Uncertainty Quantification and The Need for Better Understanding of Monte Carlo and Random Number Generation

Ali Darijani

Vision and Fusion Laboratory
Institute for Anthropomatics
Karlsruhe Institute of Technology (KIT), Germany
ali.darijani@kit.edu

Abstract

Uncertainty Quantification (UQ) is critical in scientific modeling and decision making, addressing the inherent uncertainty and lack of precision in computational modeling. Monte Carlo methods, one of the two prominent approaches in UQ, rely on random number generation to simulate complex systems and estimate probability density functions. However, the quality of these methods is naturally tied to the integrity of the random numbers used. Poorly designed random number generators (RNGs) can lead to biased or unreliable results, undermining the validity of UQ. Despite advancements in RNG algorithms, challenges remain in balancing computational efficiency, reproducibility, and randomness quality. A deeper understanding of these generators, their limitations, and their integration with Monte Carlo methods is essential to advancing UQ practices. This technical report tries to highlight keywords in the random number generation field and give a couple of examples showcasing the pitfalls.

1 Introduction

Mathematicians mostly categorize the uncertainty into epistemic, degree of knowledge when compared to an absolute truth, and aleatoric, something inherently nondeterministic[15], which is something that the physicists also support. While quantum mechanics is still young compared to its counterparts and there is still not consensus in some parts; the Bell inequality and The Einstein-Podolsky-Rosen(EPR) Argument show that quantum mechanics is nondeterministic in nature and there are no ‘‘local hidden variables’’ as opposed to what Einstein thought[14]. As models in molecular dynamics require knowledge from quantum mechanics the uncertainty will carry over and is present at the molecular scale[16]. The same happens when moving from the molecular scale to the continuum scale, and therefore we have uncertainty which we can not get rid of, in the macro scale.

Uncertainty Quantification (UQ) field is gaining attraction which focuses on characterizing, reducing, and managing uncertainty in mathematical models that are based on physical systems. In both forward and backward modes, uncertainty quantification approaches the problem of determining how uncertainties in inputs, models, or parameters affect outputs or vice versa[10].

Monte Carlo Simulation sample inputs from their distributions, run the model multiple times, and analyze the outputs to obtain insights into their distributions[13].

Random number generation from arbitrary distributions typically depends on uniform random number generators because they provide a standardized and versatile base. Transformations like the inverse transform sampling, rejection sampling, or Box-Muller method convert uniform random numbers into samples from desired distributions.

True uniform random number generation is hard to achieve, instead, Pseudo Random Number Generators (PRNGs) are used to produce sequences that approximate true randomness. PRNGs are practical, reproducible, and efficient for most applications, despite being deterministic and periodic over long sequences. Common pitfalls in using them include correlations in sequences (lack of independence), and small periodicity in pseudo-random number generators (PRNGs).

These issues can lead to inaccuracies in simulations or statistical applications if the generator is not carefully chosen.

2 Sample Generation

2.1 Uniform Random Number Generator

John von Neumann jokingly stated: ‘‘Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin’’, but still there exist the notion of a good PRNG. A good PRNG is a balance between different factors depending on the application. Some of these factors are but not limited to[13]:

- Good performance in statistical tests to justify the *independently and identically distributed* assumption which is of course too inaccurate and not feasible practically.
- Reproducible for correctness checking with a small memory footprint.
- Fast and efficient to the degree that is required of them.
- Large period to reduce the chance of correlation. The rule of thumb is that to produce N random number the period has to be at least $10N^2$.
- The ability to produce multiple independent streams.
- Cheap and easy to implement and use as opposed to having expensive physical equipments to capture nature noise as random numbers.
- Does not produce 0 and 1 for technical reasons.

Remark 2.1.1. Some might need speed over large periods. Some might need irreproducible sequences for example for Cryptography applications therefore the need for physics based generators as opposed to PRNG which brings us back to the *There is no solution that fits all*.

2.2 Sample Generation from Arbitrary Distribution

Unfortunately not every probability distribution function can be sampled easily. Different methods exist for different distributions in different dimensions and different regularities. Some with their pros and cons are:

- (Inverse-Transform Method):
  - Works if you know the inverse of the cumulative distribution function. Naturally it does not work for several variables as the probability density function is not bijective. Gets discarded if there exist no analytical inverse; for example in the case of fifth order polynomials.
- (Acceptance-Rejection Method):
  - Works if you can bound the probability density function by another probability density function with known sampling procedure. Unfortunately it is not an efficient method as a lot of samples have to be thrown away as the *Rejection* in the name implies. It is also not a black box method which makes it not usable for those who would like an abstraction layer for the random generation.
- (Composition Method):
  - The overall distribution must be written as a convex hull of distributions with known sample generation procedures. One famous example is the Gaussian Mixture.

3 The Curse of Dimensionality

Definition 3.0.1 (Curse of Dimensionality)**.** The higher the dimension the more difficult solving the problem becomes.

Remark 3.0.2. The term got coined by Richard E. Bellman[2].

Remark 3.0.3. The dimensionality is not always an adversary and there exist ‘‘Blessing of Dimensionality’’[11][4][6][7] as well.

Remark 3.0.4. The curse of dimensionality can and will happen Sampling Methods[13], Differential Equations[8][5][12], ,Tensor Based Methods[9] and many other disciplines.

3.1 Concentration of Measure

Definition 3.1.1. We call:

$$\mathrm{B}_R^n(c) = \{x, c \in \mathbb{R}^n \colon \sum_{i=1}^{n} (x_i - c_i)^2 \leq R^2\}$$

the n-ball and its points are at the distance R from the point c with respect to the Euclidean distance in the n-dimensional space. We show the volume of the said ball by $\mathrm{Vol}(\mathrm{B}_R^n(c))$.

Proposition 3.1.2.

$$\mathrm{Vol}(\mathrm{B}_\mathrm{R}^1(\mathrm{c})) = 2R.$$

Proof. Move the center of the ball to zero as the volume is invariant under translation and then:

$$x_1 = R \sin\theta, \quad -\pi/2 \leq \theta \leq +\pi/2$$

$$\int_{\mathrm{B}_R^1} \mathrm{d}x = \int_{-\pi/2}^{+\pi/2} |\det(J)| \, \mathrm{d}\theta = \int_{-\pi/2}^{+\pi/2} R\cos\theta \mathrm{d}\theta = R\sin\theta \Big|_{-\pi/2}^{+\pi/2} = 2R$$

where J is the Jacobian of the mapping:

$$J = \left(\tfrac{\partial x_1}{\partial \theta}\right) = R\cos\theta$$

□

Proposition 3.1.3.

$$\mathrm{Vol}(\mathrm{B}_\mathrm{R}^2(\mathrm{c})) = \pi R^2.$$

Proof. Move the center of the balls to zero as the volume is invariant under translation and then:

$$\begin{aligned} x_1 =& r\cos\theta, \quad 0 \leq \theta < 2\pi \\ x_2 =& r\sin\theta, \quad 0 \leq r \leq R, \end{aligned}$$

$$\int_{B_R^1} \mathrm{d}x = \int_0^R \int_0^{2\pi} |\det(J)| \,\mathrm{d}\theta \mathrm{d}r = \int_0^R \int_0^{2\pi} r \mathrm{d}\theta \mathrm{d}r = \pi R^2$$

where J is the Jacobian of the mapping:

$$J = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

□

Proposition 3.1.4.

$$\mathrm{Vol}(\mathrm{B}_\mathrm{R}^3(\mathrm{c})) = \frac{4}{3}\pi R^3.$$

Proof. Move the center of the ball to zero as the volume is invariant under translation and then:

$$\begin{aligned} x_1 &= r\sin\theta\cos\varphi, \quad 0 \le \varphi < 2\pi \\ x_2 &= r\sin\theta\sin\varphi, \quad 0 \le \theta \le \pi, \\ x_3 &= r\cos\theta, \qquad\quad 0 \le r \le R, \end{aligned}$$

$$\begin{aligned} \int_{B_R^3} \mathrm{d}x &= \int_0^R \int_0^{\pi} \int_0^{2\pi} |\det(J)| \,\mathrm{d}\varphi \mathrm{d}\theta \mathrm{d}r \\ &= \int_0^R \int_0^{\pi} \int_0^{2\pi} r^2 \sin\theta \mathrm{d}\theta \mathrm{d}r = 4/3\pi R^3 \end{aligned}$$

□

where J is the Jacobian of the mapping:

$$J = \begin{pmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} & \frac{\partial x_1}{\partial \varphi} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} & \frac{\partial x_2}{\partial \varphi} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial \theta} & \frac{\partial x_3}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}$$

Proposition 3.1.5.

$$\mathrm{Vol}(\mathrm{B}_\mathrm{R}^\mathrm{n}(\mathrm{c})) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} R^n.$$

Proof. Move the center of the ball to zero as the volume is invariant under translation and then:

$$\begin{aligned} x_1 =& r\cos\theta_1, \\ x_2 =& r\sin\theta_1\cos\theta_2, \\ &\vdots \\ x_{n-1} =& r\sin\theta_1\cdots\sin\theta_{n-2}\cos\theta_{n-1}, \\ x_n =& r\sin\theta_1\cdots\sin\theta_{n-2}\sin\theta_{n-1}, \end{aligned}$$

for:

$$r\in[0,R],\theta_1\in[0,\pi],\theta_2,\cdots,\theta_{n-1}\in[0,2\pi)$$

$$\Omega=[0,R]\times[0,\pi]\times\cdots\times[0,2\pi],\mathrm{d}\Theta=\prod_{i=1}^{i=n-1}\mathrm{d}\theta_i$$

$$\begin{aligned} \int_{\mathrm{B}_R^n}\mathrm{d}x =& \int_\Omega |\det(J)|\,\mathrm{d}r\mathrm{d}\theta_1\mathrm{d}\theta_2\cdots\mathrm{d}\theta_{n-2} \\ =& \int_\Omega r^{n-1}\sin^{n-2}\theta_1\sin^{n-3}\theta_{n-2}\cdots\sin\theta_{n-1}\mathrm{d}r\mathrm{d}\Theta \\ =& \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}R^n \end{aligned}$$

where Γ is the gamma function. □

Remark 3.1.6.

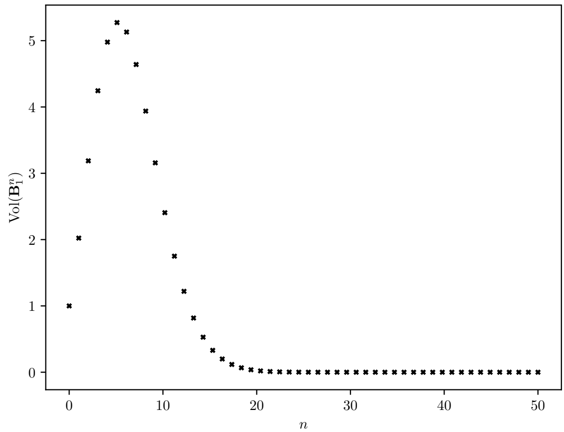
- The Gamma function[1]
- Non-geometric approach for the n-ball[3]

Proposition 3.1.7.

$$\lim_{n\to\infty}\mathrm{Vol}(\mathrm{B}_1^{\mathrm{n}}(\mathrm{c}))=0$$

Remark 3.1.8. While it is possible to prove the proposition by recalling properties of the Gamma function[1] and do asymptotic analysis, taking a look at the (3.1) is beneficial.

Figure 3.1: Volume of the unit n-ball against its dimension.



Proposition 3.1.9. *For small ε:*

$$\lim_{n \to \infty} \frac{\mathrm{Vol}(\mathrm{B}_1^n(\mathrm{c})) - \mathrm{Vol}(\mathrm{B}_{1-\varepsilon}^n(\mathrm{c}))}{\mathrm{Vol}(\mathrm{B}_1^n(\mathrm{c}))} = 1$$

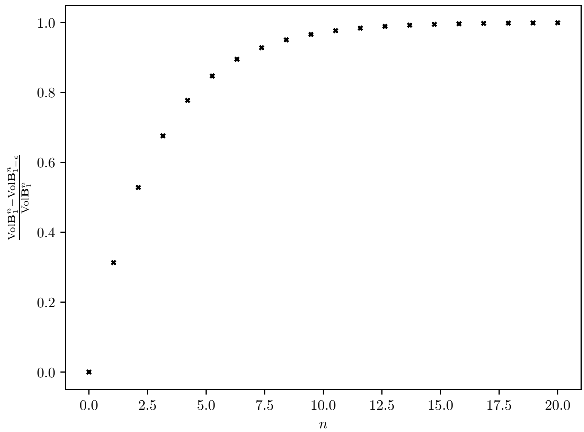
Remark 3.1.10. While it is possible to prove the proposition by recalling properties of the Gamma function[1] and do asymptotic analysis, taking a look at the (3.2) is beneficial.

Remark 3.1.11. The phenomenon is known as the Concentration of Measure.

3.2 π Calculation by Monte Carlo on the n-Ball

Monte Carlo methods can estimate π by leveraging the relationship between the volume of an n-ball and π. To estimate π, random points are generated within an n-dimensional cube that bounds an n-ball. The fraction of points that fall inside the n-ball, determined by checking if their distance from the origin is less than

Figure 3.2: Concentration of measure for the unit n-ball.



or equal to the radius, used to approximate the ratio of the volume of the n-ball to the n-cube. Results can be seen in (3.3), (3.4) show that it is quite effective in low dimensions, but unfortunately as it can be seen in (3.5) and (3.6) no matter the number of points, estimating π proves to be difficult in high dimensions.

Theorem 3.2.1 (Nyquist-Whitakker-Shannon Theorem)**.** *Let $B > 0$ and f be in $L_2(\mathbb{R}, \mathbb{C})$ be such that $\mathcal{F}f(\omega) = 0$ for $|\omega| > B$. Then f is continuous (more precisely, has a continuous representative) and is uniquely determined by the values $(f(k\pi/B))_{k\in\mathbb{Z}}$. Moreover,*

$$u(x) = \sum_{k\in\mathbb{Z}} u(\frac{k\pi}{B})\mathrm{sinc}(\frac{B}{\pi}(x - \frac{k\pi}{B})).$$

Remark 3.2.2. If the correct sampling rate indicated by the Nyquist-Whitakker-Shannon Theorem is not met it might lead to undersampling (3.7) or oversampling which are undesirable.

Figure 3.3: Calculating π using Monte Carlo on the unit 2-ball.

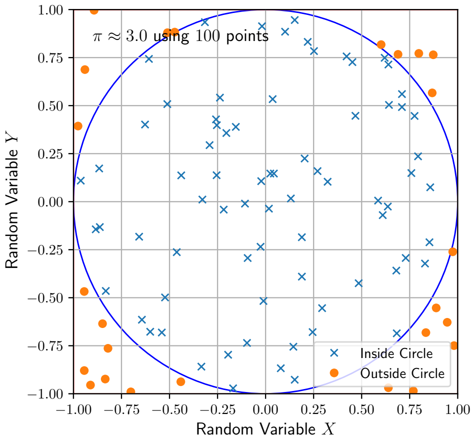


Figure 3.4: Calculating π using Monte Carlo on the unit 3-ball.

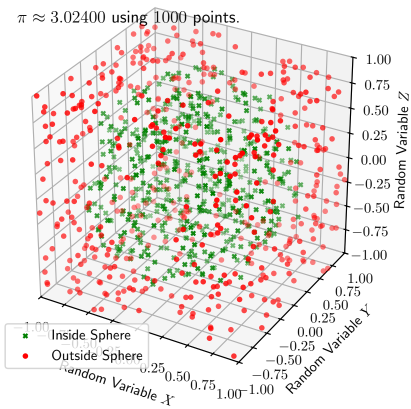


Figure 3.5: Calculating π using Monte Carlo using 100 points in different dimensions.

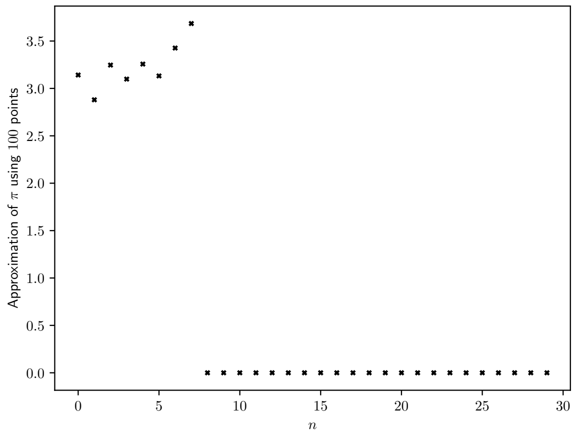


Figure 3.6: Calculating π using Monte Carlo using 1000000 points in different dimensions.

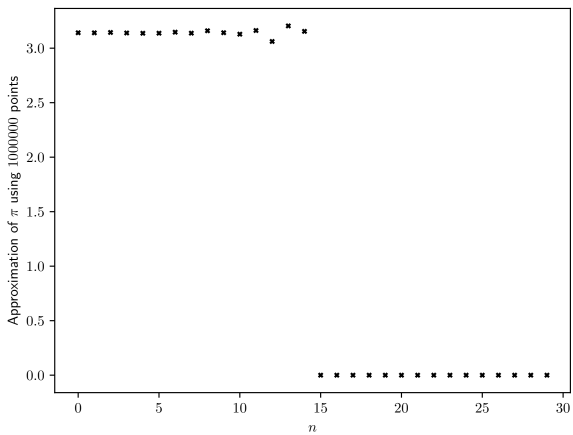
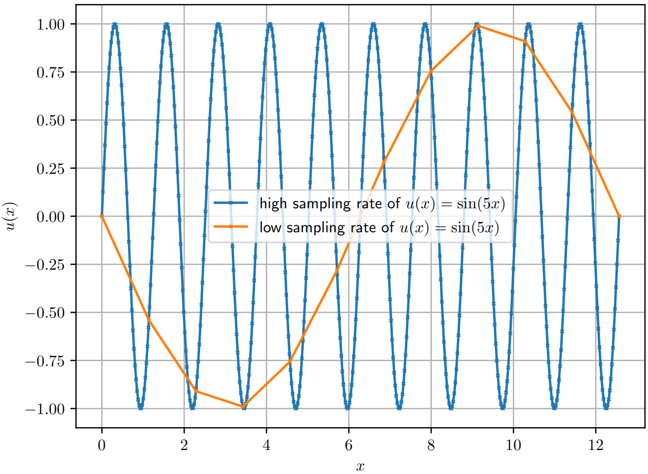


Figure 3.7: The effect of undersampling



4 My Plans

I am interested in studying Monte Carlo methods more thoroughly to gain a deeper understanding of their mathematical foundations and practical applications. By diving deep into the principles of randomness, sampling, and estimation that underpin these methods, I hope to gain the upper hand when dealing with Monte Carlo solvable problems. Additionally, I aim to explore specialized Monte Carlo techniques designed for specific scenarios, such as variance reduction methods, importance sampling, or Markov Chain Monte Carlo (MCMC), to better apply these tools to targeted challenges in fields like physics, finance, and machine learning.

References

[1] Emil Artin. *The Gamma Function*. Dover, 2015.

[2] Richard E. Bellman. *Dynamic Programming*. Princeton University Press, Dec. 2010. ISBN: 9781400835386. DOI: 10.1515/9781400835386.

[3] L. E. Blumenson. "A Derivation of n-Dimensional Spherical Coordinates". In: *The American Mathematical Monthly* 67.1 (Jan. 1960), p. 63. ISSN: 0002-9890. DOI: 10.2307/2308932.

[4] David L. Donoho. "High-Dimensional Data Analysis:The Curses and Blessings of Dimensionality". In: (2000).

[5] Edwige Godlewski and Pierre-Arnaud Raviart. *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Springer New York, 2021. ISBN: 9781071613443. DOI: 10.1007/978-1-0716-1344-3.

[6] A. N. Gorban and I. Y. Tyukin. "Blessing of dimensionality: mathematical foundations of the statistical physics of data". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2118 (Mar. 2018), p. 20170237. ISSN: 1471-2962. DOI: 10.1098/rsta.2017.0237.

[7] Alexander N. Gorban, Valery A. Makarov, and Ivan Y. Tyukin. "High-Dimensional Brain in a High-Dimensional World: Blessing of Dimensionality". In: *Entropy* 22.1 (Jan. 2020), p. 82. ISSN: 1099-4300. DOI: 10.3390/e22010082.

[8] Wolfgang Hackbusch. *Elliptic Differential Equations*. Springer Berlin Heidelberg, 2017. ISBN: 9783662549612. DOI: 10.1007/978-3-662-54961-2.

[9] Wolfgang Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Springer International Publishing, 2019. ISBN: 9783030355548. DOI: 10.1007/978-3-030-35554-8.

[10] *Handbook of Uncertainty Quantification*. 2017. DOI: 10.1007/978-3-319-12385-1.

[11] Paul C. Kainen. "Utilizing Geometric Anomalies of High Dimension: When Complexity Makes Computation Easier". In: *Computer Intensive Methods in Control and Signal Processing*. Birkhäuser Boston, 1997, pp. 283–294. ISBN: 9781461219965. DOI: 10.1007/978-1-4612-1996-5_18.

[12] Peter Knabner and Lutz Angermann. *Numerical Methods for Elliptic and Parabolic Partial Differential Equations: With contributions by Andreas Rupp*. Springer International Publishing, 2021. ISBN: 9783030793852. DOI: 10.1007/978-3-030-79385-2.

[13] Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo methods*. Wiley Series in Probability and Statistics. Includes bibliographical references and index. Hoboken, NJ: Wiley, 2011. ISBN: 0470177934.

[14] Franz Schwabl. *Quantum Mechanics*. Fourth Edition. Springer Berlin Heidelberg, 2007. ISBN: 9783540719328. DOI: 10.1007/978-3-540-71933-5.

[15] Tim Sullivan. *Introduction to Uncertainty Quantification*. Springer International Publishing, 2015. ISBN: 9783319233956. DOI: 10.1007/978-3-319-23395-6.

[16] Mark E. Tuckerman. *Statistical Mechanics*. Fourth Edition. Oxford Graduate Texts. Description based upon print version of record. Oxford: Oxford University Press, Incorporated, 2023. ISBN: 0192559613.