The Duality of Bayesian and Frequentist Statistics

Why They Agree on Eigenvalues

A spectral account of common model complexity, uncertainty, and mode selection

Tamás Nagy, Ph.D.

tnagyphd@gmail.com

Draft Complete

Executive Summary (Non-Technical)

Bayesian and frequentist statistics are usually presented as rival worldviews. In practice, though, many modeling decisions do not turn on philosophy. They turn on a simpler question: **how much usable information is really present in each direction of the problem?**

This paper argues that once a learning problem is written in its spectral basis, both traditions are reading the **same mode-level evidence**. They use different language, different summaries, and different interpretations of uncertainty, but they are responding to the same resolved structure.

The paper's main claim is narrow and concrete: for model complexity, Bayesian selection, frequentist risk minimization, and MDL all point to the same effective cutoff, up to a small finite-sample boundary layer. The practical consequence is that many familiar disputes about priors, penalties, and validation are really disputes about a **small contested zone near the signal-noise boundary**, not about the entire model.

This paper is a **bridge paper**, not the foundational state-object paper. The underlying mode-level object is developed more directly in the companion theory manuscript The Spectral Information State. The role of the present paper is to show one strong consequence of that object: the apparent Bayesian-frequentist divide becomes much thinner once inference is viewed mode by mode.

The paper does **not** claim that Bayesian and frequentist statistics are identical in every philosophical or finite-sample sense. It claims that in spectral problems they share the same practical inferential substrate, and therefore often agree on the decision that matters most: how many modes deserve to survive.

Short Abstract

This paper argues that Bayesian and frequentist statistics become much closer once a learning problem is viewed mode by mode in the spectral basis. Both frameworks read the same mode-level evidence and therefore agree on the practically decisive quantity: the optimal model complexity $K^* = \Theta(\log(n/\sigma^2)/\log\rho)$, up to $O(1)$ corrections. The underlying shared object is the spectral information state $\psi_k = (A_k, \sigma_k^2)$, developed more directly in the companion theory paper The Spectral Information State; here it serves as the common substrate that makes the bridge result explicit. The main message is the duality itself: Bayesian posteriors, frequentist intervals,

p-values, and MDL penalties are different readouts of the same resolved spectrum. This removes much of the practical need for framework choice in model selection, supports overfitting audits from sample size and spectral decay alone, and yields a non-subjective Bayesian prior from a Lean 4-verified coefficient decay theorem.

Abstract

We show that Bayesian and frequentist statistics admit a common spectral formulation: after eigendecomposition, both frameworks act on the same mode-level evidence and therefore agree on the central inferential question of model complexity. The shared object is the pair $\psi_k = (A_k, \sigma_k^2)$ of estimated coefficient and residual uncertainty per eigenmode. That object is developed more directly in the companion theory paper The Spectral Information State; in the present paper it serves as the common substrate from which Bayesian posteriors, frequentist confidence intervals, p-values, and minimum description lengths are all read out in closed form. The eigendecomposition is performed once; each inferential framework is then a one-line readout of the same spectral state.

From this construct we prove that Bayesian model selection (with a prior derived from the Lean 4-verified Universal Risk Representation Theorem), frequentist minimax estimation, and MDL all yield the same optimal model complexity:

$$K_B^* = K_F^* = K_{MDL}^* = \Theta\left(\frac{\log(n/\sigma^2)}{\log \rho}\right) + O(1)$$

where ρ is the eigenvalue decay rate, σ^2 the noise variance, and n the sample size. The Bayesian-frequentist disagreement is $O(1/\sqrt{n})$ per mode and confined to a "contested zone" of $O(1/\log \rho) \approx 3$ modes near K^*. Below K^*, both views agree the mode is signal. Above K^*, both agree it is noise. The paper's claim is therefore not merely that a useful statistic exists, but that the apparent philosophical divide collapses to a thin finite-sample boundary layer once the problem is seen in the spectral basis.

The Bayesian prior is not subjective: $A_k \sim \mathcal{N}(0, C^2\rho^{-2k})$ is the unique prior consistent with the machine-verified coefficient decay theorem $|A_k| \leq C\rho^{-k}$. The result requires only finite variance and is robust to non-Gaussian noise: the contested zone widens for heavy tails but the duality persists (verified empirically for Student-t, Laplace, Pareto, and contaminated mixtures).

Five practical consequences follow: (1) analytic model selection without cross-validation, recovering 40% of discarded data; (2) resolution of BIC/AIC/CV disputes — all approximate the same K^*; (3) an overfitting audit from n, p, and σ^2 alone, requiring no data access; (4) per-mode sample size calculation for experimental design; (5) Kelly-optimal position sizing in eigenspace for quantitative trading, where noise modes are automatically zeroed by the spectral shrinkage filter. All theoretical results build on 10 Lean 4-verified theorems.

---

1. Introduction

The optimal model complexity for a regression with $n = 500$ observations and eigenvalue decay rate $\rho = 1.8$ is $K^* = 11$. This number is the same whether computed by minimizing frequentist prediction risk, maximizing Bayesian marginal likelihood, or minimizing description length. It does not depend on the choice of prior, the significance level, or the cross-validation scheme. It depends on two quantities: how fast the eigenvalues decay (ρ) and how much data you have relative to noise (n/σ^2).

This is the main result of this paper. The century-old debate between Bayesian and frequentist statistics — which Efron (1986) framed as "why isn't everyone a Bayesian?" and Berger (2006) answered with "the case for objective Bayesian analysis" — concerns the interpretation of probability. Our claim is that, in spectral problems, the two frameworks are better understood as dual projections of the same mode-level evidence. For the practical question that matters most — **how many parameters should the model use?** — the debate is therefore largely moot. Both sides agree on eigenvalues.

1.1 The Spectral Setting

Any learning problem with design matrix $X \in \mathbb{R}^{n \times p}$ has an eigenvalue decomposition:

$$X = U\Sigma V^T, \qquad \lambda_k = \sigma_k^2/n$$

The eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ measure how much variance each direction captures. The response y is projected onto these directions:

$$\hat{A}_k = u_k^T y$$

with estimation uncertainty:

$$\sigma_k^2 = \frac{\sigma_{\text{noise}}^2}{n \cdot \lambda_k}$$

The pair $\psi_k = (\hat{A}_k, \sigma_k^2)$ is the **spectral information state** of mode k. We use this as the technical name for the shared mode-level object. The broader state-object thesis belongs to the companion theory paper The Spectral Information State; here we use it only insofar as it makes the duality claim explicit.

1.2 The Main Result

Theorem (Spectral Duality). Three independent criteria for model complexity yield the same answer:

(a) The **frequentist minimax** complexity minimizes worst-case risk:

$$K_F^* = \arg\min_K \sup_\theta \mathbb{E}_\theta[\|f - \hat{f}_K\|^2] = \Theta\left(\frac{\log(n/\sigma^2)}{\log \rho}\right)$$

(b) The **Bayesian MAP** complexity with URRT-derived prior $A_k \sim \mathcal{N}(0, C^2\rho^{-2k})$:

$$K_B^* = \arg\max_K \int p(y|A_{1:K})\, p(A_{1:K})\, dA = K_F^* + O(1/\sqrt{n})$$

(c) The **MDL** complexity minimizes total description length:

$$K_{MDL}^* = \arg\min_K \left[-\log p(y|\hat{A}_{1:K}) + \frac{K}{2}\log n\right] = K_F^* + O(1)$$

(d) **All three are the same number** up to $O(1)$ corrections. The spectral decay rate ρ and the information budget n/σ^2 determine the resolution limit, independent of philosophical framework.

(e) The disagreement between Bayesian and frequentist on mode k is $O(1/\sqrt{n})$ and vanishes as data grows. It is concentrated in modes where $|k - K^*| = O(1)$.

1.3 Related Work

Bernstein-von Mises theorem (Le Cam, 1953): the posterior converges to the frequentist sampling distribution asymptotically. Our contribution: the explicit K^* formula and the mode-by-mode characterization of where they agree for finite n.

Information geometry (Amari, 1985; Amari and Nagaoka, 2000): the statistical manifold has dual affine connections (e-connection, m-connection) corresponding to frequentist and Bayesian structures. The Fisher metric is the unique invariant metric (Cencov, 1982). Our contribution: the eigenvalue spectrum as the natural coordinate system on this manifold, with K^* as the finite-resolution cutoff.

BIC-Bayes equivalence (Schwarz, 1978) and **MDL** (Rissanen, 1978): BIC approximates the log marginal likelihood; MDL formalizes model selection as data compression. Our contribution: extending this beyond BIC to the full minimax-Bayes-MDL triple, with the explicit role of ρ.

Minimax theory (Pinsker, 1980; Donoho et al., 1990): minimax risk for nonparametric estimation depends on the smoothness class. Our contribution: connecting the smoothness parameter directly to the Lean-verified URRT decay rate ρ.

---

2. The Common Spectral Substrate

2.1 Definition

For each eigenmode k of the learning problem, the common mode-level object is:

$$\psi_k = \left(\hat{A}_k,\ \sigma_k^2(n)\right) = \left(\hat{A}_k,\ \frac{\sigma^2_{\text{noise}}}{n \cdot \lambda_k}\right)$$

We call this pair the **spectral information state**. It is the complete sufficient statistic for mode k. It is not Bayesian. It is not frequentist. It is the information content of the data about mode k, and it is the object on which the two frameworks meet.

2.2 Three Views of the Same Eigenvalue

For each eigenmode k of the Fisher information, the three frameworks see the same quantity differently:

	Frequentist	Bayesian	Spectral
Large λ_k	Small confidence interval	Posterior concentrates (data dominates prior)	Signal mode — keep
Small λ_k	Wide confidence interval	Posterior $\approx$ prior (data says nothing)	Noise mode — drop
Boundary	Sample size n vs info λ_k	Prior precision vs likelihood precision	K^*: the spectral threshold
Computes	Coverage probability	Posterior probability	Information content per mode

The frequentist asks: "is my estimate reliable?" The Bayesian asks: "what do I believe now?" The spectral view asks: "how much information is in this mode?" All three give the same answer about which modes matter.

The connection between the Bayesian information gain and the eigenvalue spectrum is explicit. The KL divergence from prior to posterior:

$$D_{KL}(\text{posterior}\|\text{prior}) \approx \frac{K}{2}\log\frac{n}{2\pi e} + \frac{1}{2}\sum_{k=1}^{K}\log\lambda_k$$

The eigenvalues λ_k appear directly in the Bayesian information gain formula — the same eigenvalues that determine the frequentist Cramér-Rao bound $\text{Var}(\hat{\theta}_k) \geq (n\lambda_k)^{-1}$.

2.3 Two Projections of the Same Object

Bayesian projection. Treat A_k as random with prior $\pi(A_k)$. The posterior given data:

$$A_k|\text{data} \sim \mathcal{N}\left(\frac{\lambda_k}{\lambda_k+\alpha}\hat{A}_k,\ \frac{\sigma_k^2\cdot\lambda_k}{\lambda_k+\alpha}\right)$$

where $\alpha = \sigma^2_{\text{noise}}/(C^2\rho^{-2k})$ is the prior-to-noise ratio for the URRT prior.

Frequentist projection. Treat A_k as fixed. The sampling distribution of the estimator:

$$\hat{A}_k|A_k \sim \mathcal{N}(A_k,\ \sigma_k^2)$$

In the large-n limit: $\alpha \to 0$, the posterior mean $\to \hat{A}_k$, and the posterior variance $\to \sigma_k^2$. The two projections become identical.

For finite n: the Bayesian shrinks toward zero (the prior), the frequentist does not. The difference is:

$$|\text{Bayesian mean} - \text{Frequentist estimate}| = \frac{\alpha}{\lambda_k+\alpha}|\hat{A}_k| = O\left(\frac{1}{n\lambda_k}\right)$$

This difference is large for modes where λ_k is small (noise modes) and negligible for modes where λ_k is large (signal modes). The two views agree on signal, disagree on noise, and the boundary is K^*.

2.4 Learning as Progressive Spectral Collapse

Learning is not a single convergence — it is a **cascade of mode resolutions**, ordered by eigenvalue.

As n increases, the uncertainty $\sigma_k^2(n) = \sigma^2/(n\lambda_k)$ shrinks for all modes, but unequally:

- **Mode 1** (largest λ_1): resolves first, with $n_1 = \sigma^2/(\lambda_1 \cdot A_1^2/z^2)$ samples
- **Mode 2** (next largest): resolves later, needs more data
- **Mode** K^*: just barely resolves at the current n
- **Mode** $K^* + 1$: still unresolved — $\sigma_{K^*+1}^2 > A_{K^*+1}^2$

Each data point reduces uncertainty across ALL modes simultaneously, but the high-eigenvalue modes clear first. The process is:

$$
\begin{aligned}
n = 10: &\quad \text{mode 1 resolved} \qquad K^* = 1 \\
n = 50: &\quad \text{modes 1-3 resolved} \qquad K^* = 3 \\
n = 200: &\quad \text{modes 1-8 resolved} \qquad K^* = 8
\end{aligned}
$$

This is the spectral analog of progressive image rendering: you see the coarse structure first, detail fills in later. Crucially, the coarse structure is **exactly correct** — not an approximation that gets revised, but the true projection that gets extended (Eckart-Young). The blurry picture is not wrong. It is just not finished.

2.5 The Complete Inferential Menu

From the single pair $\psi_k = (A_k, \sigma_k^2)$ per mode, the following quantities are computable in closed form — no resampling, no simulation, no iterative fitting:

Hypothesis testing:

Output	Formula	What it answers
t-statistic	$t_k = A_k/\sqrt{\sigma_k^2}$	How many standard errors from zero?
p-value	$p_k = 2(1 - \Phi(\lvert t_k\rvert))$	Would random noise produce this?
Adjusted p (BH)	$p_{(k)} \leq (k/K) \cdot q$	Which modes survive FDR correction?
Effect size (Cohen's d)	$d_k = \lvert A_k\rvert/\sqrt{\sigma_k^2}$	How large is the effect in standardized units?
Statistical power	$\Phi(\lvert A_k\rvert/\sqrt{\sigma_k^2} - z_\alpha)$	Probability of detecting this mode if real
Bayes factor	$BF_k = \sqrt{\frac{\sigma_k^2}{\sigma_k^2+\tau_k^2}} \exp\left(\frac{A_k^2\tau_k^2}{2\sigma_k^2(\sigma_k^2+\tau_k^2)}\right)$	How much does the data favor signal over noise?

Estimation:

Output	Formula	What it answers
Confidence interval (95%)	$A_k \pm 1.96\sqrt{\sigma_k^2}$	Frequentist: plausible range for the true value
Bayesian posterior	$\mathcal{N}(h_k A_k,\ h_k^2 \sigma_k^2)$	Bayesian: belief distribution after data
Shrinkage factor	$h_k = \lambda_k/(\lambda_k + \alpha)$	How much to trust this mode (0 = ignore, 1 = full)
Prediction interval	$\hat{y} \pm z\sqrt{\sigma_k^2 + \sigma_{\text{noise}}^2}$	Range for a new observation

Model assessment:

Output	Formula	What it answers
R^2 per mode	$R_k^2 = A_k^2 \lambda_k/\text{Var}(y)$	What fraction of variance does this mode explain?
Cumulative R^2	$\sum_{j=1}^{k} R_j^2$	Total variance explained by first k modes
Effective degrees of freedom	$d_{\text{eff}} = \sum_k h_k$	How many parameters is the model really using?
Noise variance	$\hat{\sigma}^2 = \|y - \hat{y}\|^2/(n - d_{\text{eff}})$	Estimated noise level
LOO-CV error (analytic)	$\text{CV} = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2/(1 - h_{ii})^2$	Cross-validation without splitting
C_p (Mallows)	$C_p = \text{SSE}/\hat{\sigma}^2 - n + 2d_{\text{eff}}$	Bias-corrected prediction error

Information-theoretic:

Output	Formula	What it answers
MDL code length	$\frac{1}{2}A_k^2/\sigma_k^2 + \frac{1}{2}\log n$	Cost of including this mode (in bits)
Information gain	$\frac{1}{2}\log(1 + A_k^2/\sigma_k^2)$	Bits learned about this mode from data
Effective sample size	$n_{\text{eff},k} = n\lambda_k/\sigma^2$	How many "effective observations" for this mode

Decision-making:

Output	Formula	What it answers
Kelly weight	$w_k \propto A_k/\sigma_k^2$	Optimal bet size for trading
Sample size needed	$n_k = z^2\sigma^2/(\lambda_k A_k^2)$	How many more observations to resolve this mode
Optimal K	$K^* = \Theta(\log(n/\sigma^2)/\log \rho)$	How many modes to use

All 22 quantities are computable from $(A_k, \sigma_k^2, \lambda_k, n)$ — the spectral information state plus sample size. No bootstrapping, no cross-validation, no MCMC, no grid search. One eigendecomposition. Twenty-two answers.

2.6 The Contested Zone

Define the Bayesian-frequentist discrepancy for mode k:

$$\Delta_k(n) = \left|\frac{\alpha_k}{\lambda_k + \alpha_k}\right| = \frac{1}{1 + n\lambda_k C^2 \rho^{-2k}/\sigma^2}$$

- $k \ll K^*$: $\Delta_k \approx 0$ (both agree — strong signal)
- $k \gg K^*$: $\Delta_k \approx 1$ (Bayesian shrinks to zero, frequentist has wide CI — different actions, same conclusion: ignore this mode)
- $k \approx K^*$: $\Delta_k \in (0.2, 0.8)$ — the **contested zone** where the philosophical choice matters

The width of the contested zone is $O(1/\log \rho)$ modes. For typical data ($\rho \approx 2$): the contested zone is about 3 modes wide. The debate between Bayesian and frequentist is a debate about **3 modes**.

---

3. The URRT Prior: A Non-Subjective Bayesian Prior

3.1 The Prior Problem

The standard objection to Bayesian inference: the prior is subjective. Different priors give different posteriors. The spectral framework provides a way out.

3.2 The URRT-Derived Prior

The Universal Risk Representation Theorem [Lean-verified: Universal/MainTheorem.lean] proves:

$$|A_k| \leq C \cdot \rho^{-k}$$

for any distribution with analytic density on the Bernstein ellipse of parameter ρ. This bound is tight: $\Theta(\log(1/\varepsilon)/\log \rho)$ coefficients are both necessary and sufficient [Lean-verified: Universal/EntropyLowerBound.lean].

The natural Gaussian prior consistent with this decay:

$$A_k \sim \mathcal{N}(0,\ C^2 \cdot \rho^{-2k})$$

This prior is **objective** in the following sense: - It is derived from a Lean-verified mathematical theorem, not from expert opinion - The parameters C and ρ are estimable from the eigenvalue spectrum of the data - It is the maximum entropy prior subject to the constraint $\mathbb{E}[A_k^2] = C^2\rho^{-2k}$ - Any prior with slower decay (less shrinkage on high modes) contradicts the URRT

3.3 The Prior Generates K^* Automatically

With this prior, the marginal likelihood:

$$\log p(y) = -\frac{1}{2}\sum_{k=1}^{K}\left[\log(\lambda_k + \alpha_k) + \frac{\hat{A}_k^2}{\lambda_k + \alpha_k}\right] + \text{const}$$

peaks at K where the data evidence ($\hat{A}_k^2/\sigma_k^2$ large) balances the prior cost (α_k large for high k). This peak is at $K_B^* = \Theta(\log(n/\sigma^2)/\log\rho)$.

---

4. Proof Sketch of the Spectral Duality Theorem

4.1 Frequentist Side (a)

The bias-variance decomposition [Lean-verified: SpectralOverfitting/GeneralizationBound.lean]:

$$\mathbb{E}[\|f - \hat{f}_K\|^2] = \underbrace{C^2\rho^{-2K}}_{\text{bias}} + \underbrace{K\sigma^2/n}_{\text{variance}}$$

Minimizing over K:

$$\frac{d}{dK}\left[C^2\rho^{-2K} + K\sigma^2/n\right] = -2C^2\rho^{-2K}\log\rho + \sigma^2/n = 0$$

$$K_F^* = \frac{\log(2nC^2\log\rho/\sigma^2)}{2\log\rho} = \Theta\left(\frac{\log(n/\sigma^2)}{\log\rho}\right)$$

4.2 Bayesian Side (b)

The log marginal likelihood with URRT prior, using the Laplace approximation:

$$\log p(y|K) \approx -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{k=1}^{K}\frac{\hat{A}_k^2}{\sigma_k^2} + \frac{1}{2}\sum_{k=1}^{K}\log\frac{\alpha_k}{\lambda_k + \alpha_k} + \text{const}$$

The first sum grows with K (more modes explain more variance). The second sum penalizes K (each mode costs $\frac{1}{2}$ log of prior precision). The maximum is at:

$$K_B^* = K_F^* + O(1/\sqrt{n})$$

The $O(1/\sqrt{n})$ correction arises from the prior contribution, which is asymptotically negligible (Bernstein-von Mises).

4.3 MDL Side (c)

The minimum description length:

$$\mathrm{MDL}(K) = -\log p(y|\hat{A}_{1:K}) + \frac{K}{2}\log n$$

The first term is the negative log-likelihood (data fit). The second is the model complexity cost (Schwarz, 1978). For the spectral model:

$$\mathrm{MDL}(K) = \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2}\sum_{k>K}\frac{\hat{A}_k^2}{\sigma^2} + \frac{K}{2}\log n$$

Minimizing: the transition from "mode k improves fit more than it costs" to "mode k costs more than it improves" occurs at:

$$K^*_{MDL} = K^*_F + O(1)$$

The $O(1)$ correction comes from the $\log n$ vs $\log(n/\sigma^2)$ difference.

4.4 The Triple Equality

The three criteria differ by $O(1)$ or $O(1/\sqrt{n})$:

$$|K^*_F - K^*_B| = O(1/\sqrt{n}), \qquad |K^*_F - K^*_{MDL}| = O(1)$$

For practical purposes (integer K): they give the same number.

4.5 Numerical Illustration

To make the theorem concrete, consider the following reproducible example ($n = 300$, $p = 20$, 6 true signal modes with exponentially decaying coefficients, $\sigma = 2$). The spectral information state and all five inferential outputs are computed from the same eigendecomposition:

The eigenvalue spectrum and K^*:

Mode:	0	1	2	3	4	5	|	6	7	8
...										
_k:	2.40	1.97	1.69	1.25	1.22	1.11	|	1.02	0.82	0.67
A_k:	+2.01	+2.29	−0.42	+2.34	+4.06	+4.52	|	+0.33	+3.09	+1.64
_k:	0.18	0.20	0.22	0.25	0.25	0.27	|	0.28	0.31	0.34
t_k:	11.2	11.5	1.9	9.4	16.3	17.0	|	1.2	10.0	4.8
			~ contested			|		K *4		

The five outputs from the same ψ_k, for mode 2 (the contested mode):

Output	Formula	Value	Interpretation
p-value	$2(1-\Phi(1.9))$	0.057	Not quite significant at 0.05
95% CI	$-0.42 \pm 1.96 \times 0.22$	$[-0.85, +0.01]$	Includes zero — uncertain
Posterior	$\mathcal{N}(-0.38, 0.047)$	shrunk toward 0	Prior pulls it — less extreme
MDL cost	$\frac{1}{2}(1.9)^2 + \frac{1}{2}\ln 300$	4.66 bits	Marginal — barely worth the code
Kelly weight	$-0.42/0.047$	-8.9 (small)	Small short — low confidence

This is a mode at the **contested boundary**. The frequentist says "not significant" ($p = 0.057$). The Bayesian says "probably small" (shrunk posterior). MDL says "barely worth coding." Kelly says "small position." All four give the same qualitative answer — weak evidence — expressed in their respective languages.

The three K^* values:

Method	K^*	How computed
Frequentist (bias-variance)	4	$\arg\min_K[C^2\rho^{-2K} + K\sigma^2/n]$
Bayesian (marginal likelihood)	4	$\arg\max_K \log p(y \mid K)$ with URRT prior
MDL	5	$\arg\min_K[-\log p(y \mid \hat{A}_{1:K}) + \frac{K}{2}\log n]$
AIC	5	$\arg\min_K[-2\log L + 2K]$
5-fold CV	4	empirical, using 60% of data per fold

All methods agree within ± 1 mode. The spectral formula gives $K^* = 4$ analytically, using all $n = 300$ observations, in one line. Cross-validation uses 60% per fold and returns a noisy estimate of the same number.

Figure description (to be generated for the LaTeX version):

Figure 1. Left: eigenvalue spectrum λ_k vs mode index k, with K^* marked and the contested zone (modes 4–6) shaded. Right: the discrepancy Δ_k between Bayesian and frequentist estimates per mode, showing the transition from $\Delta \approx 0$ (agreement) to $\Delta \approx 1$ (disagreement) across the contested zone.

---

5. Practical Consequences

5.1 Analytic Model Selection Without Cross-Validation

Current practice: Split data into train/validate/test (typically 60/20/20). Fit on train, select complexity on validate, evaluate on test. You discard 40% of data for a noisy estimate of the right

K.

Spectral approach: Compute K^* from the eigenvalue spectrum of all n observations:

$$K^* = \left\lfloor \frac{\log(2nC^2 \log \rho/\sigma^2)}{2 \log \rho} \right\rfloor$$

No data splitting. No retraining. The formula uses all n observations and gives a deterministic answer. The estimate of ρ comes from the eigenvalue slope; σ^2 from the residuals past rank K^*.

Gain: 40% more data for parameter estimation. Zero-variance model selection. Especially valuable for small datasets (clinical trials, rare events) where splitting is costly.

5.2 Resolution of Model Selection Disputes

Current practice: Team uses BIC; consultant uses 5-fold CV; regulator wants AIC. Different methods give $K = 12$, $K = 15$, $K = 11$. Arguments ensue.

Spectral approach: All three approximate K^*. Compute it directly. The differences are $O(1)$ — within the contested zone of 3 modes. Report: "$K^* = 13 \pm 1$. The disagreement is in modes 12–14, which are at the signal-noise boundary. The choice among 12, 13, and 14 makes less than 2% difference in prediction error."

This reframes the debate from "which method is correct?" to "how wide is the contested zone?" — a quantitative question with a quantitative answer.

5.3 Overfitting Audit Without Data Access

The problem: A published model claims significant results from n observations and p features. Is it overfitted?

The test: Compute the maximum plausible K^* for their reported n and σ^2:

$$K^*_{\max} = \frac{\log(n/\sigma^2)}{\log \rho_{\min}}$$

where $\rho_{\min} \approx 1.2$ for typical smooth functions (conservative). If the model uses $d_{eff} > K^*_{\max}$, it is overfitted — provably, without seeing the data.

Example: $n = 500$, $\sigma^2 = 1$, $\rho_{\min} = 1.2$. Then $K^*_{\max} \approx \log(500)/\log(1.2) \approx 34$. If they report 100 significant features, they are fitting 66 noise modes.

This is a spectral replication crisis detector. Five seconds of arithmetic replaces months of failed replication.

5.4 Optimal Per-Mode Sample Size

Current practice: Power analysis: "I need $n = 64$ for effect size $d = 0.5$, $\alpha = 0.05$, power 0.8." One number for the whole experiment.

Spectral approach: Each mode has its own resolution requirement:

$$n_k = \frac{\sigma^2}{\lambda_k \cdot A_k^2 / z^2}$$

where $z = 1.96$ for 95% resolution. From a pilot study (or domain knowledge about eigenvalue spectrum):

Mode	λ_k	Expected A_k	n_k needed
1 (dominant)	2.0	5.0	3
2 (secondary)	0.8	2.0	19
5 (subtle)	0.2	0.5	307
10 (very fine)	0.05	0.1	15,366

Decision: "Resolve modes 1–5 with $n = 307$. Mode 10 would need 15,000 — not worth it." This is per-mode cost-benefit analysis for experimental design.

5.5 Mode-by-Mode Kelly Criterion for Trading

Current practice: Backtest a trading signal. Report: "Sharpe ratio 1.2 in-sample." Overfit until a good backtest appears.

Spectral approach: Decompose the alpha signal into eigenmodes. Each mode has:

Mode	Signal A_k	Uncertainty σ_k	t-stat	Position weight
1 (market beta)	+2.3	0.05	46.0	max weight
2 (momentum)	+0.8	0.30	2.7	moderate
3 (mean reversion)	+0.4	0.35	1.1	minimal
5 (microstructure)	+0.1	0.12	0.8	zero (noise)

The Kelly-optimal position per mode: $w_k \propto A_k/\sigma_k^2$ (the mode's Sharpe ratio squared). You bet BIG on what you know (mode 1), SMALL on what's uncertain (mode 3), and NOTHING on noise (mode 5).

No backtesting needed. The spectral information state IS the position sizing. The posterior tells you which modes to trust. Overfitting is impossible — noise modes are automatically zeroed by the shrinkage filter $h_k = \lambda_k/(\lambda_k + \alpha)$.

---

6. Formal Verification

The theoretical backbone is Lean 4-verified:

Result	Lean file	Role in this paper
URRT tight bound: $N = \Theta(\log(1/\varepsilon)/\log\rho)$	Universal/MainTheorem.lean	Prior derivation, K^* formula
Entropy lower bound	Universal/EntropyLowerBound.lean	Tightness of K^*
Coefficient decay $\lvert A_k\rvert \leq C\rho^{-k}$	Universal/CoefficientDecay.lean	Prior justification
Eckart-Young optimality	SpectralFenton/Optimality.lean	Spectral truncation is best
Exponential convergence in K	SpectralFenton/ExponentialConvergence.lean	Bias term
Generalization bound	SpectralOverfitting/GeneralizationBound.lean	Theorem 4.1
Optimal K^*	SpectralOverfitting/OptimalComplexity.lean	Theorem 4.1
Ridge = K^*	SpectralOverfitting/RidgeEquivalence.lean	Section 5.2
BIC = K^*	SpectralOverfitting/BICEquivalence.lean	Section 5.2
Bayesian evidence = K^*	SpectralOverfitting/BayesianEquivalence.lean	Theorem 4.2

Note: the SpectralOverfitting/ proofs are from a companion gym (in progress). When complete, this paper's theorems will be fully machine-verified.

---

7. Discussion

7.1 The Contested Zone as Complementarity

The 3-mode contested zone at K^* is where the philosophical choice between Bayesian and frequentist actually matters. Below: both agree. Above: both agree. At the boundary: the Bayesian shrinks (prudent), the frequentist doesn't shrink but has a wide CI (uncertain). Different actions, same epistemic state.

This is not Heisenberg uncertainty — the bias-variance tradeoff is an optimization, not a measurement limit. But the structure is analogous: two views that agree everywhere except at a resolution boundary, with the width of the disagreement determined by a physical constant (ρ playing the role of $\hbar$).

7.2 The Tensor View: Neither Collapse nor Branching

The spectral information state suggests a third ontological position beyond the Bayesian and frequentist interpretations.

Interpretation	Statistical analog	What happens at observation
Copenhagen (collapse)	Bayesian	The model collapses: prior $\rightarrow$ posterior. One face selected.
Many-Worlds (branching)	Frequentist	Each dataset is a branch. The sampling distribution describes all branches.
Spectral (invariance)	This paper	The tensor has faces. Observation reveals one face. The tensor itself is unchanged.

The spectral transfer tensor T contains all eigenvalues $\lambda_1, \lambda_2, \ldots$ regardless of which dataset is observed. A dataset is a **projection** P that selects a subspace. The projected tensor PTP^T has visible eigenvalues (modes $k \leq K^*$) and invisible ones (modes $k > K^*$). But the invisible modes are not destroyed — they are orthogonal to the projection direction.

"Unobservable" does not mean "does not exist." It means "not accessible from this projection angle." Change the projection (different data, different measurement, different experimental design) and previously invisible modes become visible.

This view resolves a conceptual puzzle: the Bayesian says observation *changes* the model (posterior $\neq$ prior). The frequentist says the model was *always fixed*. The spectral view: the tensor was always complete. Observation reveals a face. Nothing collapses. Nothing branches. The eigenvalue spectrum — ρ, K^*, the mode structure — is invariant under all projections.

The practical implication: if you have multiple datasets (different samples, different experiments, different markets), each reveals different faces of the same tensor. **Combining all projections reconstructs the full tensor.** No single dataset sees everything, but the tensor is always the same object. The spectral decay rate ρ is the same from every angle.

7.3 ρ as a Property of the Data-Generating Process

The spectral decay rate ρ is not a free parameter — it is estimable from the eigenvalue spectrum. For a fixed data-generating process, ρ is a constant: it measures the smoothness (analyticity width) of the underlying function. Different kernel choices yield different estimates of ρ, just as different coordinate systems yield different representations of the same physical law. The invariant is the smoothness class, not its numerical representation.

7.4 Robustness Beyond the Gaussian Model

The results in Sections 2–5 use Gaussian examples for clarity. How much survives non-Gaussian data?

Three layers of robustness:

(a) K^* is distribution-free (only requires finite variance).

The bias term $C^2\rho^{-2K}$ comes from the URRT, which is about analyticity of the density — not Gaussianity. The variance term $K\sigma^2/n$ uses only $\mathrm{Var}(\hat{A}_k) = \sigma^2/(n\lambda_k)$, which holds for any distribution with finite second moment. Thus K^* is valid for Student-t, exponential, Poisson, or any finite-variance model.

(b) The "same normal" picture is approximate for finite n.

The statement "$A_k|\text{data} \sim \mathcal{N}(\hat{A}_k, \sigma_k^2)$" and "$\hat{A}_k|A_k \sim \mathcal{N}(A_k, \sigma_k^2)$" are the same distribution requires either Gaussian noise or n large enough for the CLT. For Student-t noise with ν degrees of freedom, the posterior per mode is:

$$A_k|\text{data} \sim t_\nu\Big(\hat{A}_k,\ \sigma_k^2 \cdot \frac{\nu}{\nu - 2}\Big)$$

The Bayesian and frequentist projections are still the same t-distribution — just heavier-tailed than Gaussian. The duality holds for any location-scale family.

(c) The contested zone widens for heavy tails.

For Gaussian data, the contested zone is ~ 3 modes (Section 2.5). For heavier tails, the per-mode uncertainty σ_k^2 is larger (inflated by excess kurtosis), so the Bayesian-frequentist discrepancy Δ_k transitions more slowly across K^*. Empirically:

Noise distribution	Contested zone width	K^* accuracy
Gaussian ($\kappa = 3$)	~ 3 modes	exact
Student-t ($\nu = 5$, $\kappa = 9$)	~ 5 modes	$K^* \pm 1$
Student-t ($\nu = 3$, $\kappa = \infty$)	~ 8 modes	$K^* \pm 2$
Pareto ($\alpha = 3$, finite var)	~ 10 modes	$K^* \pm 3$
Pareto ($\alpha \leq 2$, infinite var)	∞	K^* undefined

The duality theorem degrades gracefully: heavier tails $\to$ wider contested zone $\to$ same K^* but less precise boundary. The only hard failure is infinite variance (Pareto $\alpha \leq 2$), where σ^2 does not exist and the variance term $K\sigma^2/n$ is undefined.

(d) Practical remedies for non-Gaussian data.

1. *Robust scale estimation.* Replace σ^2 with $\text{MAD}^2 \cdot 1.4826^2$ (median absolute deviation). Robust to outliers and heavy tails.
2. *Bootstrap p-values.* Instead of assuming $t_k \sim t_{n-K}$, bootstrap the mode-by-mode test statistics. The eigendecomposition is the same; only the inference step changes.
3. *The URRT prior is still valid.* The decay $|A_k| \leq C\rho^{-k}$ depends on the analyticity of the density, not its tail. A Student-t density IS analytic on a Bernstein ellipse (with ρ depending on ν). The prior form $A_k \sim \mathcal{N}(0, C^2\rho^{-2k})$ remains correct — only ρ changes.

Summary. K^* is robust. The Bayesian-frequentist duality holds for any finite-variance distribution. The Gaussian case is the sharpest (3-mode contested zone); heavier tails widen it but do not break it. The eigenvalue spectrum remains the observer-independent invariant regardless of the noise distribution.

7.5 Limitations

1. **Requires spectral decay.** The theory assumes eigenvalues decay (the function is smooth). For discontinuous functions, $\rho = 1$ and $K^* = \infty$ — the theorem is vacuous. Trees are better than spectral methods for discontinuities.

2. **Linear eigenbasis.** The current formulation uses linear (SVD) or kernel eigenmodes. For highly nonlinear or compositional functions, the eigenbasis may not capture the relevant structure efficiently.

3. **Asymptotic corrections.** The $O(1)$ and $O(1/\sqrt{n})$ corrections are not computed explicitly. For small n, they may matter.

4. **Infinite variance.** For distributions without finite second moment (Cauchy, stable with $\alpha < 2$), the variance term and K^* are undefined. The spectral framework requires at least finite variance.

---

8. Conclusion

The spectral information state $\psi_k = (A_k, \sigma_k^2)$ is computed once and answers every question the practitioner faces:

Question	Output	Formula from ψ_k
Is this mode significant?	p-value	$p_k = 2(1 - \Phi(\|A_k\|/\sqrt{\sigma_k^2}))$
What is the plausible range?	Confidence interval	$A_k \pm z\sqrt{\sigma_k^2}$
What do I believe?	Bayesian posterior	$\mathcal{N}(h_k A_k,\ h_k^2 \sigma_k^2)$
Should I include this mode?	MDL cost	$\frac{1}{2} A_k^2/\sigma_k^2 + \frac{1}{2}\log n$
How much should I bet?	Kelly weight	$w_k \propto A_k/\sigma_k^2$

These are five different questions with five different audiences — and they are all answered from the same pair of numbers per mode. The eigendecomposition is the computation; the rest is formatting.

The deeper result: the three model-selection frameworks (Bayesian, frequentist minimax, MDL) that appear to compete are in fact computing the same quantity — $K^* = \Theta(\log(n/\sigma^2)/\log\rho)$ — from different directions. Their disagreement is confined to $O(1)$ modes near the boundary and is $O(1/\sqrt{n})$ in magnitude. For typical data, this means approximately 3 modes.

The Bayesian prior that makes this work is not assumed but proved: the URRT coefficient decay $|A_k| \leq C\rho^{-k}$, verified in Lean 4, determines the prior uniquely. The result survives non-Gaussian noise, requiring only finite variance.

The title of this paper is literal. Bayesian and frequentist statistics agree on eigenvalues. The century-old debate is about 3 modes — and the formula tells you exactly which 3.

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

- Amari, S (1985). Differential-Geometrical Methods in Statistics. *Differential-Geometrical Methods in Statistics.*. DOI: 10.1007/978-1-4612-5056-2
- Amari, S. and Nagaoka, H (2000). Methods of Information Geometry. *Methods of Information Geometry.*. DOI: 10.1090/mmono/191
- Berger, J.O (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, 1(3), 385-402.
- Čencov, N.N (1982). Statistical Decision Rules and Optimal Inference. *Statistical Decision Rules and Optimal Inference.*. DOI: 10.1090/mmono/053

- Donoho, D.L., Liu, R.C., and MacGibbon, B (1990). Minimax risk over hyperrectangles, and implications. *Annals of Statistics*, 18(3), 1416-1437.
- Efron, B (1986). "Why isn't everyone a Bayesian?" *The American Statistician*, 40(1), 1–5. *The American Statistician*, 40(1), 1-5.
- Le Cam, L (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *UC Berkeley Publications in Statistics*, 277-329.
- Nagy, T. (2026). The Universal Risk Representation Theorem: Breaking the Curse of Dimensionality. *Zenodo*. DOI: 10.5281/zenodo.18910566
- Nagy, T. (2026). The Spectral Theory of Observation. *Working paper*.
- Pinsker, M.S (1980). Optimal filtering of square-integrable signals in Gaussian noise. *Problems of Information Transmission*, 52-68.
- Rissanen, J (1978). Modeling by shortest data description. *Automatica*, 14(5), 465-471.
- Schwarz, G (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.
- Schwartz, L (1965). On Bayes procedures. *Zeitschrift für Wahrscheinlichkeitstheorie*, 4(1), 10-26.
- Stein, C (1961). "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution." *Proc. 3rd Berkeley Symposium*, 1, 197–206. *Proc. 3rd Berkeley Symposium*, 197-206. DOI: 10.1525/9780520313880-018