Monte Carlo Techniques for Statistical Analysis of Data

Jessica Landwersiek
Advanced Physics Laboratory, Florida International University
(Dated: June, 2021)

Monte Carlo methods were employed using acceptance/rejection and transformation techniques. The value of π was found to be 3.1404 with an uncertainty of 0.0016. Histograms of Poisson deviates were plotted with $\mu = 1$, $\mu = 10.3$, and $\mu = 102.1$, with a lower mean producing more accurate outcomes. Box-Muller transformations were used to find Gaussian deviates with reduced χ^2 of 1.097 and 1.054 for perpendicular distributions.

I. INTRODUCTION

Monte Carlo techniques serve as a tool for analyzing processes that would otherwise require difficult and inconvenient mathematical operations. In particular, integrating probability density functions will return valuable information, and these techniques serve as a way to evaluate multiple integrals by random number sampling [1]. In this experiment, computerized random number generators were used for multiple methods of simulating data in this fashion.

II. PROCEDURE

A. Finding π

A simple example of the acceptance/rejection method is to use a circle inscribed into a square and their corresponding equations to find a value for π. This is done by randomly choosing points within the square and finding the probability of those points landing within the circle. When we inscribe a circle within a square we can define the ratio of areas as:

$$\frac{A_{\text{circle}}}{A_{\text{square}}} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4} \tag{1}$$

The more iterations of random points, the closer this ratio is to the probability of points landing in the circle. This gives us:

$$\pi = 4\frac{N_{\text{circle}}}{N_{\text{total}}} \tag{2}$$

for a large number of counts N.

B. Volume of n-dimensional Hyperspheres

The volume of n-dimensional spheres/hyperspheres, from here on referred to as hyperspheres, with radius 1 can also be determined within reasonable accuracy with the acceptance/rejection method. Rather than a two-dimensional circle inscribed within a two-dimensional square, an n-dimensional hypersphere can be inscribed in an n-dimensional hypercube. As before, the probability of an event occurring (hits) can be described with the formula:

$$P(\text{hits}) = \frac{N_{\text{hits}}}{N_{\text{total}}} \tag{3}$$

where again the variable N is equal to iterations (counts). In multiple dimensions, the probability is proportional to the volume as it was with the area, but must be multiplied by 2^n, where 2 (the possible outcomes) is raised to the number of dimensions. We now have a formula for volume dependent on the iterations (hits and misses), and number of dimensions:

$$V = \frac{N_{\text{hits}}}{N_{\text{total}}} 2^n \tag{4}$$

This was used to calculate 3, 4, and 5-dimensional hypersphere volumes using 10,000 iterations for each volume.

C. Acceptance/Rejection Method

The acceptance/rejection method is used when there are two possible outcomes for an event. In this procedure, a function is defined and if the event falls within the boundaries of this function it is treated as a "hit." If it does not fall within the bounds of the curve, it is considered a "miss" [2]. Using standard geometric formulas to describe regions in space, high numbers of iterations, and probabilistic relationships, the value of π as well as the volumes of n-dimensional hyperspheres with radii of 1, were able to be determined with reasonable accuracy.

D. Transformation Method

The process for this method is to use a function of one random variable (r) and transform it into a second function of a different variable (x). Using probability density functions $p(r)$ and $P(x)$, where $p(r)$ is a uniform distribution on the interval (0,1), the Conservation of Probability defines the relationship between x and r:

$$|p(r)\Delta r| = |P(x)\Delta x| \tag{5}$$

Written in integral form:

$$\int_0^1 p(r)dr = \int_{-\infty}^{\infty} P(x)dx \tag{6}$$

Normalizing the integral for $p(r)$ and taking the definite integral of the left-hand side based on the upper limit, $r = 1$, we can see that there is a clear way to derive the new deviate:

$$\int_0^1 p(r)dr = \int_{-\infty}^{\infty} P(x)dx \tag{7}$$

Gaussian and Poisson distributions were used for $P(x)$ to find corresponding deviates. This is an important tool in analyzing data and simulating experimental results [3].

E. Poisson Deviates

To find Poisson deviates, equation (9) is used. The Poisson distribution is defined as:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{8}$$

Where x is a Poisson deviate and μ is the mean of the distribution. Inserting equation (8) into the formula:

$$r = \frac{x^{\mu x - \mu}}{x!} e \tag{9}$$

This process was repeated 10,000 times for $\mu = 1$, $\mu = 10.3$, and $\mu = 102.1$ and histograms were created to interpret the results.

III. RESULTS

A. Determining the Value of π

The derivation described above was programmed and, as expected, the experimental value of π became more accurate as the number of iterations increased. Using equations (5) and (6), the binomial uncertainty was calculated to find the error in derivation as follows:

$$\delta\pi = \sqrt{\frac{4\pi^2}{N_{\mathrm{c}}} \delta N_{\mathrm{c}}^2} \tag{10}$$

B. Box-Muller 2D Gaussian Transformation

The aim of this transformation is to generate two-dimensional Gaussian distributions (Cartesian) with mean and standard deviation $\mu = 0$ and $\sigma^2 = 1$ respectively. Computer-generated random samples (u, v) are used to generate x and y coordinates (random variables) in polar form:

$$x = \sqrt{-2\ln(u)}\cos(2\pi v) \tag{11}$$

$$y = \sqrt{-2\ln(u)}\sin(2\pi v) \tag{12}$$

In addition to binomial uncertainties, certain statistical uncertainties were desired. Using percent error and formula (15), N can be calculated directly as a function of statistical uncertainty.

$$N = (4-1)\pi^2 = (4-1)(1)^2 \tag{13}$$

TABLE I: Calculated values of π for different values of N.

N	Calculated π	$\delta\pi$
1000	2.972	0.052
10,000	3.170	0.016
100,000	3.1456	0.0052
500,000	3.1424	0.0023
1,000,000	3.1404	0.0016

C. Volume of n-dimensional Hyperspheres

The volumes of 3, 4, and 5-dimensional hyperspheres were determined using equation (7). This data was then compared to the geometric formulas for their corresponding volumes.

TABLE II: Comparison of derived hypersphere volumes and volumes calculated from geometric formulas.

Dimension	Calculated Volume	Geometric Volume
3	4.174 ± 0.040	4.189
4	4.912 ± 0.074	4.935
5	5.23 ± 0.12	5.26

D. Poisson Deviates Plots

Poisson deviates were generated for $\mu = 1$, $\mu = 10.3$, and $\mu = 102.1$. The histograms for these distributions are shown below. As expected, the distribution becomes more concentrated around the mean with higher values of μ.

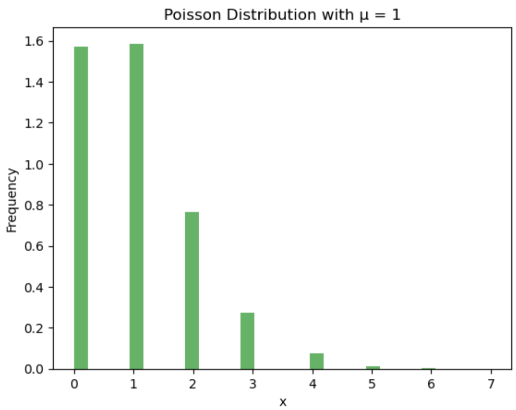

FIG. 1: Histogram of Poisson deviates for $\mu = 1$.

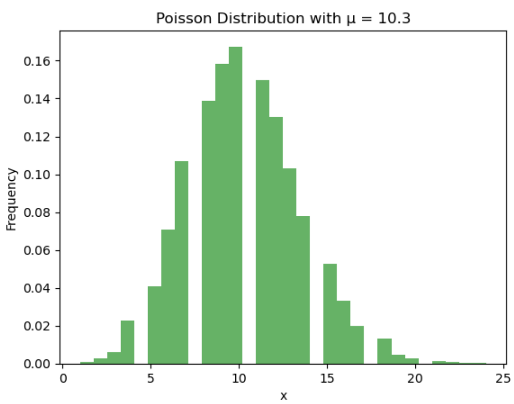

FIG. 2: Histogram of Poisson deviates for $\mu = 10.3$.

E. Gaussian Deviates Plots

The Box-Muller transformation was used to generate Gaussian deviates. The following plot shows the results for two perpendicular distributions.

IV. CONCLUSION

For cases where mathematical derivations are not attainable or convenient, Monte Carlo techniques can be employed as an effective way to interpret data. This is evident by the accuracy of results achieved in the procedure of this experiment. Mathematical models are an important tool in the evaluation of data with large numbers of events. Using Python for computer programming, random variables can be used to simulate distributions for random processes and give insight into what actual results will look like. The geometric derivations were an excellent exercise in understanding the usefulness of the acceptance/rejection technique. Using Poisson deviates for data with low values for the mean produces favorable results that become less useful as the mean increases. Gaussian deviates obtained using the Box-Muller method

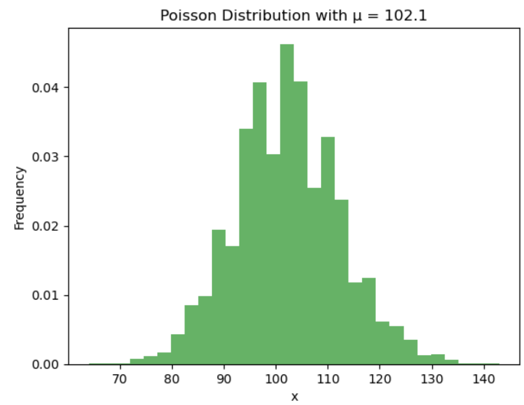

FIG. 3: Histogram of Poisson deviates for $\mu = 102.1$.

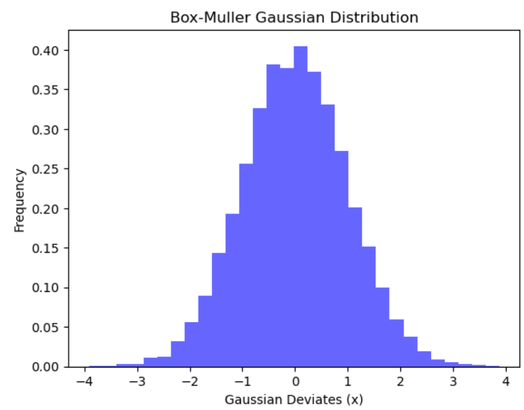

FIG. 4: Gaussian deviates (x coordinates) generated using the Box-Muller transformation.

resulted in desirable reduced χ^2 values. These procedures have various applications in the physical sciences, specifically for counting and quantum mechanical processes.

V. ERROR ANALYSIS

The accuracy of the Monte Carlo simulations presented in this experiment is subject to several sources of uncertainty, including statistical errors inherent in the random sampling process and limitations associated with the numerical methods employed.

For the calculation of π, the result of 3.1404 ± 0.0016 was obtained. The uncertainty in this value arises from the random nature of the sampling method and is inversely proportional to the square root of the number of iterations (N). As the number of iterations increases, the uncertainty decreases, as shown in Table I. Larger iterations lead to more accurate estimates of the ratio of areas in the acceptance/rejection method. The uncertainty in π can be expressed as:

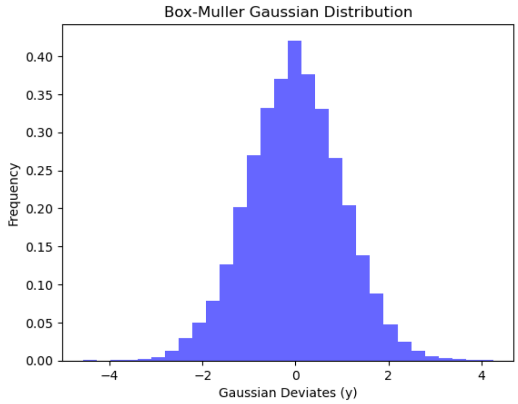

FIG. 5: Gaussian deviates (y coordinates) generated using the Box-Muller transformation.

$$\delta\pi = \frac{4\pi^2}{N\delta N^2}$$

In the volume calculations of n-dimensional hyperspheres, the results exhibited slight discrepancies from the exact geometric volumes. These differences can be attributed to the finite number of iterations used in estimating the volumes (10,000). The error in the calculated volumes was derived from the difference between the theoretical and computed values. For example, the calculated volume of the 3-dimensional hypersphere was 4.174 ± 0.040, which is reasonably close to the geometric value of 4.189. The uncertainty in the volumes can be represented as:

$$\delta V = \frac{V_{\text{calculated}} - V_{\text{geometric}}}{V_{\text{geometric}}}$$

For the Poisson deviates, simulations were run for various mean values ($\mu = 1$, $\mu = 10.3$, and $\mu = 102.1$). As expected, the distribution became more concentrated around the mean as μ increased. However, higher values of μ require larger sample sizes to maintain accuracy, and thus the statistical uncertainty becomes more significant. The precision of the Poisson distribution estimates is limited by this increased variance for larger values of the mean. The Poisson distribution is given by:

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

In the case of the Box-Muller transformation, the generated Gaussian deviates produced reduced χ^2 values of 1.097 and 1.054 for two perpendicular distributions. These values indicate a good fit to the expected Gaussian distribution, although minor deviations from the ideal fit are present. The uncertainty in the transformation is attributable to the finite number of samples used and the random nature of the underlying uniform distributions. The Box-Muller transformation for Gaussian deviates is given by:

$$x = \sqrt{-2\ln(u)}\cos(2\pi v)$$

$$y = \sqrt{-2\ln(u)}\sin(2\pi v)$$

In summary, the errors observed in the Monte Carlo simulations are predominantly due to statistical uncertainties associated with the random sampling process. These uncertainties decrease as the number of iterations increases, although practical limitations such as computational resources must be taken into account when optimizing the accuracy of the simulations.

VI. REFERENCES

1. P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (Kent A. Peterson, 2003).

2. J. K. Blitzstein and J. Hwang, *Introduction to Probability* (Taylor and Francis Group, LLC, 2019).

3. I. G. Hughes and T. P. A. Hase, *Measurements and their Uncertainties: A practical guide to modern error analysis* (Oxford University Press, Inc., New York, 2010).