Revisit to A Prominent Example of The Curse of Dimensionality: Re-investigating Anomaly Behaviour of Monte Carlo Estimate of π in Higher Dimension

Prithwish Ghosh, Shinjon Chakraborty, Debashis Chatterjee

Prithwish Ghosh
Visva Bharati University
Department of Statistics

Introduction

The Curse of Dimensionality is responsible for gradually decreasing the probability of obtaining random points generated inside the hyper-sphere relative to the hyper-cube in a higher dimension. This paper presents a rigorous statistical theory behind a popular, nontrivial example of the so-called curse of dimensionality, namely, the anomaly behavior of the Monte Carlo estimation of π using p^{th} dimensional sphere and p^{th} dimensional cube as $p \to \infty$. We illustrate the theory in dimensions $p = 2, 3, 4, 5, 6, 7, 8, 9, 10$ by simulation.

Consider a p dimensional hypersphere $\mathbb{B}(_p, r)$, where the center $_p \in \Re^p$ and radius length $r > 0$. For the purpose of this study, without loss of generality, assume center $_p$ as the p dimensional coordinate origin $\mathbf{0}$. We use the notation $\mathbb{B}_p(r) := \mathbb{B}(\mathbf{0}, r)$ (p-ball). The volume of a p-ball is

$$vol\left(\mathbb{B}_p(r)\right) = \frac{\pi^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}+1\right)} r^p. \quad (1)$$

$\Gamma(\cdot)$ is Euler's Gamma Function. For p a positive integer, we use the notation $(-a/2, a/2)^p$ for the p-dimensional hypercube with the center at the origin where $a > 0$ is the length of any side of the p-dimensional hypercube (analogous to a square (in two dimensions) and cube(in three dimensions).The detailed table for the volumes is given in Table 1. Izenman, 2008

Concept of Volume in Higher Dimension(Cont...)

Theorem (Curse of Dimensionality)

Let $C_p = (2r)^p$ be the volume of the hypercube $(-r, r)^p$ in p^{th} dimension, and S_p be the volume of the inscribed hypersphere of radius r in p^{th} dimension.

Then

1. *The ratio $\frac{S_p}{C_p} = \frac{2\pi^{p/2} r^p / [p\Gamma(p/2)]}{(2r)^p}$ is independent of location and scale.*
2. $\lim_{p\to\infty} \frac{S_p}{C_p} = \frac{2\pi^{p/2} r^p / [p\Gamma(p/2)]}{(2r)^p} = \frac{\pi^{p/2}}{2^{p-1} p\Gamma(p/2)} = 0.$

Proof of Theorem 2.1.

1. Observe that,

$$\frac{S_p}{C_p} = \frac{2\pi^{p/2} r^p / [p\Gamma(p/2)]}{(2r)^p} = \frac{\pi^{p/2}}{2^{p-1} p\Gamma(p/2)},$$

the last term being independent of location and scale.

□

Remark

Most of the volume of the high-dimensional cube is located in its corner.

Table

Number of Dimension	Volume of p-ball of radius r	$\frac{S_p}{C_p}$ = Ratio of Hypersphere and circumscribed Hypercube of side 2r
0	1	1
1	2r	1
2	πr^2	$\frac{\pi}{4}$
3	$\frac{4\pi}{3} r^3$	$\frac{\pi}{6}$
4	$\frac{\pi^2}{2} r^4$	$\frac{\pi^2}{32}$
5	$\frac{8\pi^2}{15} r^5$	$\frac{\pi^2}{60}$
6	$\frac{\pi^3}{6} r^6$	$\frac{\pi^3}{384}$
7	$\frac{16\pi^3}{105} r^7$	$\frac{\pi^3}{840}$
8	$\frac{\pi^4}{28} r^8$	$\frac{\pi^4}{7168}$
9	$\frac{32\pi^4}{945} r^9$	$\frac{\pi^4}{15120}$
10	$\frac{\pi^5}{120} r^{10}$	$\frac{\pi^5}{3840}$

Table: Table of the volumes up to the 10^{th} dimension

Curse of Dimensionality

Definition

The curse of dimensionality refers to the challenges and issues that arise when working with high-dimensional data.

- As the number of dimensions increases, the amount of data needed to cover the space increases exponentially.
- Sparsity of data points in high-dimensional space.
- Increased computational complexity and resource requirements.
- Challenges in visualization and interpretation.

Figure

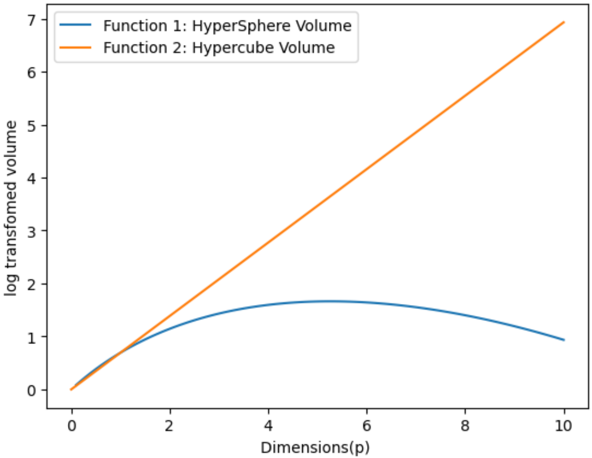

Figure: Pic till 10th dim log transformed

Monte Carlo Simulation Method & Convergence

The name Monte Carlo directs to the magnific casino in Monaco Monte Carlo. It is renowned around the world as an icon of gambling. The term "Monte Carlo" was introduced in 1947 by Nicholas Metropolis Metropolis, 1987. Sometimes referred to as stochastic simulation, Monte Carlo is a wide class of computational algorithms that depend on repeated random sampling to obtain numerical results. (seeJames, 1980; Rubinstein and Kroese, 2016; Harrison, 2010). A formal definition of Monte Carlo methods was given in Halton, 1970.

A Monte-Carlo simulation focus on constantly repeating random samples. Its theoretical convergence involves the law of large numbers integrals represented by the expected value of some random variable that can be approximated by taking the sample mean of independent samples of the variable.

Monte Carlo Simulation Method & Convergence(Cont...)

In this paper, we consider one of the simplest examples of Monte Carlo. We consider a uniform rainfall on the p^{th} dimensional hypercube experiment useful for computing π, where $p \in \mathbb{N}^+$ (set of all positive integers). Mathematically, consider random variable $Z \sim Binomial(N, q_p)$, with $N =$ the number of iterations,
$q_p = P$(drop within p^{th} dimensional hypercircle). (For instance, in two dimensions, the probability q_2 that a raindrop falls into the unit circle is $\frac{\pi}{4}$; see table 1.)

Discussion and Conclusion

We briefly restate the algorithm we adapted here to estimate π. For two dimensions, the idea is to generate random points (x, y) in the square $[-1, 1] \times [-1, 1]$ and see how many lie in the quarter circle $x^2 + y^2 < 1$. The proportion will approach $\pi/4$. The Monte-Carlo iteration results corresponding to the dart-throwing experiment (dimension $p = 2, 3$) are shown in Figure 5 and in Figure 9, respectively.

Figure

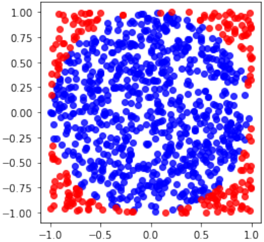

Figure: 1000 iterations

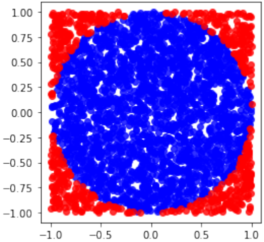

Figure: 3000 iterations

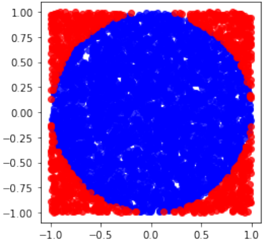

Figure: 5000 iterations

Figure: 2D Image of Monte Carlo simulation for Estimating Π

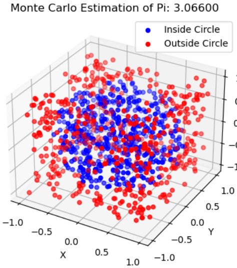

Figure: 1000 iterations

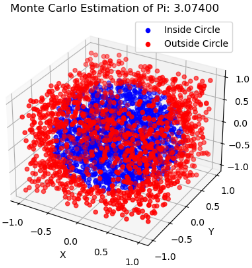

Figure: 3000 iterations

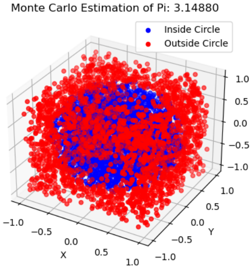

Figure: 5000 iterations

Figure: 3D Image of Monte Carlo simulation for Estimating Π

Discussion(Cont...)

We consider dimensions $p \in 1, 2, \cdots, 10$ and N number of iterations, $N = 1000, 3000, 5000$ in each p^{th} dimension, to observe the convergence. We randomly generate random points $X_1, X_2, \cdots, X_N$ uniformly from uniform distribution $(-1, 1)^p$. Consider the p dimensional hypersphere of radius 1. The ratio of the number of random points falling within the p dimensional hypersphere to the number of random points N generated on the Hypercube $(-1, 1)^p$, multiplied by a suitable constant for adjustments derived from the calculation of volume and Table 1. We get a reasonable estimate of π by allowing large enough samples (according to the Law of Large numbers); . The theory established, explains the poor estimate of π in large dimensions. As the consistency of the estimator of π is guaranteed, the anomaly behavior is clearly due to dimension curse.

Discussion(Cont...)

The estimated values of π for p dimensions (p = 2,3,4,5,6,7,8,9 and 10) correspond to 1000, 3000, and 5000 iterations is shown in figure 10, figure 11, and figure 12, respectively. The estimated values of π corresponding to 10000, 30000, and 50000 iterations, respectively, are listed in Table 2.

Table

Dim	Iterations	Est of π	Iterations	Est of π	Iterations	Est of π
2	1000	3.153692	3000	3.13257	5000	3.132347
3	1000	3.143712	3000	3.13590	5000	3.1295481
4	1000	3.156591	3000	3.13838	5000	3.1070465
5	1000	3.095292	3000	3.10380	5000	3.11321749
6	1000	3.181126	3000	3.0508	5000	3.0675510
7	1000	3.113422	3000	3.025760	5000	3.103500838
8	1000	3.414451	3000	3.2195	5000	3.20901322
9	1000	3.314694	3000	3.04191	5000	3.0859034900
10	1000	2.4051189	3000	2.42824	5000	2.41045306

Table: Table for all the dimensions for the given Number of Iterations

Figure

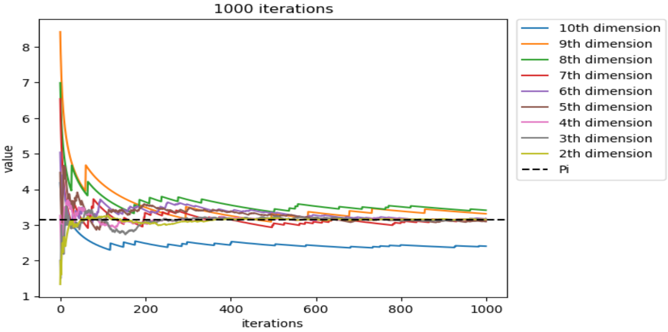

Figure: 1000 iterations

Figure

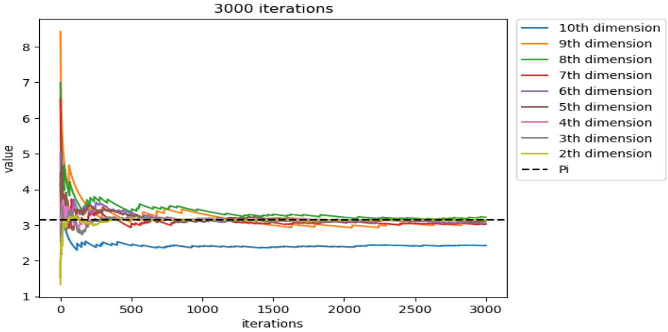

Figure: 3000 iteration

Figure

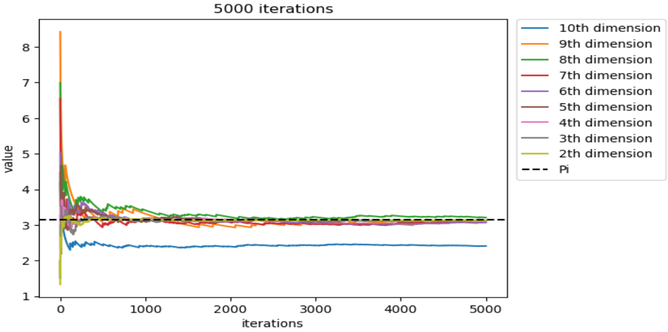

Figure: 5000 iteration

Figure 12 illustrates the rate of convergence of the simulation towards the actual value of π for 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th} and 7^{th} dimensions corresponding to 5000 iteration. For all the subsequent dimensions, the simulation is observed not to converge towards the actual value of π, despite showing initial deviation when moving towards higher iteration values. We study how much the estimated value differs from the original value of π using the mean absolute error measure (MAE) defined by:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - \pi|}{n}, \tag{2}$$

where y_i= Estimated Values, n = Sample Size. The results are tabulated in the table 3. The MAE values increase with dimension, mostly after the 5^{th} dimension.

Table

Dim	Iteration	MAE	Iteration	MAE	Iteration	MAE
2	1000	0.0120	3000	0.009	5000	0.0092
3	1000	0.0021	3000	0.0056	5000	0.012
4	1000	0.014	3000	0.0032	5000	0.0345
5	1000	0.04629	3000	0.0377	5000	0.0283
6	1000	0.0395	3000	0.0907	5000	0.07404
7	1000	0.02817	3000	0.1158	5000	0.03809
8	1000	0.272	3000	0.0779	5000	0.0674
9	1000	0.1731	3000	0.0996	5000	0.0556
10	1000	0.7364	3000	0.7133	5000	0.7311

Table: Table for MAE (1 to 10 dimensions and 1000, 3000, and 5000 iterations)

Conclusion

This paper presents a rigorous theoretical explanation behind a prominent simple example of dimension curse and illustrates the theory using simulation.

Reference

Izenman, A. J. (2008). *Modern multivariate statistical techniques* (Vol. 1). Springer.

Metropolis, N. (1987). The beginning. *Los Alamos Science, 15*, 125–130.

James, F. (1980). Monte carlo theory and practice. *Reports on progress in Physics, 43*(9), 1145.

Rubinstein, R. Y., & Kroese, D. P. (2016). *Simulation and the monte carlo method*. John Wiley & Sons.

Harrison, R. L. (2010). Introduction to monte carlo simulation. *AIP conference proceedings, 1204*, 17–21.

Halton, J. H. (1970). A retrospective and prospective survey of the monte carlo method. *Siam review, 12*(1), 1–63.