Laplace and the Origins of the Monte Carlo Method

From Buffon's needle to modern simulation

Hedibert F. Lopes

Notes after Petr Beckmann, *A History of Pi*, Ch. 15

May 11, 2026

Outline

1. Buffon's needle (1777)
2. Laplace's inversion (1812)
3. Why this is the seed of Monte Carlo
4. Connection to today's class

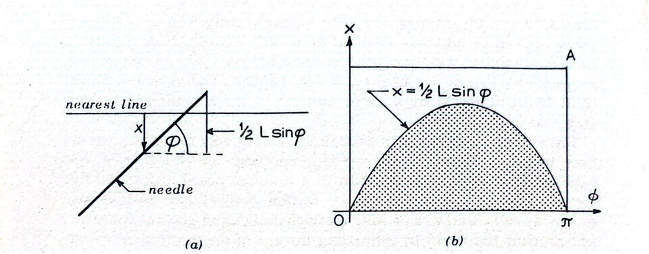
The setup

Buffon's question (1777). A needle of length L is dropped *at random* on a plane ruled with parallel lines spaced $d > L$ apart. What is the probability that the needle crosses one of the lines?

Random variables.

- x = distance from the needle's *center* to the nearest line; $x \sim \mathrm{U}(0, d/2)$.
- ϕ = orientation of the needle; $\phi \sim \mathrm{U}(0, \pi)$.
- x and ϕ are independent.

Panel (a): geometry of a single throw.
Panel (b): sample space (ϕ, x); needle crosses a line iff the point lies in the shaded region.



Buffon's problem. Reproduced from Beckmann, *A History of Pi*, Ch. 15, p. 160.

Geometry of intersection

The needle crosses the nearest line if and only if

$$\boxed{x \; < \; \tfrac{1}{2}L \, \sin\phi}$$

Geometry of intersection

The needle crosses the nearest line if and only if

$$\boxed{x \;<\; \tfrac{1}{2}L\,\sin\phi}$$

Joint density of (x, ϕ) is uniform on the rectangle

$$(x, \phi) \in (0, d/2) \times (0, \pi), \qquad f(x, \phi) = \frac{1}{(\pi d/2)}.$$

Geometry of intersection

The needle crosses the nearest line if and only if

$$\boxed{x \;<\; \tfrac{1}{2}L\,\sin\phi}$$

Joint density of (x, ϕ) is uniform on the rectangle

$$(x,\phi) \in (0, d/2) \times (0, \pi), \qquad f(x,\phi) = \frac{1}{(\pi d/2)}.$$

Probability of intersection = favourable area / total area:

$$P \;=\; \frac{\tfrac{1}{2}L\displaystyle\int_0^{\pi} \sin\phi\,d\phi}{\frac{\pi d}{2}} \;=\; \boxed{\frac{2L}{\pi d}.}$$

Recall form high school that $\int_0^\pi \sin(\phi)\,d\phi = \big[-\cos(\phi)\big]_0^\pi = -\cos(\pi) - (-\cos 0) = -(-1) + 1 = 2.$

Buffon's own experiment

- Buffon *verified* the formula by tossing a needle onto ruled paper and counting intersections.
- He was, in effect, computing

$$\widehat{P}_N = \frac{\#\{\text{crossings}\}}{N}, \qquad \widehat{P}_N \xrightarrow{N\to\infty} \frac{2L}{\pi d}.$$

- But Buffon stopped there. He did not invert the formula.
- Beckmann notes the problem then lay dormant for about 35 years.

Laplace re-reads Buffon's formula

In the *Théorie analytique des probabilités* (1812), Pierre-Simon Laplace (1749–1827) rearranges

$$P = \frac{2L}{\pi d} \quad \Longrightarrow \quad \boxed{\pi = \frac{2L}{d\,P}} \qquad (\star)$$

Laplace re-reads Buffon's formula

In the *Théorie analytique des probabilités* (1812), Pierre-Simon Laplace (1749–1827) rearranges

$$P = \frac{2L}{\pi d} \qquad \Longrightarrow \qquad \boxed{\pi = \frac{2L}{d\,P}} \qquad (\star)$$

The shift in viewpoint:

- Buffon: π is known $\Rightarrow$ predict P.
- Laplace: L and d are known, P is measured by simulation $\Rightarrow$ recover π.

Laplace re-reads Buffon's formula

In the *Théorie analytique des probabilités* (1812), Pierre-Simon Laplace (1749–1827) rearranges

$$P = \frac{2L}{\pi d} \qquad \Longrightarrow \qquad \boxed{\pi = \frac{2L}{d\,P}} \qquad (\star)$$

The shift in viewpoint:

- Buffon: π is known $\Rightarrow$ predict P.
- Laplace: L and d are known, P is measured by simulation $\Rightarrow$ recover π.

Replace an analytic constant by an empirical frequency!

Laplace's estimator of π

Throw the needle independently N times. Let

$$Y_i = \mathbf{1}\{\text{throw } i \text{ crosses a line}\}, \qquad \widehat{P}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

Laplace's estimator of π

Throw the needle independently N times. Let

$$Y_i = \mathbf{1}\{\text{throw } i \text{ crosses a line}\}, \qquad \widehat{P}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

Laplace's estimator:

$$\widehat{\pi}_N \;=\; \frac{2L}{d\,\widehat{P}_N}.$$

Laplace's estimator of π

Throw the needle independently N times. Let

$$Y_i = \mathbf{1}\{\text{throw } i \text{ crosses a line}\}, \qquad \widehat{P}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

Laplace's estimator:

$$\widehat{\pi}_N \; = \; \frac{2L}{d\,\widehat{P}_N}.$$

By the Strong Law of Large Numbers (SLLN)

$$\widehat{P}_N \xrightarrow{\text{a.s.}} P \qquad \Longrightarrow \qquad \widehat{\pi}_N \xrightarrow{\text{a.s.}} \pi.$$

By the Central Limit Theorem (CLT), the standard error of $\widehat{\pi}_N$ is of order $1/\sqrt{N}$.

Laplace's estimator of π

Throw the needle independently N times. Let

$$Y_i = \mathbf{1}\{\text{throw } i \text{ crosses a line}\}, \qquad \widehat{P}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i.$$

Laplace's estimator:

$$\widehat{\pi}_N \; = \; \frac{2L}{d\,\widehat{P}_N}.$$

By the Strong Law of Large Numbers (SLLN)

$$\widehat{P}_N \xrightarrow{\text{a.s.}} P \qquad \Longrightarrow \qquad \widehat{\pi}_N \xrightarrow{\text{a.s.}} \pi.$$

By the Central Limit Theorem (CLT), the standard error of $\widehat{\pi}_N$ is of order $1/\sqrt{N}$.

Efficiency as a π-calculator?

Beckmann is blunt: as a numerical method for π, the needle toss is very inefficient.

Example reported in Ch. 15:

> *The probability of obtaining π correct to 5 decimal places in $3{,}400$ throws is below 1.5%.*

Efficiency as a π-calculator?

Beckmann is blunt: as a numerical method for π, the needle toss is very inefficient.

Example reported in Ch. 15:

> *The probability of obtaining π correct to 5 decimal places in $3{,}400$ throws is below 1.5%.*

- Variance of $\widehat{\pi}_N$ shrinks only like $1/N$.
- Euler's series for π converges much faster.
- But the principle is what matters, not this particular use case.

The principle behind $(\star)$ - Monte Carlo integration

Rewrite $(\star)$ in modern notation:

$$\pi \;=\; \frac{2L}{d}\cdot\frac{1}{\mathbb{E}\big[\,\mathbf{1}\{X<\tfrac{1}{2}L\sin\Phi\}\,\big]},\qquad X\sim \mathrm{U}(0,d/2),\;\; \Phi\sim \mathrm{U}(0,\pi).$$

The principle behind $(\star)$ - Monte Carlo integration

Rewrite $(\star)$ in modern notation:

$$\pi \; = \; \frac{2L}{d} \cdot \frac{1}{\mathbb{E}\big[\,\mathbf{1}\{X < \tfrac{1}{2}L\sin\Phi\}\,\big]}, \quad X \sim \mathrm{U}(0, d/2),\; \Phi \sim \mathrm{U}(0, \pi).$$

A deterministic constant is expressed as a functional of a probability distribution.

The principle behind $(\star)$ - Monte Carlo integration

Rewrite $(\star)$ in modern notation:

$$\pi \;=\; \frac{2L}{d}\cdot\frac{1}{\mathbb{E}\big[\,\mathbf{1}\{X<\tfrac{1}{2}L\sin\Phi\}\,\big]},\qquad X\sim \mathrm{U}(0,d/2),\;\Phi\sim\mathrm{U}(0,\pi).$$

A deterministic constant is expressed as a functional of a probability distribution.

Estimator: sample (X_i, Φ_i) and average $\mathbf{1}\{\cdot\}$.

The principle behind $(\star)$ - Monte Carlo integration

Rewrite $(\star)$ in modern notation:

$$\pi \; = \; \frac{2L}{d} \cdot \frac{1}{\mathbb{E}\left[\, \mathbf{1}\{X < \frac{1}{2} L \sin \Phi\} \,\right]}, \quad X \sim \mathrm{U}(0, d/2), \; \Phi \sim \mathrm{U}(0, \pi).$$

A deterministic constant is expressed as a functional of a probability distribution.

Estimator: sample (X_i, Φ_i) and average $\mathbf{1}\{\cdot\}$.

This is exactly Monte Carlo integration:

$$\mathbb{E}[h(X)] \; \approx \; \frac{1}{N} \sum_{i=1}^{N} h(X_i), \qquad X_i \overset{\text{iid}}{\sim} f.$$

Beckmann's verdict

Laplace had discovered a powerful method of computation that did not come into its own until the advent of the electronic computer.

Beckmann's verdict

> *Laplace had discovered a powerful method of computation that did not come into its own until the advent of the electronic computer.*

The method that Laplace proposed consists in

finding a numerical value by realizing a random event many times and observing its outcome experimentally.

Beckmann's verdict

> *Laplace had discovered a powerful method of computation that did not come into its own until the advent of the electronic computer.*

The method that Laplace proposed consists in

finding a numerical value by realizing a random event many times and observing its outcome experimentally.

That is the modern definition of a *Monte Carlo method*.

The closing image

Beckmann ends the section memorably:

> *The man who taught us to program electronic computers in this way was Pierre Simon Laplace. His computer was neither electronic nor digital — it was an analog computer consisting of one needle and one piece of ruled paper.*

The closing image

Beckmann ends the section memorably:

> *The man who taught us to program electronic computers in this way was Pierre Simon Laplace. His computer was neither electronic nor digital — it was an analog computer consisting of one needle and one piece of ruled paper.*

Buffon supplied the identity.
Laplace ran it backwards.
That inversion is the essence of Monte Carlo.

From Laplace to our mixture problem

Every method we used today is the same idea, scaled up:

$$\mathbb{E}_{p(\theta|y)}[h(\theta)] \;\approx\; \frac{1}{N}\sum_{j=1}^{N} h(\theta^{(j)}), \qquad \theta^{(j)} \sim p(\theta \mid y).$$

From Laplace to our mixture problem

Every method we used today is the same idea, scaled up:

$$\mathbb{E}_{p(\theta|y)}[h(\theta)] \approx \frac{1}{N}\sum_{j=1}^{N} h(\theta^{(j)}), \qquad \theta^{(j)} \sim p(\theta \mid y).$$

- **SIR** — weighted resampling from a proposal.
- **Random-walk Metropolis** (componentwise / joint).
- **Gibbs sampler** with latent indicators z_i.

From Laplace to our mixture problem

Every method we used today is the same idea, scaled up:

$$\mathbb{E}_{p(\theta|y)}[h(\theta)] \;\approx\; \frac{1}{N}\sum_{j=1}^{N} h(\theta^{(j)}), \qquad \theta^{(j)} \sim p(\theta \mid y).$$

- **SIR —** weighted resampling from a proposal.
- **Random-walk Metropolis** (componentwise / joint).
- **Gibbs sampler** with latent indicators z_i.

Different machinery, same Laplacian idea: *simulate the random experiment, then average*.

A few useful links

Monte Carlo and Markov chain Monte Carlo Methods: A short note with key references
By Hedibert Lopes (2021).

Monte Carlo Integration
MCMC: Stochastic Simulation for Bayesian Inference, Sections 3.1 and 3.4, pages 81-82 & 95-97.
By Dani Gamerman and Hedibert Lopes (2006)

A Short History of MCMC: Subjective Recollections from Incomplete Data
Statistical Science, Volume 26, pages 102-115.
Christian Robert and George Casella's (2011)

An Introduction to MCMC for Machine Learning
Machine Learning, Volume 50, pages 5-43.
Christophe Andrieu, Nando de Freitas, Arnaud Doucet and Michale Jordan (2003)