O Brother, How Far Art Thou?

Computational Statistics
Instructor: Luiz Max de Carvalho

October 18, 2022

Hand-in date: 06/10/2021.

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

A large portion of the content of this course is concerned with computing high-dimensional integrals *via* simulation. Today you will be introduced to a simple-looking problem with a complicated closed-form solution and one we can approach using simulation.

Suppose you have a disc C_R of radius R. Take $p = (p_x, p_y)$ and $q = (q_x, q_y) \in C_R$ two points in the disc. Consider the Euclidean distance between p and q, $||p - q|| = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} = |p - q|$.

Problem A: What is the *average* distance between pairs of points in C_R if they are picked uniformly at random?

Part I: nuts and bolts

1. To start building intuition, let's solve a related but much simpler problem. Consider an interval $[0, s]$, with $s > 0$ and take $x_1, x_2 \in [0, s]$ *uniformly at random*. Show that the average distance between x_1 and x_2 is $s/3$.
2. Show that Problem A is equivalent to computing

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

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

   Hint: Draw a picture.
3. Compute I in closed-form.

   Hint: Look up *Crofton's mean value theorem* or *Crofton's formula.*

Part II – getting your hands dirty

Now we will move on to implementation.

Problem B: Employ a simulation algorithm to approximate I. Provide point and interval estimates and give theoretical guarantees about them (consistency, coverage, etc).

1. You have been (randomly) assigned a simulation method – see list at the end. Represent I as $\int_{\mathcal{X}} \phi(x)\pi(x) \, dx$ and justify your choice of ϕ, π 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[1] family of estimators of I and discuss whether it is (strongly) consistent and whether a central limit theorem can be established.

[1] 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.

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. estimator.

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.