Assignment 0: O Brother, How Far Art Thou?

Computational Statistics
Instructor: Luiz Max de Carvalho

September 21, 2021

Hand-in date: 06/10/2020.

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?

Questions

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 $s/3$.

   Solution. We want to calculate $\mathbb{E}[|x_1 - x_2|]$ such that $x_1, x_2 \overset{iid}{\sim} \text{Unif}(0, s)$. An extensive way to solve this problem is to introduce the random variable $Y = x_1 - x_2$ and derive its distribution and after its absolute mean. Let

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(x_1 \leq y + x_2).$$

   Suppose first that $y \geq 0$. Given x_2,

$$\mathbb{P}(x_1 \leq y + x_2 | x_2) = \begin{cases} \dfrac{y + x_2}{s} & 0 \leq x_2 \leq s - y \\ 1 & s - y < x_2 \leq s. \end{cases}$$

   Therefore,

$$\begin{aligned} F_Y(y) &= \int_0^{s-y} \frac{y + x_2}{s^2}\, dx_2 + \int_{s-y}^{s} \frac{1}{s}\, dx_2 \\ &= \frac{y(s-y) + 0.5(s-y)^2}{s^2} + \frac{y}{s} \\ &= \frac{(s-y)(y+s)}{2s^2} + \frac{y}{s} = -\frac{y^2}{2s^2} + \frac{y}{s} + \frac{1}{2}. \end{aligned}$$

   Now suppose $y < 0$. Then

$$\mathbb{P}(x_1 \leq y + x_2 | x_2) = \begin{cases} \dfrac{y + x_2}{s} & -y \leq x_2 \leq s \\ 0 & 0 \leq x_2 < -y. \end{cases}$$

   Therefore,

$$\begin{aligned} F_Y(y) &= \int_{-y}^{s} \frac{y + x_2}{s^2}\, dx_2 \\ &= \frac{ys + 0.5s^2}{s^2} - \frac{-y^2 + 0.5y^2}{s^2} \\ &= \frac{y^2}{2s^2} + \frac{y}{s} + \frac{1}{2}. \end{aligned}$$

Deriving these expressions we get the density with respect to the Lebesgue measure,

$$f_Y(y) = \begin{cases} \frac{1}{s} - \frac{y}{s^2}, & \text{if } 0 \le y \le s \\ \frac{1}{s} + \frac{y}{s^2}, & \text{if } -s \le y < 0. \end{cases}$$

Finally,

$$\begin{aligned} \mathbb{E}[|x_1 - x_2|] = \mathbb{E}[|Y|] &= \int_0^s \frac{y}{s} - \frac{y^2}{s^2}\, dy - \int_{-s}^0 \frac{y}{s} + \frac{y^2}{s^2}\, dy \\ &= \frac{s^2}{2s} - \frac{s^3}{3s^2} + \frac{s^2}{2s} - \frac{s^3}{3s^2} \\ &= \frac{s}{2} - \frac{s}{3} + \frac{s}{2} - \frac{s}{3} = \frac{s}{3}, \end{aligned}$$

as we wished to prove. Now I give a more geometric approach. Consider (x_1, x_2) uniformly distributed over $[0,1]^2$. Then, let $t \in [0, s]$,

$$\mathbb{P}(|x_1 - x_2| > t) = \mathbb{P}(x_1 > t + x_2) + \mathbb{P}(x_2 > t + x_1),$$

since they are disjoint events, since $t \ge 0$. Note that the first region is delimited by the straight line $x_1 = t + x_2$, $x_1 = s$, and $x_2 = 0$, what gives a triangle with points $(t, 0), (s, 0)$, and $(s, s-t)$ with area $(s-t)^2/2$. The second region is delimited by the straight line $x_2 = t + x_1$, $x_2 = s$, and $x_1 = 0$, what gives the triangle $(0, t).(0, s)$ and $(s - t, s)$, with area $(s - t)^2/2$. We conclude that

$$\mathbb{P}(|x_1 - x_2| > t) = \frac{(s-t)^2}{s^2} \text{ if } t \in [0, s],$$

0, if $t \ge s$ and 1 if $t \le 0$. This implies that

$$\mathbb{E}[|x_1 - x_2|] = \int_0^{+\infty} \frac{(s-t)^2}{s^2} 1_{\{t \le s\}}\, dt = -\frac{(s-t)^3}{3s^2}\bigg|_0^s = \frac{s}{3},$$

a simpler prove. □

2. Show that Problem A is equivalent to computing

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

   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*.

4. Propose a simulation algorithm to approximate I. Provide point and interval estimates and give theoretical guarantees about them (consistency, coverage, etc).