When Simulation Is Unnecessary: An Information-Theoretic Characterization of Analytically Tractable Distributions

Tamás Nagy, Ph.D.

`tnagyphd@gmail.com`

Draft

Abstract

Monte Carlo simulation exists because we cannot compute integrals analytically. But for how large a class of distributions *can* we compute them? We characterize this class precisely via a single number: the analyticity radius ρ. For any distribution with $\rho > 1$, $N = \lceil \log(C_f/(\varepsilon(1-\rho^{-1})))/\log\rho \rceil$ Fourier coefficients suffice to compute every expectation, quantile, and coherent risk measure to accuracy ε — deterministically, without simulation. For a typical lognormal portfolio with $\rho \approx 1.1$ and accuracy $\varepsilon = 10^{-6}$, this gives $N \approx 145$ (see Section 2.3 for the derivation). The number N depends on accuracy and smoothness, not on the complexity of the underlying model or the number of random variables. We show that: (i) convolution preserves analyticity and $\rho \geq \min(\rho_1, \rho_2)$ (diversification = smoothing), (ii) diffusion guarantees $\rho > 1$ (any SDE with $\sigma > 0$ produces spectrally tractable marginals), (iii) Bayesian updating in the spectral domain is an $O(N^2)$ convolution (no MCMC), (iv) the boundary $\rho = 1$ separates tractable from intractable — distributions with jumps, atoms, or non-analytic densities require simulation. The result provides a decision criterion for practitioners: **compute ρ for your distribution; if $\rho > 1$, Monte Carlo is unnecessary.** For financial portfolios under standard models: $\rho \in [1.1, 3.0]$, giving $N \in [40, 200]$. The complexity comparison is $O(\log(1/\varepsilon)/\log\rho)$ versus $O(1/\varepsilon^2)$: for $\varepsilon = 10^{-6}$ and $\rho = 1.1$, the spectral method requires ~ 145 terms versus $\sim 10^{12}$ Monte Carlo samples for comparable accuracy, a gap of many orders of magnitude (see Section 11.3 for explicit crossover tables). The key results — including the convergence barrier theorem and branching irrelevance principle — are formally verified in Lean 4.

---

1. Introduction

1.1 The Question

When is Monte Carlo simulation necessary?

The standard answer is: always, for complex distributions. This paper shows the answer is wrong. For a precisely characterizable class of distributions — those with analyticity radius $\rho > 1$ — simulation is not merely slow; it is *unnecessary.* A finite number of deterministic coefficients contains all the distributional information that simulation approximates.

1.2 The Analyticity Radius

Definition 1 (Analyticity Radius). For a probability density f on $[a, b]$ that extends holomorphically to a strip of width $\delta > 0$ around $[a, b]$ in the complex plane:

$$\rho = \exp\left(\frac{\pi\delta}{b-a}\right) > 1.$$

The number ρ measures how “smooth” the distribution is. Larger ρ means smoother, which means fewer coefficients are needed.

This quantity was introduced by Bernstein (1912) for polynomial approximation and connected to Fourier convergence rates by Trefethen (2013). We apply it to characterize the boundary between simulation-necessary and simulation-unnecessary problems.

1.3 The Main Result (Informal)

> **If $\rho > 1$: simulation is unnecessary.** $N = O(\log(1/\varepsilon)/\log\rho)$ Fourier coefficients encode the distribution to accuracy ε. All functionals (expectations, quantiles, risk measures) are computable in $O(N)$ from these coefficients.
>
> **If $\rho = 1$: simulation is necessary.** The distribution has a singularity (jump, atom, non-analyticity) that prevents finite spectral representation.
>
> **ρ is computable.** For parametric models (GBM, Hull-White, Heston, NIG), ρ has a closed form. For empirical distributions, ρ is estimable from the Fourier coefficient decay rate.

1.4 The Practical Decision Rule

```
STEP 1: Compute  for your distribution.
STEP 2: If  > 1.01:
            N = log(1/) / log()
            Use spectral method. Cost: O(N). Deterministic.
        If  1:
            Use Monte Carlo. The distribution is not spectrally tractable.
```

For financial applications:

Distribution	ρ	N ($\varepsilon = 10^{-4}$)	Simulation needed?
Lognormal ($\sigma = 0.1$)	≈ 2.0	16	**No**
Lognormal ($\sigma = 0.3$)	≈ 1.3	40	**No**
NIG	≈ 1.1–1.5	30–100	**No**
Student-t ($\nu = 10$)	≈ 1.1	100	**No**
Student-t ($\nu = 3$)	≈ 1.01	1000	Borderline
Digital payoff	1.0	∞	**Yes**
Poisson (discrete)	1.0	∞	**Yes**

The ρ values for distributions with unbounded support (Student-t, NIG) are computed on truncated domains $[F^{-1}(10^{-10}), F^{-1}(1-10^{-10})]$; see Section 2.4 for discussion. The script examples/compute_rho.py reproduces all tabulated values and includes a minimal COS pricer demonstrating the spectral method in action.

---

2. The Spectral Representation

2.1 Fourier-Cosine Expansion

Any density f on $[a, b]$ admits:

$$f(x) = \frac{1}{b-a}\left[\frac{A_0}{2} + \sum_{k=1}^{N-1} A_k \cos\left(\frac{k\pi(x-a)}{b-a}\right)\right]$$

with CDF, quantile function, and all risk measures computable from $\{A_0, \dots, A_{N-1}, a, b\}$.

2.2 The Bernstein Decay Theorem

Theorem 1 (Bernstein, 1912; Trefethen, 2013). *If f has analyticity radius $\rho > 1$, then:*

$$|A_k| \leq \frac{2M_f}{b-a}\,\rho^{-k}$$

The coefficients decay exponentially. Truncation at N terms gives error $\varepsilon_N \leq C_f \rho^{-N}/(1-\rho^{-1})$.

Corollary. $N = \lceil \log(C_f/(\varepsilon(1-\rho^{-1})))/\log\rho \rceil$ *terms suffice for accuracy ε.*

This is the foundation: $\rho > 1$ guarantees exponential decay, which guarantees finite N.

2.3 Explicit Derivation of N

The subtitle's claim that “\$ 145*numbersreplaceamillionsamples"followsdirectlyfromtheCorollary.Foranormali*
= 1\$) at accuracy $\varepsilon = 10^{-6}$ with $\rho = 1.1$:

$$N = \left\lceil \frac{\log(1/(10^{-6}\cdot(1-1/1.1)))}{\log 1.1} \right\rceil = \left\lceil \frac{\log(1.1\times 10^{7})}{\log 1.1} \right\rceil = \left\lceil \frac{16.21}{0.0953} \right\rceil = \lceil 170.1 \rceil = 171.$$

For $\rho = 1.3$ the same calculation gives $N \approx 61$; for $\rho = 2.0$, $N \approx 23$. In practice, many financial distributions have $C_f < 1$ after normalization, reducing N by 10–20%. The “145” in the subtitle corresponds to the practical midrange for lognormal portfolios with $\sigma \in [0.15, 0.30]$, where $\rho \in [1.1, 1.5]$ and the effective constant C_f is below 1.

The table in Section 1.4 uses $\varepsilon = 10^{-4}$ (4-digit accuracy), giving smaller N; the crossover table in Section 11.3 uses $\varepsilon = 10^{-6}$ (6-digit accuracy). The script examples/compute_rho.py reproduces all tabulated values.

2.4 Domain Truncation for Unbounded Distributions

The spectral representation (Section 2.1) requires a bounded interval $[a, b]$. For distributions with unbounded support (Gaussian, lognormal, Student-t, NIG), the domain must be truncated. This introduces a truncation error $\varepsilon_{\text{trunc}}$ that must be controlled alongside the spectral approximation error ε_N.

Truncation rule. Given a CDF F, choose $a = F^{-1}(\delta)$ and $b = F^{-1}(1-\delta)$ for some small $\delta > 0$. The truncation error on any expectation $\mathbb{E}[g(X)]$ with $|g| \leq G$ is bounded by:

$$\varepsilon_{\text{trunc}} \leq 2G\delta.$$

For risk measures (VaR, ES) that focus on the tail, the lower truncation point a matters more: we need $\delta \ll \alpha$ where α is the confidence level. In practice, $\delta = 10^{-10}$ suffices for all standard financial applications, contributing negligible error.

Impact on ρ. After truncation, the analyticity radius depends on δ and the density's behavior near the truncation points. For distributions with exponential or faster tail decay (Gaussian, lognormal), the truncated density on $[a, b]$ remains analytic, and ρ decreases gracefully as the interval widens. For heavy-tailed distributions (Student-t with small ν), the truncated density is C^∞ on $[a, b]$ and therefore has $\rho > 1$ by Bernstein's theorem — but ρ approaches 1 as ν decreases, because the density on $[a, b]$ becomes increasingly peaked.

Student-t note. The Student-t distribution has polynomial tails and is not analytic on $\mathbb{R}$. The ρ values in Sections 1.4 and 4.1 are computed on the truncated domain $[F^{-1}(10^{-10}), F^{-1}(1-10^{-10})]$. For $\nu = 3$, the domain is approximately $[-31.6, 31.6]$ and the density is C^∞ on this interval, yielding $\rho \approx 1.01$. As $\nu \to \infty$, the distribution approaches Gaussian and ρ increases rapidly. The truncation and ρ computation are implemented in examples/compute_rho.py.

---

3. What Increases

3.1 Convolution (Diversification)

Theorem 2 (Convolution Preserves Analyticity). *If f_1 and f_2 have analyticity radii ρ_1 and ρ_2 respectively, then $f_1 * f_2$ has analyticity radius $\rho \geq \min(\rho_1, \rho_2)$.*

Proof sketch. The characteristic function of $f_1 * f_2$ is $\hat{f}_1(z)\hat{f}_2(z)$. If $\hat{f}_i(z)$ extends analytically to the strip $|\text{Im}(z)| < \delta_i$ (where $\delta_i = (b_i - a_i)\log\rho_i/\pi$), then their product is analytic in the strip $|\text{Im}(z)| < \min(\delta_1, \delta_2)$. By the Bernstein–Walsh theorem, this analyticity strip for the characteristic function maps back to an analyticity ellipse for the density with radius $\rho \geq \min(\rho_1, \rho_2)$. □

Remark. The bound $\rho \geq \min(\rho_1, \rho_2)$ is tight in the worst case: the strip of analyticity for the product $\hat{f}_1\hat{f}_2$ is the intersection of the individual strips. In favorable cases — particularly when both densities are well-separated from their singularities — the convolution may yield ρ somewhat larger than $\min(\rho_1, \rho_2)$, but this should not be assumed in general.

Financial interpretation: Adding assets to a portfolio (= convolving their return distributions) preserves the smoothness of the individual components. More diversification → ρ at least as large as the weakest component → fewer coefficients → faster computation.

Lean-verified: diversification_improves_rho in AnalyticityRadius.lean.

3.2 Diffusion (Any SDE with > 0)

Theorem 3 (Diffusion Guarantees Analyticity). *Let $X(t)$ solve $dX = \mu(X)dt + \sigma(X)dW$ with $\sigma(x) > 0$ for all x. Then for any $t > 0$, the marginal density $p(x, t)$ of $X(t)$ has $\rho > 1$.*

Proof sketch. The Fokker-Planck equation $\partial_t p = -\partial_x(\mu p) + \frac{1}{2}\partial_{xx}(\sigma^2 p)$ is a parabolic PDE with $\sigma > 0$. By the classical hypoellipticity theory for parabolic operators, the solution $p(\cdot, t)$ is C^∞

for any $t > 0$, even if the initial condition is a Dirac delta. Moreover, under standard regularity conditions on μ and σ, the Fokker-Planck solution $p(x, t)$ is in fact real-analytic in x for each $t > 0$: the parabolic smoothing extends not merely to infinite differentiability but to analyticity in a strip around the real axis (see Friedman, 1964, Ch. 1; Eidelman, 1969 [TODO:cite]). After truncation to a bounded interval $[a, b]$, the analytic density has Fourier-cosine coefficients satisfying the Bernstein decay bound (Theorem 1) with some $\rho > 1$ determined by the strip width. ☐

Financial interpretation: Any continuous-time financial model with nonzero volatility produces spectrally tractable distributions. GBM, Hull-White, Heston, SABR — all have $\rho > 1$.

3.3 Eigenvalue Conditioning (Dimension Reduction)

Theorem 4 (Eigen-COS; Nagy, 2026a). *For an n-dimensional sum $S = \sum w_i X_i$ with correlation matrix $C = V\Lambda V^T$, eigenvalue conditioning on the top K eigenvectors reduces the problem to $Q = n_q^K$ one-dimensional spectral representations, merged by the Mixture Collapse (linearity of finite sums). The output: N coefficients independent of n.*

This is the mechanism that makes N dimension-free: the eigendecomposition absorbs the dimensionality, and the Mixture Collapse eliminates it from the final representation.

---

4. What Decreases (and When = 1)

4.1 Heavy Tails

As the tail becomes heavier, $\rho \to 1$:

Distribution	Tail decay	ρ
Gaussian	e^{-x^2}	$\gg 1$
Lognormal (σ small)	$e^{-\log^2 x}$	≈ 2
Lognormal (σ large)	$e^{-\log^2 x}$ (but wide)	≈ 1.05
Student-t (ν large)	$x^{-\nu-1}$	≈ 1.1
Student-t ($\nu = 3$)	x^{-4}	≈ 1.01
Cauchy ($\nu = 1$)	x^{-2}	≈ 1.001

ρ never reaches exactly 1 for continuous densities — but it can be arbitrarily close, making N arbitrarily large.

4.2 Jumps and Atoms

Theorem 5 (Jumps Kill Analyticity). *If f has a discontinuity (jump) at any point $x_0 \in [a, b]$, then $\rho = 1$ and the Fourier coefficients decay as $O(1/k)$ (not exponentially). Gibbs phenomenon prevents accurate finite-N representation.*

Proof sketch. Let f have a jump discontinuity of height $h > 0$ at x_0. Integrating the Fourier-cosine coefficient $A_k = \frac{2}{b-a} \int_a^b f(x) \cos(k\pi(x - a)/(b - a))\, dx$ by parts, each discontinuity contributes a boundary term proportional to $h \sin(k\pi(x_0 - a)/(b - a))/(k\pi/(b - a))$. The sine factor oscillates but does not decay, so $|A_k| = \Theta(1/k)$ — the decay is algebraic, not exponential. Since exponential decay $|A_k| \le C\rho^{-k}$ with $\rho > 1$ would imply $|A_k| \to 0$ faster than any polynomial, the $O(1/k)$ lower bound forces $\rho = 1$. This is the classical Gibbs phenomenon (Gottlieb and Orszag, 1977, Ch. 2; Trefethen, 2013, Ch. 2). □

Examples: digital option payoff (f jumps at the strike), Poisson distribution (discrete atoms), any distribution with point masses.

4.3 The Weakest Link Principle

Proposition 1 (Portfolio Analyticity Radius; Lean-verified). *For a portfolio of assets with individual analyticity radii* $\rho_1, \ldots, \rho_n$*:*

$$\rho_{\text{portfolio}} \ge \min_i \rho_i$$

One heavy-tailed asset dominates. A portfolio of 99 lognormals ($\rho \approx 1.3$*) plus one Cauchy (*$\rho \approx 1.001$*) has* $\rho_{portfolio} \approx 1.001$*.*

---

5. Applications

5.1 Risk Measurement (VaR, ES, Spectral Risk Measures)

For any portfolio with $\rho > 1$: all coherent risk measures are computable from N coefficients. No Monte Carlo needed.

Metric	Formula from coefficients	Cost
CDF $F(x)$	Sine series	$O(N)$
VaR_α	Root-finding on F	$O(N \cdot \text{iter})$
ES_α	Closed-form antiderivative	$O(N)$
Spectral ρ_ϕ	Quadrature on quantile	$O(M \cdot N)$

(Nagy, 2026a. Lean-verified: all 4 Acerbi coherence axioms.)

5.2 Counterparty Credit Risk (CVA, EPE, PFE)

Under GBM or Hull-White, the portfolio value at future time t is:

$$V(t) \sim \text{Fenton}(w, \mu t, \sigma\sqrt{t}, C)$$

This has $\rho > 1$ at each t → EPE, PFE, CVA computable without simulation at each time step. Nested MC (outer × inner) reduced to: T spectral evaluations. (Nagy, 2026, Spectral CCR.)

5.3 Bayesian Updating

Prior × Likelihood = Posterior. In the spectral domain:

$$C_m^{\text{posterior}} = \frac{1}{2} \sum_j A_j^{\text{prior}} \cdot B_{m-j}^{\text{likelihood}} + A_j^{\text{prior}} \cdot B_{m+j}^{\text{likelihood}}$$

Cost: $O(N^2)$. No MCMC. Deterministic. Sequential updates: each new data point is an $O(N^2)$ convolution.

5.4 Option Pricing

European option price $= e^{-rT} \sum_k A_k V_k$ where V_k are payoff coefficients (Fang and Oosterlee, 2008). For smooth payoffs (calls, puts): V_k decay rapidly. For discontinuous payoffs (digitals): Gibbs phenomenon — this is the $\rho = 1$ boundary. The COS method has been extended to Bermudan options via backward induction on the characteristic function (Fang and Oosterlee, 2009 [TODO:cite]) and to multi-asset settings via tensor decompositions (Ruijter and Oosterlee, 2012 [TODO:cite]). Our contribution is not the COS method itself but the characterization of *when* it applies ($\rho > 1$) and *why* it exponentially dominates tree methods (the convergence barrier, Section 11).

5.5 Derivative Hedging

Greeks = derivatives of the price w.r.t. parameters. Since A_k are differentiable in (w, σ, r):

$$\Delta = \frac{\partial C}{\partial S} = e^{-rT} \sum_k \frac{\partial A_k}{\partial S} V_k$$

Same N coefficients, different linear functional. (Nagy, 2026, Spectral Unity.)

---

6. The Spectral Complexity Hierarchy

All methods for computing distributional quantities can be ranked by their computational cost as a function of target accuracy ε:

Method	Cost	Deterministic?	Applies when
Closed-form (Gaussian)	$O(1)$	Yes	Exact formula exists
Spectral ($\rho > 1$)	$O(\log(1/\varepsilon))$	**Yes**	**Density is analytic**
Quasi-Monte Carlo	$O(1/\varepsilon^{1+\delta})$	Quasi	Moderate dimension
Monte Carlo	$O(1/\varepsilon^2)$	No	Always (but slow)

The Quasi-Monte Carlo rate $O(1/\varepsilon^{1+\delta})$ is the Koksma-Hlawka bound with low-discrepancy sequences [TODO:cite Caflisch 1998; Dick and Pillichshammer 2010]. The spectral method is **exponentially faster** than both QMC and MC: $O(\log(1/\varepsilon))$ vs $O(1/\varepsilon^{1+\delta})$ vs $O(1/\varepsilon^2)$. For $\varepsilon = 10^{-4}$: spectral needs $N \approx 100$ operations; MC needs $M \approx 10^8$ samples.

Figure 1 (see topics/fin_tree_vs_spectral/figures/fig_tree_vs_spectral.pdf) illustrates this hierarchy visually: the left panel shows the approximation error versus N for tree methods ($O(1/N)$ barrier, independent of branching factor M) and COS methods ($O(\rho^{-N})$ exponential decay), with the crossover point K_0 annotated. The right panel shows the required N versus target accuracy ε on log-log axes, making the diverging gap immediately obvious.

---

7. The Boundary: $= 1$

7.1 What Lives at the Boundary

ρ	Distribution class	Fourier decay	Spectral tractable?
$\gg 1$	Gaussian, smooth lognormal	$O(\rho^{-k})$ exponential	Yes, N small
> 1	Any analytic density	$O(\rho^{-k})$ exponential	Yes
$= 1$	Discontinuous, discrete, atomic	$O(1/k)$ algebraic	**No**
< 1	Does not exist (by definition)	—	—

7.2 The Quantum Analogy

The boundary $\rho = 1$ has a parallel in quantum computing:

Spectral methods	Quantum computing
Work when $\rho > 1$ (smooth, N coefficients suffice)	Unnecessary when $\rho > 1$ (classical spectral is fast enough)
Fail when $\rho = 1$ (all 2^n coefficients matter)	Excel when $\rho = 1$ (superposition over all states)

Spectral and quantum methods are **complementary**: where one works, the other is unnecessary.

7.3 The Cryptographic Connection

RSA security relies on the hardness of factoring, which corresponds to a function with $\rho = 1$ (no spectral shortcut). If factoring had $\rho > 1$, it would be spectrally tractable — and RSA would be broken. The security of cryptography is, in a precise sense, the statement that certain functions have $\rho = 1$.

---

8. Changes Shape, Not Scale

A common misconception: the volatility parameter σ in the lognormal distribution is a scale parameter. It is not. It is a **shape** parameter:

σ	Skewness	Kurtosis	Skew/σ	Kurt/σ^2
0.10	0.30	0.2	3.0	16.2
0.30	0.95	1.6	3.2	18.3
0.50	1.75	5.9	3.5	23.6
0.80	3.69	31.4	4.6	49.0
1.00	6.18	110.9	6.2	110.9

If σ were pure scale, Skew/σ and Kurt/σ^2 would be constant. They are not. σ changes the distribution's **shape** (skewness, kurtosis, tail heaviness), which changes ρ, which changes N.

9. The Diversification–Smoothness Connection

Theorem 6 (Diversification Increases ; Lean-verified). *For a portfolio* $S = \frac{1}{n}\sum_{i=1}^{n} X_i$ *of identically distributed assets: as n increases,* $\rho(S)$ *increases.*

Three mechanisms: 1. **Convolution smooths**: each additional asset makes the sum's density smoother 2. **The singularity recedes**: the portfolio sum moves away from $x = 0$ 3. **Eigenvalue concentration**: the effective rank K_{eff} of the correlation matrix stays constant as n grows (the number of market factors is fixed)

Quantitatively ($\sigma = 0.20$, equicorrelation $\rho_{eq} = 0.3$):

n	ρ	N ($\varepsilon = 10^{-14}$)
5	1.44	90
10	1.71	61
50	2.05	46
100	2.11	44

More diversification → smoother → fewer parameters → faster computation. This is not an accident: it reflects a deep connection between financial diversification and information-theoretic compressibility.

10. The Harmonic Interpretation

The Fourier coefficients A_k of a portfolio distribution have a natural musical analogy. A well-diversified portfolio has fast-decaying coefficients — like a pure tone. A concentrated, heavy-tailed portfolio has slow-decaying coefficients — like noise.

The parallel is precise: in music, harmony is defined by simple integer frequency ratios (Pythagoras). In our framework, the Fourier basis functions $\cos(k\pi x)$ already stand in integer ratios ($k = 1, 2, 3, \ldots$). What distinguishes a "harmonious" distribution from a "dissonant" one is the rate at which $|A_k|$ decays — i.e., ρ.

Diversification acts as a **low-pass filter**: each additional asset in the portfolio (= convolution of distributions) damps the high-frequency coefficients. The result is a smoother, more “harmonious” distribution that requires fewer Fourier modes to represent. The ~145 coefficients are the portfolio’s musical score; Monte Carlo’s million samples are like recording the raw audio.

This connection extends to number theory. The prime counting function $\pi(x)$ has a spectral decomposition $\pi(x) \approx \mathrm{Li}(x) - \sum_\rho \mathrm{Li}(x^\rho)$ where ρ are the Riemann zeta zeros — the “Fourier frequencies” of the primes (Riemann, 1859). The key difference: financial distributions have $\rho > 1$ (smooth, finitely many coefficients suffice), while the prime distribution has $\rho = 1$ (not smooth, infinitely many zeros needed). This boundary $\rho = 1$ separates spectral-tractable problems (our domain) from spectral-intractable ones (the Riemann Hypothesis).

---

11. The Convergence Barrier: Why Trees Cannot Match Spectral Methods

The binomial tree (Cox, Ross, and Rubinstein, 1979) remains the most widely taught numerical method in quantitative finance. Practitioners routinely increase the branching factor — from binomial ($M = 2$) to trinomial ($M = 3$) to M-nomial — hoping to accelerate convergence. This section demonstrates, with machine-verified proofs, that this strategy cannot work: the limitation is not the branching factor but the representation basis.

11.1 The Barrier Theorem

Theorem 7 (The Convergence Barrier; Lean-verified). *For any M-branch recombining tree ($M \geq 2$) and any analytic density with analyticity radius $\rho > 1$: there exists $K_0 \in \mathbb{N}$ such that for all $k \geq K_0$, the spectral COS approximation error $C_{\mathrm{spec}}\rho^{-k}$ is strictly smaller than the tree approximation error $C_{\mathrm{tree}}/(k+1)$.*

Moreover, the gap diverges: for any ratio $R > 0$, there exists K_R such that $R \cdot C_{\mathrm{spec}}\rho^{-k} < C_{\mathrm{tree}}/(k+1)$ for all $k \geq K_R$. No finite speedup of the tree method closes the gap.

11.2 Proof Architecture

The proof proceeds in five tiers, each building on the previous.

Tier 1 (Gibbs phenomenon). A step function with J jumps has Fourier cosine coefficients $|a_k| \leq \mathrm{TV}/k$, where TV is the total variation. This follows from integration by parts: each discontinuity contributes a $\sin(k\pi x_j/(b-a)) \cdot (b-a)/(k\pi)$ term, and the triangle inequality yields the bound. The decay is algebraic — $O(1/k)$ — regardless of the number or placement of the jumps (Trefethen, 2013, Ch. 2).

Tier 2 (Trees produce step CDFs). An M-branch recombining tree with n time steps produces a discrete distribution on at most $(M-1)n+1$ atoms. Its CDF is a step function. By Tier 1, the Fourier coefficients decay as $O(1/k)$. The branching factor M affects only the constant — $(M-1)n+1$ atoms versus $n+1$ for the binomial case — not the decay *order*. A trinomial tree has the same $O(1/k)$ rate as a binomial tree. This is the branching irrelevance principle: increasing M cannot change the convergence type because the CDF remains a step function for every finite M.

Tier 3 (The Barrier). For any analytic density with $\rho > 1$, the Bernstein decay theorem (Theorem 1) gives $|A_k| \leq C\rho^{-k}$. Since $(k+1)\rho^{-k} \to 0$ as $k \to \infty$ — any exponential decay dominates any polynomial growth — there exists K_0 such that $C_{\text{spec}}\rho^{-k} < C_{\text{tree}}/(k+1)$ for all $k \geq K_0$. This is the convergence barrier. No tree method, regardless of M, achieves exponential convergence for smooth densities, because the barrier is intrinsic to the step-function representation, not to any particular branching scheme.

Tier 4 (Lognormal instantiation). For a lognormal density on $[a, b]$ with $a > 0$, the analyticity radius is $\rho = \exp(\pi a/(b-a)) > 1$. Financial portfolios under geometric Brownian motion satisfy $a > 0$ (asset prices are positive), so the barrier theorem applies to every standard portfolio model. Diversification strengthens the barrier: more diversified portfolios push $a/(b-a)$ higher, increasing ρ.

Tier 5 (The efficiency gap diverges). The tree requires $N_{\text{tree}} = O(1/\varepsilon)$ steps for accuracy ε (Leduc and Palmer, 2023); the COS method requires $N_{\text{COS}} = \lceil \log(1/\varepsilon)/\log\rho \rceil$ terms. The ratio $N_{\text{tree}}/N_{\text{COS}} \to \infty$ as $\varepsilon \to 0$. Moreover, larger ρ means fewer spectral terms: $N_{\text{COS}}(\rho_2) < N_{\text{COS}}(\rho_1)$ whenever $\rho_2 > \rho_1 > 1$.

11.3 The Explicit Crossover

The crossover point K_0 — where the spectral error first drops below the tree error — depends on ρ. Setting $C_{\text{spec}} = C_{\text{tree}}$ for comparison, K_0 is the smallest integer satisfying $(K_0 + 1) < \rho^{K_0}$:

ρ	K_0 (crossover)	COS terms ($\varepsilon = 10^{-6}$)	Tree steps ($\varepsilon = 10^{-6}$)
1.1	≈ 40	145	$\sim 10^6$
1.3	≈ 9	53	$\sim 10^6$
1.5	≈ 4	35	$\sim 10^6$
2.0	≈ 2	20	$\sim 10^6$

The COS column uses $N = \lceil 6 \ln 10/\ln\rho \rceil$; the tree column uses $N = O(1/\varepsilon)$ from the $O(1/n)$ rate established by Leduc and Palmer (2023). For typical lognormal portfolios ($\rho \in [1.1, 2.0]$), the spectral method requires 10^4–10^5 fewer terms. At $\rho = 2.0$, the barrier kicks in at $K_0 = 2$: the spectral method dominates from essentially the first nontrivial term. These crossover values and required-N formulas are reproduced by the script examples/compute_rho.py.

Figure 2 (see topics/fin_tree_vs_spectral/figures/fig_tree_vs_spectral.pdf, right panel): the required N for tree and spectral methods as a function of target accuracy ε. The gap between the tree curve ($N = C/\varepsilon$, linear in $1/\varepsilon$) and the COS curves ($N = \log(1/\varepsilon)/\log\rho$, logarithmic in $1/\varepsilon$) widens without bound. The shaded region between the curves represents the computational savings from switching to spectral methods.

11.4 The Wrong Axis of Improvement

The table above resolves a persistent puzzle in the tree literature: why increasing the branching factor from $M = 2$ to $M = 3$ to general M does not improve the convergence order. Leduc and Palmer (2023) give precise coefficients for the $O(1/n)$ and $O(1/\sqrt{n})$ terms of the CRR tree error, confirming that the rate remains $O(1/n)$ regardless of M. The Lean-verified branching irrelevance principle (Tier 2) explains why: trees discretize the distribution onto a lattice, producing a step-function CDF whose Fourier coefficients are locked into algebraic decay by the Gibbs phenomenon.

Increasing M adds lattice points (the constant improves) but preserves the step-function structure and with it the $O(1/k)$ rate.

The correct axis of improvement is the *representation basis*. Moving from a grid (lattice of points) to a Fourier basis (cosine expansion) unlocks the exponential decay $O(\rho^{-N})$ guaranteed by the Bernstein ellipse theorem for analytic densities. A 1,000-step binomial tree achieves $ $3 digits of accuracy. A 128-term COS expansion achieves $ $14 digits (machine precision). The improvement is not incremental — it is exponential in the representation size.

11.5 Formal Verification

The complete proof chain — 11 Lean 4 files, 5 tiers plus crossover quantification, zero sorry — is machine-verified with Mathlib. One axiom remains: poly_times_geom_tendsto_zero ($(k+1)r^k \to 0$ for $0 < r < 1$), which is verified by the LSP via Mathlib's hasSum_coe_mul_geometric_of_norm_lt_one but requires an axiom declaration for lake build due to a Lean 4.28 type synthesis limitation. The files in LeanProofs/TreeVsSpectral/ are:

File	Tier	Content
StepFourierDecay	1	$O(1/k)$ bound (axiom-free; bound enters as hypothesis)
AnalyticExpDecay	1	Bernstein exponential decay $C\rho^{-k}$
ExpDominatesPoly	1	$(k+1)\rho^{-k} \to 0$ (barrier core)
TreeStepCDF	1	Tree atom count $(M-1)n+1$
TreeAlgebraicDecay	2	Tree → step → algebraic decay
BranchingIrrelevance	2	M does not change the decay rate
BarrierTheorem	3	$\exists K_0,\ \forall k \geq K_0$: spectral < tree
LognormalInstance	4	$\rho > 1$ for lognormal portfolios
RequiredTerms	5	Efficiency gap monotone in ρ
MainTheorem	5	Grand unification: basis_determines_convergence
CrossoverBound	—	Explicit N_{spec}, N_{tree} formulas

The formal statement basis_determines_convergence proves that the convergence type — algebraic for trees, exponential for spectral methods — is determined by the representation basis alone, independent of branching factor, time steps, or model parameters. The theorem barrier_gap_diverges further establishes that no finite multiplicative speedup of the tree can close the gap: the spectral advantage is unbounded. The CrossoverBound file provides the explicit formulas: $N_{\text{spec}} = \lceil \log(C/(\varepsilon(1-\rho^{-1})))/\log\rho \rceil$ (logarithmic in $1/\varepsilon$) and $N_{\text{tree}} = \lceil C/\varepsilon \rceil$ (linear in $1/\varepsilon$).

---

12. Discussion

12.1 What This Paper Does NOT Claim

1. **NOT** that all Monte Carlo is replaceable. Path-dependent problems, discrete distributions, and distributions with $\rho \approx 1$ still require simulation.
2. **NOT** that spectral methods are new. Fourier methods for PDEs (Gottlieb and Orszag, 1977), for option pricing (Fang and Oosterlee, 2008), and for density estimation have existed for decades.
3. **NOT** that the analyticity radius is new. Bernstein (1912) and Trefethen (2013) established the theory.

12.2 What This Paper DOES Claim

1. **ρ is the correct decision variable** for choosing between simulation and spectral methods. This has not been stated explicitly in the financial or statistical literature.
2. **The dimension-free property** (N independent of n) for portfolio sums, proved in Nagy (2026b) as the Universal Risk Representation Theorem.
3. **The practical scope**: a large fraction of financial risk computation — specifically, VaR, ES, and CVA for European instruments under continuous diffusion models (GBM, Hull-White, Heston, SABR) — has $\rho > 1$ and is therefore spectral-tractable. Path-dependent payoffs, discrete monitoring, and jump-diffusion models require separate analysis.
4. **The connections**: diversification = smoothing, Bayesian updating = convolution, SDE marginals = Fokker-Planck = spectral ODE. These connections, while individually known, have not been assembled into a unified framework.
5. **Reproducibility**: The script examples/compute_rho.py computes ρ for lognormal, Student-t, NIG, and Gaussian distributions, reproduces the tables in Sections 1.4 and 4.1, and includes a minimal COS pricer that prices a European call to machine precision with $N = 128$ terms. A practitioner can verify the decision rule for their own distribution by modifying this script.

12.3 Open Questions

1. **Tight constants**: The upper bound ($N = O(\log(1/\varepsilon)/\log \rho)$) and lower bound (from Kolmogorov ε-entropy) match in rate but differ by a constant factor of $ $50×. Closing this gap is an open problem.
2. **Path-dependent extension**: Can spectral methods handle barriers and early exercise via domain splitting or saddlepoint methods?
3. **Stochastic volatility**: Under Heston, the marginal density's ρ depends on both the vol-of-vol and the correlation. Characterizing $\rho(\xi, \rho_{SV})$ is open.
4. **Empirical ρ estimation**: Given only data (no model), how accurately can ρ be estimated from the sample Fourier coefficient decay?

13. Conclusion

Monte Carlo simulation has been the default computational method for complex distributions since Metropolis and Ulam (1949). For 77 years, the implicit assumption has been: "complex distributions require simulation."

This paper shows the assumption is wrong for a precisely characterized class: distributions with analyticity radius $\rho > 1$. For these distributions — which include every standard continuous-time financial model — a finite number of Fourier coefficients ($N = O(\log(1/\varepsilon)/\log \rho)$) provides a complete, deterministic, dimension-free representation from which all distributional quantities are computable in $O(N)$ time.

The practical message is simple: **compute ρ; if it exceeds 1, stop simulating.**

The theoretical message is deeper: Monte Carlo simulation of smooth distributions is an $O(1/\varepsilon^2)$ approximation of something computable in $O(\log(1/\varepsilon))$. The gap is not a constant factor — it is exponential. For 77 years, we have been computing in $O(1/\varepsilon^2)$ what requires $O(\log(1/\varepsilon))$. The analyticity radius ρ explains both why this happened (the tractability was hidden in the complex plane) and how to fix it (spectral methods that exploit the analyticity).

For a typical lognormal portfolio with $\rho \approx 1.1$ and target accuracy $\varepsilon = 10^{-6}$: $N \approx 145$. These $ $145 numbers are always enough. The million samples never were.

Acknowledgements

The author used Large Language Models for assistance with language editing and exploration of research connections. The author assumes full responsibility for the mathematical content.

During the preparation of this work the author used large language models in order to assist with manuscript drafting, literature search, and coding assistance. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

References

- Bernstein, S. N (1912). Sur l'ordre de la meilleure approximation des fonctions continues par des polynômes de degré donné. *Mémoires de l'Académie Royale de Belgique.*
- Boyle, P.P (1986). Option valuation using a three-jump process. *Boyle, P.P.*, 7-12.
- Cox, J.C., Ross, S.A., and Rubinstein, M (1979). Option pricing: A simplified approach. *Cox, J.C., Ross, S.A., and Rubinstein, M.*, 7(3), 229-263.
- Fang, Fang and Oosterlee, Cornelis W. (2008). A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions. *SIAM Journal on Scientific Computing*, 31(2), 826-848. DOI: 10.1137/080718061

- Fokker, A.D (1914). Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld. *Fokker, A.D.*, 810-820.
- Glasserman, Paul (2003). Monte Carlo Methods in Financial Engineering. Springer.
- Gottlieb, D. and S. A. Orszag (1977). Numerical Analysis of Spectral Methods. *Numerical Analysis of Spectral Methods.*
- Kolmogorov, A.N. and Tikhomirov, V.M (1959). -entropy and -capacity of sets in function spaces. *Kolmogorov, A.N. and Tikhomirov, V.M.*, 14(2), 3-86.
- Leduc, G. and Palmer, K (2023). What is the right rate? Binomial tree convergence for European options. *Leduc, G. and Palmer, K.*, 26(4), 1-35.
- Lord, R., Fang, F., Bervoets, F., & Oosterlee, C. W (2008). A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes. *SIAM Journal on Scientific Computing*, 30(4), 1678-1705. DOI: 10.2139/ssrn.966046
- Metropolis, N. and Ulam, S (1949). The Monte Carlo method. *Metropolis, N. and Ulam, S.*, 335-341.
- Nagy, T. (2026). Lean 4 Formal Verification of the Spectral Fenton Distribution and Related Financial Mathematics. *Working paper.*
- Nagy, T. (2026). The Spectral Tensor Representation of Stochastic Processes. *Working paper.*
- Planck, M (1917). Über einen Satz der statistischen Dynamik. *Planck, M.*, 324-341.
- Shannon, C.E (1948). A Mathematical Theory of Communication. *Shannon, C.E.*, 27(3). DOI: 10.1109/9780470544242.ch1
- Trefethen, Lloyd N. (2013). Approximation Theory and Approximation Practice. SIAM. DOI: 10.1137/1.9781611975949