GALTON BOARD

Monte Carlo Demonstrator

COMPLETE TEACHER'S GUIDE

Grades 3 through 12

Grade	Focus	Color
3rd Grade	Chance, Patterns & First Steps in Monte Carlo	Blue
6th Grade	Probability Fractions, Binomial Distribution & Empirical Rule	Green
9th Grade	Probability Axioms, Hypothesis Testing & Random Walks	Purple
12th Grade	Measure Theory, CLT Proof, Bayesian Inference & MCMC	Orange-Red

Index Fund Advisors (IFA) • ifa.com • In collaboration with Los Alamos National Laboratory
US Patent No. 12,268,971 B1 • US Des. Patent No. D784,449

How to Use This Guide

This manual contains four complete teacher's guides — one for each grade band — organized from simplest to most advanced. Each section can be used independently or as part of a multi-year progression. The Galton Board: Monte Carlo Demonstrator is the same physical device at every grade level; what changes is the mathematical depth with which students explore it.

Section	Prerequisite Knowledge	Duration
Grade 3	None — appropriate for all elementary students	4–5 class periods (~40 min each)
Grade 6	Fractions, ratios, basic graphing	6–7 class periods (~50 min each)
Grade 9	Algebra I, basic combinatorics, summation notation	7–8 class periods (~50 min each)
Grade 12	Single-variable calculus, linear algebra, Grade 9 content	8 lessons (~60 min each) + project

About the Galton Board: Monte Carlo Demonstrator

The board contains 4,280 steel beads + 1 golden bead, 105 hexagonal pegs in 14 rows, and 15 collection bins (0–14). Each bead makes 14 independent left-or-right random decisions as it falls — equivalent to flipping a coin 14 times. Together, 4,280 beads simulate 59,920 coin tosses in about 2 seconds. The resulting hill-shaped distribution is both the binomial distribution and (for large n) the bell curve (normal distribution). This was co-developed by Index Fund Advisors with Art Forster and Joel Kulesza of Los Alamos National Laboratory.

GRADE 3

Chance, Patterns & First Steps in Monte Carlo

Welcome, Teacher! — 3rd Grade

This guide weaves two exciting stories together. The first is the Galton Board story — a beautiful device showing how randomness creates patterns. The second is the story of Monte Carlo methods, invented at Los Alamos National Laboratory, where scientists discovered you can solve incredibly hard problems by repeating simple random experiments. The historical stories of Buffon's needle and Stanislaw Ulam's solitaire game are told in an age-appropriate way that will captivate students.

What 3rd Graders Will Learn

- Some things are fair (equally likely) and some are not
- When you repeat a random event many times, a reliable pattern appears
- The Galton Board shows this pattern as a hill shape called the bell curve
- Each bead is like one scientist's 'guess' — one Monte Carlo trial
- The more guesses you make, the closer you get to the right answer
- Pascal's Triangle connects to the board's numbers
- Scientists at Los Alamos use these same ideas to solve real science problems every day

Pacing

4–5 class periods of ~40 minutes. Lessons 1–4 cover probability and the board. Lesson 5 introduces Monte Carlo. Lesson 6 covers Pascal's Triangle. Lesson 7 connects to the real world. Each lesson can stand alone.

🎲 LESSON 1 — What Is Chance?

The Big Question

Can you ever KNOW for sure what will happen when you flip a coin? No! And that's what makes it interesting. When we can't predict what will happen, we call it chance (or luck, or probability). But here is the amazing part: even though we can't predict any single flip, when we flip enough times we can predict the PATTERN. Scientists have studied this idea for hundreds of years.

Fair and Unfair

Something is FAIR when all outcomes are equally likely. Something is UNFAIR when some outcomes are more likely than others.

Situation	Fair?	Why?
Flip a regular coin	☑ Fair	Heads and tails are equally likely

Roll a regular 6-sided die	☑ Fair	All six numbers have the same chance
Pick from 9 red marbles + 1 blue	✕ Unfair	Red is 9 times more likely than blue
A coin with heads on both sides	✕ Unfair	You can never get tails!

Two Coins Together

Flip two coins at once. List all four equally likely outcomes, then count:

Ways to get 2 Heads	Ways to get 1 of each	Ways to get 2 Tails
1 way (HH)	2 ways (HT or TH)	1 way (TT)
Probability: 1/4	Probability: 2/4 = 1/2	Probability: 1/4

Pattern Alert!

Write these down: 1 — 2 — 1. You will see these EXACT numbers again when we look at the board's Pascal's Triangle!

LESSON 2 — The Galton Board

Who Made This?

Sir Francis Galton (1822–1911) invented the first version of this board in 1873. He was captivated by the fact that thousands of randomly-bouncing beads always produce the same beautiful hill shape. He called it 'order in apparent chaos.' Today's version — the Galton Board: Monte Carlo Demonstrator — was made by Index Fund Advisors (IFA) together with scientists Art Forster and Joel Kulesza from Los Alamos National Laboratory.

Each Bead Makes a Journey

- The bead enters at the very top of the board.
- It hits the first hexagon peg and bounces either LEFT or RIGHT — by complete chance.
- This happens 14 times — one bounce per row.
- The bead falls into whichever bin it ends up above.
- Most beads go left sometimes and right sometimes, so most land in the middle.

The Bell Curve Appears!

Every single flip, the beads make a hill — tall in the middle, short at the edges. Mathematicians call this the BELL CURVE (or normal distribution). Los Alamos scientists see this same shape every time their Monte Carlo simulations finish. Order always emerges from chaos!

LESSON 3 — What Is Monte Carlo?

A Big Idea with a Fun Name

Monte Carlo is the name of a famous place in Europe known for its casinos. Scientists borrowed this name because their method also uses randomness!

Do many random trials → Look at the pattern → Find the answer!

That is EXACTLY what the Galton Board does! Each bead is one random trial. 4,280 beads are 4,280 trials. The bell curve pattern that appears is the answer the board reveals.

A Story About Needles (1777)

A French scientist named Buffon (Boo-FAWN) had a clever idea. He drew parallel lines on a floor, then dropped a needle onto it many times. By counting how often the needle crossed a line across MANY drops, he could calculate the value of π (pi ≈ 3.14159). He was using many random trials to discover a mathematical answer — the first Monte Carlo experiment!

A Story About Solitaire (1946, Los Alamos)

A brilliant mathematician named Stanislaw Ulam was recovering from an illness at home, playing card solitaire. He wondered: what are my chances of winning? Instead of calculating the exact math (very hard!), he thought: 'What if I just played 100 games and counted how many I won?' That count divided by 100 would give a reliable estimate. He realized this idea could solve MUCH bigger problems at Los Alamos. Together with John von Neumann, this became the modern Monte Carlo method — named in 1946!

> **You Are Scientists!**
> By flipping the Galton Board, predicting where the golden bead lands, and reading Pascal's Triangle, you are thinking like real scientists at Los Alamos National Laboratory!

LESSON 4 — The Bell Curve

Why Does the Middle Fill Up?

To land in the edge bins (0 or 14), a bead must bounce the SAME direction every single time for all 14 rows. That is like flipping heads 14 times in a row — possible, but almost never happens! To land near the middle, a bead just needs to go left and right roughly the same number of times — easy, because there are many different combinations of lefts and rights that still average out to the center.

Where the Bead Lands	How Hard Is It?	How Often?
Bin 7 (center)	Easy — lots of paths lead here	Most often (≈21%)
Bins 5, 6, 8, 9	Pretty easy	Often
Bins 3, 4, 10, 11	A bit harder	Sometimes
Bins 1, 2, 12, 13	Hard — few paths	Rarely
Bin 0 or 14	Almost impossible (same way 14 times!)	Almost never!

The Bell Curve in Real Life

- Heights of people in your school — most near average, few very tall or very short
- Size of apples on a tree — most medium, very few tiny or giant
- How far darts land from the bullseye — most close, few far away
- How neutrons (tiny particles) travel inside a reactor at Los Alamos — they spread in a bell pattern!

LESSON 5 — Pascal's Triangle

One Simple Rule

The hexagons on the board display Pascal's Triangle — named after Blaise Pascal (1623–1662), though mathematicians in Persia, India, and China discovered it centuries earlier. One simple rule builds the whole triangle:

★ Each number = the TWO numbers directly above it added together! ★

Each hexagon's number tells you how many different paths a bead can take to arrive at that hexagon from the top. MORE PATHS → MORE BEADS → TALLER BIN → BELL CURVE! Pascal's Triangle is the mathematical explanation for why the bell curve appears.

Pattern	Where to Find It	Example
Powers of 2	Right side of board	Row 4 sums to 1+4+6+4+1 = 16 = 2^4
Powers of 11	Left side of board	Row 3: 1,3,3,1 → 1331 = 11^3
Fibonacci numbers	Diagonal lines on right side	1,1,2,3,5,8,13,21,34,55,89,144...
Symmetry	Every row reads same forwards and backwards	1,3,3,1 → reverse: 1,3,3,1
Row totals double	Right edge	Each row's total = exactly double the row above

LESSON 6 — Reading the Board and Real-World Connections

Standard Deviation Lines

Region	Between These Lines	Expected % of Beads
1 standard deviation	$\mu-1\sigma$ to $\mu+1\sigma$	68.26%
2 standard deviations	$\mu-2\sigma$ to $\mu+2\sigma$	95.44%
3 standard deviations	$\mu-3\sigma$ to $\mu+3\sigma$	99.72%
Beyond 4 standard devs	Bins 0 and 14	Only 0.26% of all beads!

Monte Carlo in the Real World

Field	How Monte Carlo Helps
Weather forecasting 🌧	Run thousands of random weather simulations. Count how many show rain. That count ÷ total runs = the '70% chance of rain' forecast!
Space exploration 🚀	Simulate many possible paths an asteroid might take to calculate the probability of it approaching Earth.
Nuclear physics ☢ (Los Alamos!)	Simulate millions of individual particle journeys through materials to determine safety.
Medicine 💊	Simulate many possible molecular combinations to find which drugs might work.
Computer graphics 🎮	Simulate how light bounces off surfaces millions of times to render realistic shadows in movies and games.

Grade 3 Word Wall

Word	What It Means
Chance	When we can't know for sure what will happen next.
Fair	When all outcomes are equally likely.
Probability	How likely something is to happen. Heads has probability 1/2.
Trial	One single random experiment. One flip of the Galton Board is one trial.
Monte Carlo	A method for solving hard problems by doing many random trials and watching the pattern. Invented at Los Alamos in the 1940s.
Bell Curve	A hill-shaped graph where most results are in the middle and fewer are at the edges.
Pascal's Triangle	A triangle of numbers where each number equals the two above it added together.
Golden Bead	The one special bead among the 4,280. It represents a single Monte Carlo trial.
Standard Deviation (σ)	A measure of how spread out data is from the average.
Central Limit Theorem	The mathematical law that says: when you repeat many small random events, the results always pile up in a bell curve.

Grade 3 Assessment Ideas

Quick Checks

- Thumbs up/sideways/down: 'Is this fair?' (show scenarios)
- Exit ticket: 'Draw the Galton Board after a flip. Label the bin with the most beads.'

- Think-pair-share: 'Why does the middle bin get the most beads?'
- Monte Carlo check: 'Should you flip the board 1 time or 20 times to find where the golden bead usually lands? Why?'

End-of-Unit Written Check

- Draw a coin. What are the two outcomes? Which is more likely?
- If you flip a coin 20 times, about how many heads do you expect? Why?
- What is Monte Carlo? Explain it in your own words.
- Draw the shape that the Galton Board beads make. What is it called?
- Why is the golden bead harder to predict than all 4,280 beads together?

Creative Extension

Have students write a 'Monte Carlo Scientist's Journal' following one imaginary golden bead through all 14 bounces. For each row, flip a real coin: heads = go right, tails = go left. Draw the path! Which bin does the bead land in? Write a short story from the bead's point of view.

GRADE 6

Probability Fractions, Binomial Distribution & Empirical Rule

Welcome, Teacher! — 6th Grade

This guide uses the Galton Board to build rigorous 6th-grade probability and statistics through hands-on investigation. Students move from intuitive chance to formal probability fractions, binomial distributions, the normal distribution, the empirical rule, and an introduction to Monte Carlo simulation.

6th Grade Learning Objectives

- Express probability as a fraction, decimal, and percent; calculate theoretical vs. experimental probability
- Interpret and extend Pascal's Triangle; connect row entries to binomial coefficients C(n,k)
- Read a binomial distribution table and explain why it approximates the normal distribution
- Apply the empirical rule (68–95–99.7) to interpret standard deviation lines on the Galton Board
- Explain in writing what a Monte Carlo simulation is and why more trials increase reliability
- Calculate sample mean and sample standard deviation from Galton Board flip data
- Construct a properly labeled histogram and overlay a smooth bell curve
- Identify the Fibonacci sequence in Pascal's Triangle and approximate the golden ratio

Pacing

6–7 class periods of ~50 minutes. L1–2: probability foundations. L3: binomial distribution. L4: normal distribution and empirical rule. L5: Pascal's Triangle. L6: Monte Carlo. L7: culminating data investigation.

LESSON 1 — Probability as a Fraction

From Chance to Numbers

P(event) = (number of favorable outcomes) / (total number of equally likely outcomes)

Probability	As a Fraction	As a Decimal	As a Percent	What It Means
Impossible	0/1	0.00	0%	Can never happen
Very unlikely	1/10	0.10	10%	Happens about 1 in 10 tries
Even odds	1/2	0.50	50%	Happens about 1 in 2 tries
Likely	3/4	0.75	75%	Happens about 3 in 4 tries
Certain	1/1	1.00	100%	Always happens

Single-Bead Probability on the Galton Board

When a bead hits any hexagon, it bounces left or right with equal probability. After 14 rows, a bead has made 14 independent binary decisions. The probability of landing in bin k after 14 rows is:

$P(\text{bin } k) = C(14,k) \times (1/2)^{14} = C(14,k) / 16{,}384$

Here C(14,k) is the number of distinct paths leading to bin k — exactly the Pascal's Triangle entry in row 14, position k.

LESSON 2 — The Galton Board as a Physical Model

Every Feature Maps to Mathematics

Physical Feature	Mathematical Concept	Key Number
4,280 steel beads	Independent identically distributed random variables	n = 4,280 trials per flip
14 rows of hexagons	Number of Bernoulli trials per bead	n = 14 trials per bead
15 bins (0–14)	Discrete random variable values	k = 0, 1, 2, ..., 14
1 golden bead	A single Monte Carlo trial	1 sample from distribution
Pascal's Triangle overlay	Binomial coefficients C(14,k)	C(14,7) = 3,432 at center
Bell curve overlay	Normal distribution approximation	$\mu = 7$, $\sigma = \sqrt{3.5} \approx 1.87$
Standard deviation lines	Empirical rule: 68–95–99.7%	$\mu \pm 1\sigma, 2\sigma, 3\sigma, 4\sigma$
59,920 coin tosses per flip	Total Bernoulli trials: 4,280 × 14	59,920 per flip

LESSON 3 — The Binomial Distribution

The Binomial Formula

$P(X = k) = C(n,k) \times p^k \times (1-p)^{n-k}$

For the Galton Board with p = 0.5, this simplifies to: P(bin k) = C(14,k) / 16,384

Full Binomial Distribution Table

Bin (k)	C(14,k)	P as %	Expected beads (of 4,280)
0	1	0.006%	0.3
1	14	0.085%	3.6
2	91	0.555%	23.8
3	364	2.222%	95.1

4	1,001	6.116%	261.8
5	2,002	12.207%	522.5
6	3,003	18.311%	783.7
7	3,432	20.947%	896.5
8	3,003	18.311%	783.7
9	2,002	12.207%	522.5
10	1,001	6.116%	261.8
11	364	2.222%	95.1
12	91	0.555%	23.8
13	14	0.085%	3.6
14	1	0.006%	0.3

Check Your Work!

The probabilities must sum to 1. Add all P(bin k) values: (1+14+91+364+1001+2002+3003+3432+3003+2002+1001+364+91+14+1) / 16,384 = 16,384 / 16,384 = 1. ✓

LESSON 4 — The Normal Distribution & Empirical Rule

From Binomial to Normal

The binomial distribution B(n, 0.5) becomes more bell-shaped as n increases. At n = 14, it is already a very good approximation of the normal distribution, owing to the Central Limit Theorem. For the Galton Board:

$\mu = n \times p = 14 \times 0.5 = 7.0$

$\sigma = \sqrt{(n \times p \times (1-p))} = \sqrt{(14 \times 0.5 \times 0.5)} = \sqrt{3.5} \approx 1.871$

The Empirical Rule (68–95–99.7 Rule)

Range	% of values	For Galton Board	Expected beads
$\mu \pm 1\sigma$	68.27%	Roughly bins 5–9	≈2,919 beads
$\mu \pm 2\sigma$	95.45%	Roughly bins 3–11	≈4,085 beads
$\mu \pm 3\sigma$	99.73%	Roughly bins 1–13	≈4,268 beads
$\mu \pm 4\sigma$	99.994%	All bins 0–14	≈4,280 beads

Why the Empirical Rule Matters

The 68-95-99.7 rule tells scientists how to interpret any normally-distributed result. A Monte Carlo simulation at Los Alamos reports a result μ ± σ. Scientists know immediately that the true answer has a 68.27% chance of being within 1σ of μ, and a 95.45% chance of being within 2σ. This is how all scientific uncertainty is communicated.

LESSON 5 — Pascal's Triangle: Patterns & Proofs

Pascal's Rule

C(n,k) = C(n-1,k-1) + C(n-1,k)

Ten Patterns in Pascal's Triangle

Pattern	Location on Board	Example / Formula
Row totals = powers of 2	Right margin	Row 4: 1+4+6+4+1 = 16 = 2^4
Powers of 11	Left margin	Row 3: 1331 = 11^3
Symmetry	Mirror image left-right	C(n,k) = C(n,n-k)
Fibonacci numbers	Diagonal sums on right	1,1,2,3,5,8,13,21,34,55,89,144
Triangular numbers	2nd diagonal	T_n = n(n+1)/2
Hockey stick pattern	Sum a diagonal = next diagonal below	1+3+6+10 = 20 = C(6,3)
Sierpiński triangle	Shade odd numbers	Fractal self-similarity
Binomial theorem	Entire triangle	$(a+b)^n$ = sum C(n,k) a^k b(n-k)

Fibonacci and the Golden Ratio

The Fibonacci numbers on the board's diagonals converge to the golden ratio φ = (1 + √5) / 2 ≈ 1.61803...

Fibonacci Ratio	Calculation	Value	Difference from φ
F(8)/F(7)	21/13	1.6154	0.0027
F(10)/F(9)	55/34	1.6176	0.0004
F(12)/F(11)	144/89	1.6180	0.00004
F(14)/F(13)	377/233	1.61803	0.000003

LESSON 6 — Monte Carlo Methods

The Core Idea

A Monte Carlo method solves a problem by running many random simulations and aggregating the results. The accuracy improves with the number of trials N. Specifically:

Error ∝ 1/√N

To halve the error, you need 4 times as many trials. To reduce error by a factor of 10, you need 100 times as many trials. This is why Monte Carlo computers run millions of simulations.

The 1/√N Law in Action

Number of flips (N)	Approximate error in P(bin 7)	Example: if true P = 0.209

1 flip	≈0.62%	Your estimate: 0.209 ± 0.006
4 flips	0.31%	Your estimate: 0.209 ± 0.003
9 flips	0.21%	Your estimate: 0.209 ± 0.002
100 flips	0.062%	Your estimate: 0.209 ± 0.001
10,000 flips	0.0062%	Very precise estimate

Historical Development at Los Alamos

1777 — Buffon's Needle: first probabilistic estimation, calculating π by dropping a needle on a lined floor.

1946 — Stanislaw Ulam invents Monte Carlo while playing solitaire. Shares with John von Neumann.

1947 — Metropolis suggests the code name 'Monte Carlo' after Ulam's uncle.

1948 — ENIAC (first electronic computer) runs early Monte Carlo calculations at LANL.

1949 — Metropolis and Ulam publish 'The Monte Carlo Method' in JASA, establishing the field.

1977 — MCNP® code first developed at LANL. Now simulates nuclear reactors, medical radiation therapy, and space radiation shielding.

Grade 6 Key Vocabulary

Term	Definition
Bernoulli trial	A single binary random experiment. Each hexagon bounce is a Bernoulli trial with $p = 0.5$.
Binomial distribution B(n,p)	Distribution of count of successes in n independent Bernoulli trials. $P(X=k) = C(n,k)\, p^k (1-p)^{n-k}$.
Binomial coefficient C(n,k)	n! / (k!(n-k)!). Counts the number of ways to choose k items from n. Appears in Pascal's Triangle.
Normal distribution	A continuous bell-shaped probability distribution described by mean μ and standard deviation σ.
Mean μ (mu)	The average (expected value). For B(14, 0.5): $\mu = 14 \times 0.5 = 7.0$.
Standard deviation σ	Measures spread. For B(14, 0.5): $\sigma = \sqrt{3.5} \approx 1.87$ bins.
Empirical rule	For normal distributions: ~68% within ±1σ, ~95% within ±2σ, ~99.7% within ±3σ.
Central Limit Theorem	Sum of many independent random variables tends to a normal distribution.
Law of Large Numbers	As trials increase, experimental probability converges to theoretical probability.
Monte Carlo method	Solving problems by running many random simulations and averaging results. Error $\propto 1/\sqrt{N}$.
z-score	z = (observed - expected) / σ. Measures how many standard deviations an observation is from the mean.

Pascal's Rule	C(n,k) = C(n-1,k-1) + C(n-1,k). The fundamental recurrence that builds Pascal's Triangle.
Golden ratio φ	(1+√5)/2 ≈ 1.618. The limit of consecutive Fibonacci ratios.
MCNP®	Monte Carlo N-Particle code. First developed at LANL in 1977. Used to simulate neutron, photon, and electron transport.

Grade 6 Assessment Bank

End-of-Unit Assessment (Part A: Calculations)

- Calculate C(14,4) using the formula n!/(k!(n-k)!). Then verify using Pascal's Rule.
- A bead falls through the Galton Board. Calculate P(bin 4). Express as a fraction, decimal, and percent.
- For B(14, 0.5), compute μ and σ. Show all steps.
- Using the empirical rule, estimate the probability that a bead lands within 2 standard deviations of the mean.
- If a Monte Carlo simulation with N = 400 trials gives error 5%, what N gives error 2.5%? Error 1%?

Part B: Conceptual Questions

- Explain in two sentences why the center bins always receive more beads than the edge bins.
- What is the Central Limit Theorem? How does the Galton Board demonstrate it physically?
- A Los Alamos scientist runs a Monte Carlo simulation with 10,000 trials and reports 2.54 ± 0.03. What does the ±0.03 represent? How could she reduce the uncertainty to ±0.01?

Challenge Problem

The Galton Board uses p = 0.5 at each hexagon. Suppose we modified the board so that p = 0.6 (60% chance of going right). Compute: (a) the new mean μ = np; (b) the new standard deviation σ = √(np(1-p)); (c) P(bin 7) under the new distribution; (d) which bin would now be most likely?

GRADE 9

Probability Axioms, Statistical Inference & Random Walks

Welcome, Teacher! — 9th Grade / Algebra II

This guide connects the Galton Board to rigorous 9th-grade statistics. Students derive results from first principles, run formal hypothesis tests, and engage with the same probabilistic reasoning used by scientists at Los Alamos. Prerequisites: Algebra I, basic combinatorics, summation notation.

9th Grade Learning Objectives

- State Kolmogorov's three axioms and derive elementary probability rules from them
- Prove the binomial theorem by induction and connect coefficients to Pascal's Triangle
- Derive $\mu = np$ and $\sigma^2 = np(1-p)$ for $B(n,p)$ from the definition of expectation
- Sketch the de Moivre–Laplace theorem argument connecting $B(n,p)$ to $N(\mu,\sigma^2)$
- Perform a chi-square goodness-of-fit test on Galton Board data and interpret the result
- Construct a confidence interval for a binomial proportion from flip data
- Model the bead path as a random walk and connect it to Brownian motion and financial markets
- Explain Monte Carlo variance reduction techniques with quantitative error bounds
- Describe the Metropolis algorithm and its role in computational physics

Pacing

7–8 class periods of 50 minutes. Lessons 7 (random walk / EMH) and 8 (Metropolis algorithm) are enrichment lessons suitable for advanced classes. The chi-square activity in Lesson 5 requires a χ^2 table (provided in the Assessment Appendix).

Ω LESSON 1 — Probability Axioms & Sample Spaces

Kolmogorov's Axiomatic Foundation (1933)

Modern probability theory rests on three axioms. All probability results — including everything on the Galton Board — can be derived from these three statements alone.

Kolmogorov's Three Axioms

Axiom 1 (Non-negativity): $P(A) \geq 0$ for every event A. Axiom 2 (Normalization): $P(\Omega) = 1$. Axiom 3 (Countable additivity): If $A_1, A_2, \ldots$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

Derived Rules (proven from the axioms alone)

Rule	Statement	Proof Sketch
Complement	$P(A^c) = 1 - P(A)$	A and A^c are disjoint; $A \cup A^c = \Omega$; apply Axioms 2 & 3

Impossible event	$P(\emptyset) = 0$	$P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 0$
Monotonicity	$A \subseteq B \Rightarrow P(A) \le P(B)$	$B = A \cup (B \cap A^c)$; apply Axiom 3 and non-negativity
Addition rule	$P(A \cup B) = P(A)+P(B)-P(A \cap B)$	$A \cup B = A \cup (B \cap A^c)$; careful application of Axiom 3
Bayes' theorem	$P(B \mid A) = P(A \mid B)P(B) / P(A)$	Rewrite $P(A \cap B)$ two ways

Sample Space for the Galton Board

For a single bead traversing 14 rows, the sample space is all sequences of L (left) and R (right) of length 14:

$\Omega = \{ (d_1, d_2, \ldots, d_{14}) : d_i \in \{L, R\} \}$ $\quad |\Omega| = 2^{14} = 16{,}384$

$P(X = k) = C(14,k) / 16{,}384$

> **Verify Axiom 2**
>
> The probabilities must sum to 1. By the Binomial Theorem: $\Sigma_{k=0}^{14} C(14,k) = (1+1)^{14} = 2^{14} = 16{,}384$. So $\Sigma P(X=k) = 16{,}384/16{,}384 = 1$. ✓ Axiom 2 forces the binomial coefficients to sum to a power of 2.

∑ LESSON 2 — Combinatorics & the Binomial Theorem

Pascal's Rule: An Algebraic Proof

> **Proof of Pascal's Rule: $C(n,k) = C(n-1,k-1) + C(n-1,k)$**
>
> $C(n-1,k-1) + C(n-1,k) = (n-1)!/((k-1)!(n-k)!) + (n-1)!/(k!(n-k-1)!)$ Factor out $(n-1)! / (k!(n-k)!)$: $= [(n-1)!/(k!(n-k)!)] \times [k + (n-k)] = [(n-1)!/(k!(n-k)!)] \times n = n! / (k!(n-k)!) = C(n,k)$ ■

The Binomial Theorem

$(a + b)^n = \Sigma_{k=0}^{n} C(n,k) \times a^{n-k} \times b^k$

Key Corollaries from the Galton Board

- Set a=b=1: $(1+1)^n = \Sigma C(n,k) = 2^n$. Row sums of Pascal's Triangle are powers of 2. ✓
- Set a=1, b=-1: $0 = \Sigma (-1)^k C(n,k)$. Alternating sum of each row = 0. ✓
- Differentiating informally: $n \cdot 2^{n-1} = \Sigma k \cdot C(n,k)$. This gives $E[X] = n/2$ for the board!

μ LESSON 3 — Expectation, Variance & Deriving μ = np, σ² = np(1-p)

Deriving E[X] = np

> **Derivation: $E[X] = np$ for $X \sim B(n,p)$**
>
> $E[X] = \Sigma_{k=0}^{n} k \times C(n,k)\, p^k (1-p)^{n-k}$ The k=0 term is 0. Factor: $k \cdot C(n,k) = n \cdot C(n-1,k-1)$: $E[X] = np \times \Sigma_{k=1}^{n} C(n-1,k-1)\, p^{k-1} (1-p)^{n-k}$ Substitute j = k-1: $E[X] = np \times \Sigma_{j=0}^{n-1} C(n-1,j)\, p^j (1-p)^{n-1-j}$ The sum = (p + (1-

p))$^{n-1}$ = 1 by the Binomial Theorem. Therefore: E[X] = np ■ For the Galton Board: E[X] = 14 × 0.5 = 7.0

Deriving Var(X) = np(1-p)

Using $E[X^2] = E[X(X-1)] + E[X]$ and the same factoring technique:

- $E[X(X-1)] = n(n-1)p^2$
- $E[X^2] = n(n-1)p^2 + np$
- $Var(X) = E[X^2] - (E[X])^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$ ■
- For the Galton Board: Var(X) = 14×0.5×0.5 = 3.5, σ = √3.5 ≈ 1.871 bins

~ LESSON 4 — The Central Limit Theorem: From Binomial to Normal

The de Moivre–Laplace Theorem

de Moivre–Laplace Theorem

Let X ~ B(n, p). Define the standardized variable: Z_n = (X - np) / √(np(1-p)) Then for any fixed a < b: lim_{n→∞} P(a ≤ Z_n ≤ b) = Φ(b) - Φ(a) where Φ(x) is the standard normal CDF. In words: the standardized binomial converges in distribution to N(0,1) as n → ∞.

Using the Normal Approximation with Continuity Correction

When using the normal approximation for a discrete distribution, replace P(a ≤ X ≤ b) with P(a-0.5 ≤ X ≤ b+0.5). This accounts for the fact that a discrete bar of width 1 is being approximated by a continuous curve.

n	Exact P(center bin)	Normal Approx	Error
4	C(4,2)/16 = 37.50%	≈38.29%	0.79%
8	C(8,4)/256 = 27.34%	≈27.09%	0.25%
14	C(14,7)/16384 = 20.95%	≈21.28%	0.33%
100	≈7.96%	≈7.98%	0.02%

χ² LESSON 5 — Chi-Square Goodness-of-Fit Test

The Chi-Square Statistic

$\chi^2 = \Sigma_k (O\_k - E\_k)^2 / E\_k$

For each bin k, O_k = observed bead count and E_k = expected count (4280 × P(bin k)). Under H_0 (the board is fair), this statistic follows a chi-square distribution with df = 14 (or 12 after merging low-count edge bins).

Step-by-Step Procedure

- State hypotheses: H_0: Board is fair (p=0.5). H_1: Board is biased.
- Choose significance level $\alpha = 0.05$.
- Find critical value: for χ^2 with df=12 at α=0.05, the critical value is $\chi^2\_crit = 21.026$.
- Flip the board once. Record counts O_k for each bin.
- Compute expected counts $E\_k = 4280 \times C(14,k)/16384$.
- Compute $\chi^2 = \Sigma (O\_k - E\_k)^2 / E\_k$.
- Decision: if $\chi^2 > 21.026$, reject H_0.

χ^2 range (df=12)	Interpretation	p-value range
< 4.40	Suspiciously good fit	p > 0.97
4.40 – 18.55	Normal variation, no concern	0.10 < p < 0.97
21.03 – 26.22	Significant evidence of bias (reject H_0)	0.01 < p < 0.05
> 26.22	Strong evidence of bias	p < 0.01

LANL Connection

Code verification at LANL uses chi-square tests on Monte Carlo output. When a new version of MCNP® is released, its results on benchmark problems are compared to known analytic solutions using statistical tests. The Galton Board teaches exactly the statistical reasoning that underpins this validation workflow.

LESSON 6 — Confidence Intervals for Proportions

The Wald Confidence Interval

$\hat{p} \pm 1.96 \times \sqrt{(\hat{p}(1-\hat{p})/n)}$

From n flips of the golden bead, observing k landings in bin 7: $\hat{p} = k/n$. The 95% CI above uses z = 1.96 from the standard normal, since $P(-1.96 \leq Z \leq 1.96) = 0.95$.

Crucial Distinction

The 95% refers to the PROCEDURE, not to any single interval. The true parameter is fixed (not random). The interval is random — it varies across samples. A 95% CI does NOT mean 'there is a 95% probability the true value is in this interval' for any one specific interval.

n (flips)	95% CI width (if $\hat{p}$=0.20)	Includes 0.2095?
10	±0.248 → width 0.496	Almost certainly
20	±0.175 → width 0.350	Almost certainly
50	±0.111 → width 0.222	Yes
100	±0.078 → width 0.157	Yes
400	±0.039 → width 0.078	Yes
1000	±0.025 → width 0.050	Tight!

LESSON 7 — Random Walk & the Efficient Market Hypothesis

The Galton Board as a Random Walk

Track the bead's horizontal position as it moves through successive rows. At each row, position changes by +1 (right) or -1 (left) with equal probability. Define the bead's position after k rows as $S_k = X_1 + X_2 + ... + X_k$ where X_i = +1 or -1 each with probability 1/2.

$E[S_k] = 0$ $\quad$ $Var(S_k) = k$ $\quad$ $\sigma_k = \sqrt{k}$

The $\sqrt{k}$ growth in spread is fundamental. It appears in Brownian motion (Einstein, 1905), stock price models (Black-Scholes, 1973), neutron random walks in reactor cores at Los Alamos, and polymer chain statistics.

Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH, Fama 1970) states that asset prices fully reflect all available information. A key implication: price changes should be unpredictable — they should form a random walk. If prices were predictable, arbitrageurs would immediately trade away the predictability.

Market Model	Price Process	Galton Board Analogy
Random walk	$P(t) = P(0) + \Sigma X_i$, $X_i = \pm\varepsilon$	Bead path: $S_k = \Sigma (\pm 1)$ random steps
Geometric random walk	$P(t) = P(0) \times \Pi (1 + r_i)$	Multiplicative shocks, normal returns
Brownian motion	$dP = \mu\, dt + \sigma\, dW$	Continuous-time limit of the walk
Black-Scholes	$dS = \mu S\, dt + \sigma S\, dW$	Used to price options (Nobel 1997)

IFA's Connection

Index Fund Advisors (IFA), which co-created this Galton Board, was founded on evidence-based passive investing directly inspired by the EMH and random walk research. Mark Hebner's book uses the Galton Board to illustrate why predicting which stocks will outperform is as futile as predicting where the golden bead will land.

LESSON 8 — Advanced Monte Carlo: The Metropolis Algorithm

The Metropolis Algorithm (1953, Los Alamos)

The Metropolis Algorithm

Goal: sample from a target distribution $\pi(x)$. 1. Start at any state x_0. 2. At each step t, propose a new state x' from a symmetric proposal distribution $q(x'|x_t)$. 3. Compute acceptance ratio: $\alpha = \min(1, \pi(x') / \pi(x_t))$ 4. Accept x' with probability α; otherwise stay at x_t. 5. The sequence $x_0, x_1, x_2, ...$ converges to $\pi(x)$. Key property: π is sampled correctly in the long run, guaranteed by the detailed balance condition: $\pi(x)q(x'|x) = \pi(x')q(x|x')$.

Method	Galton Board (Direct)	Metropolis (MCMC)
Sample generation	Physical: bead falls through pegs	Computational: accept/reject chain
Requires knowing distribution?	No — physics generates it	No — only ratios $\pi(x')/\pi(x)$ needed
Samples independent?	Yes — each bead is fresh	No — chain has autocorrelation
High-dimensional problems?	No — 1D only	Yes — millions of dimensions
Modern applications	Education, finance illustration	Bayesian inference, drug discovery, LLM training

From MANIAC to GPT: The Metropolis Legacy

1953: First MCMC simulation on the MANIAC at Los Alamos — 224 hard-sphere molecules. 1970: Hastings generalizes to non-symmetric proposals. 1990s: MCMC revolutionizes genetics, epidemiology, climate modeling. 2020s: MCMC variants power protein structure prediction (AlphaFold2) and are used in MCNP® for nuclear criticality safety analysis at LANL. The algorithm that ran on 1,000 vacuum tubes in 1953 now runs on GPU clusters with billions of parameters. The mathematics is unchanged.

Grade 9 Assessment Bank

Proof-Based Questions

- Prove that $P(\emptyset) = 0$ using only Kolmogorov's three axioms.
- Prove by induction that the sum of all entries in row n of Pascal's Triangle equals 2^n.
- Derive $E[X] = np$ for $X \sim B(n,p)$ using the indicator variable decomposition $X = X_1+...+X_n$.
- Show that $Var(X_1+X_2) = Var(X_1)+Var(X_2)$ when X_1 and X_2 are independent. Why does independence matter?

Computation Questions

- A bead has passed through 10 rows and is in column 6. What is P(it lands in bin 7)? Bin 10?
- Construct a 95% Wilson confidence interval for P(bin 5) using n=20 flips and k=3 bin-5 landings.
- For a random walk with 100 steps, compute $E[S_{100}]$, $Var(S_{100})$, and $P(|S_{100}| \leq 10)$.
- How many Galton Board flips are needed to estimate P(bin 7) with a 95% CI of width ≤ 0.02?

Chi-Square Table (Critical Values)

df	α = 0.10	α = 0.05	α = 0.01	α = 0.001
1	2.706	3.841	6.635	10.828
6	10.645	12.592	16.812	22.458
12	18.549	21.026	26.217	32.909
14	21.064	23.685	29.141	36.123
20	28.412	31.410	37.566	45.315

For the Galton Board chi-square test: use df = 12 (after merging low-count edge bins). At $\alpha = 0.05$, critical value = 21.026. Reject H_0 if $\chi^2 > 21.026$.

GRADE 12

Measure Theory, CLT Proof, Bayesian Inference & MCMC Convergence

Welcome, Teacher! — 12th Grade / AP Statistics / AP Calculus BC

This guide treats the Galton Board as a gateway to rigorous university mathematics: measure-theoretic probability, characteristic functions, the Gaussian integral, Stirling's formula, generating functions, Bayesian inference, statistical mechanics, and the theoretical foundations of Monte Carlo convergence. Prerequisites: single-variable calculus, linear algebra, and the 9th-grade content of this series.

12th Grade Learning Objectives

- State the measure-theoretic definition of probability and explain why it is needed beyond Kolmogorov's axioms
- Prove the Gaussian integral $\int e^{-x^2}\,dx = \sqrt{\pi}$ using polar coordinates
- Derive Stirling's approximation $n! \approx \sqrt{2\pi n}(n/e)^n$ and apply it to $C(2n,n)$
- Compute moment generating functions and characteristic functions; use them to prove the CLT
- Prove Vandermonde's identity and derive the probability generating function for the Galton Board
- Construct Bayesian credible intervals using the Beta-Binomial conjugate model
- Connect the binomial distribution to the Boltzmann distribution via maximum entropy
- Derive Shannon entropy and show the normal distribution maximizes entropy for fixed variance
- Explain Markov chain ergodicity and prove detailed balance implies stationarity
- Compute Value at Risk (VaR) and Expected Shortfall from a normal distribution

Pacing

8 lessons of 50–60 minutes, plus an open-ended capstone project. Lessons 1–3 are the most mathematically demanding and may require 2 periods each. Lessons 6–8 are enrichment for motivated students. The capstone project asks students to write a 3–5 page paper connecting the Galton Board to a domain of their choice.

Σ LESSON 1 — Measure-Theoretic Probability & The Gaussian Integral

Why Measure Theory?

Kolmogorov's axioms leave an important question unanswered: what is the collection F of events to which we assign probabilities? For continuous distributions — like the normal distribution printed on the board — we cannot assign probabilities to every subset of $\mathbb{R}$ without contradictions (the Banach-Tarski paradox). The resolution is measure theory: we restrict attention to a σ-algebra of 'measurable' sets.

Formal Probability Space (Ω, F, P)

Ω: sample space (any non-empty set) F: a σ-algebra — closed under complement and countable union P: F → [0,1] satisfying $P(\Omega) = 1$ and countable additivity A random variable $X: \Omega \to \mathbb{R}$ is a measurable function: $\{X \le x\} \in F$ for all $x \in \mathbb{R}$.

The Gaussian Integral: Proving ∫ e^(-x²)dx = √π

Proof via Polar Coordinates

Let $I = \int_{-\infty}^{\infty} e^{-x^2}\, dx$. Consider I^2: $I^2 = \iint e^{-(x^2+y^2)}\, dx\, dy$ Switch to polar: $x = r\cdot\cos\theta$, $y = r\cdot\sin\theta$, $dx\, dy = r\, dr\, d\theta$ $I^2 = \int_0^{2\pi}\int_0^{\infty} e^{-r^2}\, r\, dr\, d\theta = 2\pi \int_0^{\infty} r\, e^{-r^2}\, dr$ Substitute $u = r^2$: $I^2 = 2\pi \times (1/2) \int_0^{\infty} e^{-u}\, du = \pi \times 1 = \pi$ Therefore $I = \sqrt{\pi}$. ■ Consequence: the normalization constant $1/(\sigma\sqrt{2\pi})$ ensures $\int f(x)dx = 1$.

Stirling's Approximation

Stirling's Approximation: n! ≈ √(2πn)(n/e)ⁿ

Using the Gamma function and Taylor-expanding log t around the mode t=n: The integrand becomes approximately $n^n e^{-n} \times \exp(-s^2/2) \times \sqrt{n}$ So: $n! \approx n^n e^{-n} \sqrt{n} \times \sqrt{2\pi} = \sqrt{2\pi n}(n/e)^n$ ■ Error: $|n! - \text{Stirling}| / n! \approx 1/(12n)$. At n=14: Stirling gives 8.71×10^{10} vs exact 8.72×10^{10}. For the center bin: $C(14,7)/2^{14} \approx 1/\sqrt{7\pi} \approx 0.2132$. Exact: 3432/16384 = 0.2095. Error: ~1.7%

The Deep Connection

Stirling's approximation shows that the Gaussian integral $\int e^{-x^2}\, dx = \sqrt{\pi}$ is not just the normalization of the normal distribution — it is the reason that factorials (and therefore binomial coefficients) look Gaussian for large n. The bell curve on the Galton Board emerges from the behavior of n! as $n\to\infty$.

φ LESSON 2 — Moment Generating Functions & the CLT Proof

Moment Generating Functions (MGFs)

$M_X(t) = E[e^{tX}]$ $M_X^{(k)}(0) = E[X^k]$

The MGF of B(n,p) is $M_X(t) = (pe^t + (1-p))^n$. Uniqueness property: if two random variables have the same MGF in a neighborhood of 0, they have the same distribution.

The CLT via MGFs: A Complete Proof

CLT Proof via MGF Convergence

Let $X_1,\ldots,X_n$ be i.i.d. with $E[X_i]=p$, $Var(X_i)=p(1-p)$. Let $Z_n = (S_n - np) / \sqrt{np(1-p)}$. Let $\tau = t/\sqrt{np(1-p)}$. Taylor expand $M_{X_i}(\tau)$ about $\tau=0$: $M_{X_i}(\tau) = 1 + p\tau + E[X_i^2]\tau^2/2 + O(\tau^3)$ After centering: $\log M_{Z_n}(t) = n \log(1 + t^2/(2n) + O(n^{-3/2})) \to t^2/2$ as $n\to\infty$ Therefore: $M_{Z_n}(t) \to e^{t^2/2}$ — the MGF of N(0,1)! By the continuity theorem for MGFs: $Z_n \to N(0,1)$ in distribution. ■

Characteristic Functions: The Full Generality

The characteristic function $\varphi_X(t) = E[e^{itX}]$ always exists (even for heavy-tailed distributions where MGFs fail). For the standard normal, $\varphi_Z(t) = e^{-t^2/2}$. The CLT proof via characteristic functions and the Lévy continuity theorem covers all distributions with finite mean and variance.

Distribution	MGF M(t)	Mean	Variance
Bernoulli(p)	$(1-p)+pe^t$	p	p(1-p)
B(n,p)	$((1-p)+pe^t)^n$	np	np(1-p)
$N(\mu,\sigma^2)$	$e^{\mu t+\sigma^2 t^2/2}$	μ	σ^2
Poisson(λ)	$e^{\lambda(e^t-1)}$	λ	λ
Cauchy	does not exist	undefined	undefined

∑ LESSON 3 — Generating Functions & Advanced Combinatorial Identities

Probability Generating Functions

For $X \sim B(14, 1/2)$, the PGF is $G_X(z) = ((1+z)/2)^{14}$. The coefficient of z^k is $C(14,k)/2^{14} = P(X=k)$. This is exactly the Binomial Theorem in generating function form.

Vandermonde's Identity

Vandermonde's Identity: $C(m+n,k) = \Sigma_{j=0}^k C(m,j)\cdot C(n,k-j)$

Proof via generating functions: $(1+z)^m \times (1+z)^n = (1+z)^{m+n}$ Coefficient of z^k on the left (Cauchy product): $\Sigma_{j=0}^k C(m,j)\, C(n,k-j)$ Coefficient of z^k on the right: $C(m+n,k)$ Equating: $C(m+n,k) = \Sigma_{j=0}^k C(m,j)C(n,k-j)$ ■ Galton Board interpretation: $C(14,7) = \Sigma_{j=0}^7 C(7,j)^2$ — paths through each half of the board.

Closed Form for Fibonacci: Binet's Formula

The Fibonacci numbers on the board's diagonals can be recovered from a single generating function. The recurrence $F_n = F_{n-1} + F_{n-2}$ gives:

$F(z) = z / (1 - z - z^2)$ → $F_n = (\varphi^n - \psi^n) / \sqrt{5}$ (Binet's formula)

where $\varphi = (1+\sqrt{5})/2 \approx 1.618$ and $\psi = (1-\sqrt{5})/2 \approx -0.618$. Since $|\psi| < 1$, $\psi^n \to 0$ and $F_n \approx \varphi^n/\sqrt{5}$ for large n, confirming $F_n/F_{n-1} \to \varphi$.

β LESSON 4 — Bayesian Inference & the Beta-Binomial Model

Bayesian vs. Frequentist

In the Bayesian view, p is itself a random variable with a prior distribution reflecting our beliefs before observing data. After observing data, we update to a posterior via Bayes' theorem:

Posterior ∝ Likelihood × Prior

Beta-Binomial Conjugacy

Conjugate Update: Beta(α,β) + B(n,p) likelihood → Beta(α+k, β+n-k)

Prior: p ~ Beta(α, β) Likelihood: k | p ~ B(n,p) Posterior: p|k ~ Beta(α+k, β+n-k) ■ Intuitive rule: α += k (add observed successes); β += n-k (add observed failures) Posterior mean: E[p|k] = (α+k)/(α+β+n)

Bayesian Analysis of the Galton Board

n flips	k in bin 7	Posterior	Posterior Mean	95% Credible Interval
0 (prior)	0	Beta(1,1)	0.500	(0.025, 0.975)
10	2	Beta(3,9)	0.250	(0.071, 0.524)
50	10	Beta(11,41)	0.212	(0.116, 0.335)
100	21	Beta(22,80)	0.216	(0.143, 0.299)
1000	210	Beta(211,791)	0.210	(0.184, 0.238)

Credible Interval vs. Confidence Interval

Frequentist 95% CI: the procedure [L(data), U(data)] will contain the true (fixed) p in 95% of repetitions. Any particular computed interval either contains p or doesn't. Bayesian 95% Credible Interval: given the observed data, P(p ∈ [L,U]) = 0.95. The parameter p is treated as random (has a posterior distribution). With a flat prior and large n, both intervals are approximately equal.

LESSON 5 — Statistical Mechanics, Maximum Entropy & Boltzmann

Shannon Entropy

$H = -\Sigma_k\, p_k \log_2(p_k)$ (measured in bits)

The Galton Board distribution has H ≈ 3.14 bits, compared to $\log_2(15) \approx 3.91$ bits for a uniform distribution. The Galton Board has lower entropy because it concentrates mass near the center.

Maximum Entropy Principle

Constraints	Maximum Entropy Distribution	Connection to Board
No constraints	Uniform	Would give equal beads in all bins
Fixed mean μ (continuous, [0,∞))	Exponential	Neutron lifetimes in reactor
Fixed mean μ and variance σ^2 (ℝ)	Normal $N(\mu,\sigma^2)$	The bell curve on the board!
Fixed mean k (binary), n trials	Binomial	Exact Galton Board distribution

Normal Distribution Maximizes Entropy for Fixed Mean and Variance

Among all continuous distributions with mean μ and variance σ^2, $N(\mu, \sigma^2)$ has the maximum differential entropy $h(X) = -\int f \log f\, dx$. Proof: use Lagrange multipliers on $L[f] = -\int f \log f + \lambda_0\int f + \lambda_1\int xf + \lambda_2\int x^2 f$. Setting $\delta L/\delta f = 0$ gives $\log f(x) = \lambda_0 - 1 + \lambda_1 x + \lambda_2 x^2$, which is Gaussian. Maximum entropy: $h(N(\mu,\sigma^2)) = (1/2)\log(2\pi e\sigma^2)$. ■

🔗 LESSON 6 — Markov Chain Theory & MCMC Convergence

Stationary Distributions & Ergodicity

Stationary Distribution and Ergodic Theorem

A distribution π is stationary if: π P = π (π_j = Σ_i π_i P_{ij}) A chain is ergodic if it is (1) irreducible and (2) aperiodic. Ergodic Theorem: if ergodic with stationary π, for any starting distribution: P(X_n = j) → π_j as n → ∞ (for all j) AND the time average converges to the space average: (1/N) Σ_{n=0}^{N-1} f(X_n) → E_π[f(X)] a.s. as N→∞ This is why MCMC works: time averages estimate E_π[f].

Detailed Balance ⇒ Stationarity

Proof: Detailed Balance Implies Stationarity

If π_i P_{ij} = π_j P_{ji} for all states i,j, then: (π P)_j = Σ_i π_i P_{ij}   = Σ_i π_j P_{ji}   [by detailed balance]   = π_j Σ_i P_{ji}   = π_j × 1   [rows of P sum to 1]   = π_j ■ Interpretation: probability flux from i to j equals flux from j to i at stationarity.

Spectral Gap and Mixing Time

For a reversible Markov chain with eigenvalues $1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda\_m \geq -1$, the spectral gap is $\gamma = 1 - \lambda_2$. The mixing time satisfies:

τ_mix ≤ (1/γ) × log(1/(επ_min))

Chain property	Effect on mixing	MCMC design implication
Large spectral gap γ ≈ 1	Fast mixing	Few steps needed; estimates converge quickly
Small spectral gap γ ≈ 0	Slow mixing	Many steps needed; high autocorrelation
Multimodal target	Chain trapped in one mode	Parallel tempering, simulated annealing

📈 LESSON 7 — Factor Models, Value at Risk & Portfolio Theory

The Fama-French Three-Factor Model

R_i - R_f = α_i + β_i(R_m-R_f) + s_i·SMB + h_i·HML + ε_i

Factor	Symbol	Economic Interpretation	Annual Premium (historical)
Market	R_m - R_f	Compensation for bearing systematic risk	~5-7% above risk-free
Size	SMB	Small caps outperform large caps on average	~2-3%
Value	HML	High book/market (cheap) outperform growth	~3-5%
Profitability	RMW	Profitable firms outperform unprofitable	~3-4%
Investment	CMA	Conservative investors outperform aggressive	~2-3%

Value at Risk (VaR) and Expected Shortfall

$VaR_\alpha = \mu + \sigma \times \Phi^{-1}(\alpha)$

The standard deviation lines on the Galton Board are exactly VaR lines. Expected Shortfall (ES) fixes the limitation of VaR by computing the average loss given that it exceeds VaR:

$ES_\alpha = \mu + \sigma \times \varphi(\Phi^{-1}(\alpha)) / (1-\alpha)$

LESSON 8 — Advanced Monte Carlo: Importance Sampling & HMC

Importance Sampling

$E_p[f(X)] = E_q[f(X)\,w(X)]$ where $w(x) = p(x)/q(x)$

By sampling from an alternative distribution q(x) instead of the target p(x) and correcting with importance weights, we can estimate rare-event probabilities efficiently.

Hamiltonian Monte Carlo (HMC)

HMC Leapfrog Integrator

Augment target $\pi(q)$ with momentum $p \sim N(0, M)$. Hamiltonian: $H(q,p) = -\log\pi(q) + (1/2)p^T M^{-1} p$ Leapfrog step (step size ε): $p_{t+\varepsilon/2} = p_t + (\varepsilon/2)\, \nabla_q \log\pi(q_t)$ [half-step in p] $q_{t+\varepsilon} = q_t + \varepsilon M^{-1} p_{t+\varepsilon/2}$ [full step in q] $p_{t+\varepsilon} = p_{t+\varepsilon/2} + (\varepsilon/2)\, \nabla_q \log\pi(q_{t+\varepsilon})$ [half-step in p] Accept with probability $\min(1, \exp(-H(q^*,p^*)+H(q,p)))$. Key: leapfrog preserves H exactly for small ε; acceptance rate → 1 as ε → 0.

Algorithm	Proposal Type	Dimension Scaling	Modern Use
Random Walk MH	Random normal jump	$O(d^{-1})$ step size	Simple 1D problems
HMC	Hamiltonian dynamics	$O(d^{-1/4})$ step size	High-dim Bayesian models
NUTS	No-U-Turn Sampler (adaptive HMC)	$O(d^{-1/4})$	Stan, PyMC, NumPyro

MCNP®	Analog particle physics	N/A (physics-driven)	Nuclear transport at LANL

The NUTS Sampler

The No-U-Turn Sampler (Hoffman & Gelman, 2014) automatically tunes HMC step size and trajectory length. It is the engine behind Stan and PyMC — the most widely used probabilistic programming languages in science and industry. When a biologist estimates phylogenetic trees, when an epidemiologist models disease spread, or when a central bank calibrates economic models, they are almost certainly using NUTS — the direct algorithmic descendant of the 1953 Metropolis paper at Los Alamos.

Capstone Project: The Galton Board in Your Domain

Students write a 3–5 page paper demonstrating that the mathematical content of this guide appears in a domain of their choice. The paper must include: one rigorous proof, one data analysis using real or simulated data, one connection to Monte Carlo simulation, and one connection to Los Alamos National Laboratory or a modern application.

Option	Domain	Key Connection
A	Genetics & Evolution	Hardy-Weinberg equilibrium, genetic drift, BEAST MCMC for phylogenetics
B	Climate Science	Normal distribution of temperature anomalies, GCM Monte Carlo uncertainty
C	Nuclear Engineering	Boltzmann transport equation, MCNP® neutron random walks, criticality
D	Quantitative Finance	Fama-French five-factor model, VaR, fat tails, leptokurtosis
E	Machine Learning	Mini-batch gradient CLT, Langevin dynamics, Bayesian neural networks

Grade 12 Assessment Bank

Proof Problems

- Prove that for any probability space (Ω, F, P), if $A \subseteq B$ then $P(A) \le P(B)$. Use only Kolmogorov's axioms.
- Use the Gaussian integral $I = \sqrt{\pi}$ to show that the normal PDF integrates to 1 over $\mathbb{R}$.
- Prove Vandermonde's identity using both a combinatorial argument and the generating function proof.
- Derive the MGF of the Binomial distribution. Use it to compute $E[X]$ and $E[X^2]$ by differentiation.
- Prove that detailed balance implies stationarity for a finite Markov chain.
- Show the Beta(α,β) distribution is the conjugate prior for the Binomial likelihood. Derive the posterior in closed form.

Computation and Analysis

- Using Stirling's approximation, estimate $C(30,15)/2^{30}$. Compare to the exact value and compute the relative error.
- Compute VaR at the 1%, 5%, and 10% levels for $\mu = 8\%$, $\sigma = 15\%$. Express in dollar terms for a $1M portfolio.
- A flat prior Beta(1,1) is updated by observing the golden bead land in bin 7 exactly 4 times in 25 flips. State the posterior, compute its mean and variance, and construct a 95% credible interval.

Grand Challenge Problem

Prove the Central Limit Theorem for the Galton Board in full generality: let $X_1,\ldots,X_n$ be i.i.d. Bernoulli(1/2), and let Z_n = $(X_1+\ldots+X_n - n/2)/\sqrt{(n/4)}$. Show that the characteristic function of Z_n converges pointwise to e^{-t²/2}, and invoke the Lévy continuity theorem to conclude Z_n → N(0,1) in distribution. You may state the Lévy theorem without proof.

Appendix: Standard Normal CDF Φ(z)

Values of $\Phi(z) = P(Z \le z)$ for $Z \sim N(0,1)$. For negative z: $\Phi(-z) = 1 - \Phi(z)$.

z	Φ(z)	z	Φ(z)	z	Φ(z)
0.00	0.5000	1.00	0.8413	2.00	0.9772
0.50	0.6915	1.28	0.8997	2.33	0.9901
0.67	0.7486	1.50	0.9332	2.58	0.9951
0.84	0.7995	1.64	0.9495	3.00	0.9987
1.04	0.8508	1.96	0.9750	4.00	0.99997

Key values: $\Phi(1.96) = 0.9750$ (95% CI); $\Phi(2.576) = 0.9950$ (99% CI).

For the Galton Board SD lines: $P(\mu-k\sigma \le X \le \mu+k\sigma) = 2\Phi(k) - 1$ = 68.27% (k=1), 95.45% (k=2), 99.73% (k=3), 99.994% (k=4).