Random Number Generation and Monte Carlo Simulation

by Naomi Getzler

Submission date: 05-Apr-2019 02:18PM (UTC+0100)
Submission ID: 104202776
File name: Ex_3_Random_Number_Generation_and_Monte_Carlo_Simulation.pdf (641.87K)
Word count: 2244
Character count: 11511

Random Number Generation and Monte Carlo Simulation

Naomi Getzler
Computational Physics, School of Physics, University of Bristol.
(Dated: April 5, 2019)

The aim of this code was to use the inverse transformation method and the reject-accept method to convert random numbers in a uniform distribution to a non-uniform distribution specified by a function. It was found that the inverse transformation method was a more effective method. Hence it was used for a particle experiment simulation. A Monte Carlo Toy simulation of a collider experiment looking for a hypothesised new particle X was used to determine the cross section as 0.406 nb.

RANDOM NUMBER GENERATION

Random Number Generators (RGNs) can be divided into two categories: Pseudo-Random Number Generators (PRNGs) and True Random Number Generators (TRNGs). [1] PRNGs are deterministic algorithms that generate numbers such that the previous numbers establish the next. By contrast, TRNGs are non-deterministic systems that requires a hardware component to pass a non-deterministic source of data into the system. It is therefore impossible to write an algorithm for TRNGs that works in isolation, hence the focus of the random number generation will be PRNGs exclusively.

Python uses the core generator, the Mersenne twister, in order to produce PRNGs. [2] It is the most extensively tested random number generator to date and has a 53-bit precision float.[3] Due to its deterministic nature, it is not appropriate for cryptographic purposes but it is suitable for this investigation.

Analytical method: Inverse transform sampling

The objective of this algorithm was to take a uniform distribution of random numbers [0,1] and convert it into a sinusoidal distribution $[0,\pi]$. This was achieved via two methods, the inverse transform sampling method and the reject-accept method. To produce a continuous sinusoidal distribution using the inverse transform sampling method, a uniform probability density function, P(x), must be generated between 0 and 1. To convert the uniform distribution P(X) to a non uniform distribution P'(x') = sin(x') a coordinate transformation is achieved using the cumulative distribution. It follows that:

$$x'_{gen} = F^{-1}(x'_{req}) = Q(x'_{req}) \tag{1}$$

where F(x) is the cumulative distribution and Q(x) is the quantile fucntion (the inverse of the cumlative distribution).

The cumulative frequency distribution, F(x') between 0 and x':

$$F(x'_{req}) = \frac{1}{2}\int_0^{x'_{req}} sin(x')dx' = \frac{1-cos(x')}{2} \tag{2}$$

$$Q(x'_{req}) = x_{gen} = cos^{-1}(1-2x'_{req}) \tag{3}$$

To implement the analytical method, the inverse of the function for the cumulative frequency F(x) must be possible to compute. [4] In this case, it is possible to take the inverse of F(x) but in other cases the implementation of the inverse transformation method is possible. A histogram of the probability density function of the deviates was generated using the inverse transformation method with 10^5 points:

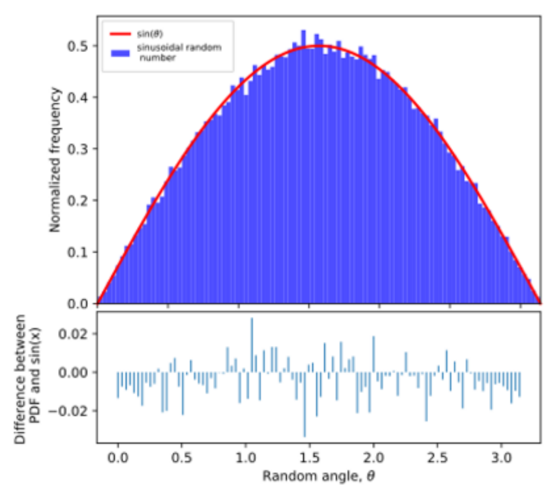

FIG. 1: Probability density function with 100 bins with an area normalised to 1. Subplot of the difference between the pdf of the generated deviates and sin(x)

Reject-Accept

The Reject-Accept method for pseudo-random numbers is often referred to as a universal sampler of the target distribution due to its versatile applications. For this reason, the rejection class of sampling has a functional benefit over direct methods such as the inverse transformation method. [5] In the case where a particular function does not have an explicit inverse, a finite comparison function f(x) is used. For P'(x') = sin(x'), a uniform probability distribution of x: $[0, \pi]$ and y: [0,1] forms an enclosing rectangle around the probability distribution. The area of the desired function P'(x') is in the range of x that corresponds to the probability generated. Uniform random two-dimensional coordinates are taken in the area under the comparison function. If the coordinate lies outside the area of the desired probability distribution, P'(x'), it is rejected. Whenever the point lies within P'(x') it is accepted. This process must be repeated multiple times until the final deviate distribution is obtained. The probability distribution generated using the reject-accept sampling method with 10^5 points is shown below in FIG 2.:

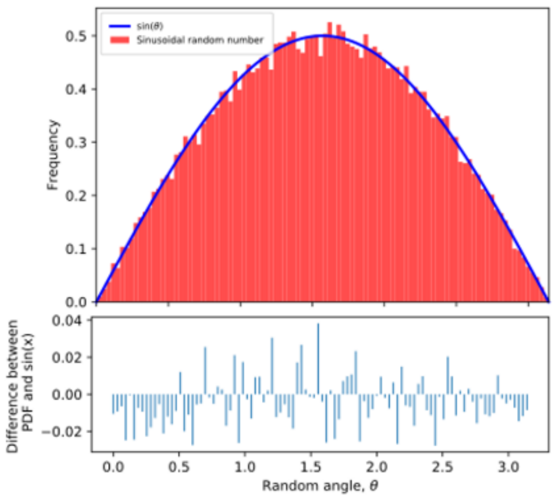

FIG. 2: Probability density function of the deviates generated with 100 bins fitted with an area normalised to 1. Difference between the pdf of the generated deviates for the 100 bins and the sin(x) function

The function's acceptance rate was tested and it was found that when using 10^5 points for the reject-accept method, 63.705% of the points where accepted. This highlights the flaws in this method as a low acceptance rate requires a high computational cost. Consequently, many adaptive methods have been proposed to improve the optimisation of this algorithm including a notable piece work: 'Adaptive rejection sampling for Gibbs sampling' by W R Gilks et al. [6]

Method Comparison

The statistical moments of the distribution were taken for each method in order to compare the methods. The following results were found as shown in TABLE 1.

Both methods produced the expected value of 0 by definition for the mean of the distribution. The variance that represents the spread of the distribution of numbers are both approximately 0.47. However, the inverse transformation method has a lower skewness than the reject-accept method.

Further tests were conducted to measure the execution time of each method as shown in FIG. 3.

Moment	Measure	Inverse Transform	Reject-accept
1	Mean	0.0	0.0
2	Variance	0.46721	0.46814
3	Skewness	0.00096	-0.00223
4	Kurtosis	0.47891	0.480452

TABLE I: The first 4 statistical moments of the sin(x) distribution for the inverse transformation and the reject accept method with a limit factoring of bin size 100.

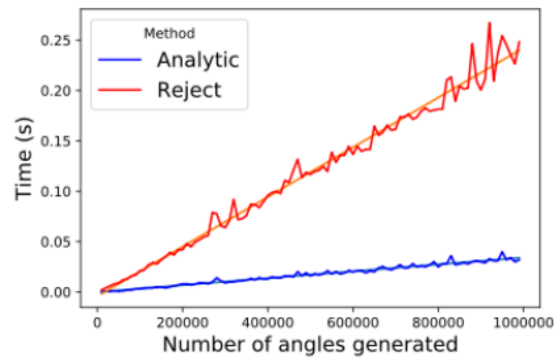

FIG. 3: Time taken per number of points to generate a non uniform distribution of random numbers for sin(x) the inverse-transformation method and the reject-accept method

The reject-accept method took a considerably longer time to execute for the number of points than the inverse transform method. At 10^6 number of points, it was found that the reject-accept method took 5.5 times as long as the inverse-transform method. The reason for the slow execution time is that reject-accept method must carry out two transformations to find the deviate coordinate whereas the inverse transformation method must only requires one. Secondly, as stated previously, the acceptance rate was found to be $\sim$ 60% hence the algorithm is poorly optimised.

MONTE CARLO SIMULATION OF A GAMMA RAYS DISTRIBUTION

Monte Carlo methods have a wide applications in particle physics ranging from statistical methods to high energy particles. This experiment utilised Monte Carlo simulation in a computational particle experiment whereby a beam of unstable nuclei were fired at a distance of 2m from a detector. The mean lifetime of nuclei decay = 550 μs was modelled using an exponential distribution and the position of decay from the detector was found using speed = distance x time.

Following the comparison of the two sampling methods, the inverse-transformation method was found to produce a smaller skew and more importantly, had a considerable time

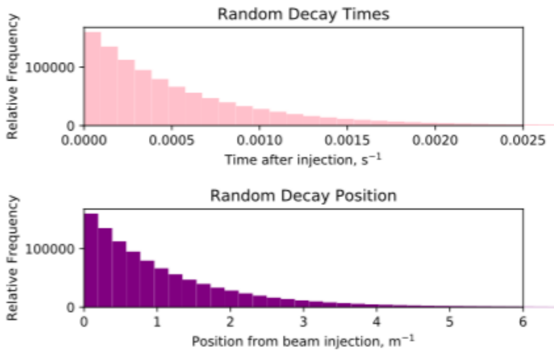

FIG. 4: Exponential distribution of the random decay times and the random decay positions with end decay tail cut from 0.008 s^{-1} and $16m^{-1}$ respectively

efficiency benefit. The inverse transformation non-uniform distribution is preferred as it generates a sphere of isotropic random angles using polar coordinates. The uniform distribution produces angles with higher density at either poles of the sphere as shown in the unit sphere where $\rho = 1$:

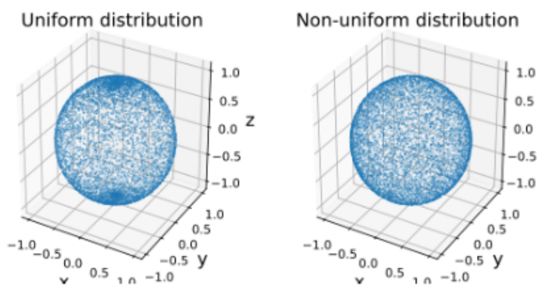

FIG. 5: Unit sphere of randomly distribution angles with a uniform distribution and non-uniform distribution using the inverse transformation method.

To generate the hit position on the detector, the trajectory of the particle decays are extrapolated from the decay positions. The detector array has a resolution of 10cm in x and 30 cm in y. Therefore, a Gaussian distribution must be added to each x and y coordinate with the respective smearing value as shown in FIG. 6.

The hit maps of the gamma decays generated by this experiment were as expected with a roughly Gaussian distribution as shown in FIG.6 without the tails in both the x and the y positions of the screen hits due to the 1x1m detector. The difference in standard deviation between the x and y position was 40 cm. This was an unexpected result as the difference was expected to be 20 cm due to the resolution difference in the x and y direction.

STATISTICAL EXPERIMENT

Toys simulation (ToySim) is used to acquire a basic understanding of a process with only the most fundamental characteristics without any external complications. They are mostly used to demonstrate the large physical effects in a measurement and as a statistical tool for hypothesis testing. [7] In ToySim, the sample is normally a collection of particles with their corresponding four-vector position and momentum. However, in this experiment the ToySim uses three distributions to model the following pseudo-experiment. A preliminary background prediction is simulated using a normal distribution with $\mu = 5.7$ and $\sigma = \pm 0.4$. Two further Poisson distributions are made in order to discretize the variables. The first Poisson uses the values generated by the background prediction distribution as the mean. The second uses the product of integrated luminosity, $L = 12/nb$ and the production cross section, $\langle X \rangle = L\sigma$:

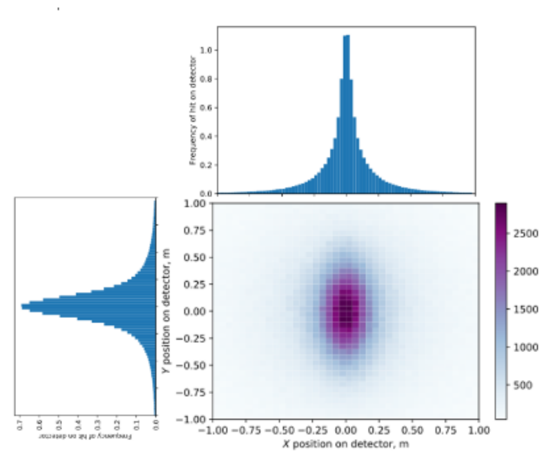

FIG. 6: Hitmap and corresponding x and y position frequency of the gamma decay position on a detector with 10^6 random angle points and 40 bins

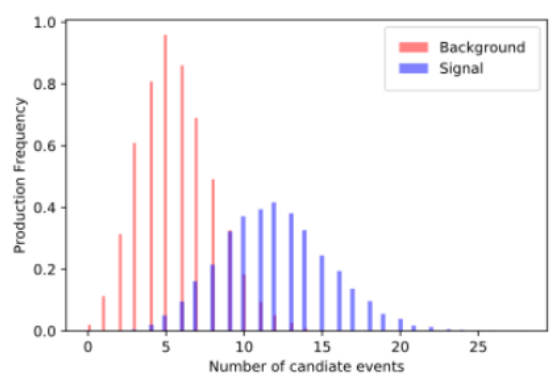

FIG. 7: Background and signal production modelled using Poisson distributions with 10^4 repeats and 100 bins

If X and Y variable are independent it holds that the Poisson distribution can be combined whereby $X + Y \sim P(\mu + \lambda)$.

Therefore the combined production can be modelled using one Poisson distribution:

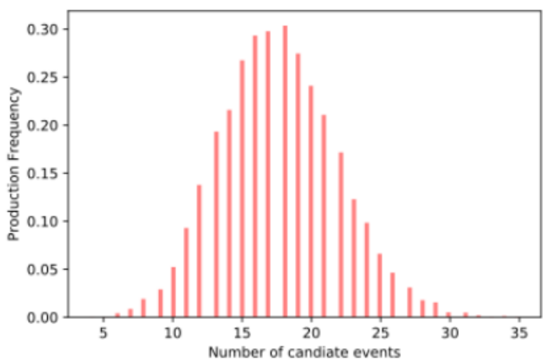

FIG. 8: Combined background and signal production modelled using one Poisson distribution with 10^4 repeats and 100 bins

The experiment produced the result that the total number of events observed was 5. It was therefore found that the first cross section, σ, with a 95% confidence level was 0.406 nb.

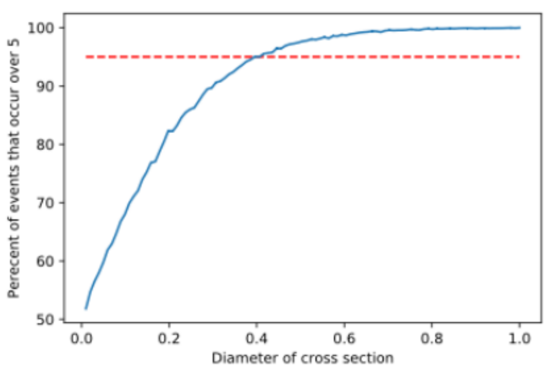

FIG. 9: Confidence level over 5 events for a range of cross sections, the red dashed line marks a 95% confidence interval

The aim of this experiment was to find a new hypothesised particle X. In 'Z and W production associated with quark-antiquark pair in /r-factorization at the LHC' by Michal Deak et. al. [8] it was found that $Zb\bar{b}$ had a total $\sigma = 0.406nb$ in k_T factorisation.

final state	$Zc\bar{c}$	$Zb\bar{b}$	$Zt\bar{t}$	$W^+s\bar{c}$, $W^-c\bar{s}$
σ_{tot} [nb]	0.430	0.406	$0.525 \cdot 10^{-3}$	1.92

FIG. 10: Total cross sections for different final states calculate in k_T-factorisation by Michal Deak et. al. [8]

Additional uncertainties arise in the calculation of the expectation value of the signal production using the integrated luminosity. In order to model this simulation with fewer uncertainties, a normal distribution of the luminosity was included with a standard deviation, $\sigma = 0.5$. Increasing the standard deviation on the luminosity was found to increase the value of the cross section. For example, when $\sigma = 1$, the cross section was found to be 0.4159 nb.

CONCLUSION

After a comparison of the two sampling methods, it was found that although the inverse transformation method is less versatile, it was advantageous over the reject-accept method for the purpose of this investigation. The inverse transformation method had a smaller skewness and was 5.5 times faster at 10^6 points. It was also found that the reject-accept method was at 63% optimisation efficiency due to the large number of points it rejects.

In consideration of these findings, the inverse transformation method was used to carry out the particle decay experiment. This experiment produced a hit map of decay with a skew in the y direction due to the low resolution of the y array of the detector. Both the x and y positions produced a roughly Gaussian distribution of hits on the detector screen.

The final part of the investigation was to find the lower limit on the cross section for a particle with observed events of 5 with a confidence level of 95%. This toy Monte Carlo simulation was achieved using a Gaussian distribution for the background noise and a Poisson distribution for the combined mean of the background and the signal. The cross section was found to be 0.406 nb which was the same cross section as the $Zb\bar{b}$ vertex.

REFERENCES

[1] Johnston, D. *Random Number GeneratorsPrinciples and Practices: A Guide for Engineers and Programmers* (De Gruyter, 2018).

[2] Matsumoto, M. & Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **8**, 3–30 (1998).

[3] De Schryver, C. *et al.* A hardware efficient random number generator for nonuniform distributions with arbitrary precision. *International Journal of Reconfigurable Computing* **2012**, 12 (2012).

[4] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T. *et al.* Numerical recipes (1989).

[5] Martino, L., Luengo, D. & Míguez, J. *Independent random sampling methods* (Springer, 2018).

[6] Gilks, W. R. & Wild, P. Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41**, 337–348 (1992).

[7] Elvira, V. D. Impact of detector simulation in particle physics collider experiments. *Physics Reports* **695**, 1–54 (2017).

[8] Deák, M. & Schwennsen, F. Z and w±production associated with quark-antiquark pair in kt-factorization at the lhc. *Journal of High Energy Physics* **2008**, 035 (2008).