Probability for Data Science – Master's Degree in Data Science – A.Y. 2023/2024
Academic Staff: Francesca Collet, Paolo Dai Pra

PROJECT 3: RANDOM ASSIGNMENT PROBLEM

Abstract. The random assignment problem consists of allocating n jobs to an equal number of machines to minimize a random total cost. We aim to estimate the expected cost value associated with the optimal solution.

Assignment problem

Consider the task of choosing an assignment of n jobs to n machines in order to minimize the total cost of performing the n jobs. The basic input for the problem is an $n \times n$ matrix $C = (c(i,j))_{i,j=1}^n$, where $c(i,j)$ is viewed as the cost of performing job i on machine j, and the assignment problem is to determine the permutation σ on $\{1, 2, \ldots, n\}$ that minimizes the total cost

$$A_n(\sigma) = \sum_{i=1}^{n} c(i, \sigma(i)).$$

In particular, this project will focus on the solution of the random assignment problem, where the costs $c(i,j)$ are i.i.d. random variables with distribution $\mathrm{U}(0,1)$.

Metropolis-Hastings algorithm

Let $\beta > 0$ be a fixed real parameter. We construct the Metropolis-Hastings (discrete-time) Markov chain on the state space $S = \{\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\} | \sigma \text{ is a permutation}\}$, with stationary distribution

$$\pi_\beta(\sigma) = \frac{e^{-\beta A_n(\sigma)}}{Z_\beta}, \qquad \text{with} \qquad Z_\beta = \sum_{\sigma \in S} e^{-\beta A_n(\sigma)}.$$

Observe that the probability distribution π_β concentrates on the permutation achieving the minimal total cost as $\beta \to +\infty$. Therefore, if we choose β sufficiently large and we run the chain for a large number N of steps, we can take the state visited at time N as the optimal solution of the random assignment problem.

The following algorithm produces the first N steps $\sigma_1, \ldots, \sigma_N$ of the Metropolis-Hasting chain on S.

Input: value of the parameter β;
number of steps N;
initial state $\bar{\sigma} \in S$;

Output: trajectory of the Metropolis-Hastings chain starting at $\bar{\sigma}$;

<u>Procedure</u>

Step 1. Set $\sigma_0 = \bar{\sigma}$.

Step 2. For $t = 1, 2, \ldots, N-1$:

1. pick σ' uniformly at random in S;
2. set

$$\sigma_t = \begin{cases} \sigma' & \text{with probability } \min\left\{1, \dfrac{e^{-\beta A_n(\sigma')}}{e^{-\beta A_n(\sigma_{t-1})}}\right\} \\[2ex] \sigma_{t-1} & \text{with probability } 1 - \min\left\{1, \dfrac{e^{-\beta A_n(\sigma')}}{e^{-\beta A_n(\sigma_{t-1})}}\right\}. \end{cases}$$

Project

By implementing the Metropolis-Hastings algorithm above, we determine a minimizer of the function A_n, for any given realization of the matrix C. We want to study the asymptotic (in n) behavior of the average total cost associated with the optimal solution.
By running several simulations and collecting the results in appropriate plots, show that the expectation $E(A_n)$ approaches the value $\frac{\pi^2}{6}$, as n grows large. Perform the analysis for dimensions of the form $n = 5\alpha$, with $\alpha \in \{1, 2, \ldots, 10\}$ (integer numbers from 1 to 10).

Remark. The expected value $E(A_n)$ can be estimated by exploiting the law of large numbers. Let M denote the number of independent realizations of C. Moreover, let $A_n^{(j)}$ be the minimal total cost obtained by the j-th run of the Metropolis-Hastings algorithm, given $C^{(j)}$. If M is *sufficiently large*, then we have the approximation

$$E(A_n) \approx \frac{1}{M} \sum_{j=1}^{M} A_n^{(j)}.$$

References

[1] Levin D.A. and Peres Y., *Markov chains and mixing times*, Volume 107, American Mathematical Society, 2017

[2] Mézard M. and Parisi G., On the solution of the random link matching problems, *Journal de Physique*, 48(9):1451–1459, 1987

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