mathcal{X}} \phi(x) \pi(x) dx$  and justify your choice of  $\phi$ ,  $\pi$  and  $\mathcal{X}$ . Recall that these choices are arbitrary up to a point, but they might lead to wildly different empirical performances **and** theoretical properties for estimators of  $I$ . **Justify** your choices in light of the method you have been given to work with. Choose wisely and be rigorous in your justifications.
2. Again, starting from the eventual samples you will obtain with your method, construct a non-empty<sup>1</sup> family of estimators of  $I$  and discuss whether it is (strongly) consistent and whether a central limit theorem can be established.
3. Detail a suite of diagnostics that might be employed in your application to detect convergence or performance problems. Extra points for those who design algorithms that exploit the structure of this particular integration problem.
4. For each  $R \in \{0.01, 0.1, 1, 10, 100, 1000, 10000\}$ , perform  $M = 500$  runs from your simulation method and compute: (i) variance (ii) bias (iii) standard deviation of the mean (MCSE).
5. Can you identify one key quantity missing from the previous item? *Hint*: it bears relevance to the real world application of any computational method.

**Warning:** the questions in this assignment might seem deceptively simple; do not be fooled. I expect a lot of effort from you in making your method work the best it can. This entails loads of failed derivations and experiments, which you are encouraged to report in order to document the discovery process. Also, feel free to include answers to questions that have not been asked, if you feel they are relevant. Make loads of figures and tables and let your scientific imagination run wild! Good luck!

---

<sup>1</sup>This is a joke. It means you should come up with at least one estimator. But you might, and are even encouraged to, entertain more than one estimator.