Monte Carlo Techniques for Statistical Analysis of Data

Jessica Landwersiek[1]
Advanced Physics Laboratory, Florida International University

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 μ= 1, μ=10.3, and μ=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.

A. 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" [1]. 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.

B. Transformation Method

The process for this method is to use a function of one random variable (r) and transform in to 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:

$$|p(r)\triangle r| = |P(x)\triangle x| \tag{1}$$

Written in integral form:

$$\int_{r=-\infty}^{r} p(r)dr = \int_{x=-\infty}^{x} P(x)\,\mathrm{d}\,x \tag{2}$$

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:

$$\int_0^1 1\,\mathrm{d}\,r = r = \int_{x=-\infty}^{x} P(x)\,\mathrm{d}\,x \tag{3}$$

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 [1].

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:

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

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

$$\pi = 4\frac{N_{circle}}{N_{total}} \tag{5}$$

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:

$$P(hits) = \frac{N_{hits}}{N_{total}} \tag{6}$$

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:

$$V \; = \; 2^n\frac{N_{hits}}{N_{total}} \tag{7}$$

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

C. Poisson Deviates

To find Poisson deviates equation (3) is used. The Poisson distribution is defined as:

$$P(x) = \sum_{x=0}^{x} \frac{\mu^x}{x!} e^{-\mu} \tag{8}$$

Where x is a Poisson deviate and μ is the mean of the distribution. Inserting (8) into (3):

$$r = \sum_{x=0}^{x} \frac{\mu^x}{x!} e^{-\mu} \tag{9}$$

What this gives us is an upper limit for our random variable r. Given that the values for x in the Poisson function are discrete, a computer program can be written to find the value for x that makes the right-hand side of (9) just larger than a random variable r. Given that the x values in the Poisson function are discrete, X = x-1 is the Poisson deviate. This process was repeated 10,000 times for $\mu = 1, 10.3, \textit{ and } 102.1$ and histograms were created to interpret the results.

D. 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.

$$x = \sqrt{-2\ln(u)}\cos(2\pi v) = R\cos\theta \tag{10}$$

$$y = \sqrt{-2\ln(u)}\sin(2\pi v) = R\sin\theta \tag{11}$$

For random independent variables the combined probability density function is the product of the probability density of each function:

$$P(x,y) = P(x)P(y) \tag{12}$$

In this case P(x) and P(y) correspond to the Gaussian distribution dependent on each variable. The Gaussian function is defined as follows:

$$G(a) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(a-\mu)^2}{2\sigma^2}} \tag{13}$$

The joint probability density distribution with $\mu = 0$ and $\sigma^2 = 1$ can now be defined using (12) and (13):

$$P(x,y) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}} \rightarrow P(r,\theta) = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \tag{14}$$

where r is the 2D polar coordinate derived from the Pythagorean Theorem, $x^2 + y^2 = r^2$. This process gives two independent, perpendicular Gaussian distributions. The importance of translating and using the modulus squared (r^2), is that it gives a chi-squared (χ^2) distribution and therefore a level of fitness. Due to the fact that two deviates are being used, the χ^2distribution will have two degrees of freedom and therefore the reduced $\chi^2 = \frac{\chi^2}{2}$. The closer this value is to 1, the better the fit.

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 equation (5) and (6), the binomial uncertainty was calculated to find the error in derivation as follows:

$$\delta\pi = \sqrt{(\frac{\partial\pi}{\partial N_c})^2 \delta^2 N_c} = \sqrt{\frac{4\pi - \pi^2}{N}} \tag{15}$$

where N and N_c are the total number of iterations and the number of iterations inside the circle, respectively. Table 1. contains derivations of π based on different values for N.

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

N	$Calculated\ \pi$	$\delta\pi$
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

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.

$$N = (\frac{4}{\pi} - 1)(\frac{\pi}{\delta\pi})^2 = (\frac{4}{\pi} - 1)(\frac{1}{error})^2 \tag{16}$$

TABLE II. Values of N based on statistical uncertainty. The published value of π to six decimal places is 3.141593.

Error	N	Calculated π
0.01	2,732	3.141907
0.0001	2.732e11	3.141596
0.0000001	2.732e13	3.141593

B. Volume of n-dimensional hyperspheres

The volumes of 3, 4, and 5 dimensional hyperspheres were determined using equation (7). This data was then

TABLE III. Comparison of derived hypersphere volumes and volumes calculated from geometric formulas with radius (r) equal to 1.

Dimension	Geometric Formula	Calculated Volume	Geometric Volume
3	$\frac{4}{3}\pi r^3$	4.174 ± 0.040	4.189
4	$\frac{1}{2}\pi^2 r^4$	4.912 ± 0.074	4.935
5	$\frac{8}{15}\pi^2 r^5$	5.23 ± 0.12	5.26

compared to the geometric formulas for their corresponding volumes.

Although the derived volumes were within the range of uncertainties for each hypersphere, as the dimensions increased so did the uncertainty. To achieve a better result more iterations are needed for each dimensional step.

C. Histograms of Poisson Deviates

Generating histograms for Poisson deviates, the results correlated with expected Poisson distributions. Poisson distributions retain their Poisson (fish) like shape only for small values of μ. As μ of a Poisson distribution increases, it begins to present as a Gaussian distribution as it becomes less skewed.

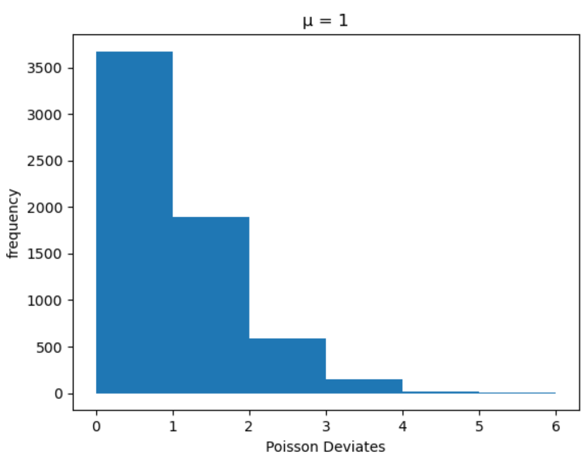

FIG. 1. The distribution of positive poisson deviates with a mean of 1. The fish-like shape is evident when interpreting poisson deviates with a small mean.

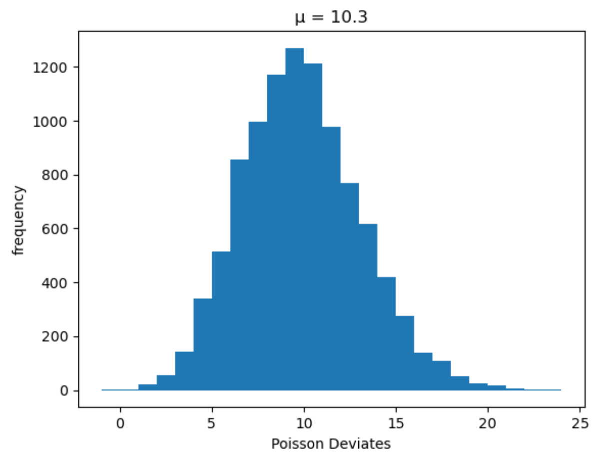

FIG. 2. As the mean increases to 10.3, the distribution becomes less skewed, although it is still skewed slightly right.

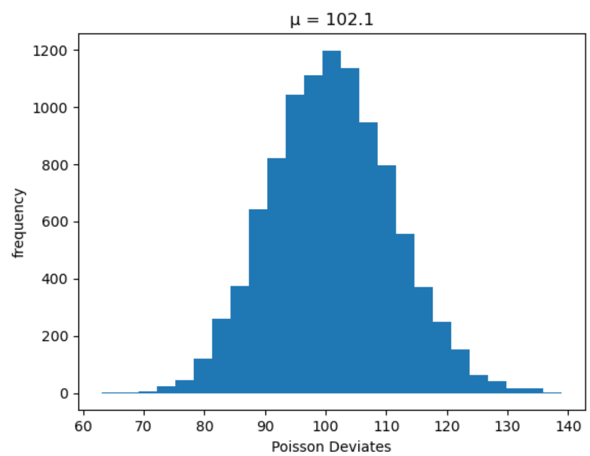

FIG. 3. The distribution of poisson deviates with a mean of 102.1 is more symmetric, visually presenting as a normal distribution.

D. Box-Muller Transform

The generated deviates have been plotted as histograms and fit with Gaussian functions to usingχ^2 to describe goodness of fit. For the deviates (x), a reduced χ^2 of 1.097 was produced and for the deviates (y), a reduced χ^2 of 1.054 was produced. These values for the reduced χ^2 are very close to 1 and therefore are good models for the data.

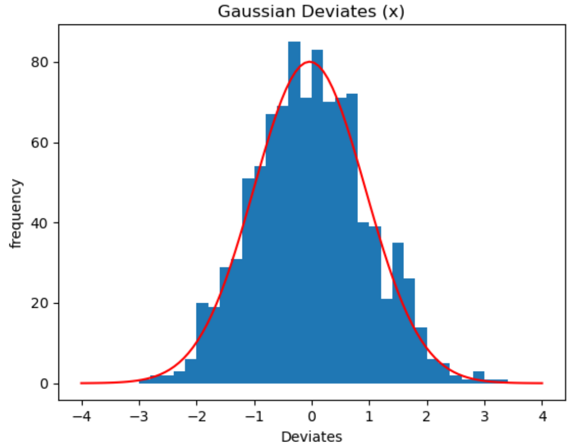

FIG. 4. Gaussian deviates from equation (10). The fit produces a reduced χ^2 of 1.097.

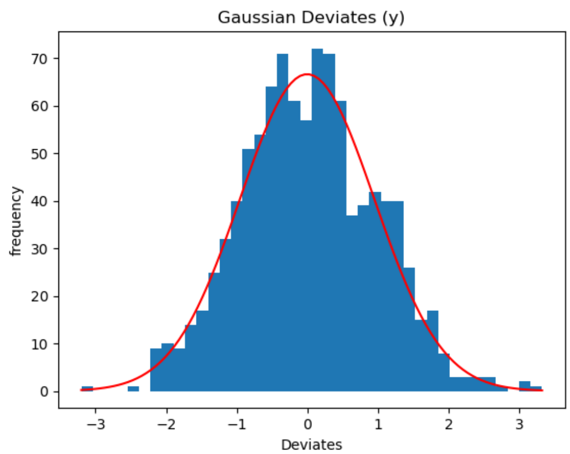

FIG. 5. Gaussian deviates from equation (11). The fit produces a reduced of 1.054.

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 evaluation of data with large numbers of events. Using python for computer programming, random variables can be used to simulate distributions for random processed 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 resulted in desirable reduced χ^2 values. These procedures have various applications in the physical sciences, specifically for counting and quantum mechanical processes.

V. REFERENCES

[1–3]

[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).