Probability for Data Science – Master's Degree in Data Science – A.Y. 2022/2023
Academic Staff: Francesca Collet, Paolo Dai Pra

PROJECT 1: ESTIMATING π

Abstract. The scope of the present project is to illustrate how Monte Carlo simulation can be used to estimate the value of π.

Estimate of π by means of a Poisson process on the plane

The mathematical constant π represents the ratio of a circle circumference to its diameter. We want to find an expression to evaluate in our simulation to approximate π. We know that by dividing the area of the circle by the area of its bounding square we get $\frac{\pi}{4}$. By simulating random numbers within the latter square, we get approximated values for the area of the circle and for the area of the square. Hence, we can estimate π by taking the ratio of simulated points that fall within the circle and the total number of simulated points and then multiplying by 4. In other words, we can estimate π by using the formula

$$\pi \approx 4 \cdot \frac{\text{number of simulated points within the circle}}{\text{total number of simulated points}}.$$

How do we generate random points in the square? We follow a Poisson process on the plane. A random set of points $X \subset \mathbb{R}^2$ is said to be a **Poisson process of density** $\lambda > 0$ on the plane if it satisfies the following conditions:

- Let $X(D)$ denote the random number of points of X inside a domain D. For mutually disjoint sets $D_1, D_2, \dots, D_k \subset \mathbb{R}^2$, the random variables $X(D_1), X(D_2), \dots, X(D_k)$ are independent.
- Let $|D|$ denote the area of a domain D. For any bounded $D \subset \mathbb{R}^2$ and any $k \geq 0$, it holds

  $$P\left(X(D) = k\right) = e^{-\lambda|D|} \frac{\left(\lambda|D|\right)^k}{k!},$$

  that is $X(D) \sim \mathrm{Poi}(\lambda|D|)$. In particular, as a consequence, the mean number of points of X in D is $E[X(D)] = \lambda|D|$.

Algorithm to simulate a Poisson process on a square

We describe a simple algorithm to simulate a Poisson process X of density $\lambda > 0$ on the squared domain $\Lambda = [0, T] \times [0, T]$, with $T > 0$. The algorithm is based on two key facts:

- the number of points in Λ has distribution $\mathrm{Poi}(\lambda T^2)$;
- conditional on $X(\Lambda)$ (number of points in Λ), locations of points $x \in X$ are independent and uniformly distributed on the domain Λ (see [2, p. 359]).

Input: size of the squared domain T;
density of the Poisson process λ;

Output: Euclidean coordinates of N points of the Poisson process;

Procedure

Step 1. Generate $N \sim \mathrm{Poi}(\lambda T^2)$.

Step 2. If $N = 0$, then stop; there are no points in Λ. Otherwise, generate $2N$ independent random numbers in $(0, 1)$: $U_1, U_2, \dots, U_N, V_1, V_2, \dots, V_N$.

Step 3. The Euclidean coordinates of the N points of the Poisson process are (U_iT, V_iT), for $i = 1, \dots, N$.

Remark. Step 2 above relies on the fact that we are dealing with a squared domain. In general, different domain shapes require different algorithms. For instance, in [1, p. 279], an algorithm to generate a homogeneous Poisson process on a circle is given.

Project

Set $\Lambda = [0,1] \times [0,1]$ (squared domain) and let Γ be the circle with center $\left(\frac{1}{2}, \frac{1}{2}\right)$ and radius $\frac{1}{2}$ (inscribed circle). Fix a value of $\lambda > 0$ and consider a Poisson process X of density λ on the square Λ. An estimator for π is given by

$$\widehat{\Pi} = 4 \cdot \frac{X(\Gamma)}{X(\Lambda)}.$$

The value of π can be estimated by exploiting the law of large numbers. Let M denote the number of independent realizations of a Poisson process of density λ on the square Λ and let $\widehat{\Pi}_j$ be the estimate of π relative to the j-th simulation. If M is *sufficiently large*, then we have the approximation

$$\pi \approx \frac{1}{M} \sum_{j=1}^{M} \widehat{\Pi}_j.$$

By running several simulations and collecting the results in appropriate plots, investigate the following problems:

- How the approximate value of π depends on λ. Perform the analysis for densities of the form $\lambda = 5^k$, with $k \in \{1, 2, \dots, 10\}$ (integer numbers from 1 to 10).
- How the standard deviation depends on λ. Show that the sample standard deviation

  $$S_M = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} \left(\widehat{\Pi}_j - \pi\right)^2}$$

  decreases to zero roughly as $\frac{c}{\sqrt{\lambda}}$, for some constant $c > 0$. Perform the analysis for densities of the form $\lambda = 5^k$, with $k \in \{1, 2, \dots, 10\}$ (integer numbers from 1 to 10).

References

[1] S.M. Ross. *Simulation.* Academic Press, 2006

[2] S.I. Resnick. *Adventures in stochastic processes.* Birkhäuser Boston, 1992