Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

One Truncated Likelihood Expansion: Estimation, Testing, and Classification as a Single Captured-Fraction Functional

Serhii Zabolotnii

zabolotnii.serhii@csbc.edu.ua

Cherkasy State Business College https://orcid.org/0000-0003-0242-2234

Research Article

Keywords: Stochastic polynomials, polynomial maximization method captured-fraction functional, asymptotic relative efficiency, sequential change-point detection, heavy-tailed inference

Posted Date: July 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10226291/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

One Truncated Likelihood Expansion: Estimation, Testing, and Classification as a Single Captured-Fraction Functional

Serhii Zabolotnii
Department of Information, Multimedia Technologies and Design,
Cherkasy State Business College, Cherkasy 18028, Ukraine
State Scientific Research Institute of Armament and Military Equipment
Testing and Certification, Cherkasy, Ukraine
Department of Cybernetics and Applied Mathematics,
Uzhhorod National University, Uzhhorod, Ukraine
ORCID: 0000-0003-0242-2234
zabolotnii.serhii@csbc.edu.ua

2026

Abstract

Parameter estimation, hypothesis testing, and signal classification are usually studied with separate machinery and separate notions of quality (estimator efficiency, error probability, reconstruction error). We show that, when each task is built from the truncated expansion of a log-likelihood in a fixed basis, all three are read off by *one* functional: the fraction of the relevant object captured by an order-m projection. Writing $\kappa_m(z;\nu) = \|\operatorname{proj}_{V_m} \mathring{z}\|_\nu^2 / \|\mathring{z}\|_\nu^2$ for the captured fraction of an object z in $L^2(\nu)$, estimation reads κ_m on the score, detection on the log-likelihood ratio (its sequential change-point form reads the same object as a normalized information functional), and decomposition on the log-density. These are the *same functional on different objects*, not a single number; they reach the value one *simultaneously* exactly when the truncated series captures the centered log-density itself—the sufficient statistic *and* the carrier ($\mathring{\log p} \in V_m$)—and otherwise lie strictly below one and increase as the basis is enriched. The paper gives a self-contained projection-efficiency identity for the estimation branch; a per-branch termination theorem on exponential families; a completeness theorem showing that, under a heavy tail, the non-orthogonal fractional basis is dense in L^2 exactly where the polynomial/Hermite

system is not even square-integrable, so the basis *restores* the closure the orthogonal system loses; a bridge identifying the captured fraction of the score with the asymptotic relative efficiency and local power of both the estimation and detection branches; a local cross-branch equality ($c(\delta) \to g$ as a two-hypothesis gap $\delta \to 0$ for location families); and a Gaussian covariance-shift anchor whose order-two projection is exact, so the sequential statistic equals the exact likelihood-ratio and classical change-point optimality applies directly. The detection identities of that anchor are machine-checked in Lean 4/Mathlib. A controlled study (fifteen pre-specified checks, all passing) confirms simultaneous termination on the Gaussian, joint rise of all readings on the Laplace, the local equality numerically, and—on Student-t and Cauchy laws—that the non-orthogonal fractional element captures Fisher information where a polynomial/Hermite basis is not even square-integrable. The same construction yields the reduction principle as a corollary.

Keywords: Stochastic polynomials; polynomial maximization method; captured-fraction functional; asymptotic relative efficiency; sequential change-point detection; heavy-tailed inference.

Mathematics Subject Classification (2020): 62F12 (primary); 62F03; 62L10; 62H30; 42C05.

1 Introduction

Three of the most basic tasks of statistical inference—estimating a parameter, deciding between two hypotheses, and classifying a signal by how well it can be reconstructed—are normally treated by separate theories, each with its own measure of success: the relative efficiency of an estimator, the error probability or detection delay of a test, the reconstruction error of a decomposition. A single object that produced all three, with a single measure of how complete it is, would tie these theories together and make their shared limits visible. This paper exhibits such an object for inference built on *stochastic polynomials*: the truncated expansion of a log-likelihood in a fixed, possibly non-orthogonal basis.

Concretely, fix a basis $\{\varphi_i\}$ and expand the log-likelihood $\log p(x;\theta)$ in it, truncating at order m. Three elementary operations on this one series produce the three tasks. Differentiating in the parameter, ∂_θ, gives the score and the polynomial-maximization estimator. Differencing between two hypotheses, Δ, gives the log-likelihood ratio and the Neyman–Pearson test; composing the same difference along time and feeding the increments to a stopping rule gives the sequential change-point statistic. Taking the norm

of the truncation residual gives the reconstruction (decomposition) error. This basis and these operations are the apparatus of Kunchenko stochastic polynomials, whose three lines— polynomial parameter estimation, polynomial detection criteria, and decomposition in a space with a generating element—have developed largely separately [20, 37].

Thesis. We argue that these are not three methods but one object read three ways, and that their quality is one functional read on three different objects. For a reference measure ν and the centered order-m subspace V_m, let $\kappa_m(z;\nu) = \|\operatorname{proj}_{V_m} \mathring{z}\|_\nu^2 / \|\mathring{z}\|_\nu^2 \in [0,1]$ be the fraction of an object z captured by the projection (the coefficient of determination of z on the basis). Then the estimation efficiency g, the detection captured-ratio c, the sequential information ratio $J(s)/\|z\|^2$, and the decomposition R^2 are *the same functional κ_m evaluated on different objects in different reference measures*—the score, the likelihood ratio, the change-point drift, the log-density. They are emphatically *not the same number*: on the Laplace law below, g rises from 0.50 to 0.98 while c rises from 0.71 to 0.94. What is true—and is the content of this paper—is that all of them equal one *simultaneously* exactly when the truncated series captures the centered log-density—the sufficient statistic *and* the carrier, $\log \mathring{p} \in V_m$ (Theorem 1)—all of them rise as the basis is enriched (Lemma 1), and the estimation and detection readings genuinely *coincide* in the local limit of a vanishing hypothesis gap (Proposition 2).

Relation to prior work (stated up front). Expanding a likelihood in a basis is classical—Gram–Charlier and Edgeworth series, Hermite expansions of the likelihood ratio, Neyman's smooth test [27], score matching [19], estimating functions and the generalized method of moments [15, 17], exponential-family sufficiency [7], and density approximation by exponential families [4]. The captured-fraction functional itself is the textbook coefficient of determination, and the estimation instance g is the classical efficiency-as-squared-correlation of an estimating function [14]. *We claim none of these.* The contribution is their *conjunction*: (i) the non-orthogonal fractional/PATP Kunchenko generating element $\rho_a(x) = \operatorname{sgn}(x)|x|^a$, which remains square-integrable on heavy-tailed laws where a Hermite system diverges [35, 36]; (ii) the identification of the three Kunchenko tasks as ∂/Δ/residual of *one* truncated series (with the sequential change-point line as Δ composed with time); (iii) the recognition that one functional κ_m governs all of them, with the precise statements—simultaneous termination, joint shrinkage, and the local coincidence—that replace the (false) claim that the four are one number; and (iv) an invariance corollary. A prior re-

view [37] already describes the three Kunchenko uses as a "coherent family of moment–cumulant procedures"; it does *not* reduce them to functionals of one series, nor identify a shared captured-fraction reading, nor characterize termination on sufficiency (§2).

Contributions. We separate what is genuinely new from what is recalled and unified.

C1. The master object (Definition 1, well-posed via the normal system and explicit integrability conditions) and the mapping of the three tasks onto $\partial/\Delta/$ residual of one series, read by one functional κ_m (§3–§4).

C2. *New—completeness by basis switch.* Under a heavy tail the admissible non-orthogonal fractional system is *complete* in $L_0^2(\nu)$ once its exponents accumulate inside $(0, \alpha/2)$ (Theorem 3; Proposition 4 is the light-tail case), restoring the L^2 closure the polynomial/Hermite system loses there. This is the precise sense in which the non-orthogonal basis is necessary, not merely convenient.

C3. *New—the captured fraction is the operating characteristic.* On the estimation and detection branches $\kappa_m(S) = g$ is the asymptotic relative efficiency, with explicit local power $\Phi(\sqrt{gI}\, h - z_\alpha)$ and, locally, error-exponent ratio g (Proposition 3, Corollary 3); on heavy tails this is a power statement the orthogonal basis cannot even make.

C4. *Recalled and unified.* A self-contained projection-efficiency identity $g = b^\top G^{-1} b/I$ (Proposition 1; the classical efficiency-as-squared-correlation, with the symmetric oPMM/PMM coefficient as Corollary 1); a per-branch termination theorem on exponential families (Theorem 1); monotone shrinkage on a nested basis (Lemma 1, standard R^2 monotonicity); and the local cross-branch equality $c(\delta) \to g$ (Proposition 2, the captured-fraction form of LAN).

C5. The detection identity $\Lambda = \Delta(\frac{1}{2}D^2)$ and a Gaussian covariance-shift anchor on which the order-two projection is exact, so the sequential statistic *is* the exact likelihood-ratio and classical CUSUM optimality applies directly (Theorem 2, Corollary 2); these identities are machine-checked in Lean (scope delimited in §5).

C6. A controlled study (§6, pre-specified checks with reported uncertainty) confirming termination, joint rise, the local equality, the operating-characteristic and completeness predictions, and the heavy-tailed advantage; and the recovery of a reduction principle as an invariance

corollary (§7), with related program results recovered as consequences (§8).

2 Related work

We situate the contribution against the relevant prior work before developing it; references to our own results below point forward to the formal statements in §5–§6.

Likelihood expansions and smooth tests. Gram–Charlier/Edgeworth series and Hermite expansions of the log-ratio [25] use a fixed orthogonal system; Neyman's smooth test [27, 32] expands the log-density in a basis around the null and tests the projection coefficients—the closest detection-side precedent to our order-m log-ratio projection. Our basis is the non-orthogonal fractional element, square-integrable on heavy tails where Hermite diverges, and we read the same projection across three tasks rather than testing one. Moreover, where the orthogonal systems lose L^2 completeness under a heavy tail, the non-orthogonal fractional system remains dense (Theorem 3)—a Müntz–Szász-type phenomenon [6].

Estimating functions, GMM, efficiency. $g = b^\top G^{-1} b/I$ is the efficiency-as-squared-correlation of an optimal estimating function [14, 15] and the GMM-vs-MLE efficiency [17]; score matching [19] fits unnormalized models by the score. These concern the estimation branch alone; the contribution is sharing that reading with detection and decomposition under one functional.

Local asymptotics, efficiency, and error exponents. Our local cross-branch equality (Proposition 2) is the captured-fraction form of the local equivalence of estimation and testing under quadratic-mean differentiability [21, 33], and the degree-two structure of Corollary 2 is the local quadratic of LAN. We cast the captured fraction as a Pitman asymptotic relative efficiency [28, 29] (Proposition 3), and locally as a Chernoff/Bhattacharyya error-exponent ratio [8, 9]; this is what ties the functional to operating characteristics rather than to a descriptive R^2. The captured fraction and the termination-on-sufficiency picture are the m-projection/Pythagorean view of information geometry [2, 10]; what is added is the non-orthogonal fractional basis, the cross-branch sharing, and the completeness theorem (Theorem 3).

Robustness, density ratios, sufficiency. The fractional-basis estimator is a structured M-estimator [16, 18] aligned with a heavy-tailed, non-smooth score; the detection branch is a basis-projected log-density-ratio, related to but distinct from generic density-ratio fitting; the empirical characteristic function offers a moment-free twin [12]. Exponential families and sufficiency [7, 22] give the termination case (Theorem 1); the work is the truncation regime off the family.

Fisher information as a single cross-task scalar. The closest precedent to a single cross-task scalar is the classical fact that, for weak-signal processing, the Fisher information alone sets estimator efficiency, detection efficacy, and minimum mean-squared error [11]. We differ in three ways: our reading is the captured fraction of a finite *non-orthogonal* series ($\|\,\mathrm{proj}\,\|^2/\mathrm{total}$), not the full Fisher information of one operation; it includes the decomposition branch; and it is tied to the termination-on-sufficiency characterization (Theorem 1) and the explicit detection identity (Theorem 2).

The Kunchenko apparatus. Polynomial maximization [20], the fractional line [35, 36], and a related review [37] are the substrate. That review frames the three uses as a "coherent family of moment–cumulant procedures"; that is a shared historical apparatus, not the present thesis. We add that the three are ∂/Δ/residual of *one* truncated series, read by *one* captured-fraction functional, with a termination criterion and a local cross-branch equality—and the invariance corollary.

3 The master object

Let $p(x;\theta)$ be a density and let $\{\varphi_i\}$ be a generating-element basis— polynomial, fractional, or PATP (the continuous-α family $\rho_a(x) = \mathrm{sgn}(x)|x|^a$). Fix a reference probability measure ν, write $\langle f, g\rangle_\nu = \mathbb{E}_\nu[fg]$, $\mathring{f} = f - \mathbb{E}_\nu f$, and center each basis function by $\psi_i = \mathbb{E}_\nu[\varphi_i]$.

Definition 1 (Truncated Kunchenko likelihood series)**.** *For an order m,*

$$\log p(x;\theta) = \psi_0(\theta) + \sum_{i\le m} k_i(\theta)\,(\varphi_i(x) - \psi_i) + R_m(x;\theta),$$

where the coefficients $K = (k_i)$ solve the normal system $F\,K = B$ with $F_{ij} = \mathrm{Cov}_\nu(\varphi_i, \varphi_j)$ and $B_i = \mathrm{Cov}_\nu(\log p, \varphi_i)$, and R_m is the $L^2(\nu)$-orthogonal trun-

cation residual. The branch variants replace the target $\log p$ *by the score* S *(estimation) or the log-ratio* Λ *(detection).*

Well-posedness (Lemma, standard). F is symmetric positive semidefinite, and positive definite—hence $K = F^{-1}B$ exists and is unique—iff $\{\mathring{\varphi}_1, \ldots, \mathring{\varphi}_m\}$ are linearly independent in $L^2(\nu)$ (no nontrivial a.s. affine relation); these are the normal equations of the L^2 orthogonal projection [24]. Finiteness at finite m requires only $\varphi_i, \log p \in L^2(\nu)$. For the fractional element, $\rho_a \in L^2(\nu)$ iff $\mathbb{E}_\nu |X|^{2a} < \infty$; under a regularly varying tail of index α this is guaranteed when $2a < \alpha$ and fails when $2a > \alpha$ (the boundary $2a = \alpha$ depends on the slowly varying factor). Hence the whole fractional set is admissible whenever $2 \max_i a_i < \alpha$—a finite constraint that is exactly the heavy-tail regime in which the Hermite system, requiring all polynomial moments, leaves L^2 (§6). Far from a gap, this is precisely where the non-orthogonal basis is needed: under a heavy tail the admissible fractional system is *complete* in $L^2_0(\nu)$ as soon as its exponents accumulate inside $(0, \alpha/2)$ (Theorem 3), restoring the L^2 closure that the polynomial/Hermite system loses there.

Termination on exponential families. A distinguished case is the exponential family $p_\theta(x) = h(x) \exp(\langle \eta(\theta), \varphi(x) \rangle - A(\theta))$ with the basis φ a sufficient statistic: there the series *terminates*, $R_m \equiv 0$ at finite m, and the expansion is an identity rather than an approximation. This is the anchor against which the general case is measured (Theorem 1).

4 One functional, read on three objects

Definition 2 (Captured-fraction functional)**.** *For* $z \in L^2(\nu)$ *with* $\mathring{z} \neq 0$*, and* $V_m(\nu) = \operatorname{span}\{\mathring{\varphi}_1, \ldots, \mathring{\varphi}_m\}$ *with orthogonal projection* P_{V_m}*,*

$$\kappa_m(z; \nu) := \frac{\|P_{V_m} \mathring{z}\|^2_\nu}{\|\mathring{z}\|^2_\nu} \in [0, 1].$$

This is the coefficient of determination of z *on the (centered) basis;* $\kappa_m \leq 1$ *is the projection theorem.*

The three tasks report completeness through κ_m applied to three different objects, in three reference measures, summarized in Table 1. The estimation efficiency is $g = \kappa_m(S; p_\theta)$ for the score $S = \partial_\theta \log p$; the detection captured-ratio is $c = \kappa_m(\Lambda; \bar{p})$ for the log-ratio $\Lambda = \log(p_1/p_0)$ under the mixture $\bar{p} = \frac{1}{2}(p_0 + p_1)$; the decomposition is $R^2 = \kappa_m(\log p; \nu)$. We call

Table 1: The three tasks as ∂/Δ/residual of one truncated log-likelihood series, each read by the same functional κ_m on a different object. The detection task has a static form (Δ) and a sequential form (the change-point line, Δ composed with time); the latter shares the detection object and reads κ_s as a normalized information functional. All readings use centered $L^2(\nu)$.

Task (operation)	Object z (measure ν)	Reading $= \kappa_m(z;\nu)$
Estimation (∂_θ)	score S (model p_θ)	$g = \|\operatorname{proj} S\|^2/I$
Detection (Δ of hypotheses)	log-ratio Λ (mixture $\bar{p}$)	$c = \|\operatorname{proj}\Lambda\|^2/\|\Lambda\|^2$
$\hookrightarrow$ sequential (change-point)	projected Λ-increments	$J(s)/\|z\|^2$
Decomposition (residual norm)	log-density $\log p$ (data ν)	$R^2 = 1 - \mathrm{MSED}/\mathrm{Var}$

g, c, R^2 the three *readings* (each the captured fraction of its own object), and a model on which a branch reaches $\kappa_m = 1$ at finite order an *anchor*; each task's natural inner product fixes its reference measure ν, which is why the same functional is comparable across branches yet in general returns different numbers. The sequential change-point line is not a fourth object: it projects the log-ratio increment onto the basis, runs a stopping rule on the projected increments, and measures the projection by the normalized information functional $J(s)/\|z\|^2$, which is κ_s on the change-point drift.

What is and is not claimed. Under this reading the four quantities are, in general, *different numbers*, not one hidden scalar: they coincide only on a terminating anchor, where all reach one (Theorem 1); the estimation and detection readings agree off it only in the limit of a vanishing hypothesis gap (Proposition 2); and each is separately non-decreasing in the basis (Lemma 1). The claim is thus structural, not numerical—off sufficiency every reading stays below one and rises with m (§6). The one classical precedent for a single cross-task scalar, the Fisher information [11], differs in being a full information rather than the captured fraction of a finite *non-orthogonal* series shared with decomposition (§2).

5 Core identities and termination

We give one identity per branch and the two theorems—termination and the local equality—that carry the unification, with proofs. The detection identities of §5.3 are machine-checked; we delimit that scope precisely.

5.1 Estimation: projection efficiency

Proposition 1 (Projection efficiency)**.** *Let $\nu = p_\theta$, $S = \partial_\theta \log p_\theta \in L^2(p_\theta)$ with the standard regularity $\partial_\theta \int p_\theta = \int \partial_\theta p_\theta$ (so $\mathbb{E}_{p_\theta} S = 0$ and $0 < I = \mathbb{E}_{p_\theta}[S^2] < \infty$), $\varphi_i \in L^2(p_\theta)$, and centered Gram $G_{ij} = \langle \mathring{\varphi}_i, \mathring{\varphi}_j \rangle$ nonsingular. Then*

$$g = \kappa_m(S; p_\theta) = \frac{\|P_{V_m} S\|^2}{I} = \frac{b^\top G^{-1} b}{I}, \qquad b_i = \langle \mathring{\varphi}_i, S \rangle = \langle \varphi_i, S \rangle,$$

with $0 \le g \le 1$ and $g = 1$ iff $S \in V_m$; the maximal squared correlation $\max_{\psi \in V_m} \operatorname{corr}(\psi, S)^2 = g$ is attained on the ray $k \propto G^{-1} b$, with projection representative $K = G^{-1} b$.

Proof. Regularity gives $\mathbb{E} S = 0$, so $\mathring{S} = S$ and $b_i = \langle \varphi_i, S \rangle$. The projection $P_{V_m} S = \sum_j k_j \mathring{\varphi}_j$ is fixed by residual orthogonality $\langle \mathring{\varphi}_i, S - \sum_j k_j \mathring{\varphi}_j \rangle = 0$, i.e. $Gk = b$, so $k = G^{-1} b$. Since $S - P_{V_m} S \perp V_m$, $\|P_{V_m} S\|^2 = \langle P_{V_m} S, S \rangle = k^\top b = b^\top G^{-1} b$, and $\|S\|^2 = I$. Pythagoras gives $0 \le g \le 1$ and $g = 1 \iff S \in V_m$. For any $\psi = \sum_i k_i \mathring{\varphi}_i$, $\operatorname{corr}(\psi, S)^2 = (b^\top k)^2 / (k^\top G k \, I)$ is maximized over k by Cauchy–Schwarz in the G-inner product at $k \propto G^{-1} b$, value $b^\top G^{-1} b / I = g$. □

Proposition 1 is the L^2 content; its reading as a relative *efficiency* of a one-step estimator—that g equals the asymptotic CRB/variance ratio—requires the estimating-function sensitivity (Godambe–Bartlett) identity and is the classical efficiency-as-squared-correlation of an optimal estimating function [14, 15], of which the polynomial-maximization estimator is the Kunchenko-basis instance; it is also the generalized-method-of-moments efficiency relative to maximum likelihood [17]. The numerical study (§6, T2) confirms the two readings agree to 0.014. For the symmetric power basis this admits a closed form.

Corollary 1 (Symmetric degree-3 reduction)**.** *For a symmetric standardized location family with finite sixth moment and cumulants γ_4, γ_6, the even element x^2 is orthogonal to the odd score and to $\{x, x^3\}$, so V_m at order three is* $\operatorname{span}\{x, x^3\}$ *and the order-3/order-1 variance ratio is*

$$g_{3|1} = \frac{g_1(S)}{g_3(S)} = 1 - \frac{\gamma_4^2}{6 + 9\gamma_4 + \gamma_6},$$

where $g_j(S)$ denotes the order-j captured-Fisher fraction. (For standardized t_{10}, $\gamma_4 = 1, \gamma_6 = 10$ give $g_{3|1} = 0.96$ exactly.)

This is the symmetric oPMM/PMM coefficient [35]; we record it for context. Note $g_{3|1}$ is a variance-*ratio* between orders, distinct from the captured fraction g itself (which carries the $1/I$ factor and has no cumulant-only closed form): the two must not be conflated.

5.2 The termination theorem

Theorem 1 (Simultaneous termination, per branch)**.** *Let* $p_\theta = h\exp(\langle\eta(\theta), T\rangle - A(\theta))$ *be a full-rank exponential family with* $T = (\varphi_1, \ldots, \varphi_m)$ *the basis. Then:*

(a) Estimation *terminates for any such family:* $S_\theta = \langle\dot\eta, \mathring{T}\rangle \in V_m$*, so* $g = 1$ *(the carrier* $\log h$ *drops,* $\partial_\theta \log h = 0$*).*

(b) Detection *terminates for any such family with shared carrier:* $\Lambda = \langle\eta_1 - \eta_0, T\rangle - (A_1 - A_0) \in \mathbb{R} \oplus V_m$*, so* $c = 1$ *(*$\log h$ *cancels in the ratio).*

(c) Decomposition *terminates iff additionally* $\log h \in \operatorname{span}\{1, \varphi\}$*: only then is* $\log p_\theta$ *itself in* $\mathbb{R} \oplus V_m$*, giving* $R^2 = 1$*.*

(d) For the full family, sufficiency of the basis is equivalent to (a)–(b).

Consequently all readings reach one simultaneously *exactly on a family whose carrier is captured by the basis (e.g. the Gaussian).*

Proof. $\dot A = \langle\dot\eta, \mathbb{E}_{p_\theta} T\rangle$ gives $S_\theta = \langle\dot\eta, T - \mathbb{E}T\rangle = \langle\dot\eta, \mathring{T}\rangle \in V_m$, proving (a). For (b), $\log h$ is common to p_0, p_1 and cancels, leaving Λ affine in T, hence in $\mathbb{R} \oplus V_m$; centering removes the constant, so $c = 1$. For (c), $\log p_\theta = \log h + \langle\eta, T\rangle - A$, so $\mathring{\log} p_\theta \in V_m$ iff $\mathring{\log} h \in V_m$, i.e. $\log h \in \operatorname{span}\{1, \varphi\}$; the Gaussian carrier $\log h = -x^2/2$ lies in $\operatorname{span}\{1, x^2\}$ and terminates, a generic carrier does not. Part (d) is the equivalence between sufficiency and exact linearity of the log-likelihood in the statistic. □

The Gaussian known-variance mean family illustrates the separation: with basis $\varphi = \{x\}$, estimation and detection terminate ($g = c = 1$, score and log-ratio affine in x), yet the carrier $-x^2/(2\sigma^2) \notin \operatorname{span}\{1, x\}$, so decomposition does *not* terminate at order one—it needs x^2. Parts (a)–(d) are classical exponential-family facts [7, 22]; the contribution is reading them through the single functional and the *simultaneity*. The non-exponential case is exactly where the work lies: there no finite-order termination occurs, and the readings merely rise.

Lemma 1 (Monotone shrinkage on a nested basis)**.** *For a probability measure ν, a fixed $z \in L^2(\nu)$ with $\mathring{z} \neq 0$, and a nested sequence $V_1 \subseteq V_2 \subseteq \cdots$ obtained by appending one $\varphi_i \in L^2(\nu)$ at a time, $m \mapsto \kappa_m(z;\nu)$ is non-decreasing and ≤ 1, with $\kappa_m = 1$ at finite m iff $\mathring{z} \in V_m$. Equivalently the residual fraction $1 - \kappa_m$ is non-increasing.*

Proof. For closed $U \subseteq W$, $P_W = P_U + P_{W \cap U^\perp}$, so $\|P_W \mathring{z}\|^2 \geq \|P_U \mathring{z}\|^2$ by Pythagoras; nesting gives the claim. This is the standard monotonicity of R^2 under adding regressors [24]. □

Lemma 1 is per-branch (fixed object and measure); it does not by itself relate g, c, R^2 *across* branches, which use different (z, ν). The sequential development already machine-checks the matching $J(s) \leq \|z\|^2$, monotone, $\to \|z\|^2$ [34]; Lemma 1 is the projection statement behind it.

5.3 Detection: the bridge, the anchor, and the local equality

Theorem 2 (Detection bridge)**.** *For a Gaussian-in-basis model with per-class Mahalanobis fields D_c and log-determinants $\log|F_c|$, the log-ratio is $\Lambda = -(\frac{1}{2}D_1 - \frac{1}{2}D_0) - \frac{1}{2}(\log|F_1| - \log|F_0|)$; when $\log|F_1| = \log|F_0|$ it is exactly the QDA discriminant. The argmax-posterior label minimizes the per-class penalty $\frac{1}{2}D_c + \frac{1}{2}\log|F_c|$ (the MAP/ML decision under equal priors), so optimality of that rule transfers to the decomposition decision.*

The D_c are the per-class reconstruction (decomposition) penalties, so detection and decomposition read the same statistic on this model. On a Gaussian covariance shift the log-ratio is an explicit quadratic, which makes the order-two projection exact.

Corollary 2 (Gaussian anchor = Ghosh = Bayes)**.** *For $H_0 = \mathcal{N}(\mu_0, v_0)$, $H_1 = \mathcal{N}(\mu_1, v_1)$ with $v_0, v_1 \neq 0$, $\Lambda(x) = \left(\frac{1}{2v_0} - \frac{1}{2v_1}\right)x^2 + \left(\frac{\mu_1}{v_1} - \frac{\mu_0}{v_0}\right)x +$ const $\in \operatorname{span}\{1, x, x^2\}$; equal variances kill the quadratic term (the S=1 rung). Hence the order-two projection returns Λ in any sampling measure, the information functional saturates ($J(s) = \|z\|^2$), and Λ is an affine function of the two Mahalanobis heads—the Mahalanobis-distance classifier of [13] with identity link—whose sign is the Bayes decision. CUSUM on the order-two projected increments is therefore CUSUM on the exact log-ratio, and the classical optimality [23, 26, 30] applies directly.*

Scope of the machine-checked core. Theorem 2 and Corollary 2 are formalized in Lean 4/Mathlib (16 theorems, zero `sorry`, compiled against

Mathlib v4.26.0; the lemma inventory and build instructions are in the reproducibility appendix). What the Lean core certifies is precisely the *algebraic exactness on the Gaussian terminating anchor*—that the Gaussian log-ratio is a degree-two polynomial, equals a difference of Mahalanobis scores, has Bayes-decision sign, and that a finite orthonormal system saturates Parseval on its span. It does *not* certify the off-family, non-orthogonal, or general-basis statements (Propositions 1, 2, Theorem 1), which rest on the proofs above and the numerical study.

Proposition 2 (Local cross-branch equality)**.** *For a location family $p(x-\theta)$ such that $\theta \mapsto \log p(x-\theta)$ is $L^2(p_0)$-differentiable at 0 with score $S = -p'/p \in L^2(p_0)$, $0 < I < \infty$, and $\varphi_i \in L^2(p_0)$, the two-hypothesis log-ratio satisfies $\Lambda_\delta(x) = \log p(x-\delta) - \log p(x) = \delta S(x) + R_\delta$ with $\|R_\delta\|_{L^2(p_0)} = o(\delta)$. Hence, computing both readings under p_0 with a matched centered basis,*

$$c(\delta) = \kappa_m(\Lambda_\delta; p_0) \xrightarrow[\delta \to 0^+]{} \kappa_m(S; p_0) = g.$$

Proof. $\Lambda_\delta = \delta S + R_\delta$ with $R_\delta/\delta \to 0$ in $L^2(p_0)$ by the assumed differentiability. κ_m is scale-invariant and continuous in its argument where the denominator is bounded away from 0 (here $\|\mathring{S}\| = \|S\| > 0$); applying it to $\Lambda_\delta/\delta \to S$ gives $c(\delta) \to \kappa_m(S; p_0) = g$. □

Proposition 2 is the captured-fraction reflection of the classical local equivalence of estimation and testing under quadratic-mean differentiability [33]; it is the one place the estimation and detection readings genuinely coincide *off* the terminating anchor (on the anchor they are both one already).

Remark (sequential operating characteristics). On the Gaussian covariance-shift anchor (Corollary 2) the order-two projected increments *equal* the exact log-ratio increments, so CUSUM/Shiryaev–Roberts on them is the exact-likelihood-ratio procedure and the minimax delay optimality of [23, 26, 31] transfers verbatim. Off the anchor the projected increment $Z = P_{V_m}\Lambda$ is a quasi-log-ratio with H_1-drift $\mu_1 = \mathbb{E}_{H_1}[Z]$ (the *captured drift*); by Wald's identity a one-sided rule with fixed threshold b has first-order delay ADD $\approx b/\mu_1$. Because Z is not a genuine log-ratio, the threshold matching a target false-alarm run length is set by the H_0 large-deviation rate of Z and need not equal the exact-ratio threshold; a matched-alarm delay ratio, and any minimax statement, are therefore deferred (an inherited open problem of the change-point line, §9).

5.4 Operating characteristics: efficiency, power, and error exponent

On the estimation and detection branches the captured fraction is the asymptotic relative efficiency, with an explicit local-power curve.

Proposition 3 (Captured fraction as operating characteristic)**.** *Under the local setting of Proposition 2 (location family, quadratic-mean differentiable at* 0*,* $S \in L^2(p_0)$*,* $0 < I < \infty$*,* $\varphi_i \in L^2(p_0)$*):*

(i) Estimation. *The largest estimating-function efficiency* $\langle \psi, S\rangle^2/\|\psi\|^2$ *over* $\psi \in V_m$ *equals* $\|P_{V_m}S\|^2 = gI$*, so the order-m estimator has asymptotic relative efficiency* $g = \kappa_m(S; p_0)$ *against maximum likelihood.*

(ii) Testing. *For* $H_0 \colon \theta = 0$ *against local alternatives* $\theta_n = h/\sqrt{n}$ *(*$h > 0$*), among level-*α *tests with statistic* $n^{-1/2}\sum_i \psi(X_i)$*,* $\psi \in V_m$*, the most powerful within the order-m class has Pitman efficacy* $\|P_{V_m}S\|^2 = gI$ *and local asymptotic power* $\pi_m(h) = \Phi(\sqrt{gI}\, h - z_\alpha)$*, against the optimal score test's* $\pi^*(h) = \Phi(\sqrt{I}\, h - z_\alpha)$*; its Pitman relative efficiency to the score test is exactly* $g = \kappa_m(S; p_0)$*.*

Hence, in the contiguous limit*, both branches are read by the same efficacy and a single g is the order-m relative efficiency of each, with efficiency gap* $1 - g$ *(sample-size inflation* $1/g$*).*

Proof. For centered ψ, quadratic-mean differentiability gives $\frac{d}{d\theta}\mathbb{E}_\theta[\psi]\big|_0 = \int \psi(x)(-p'(x))\,dx = \langle \psi, S\rangle_{p_0}$, equal to $\mathrm{Cov}_{p_0}(\psi, S)$ since $\mathbb{E}_{p_0}S = 0$. *(ii)* Under $\theta_n = h/\sqrt{n}$, Le Cam's third lemma [21] gives $U_n := n^{-1/2}\sum_i \psi(X_i) \Rightarrow \mathcal{N}(\langle \psi, S\rangle h, \|\psi\|^2)$ (the drift is the asymptotic covariance with the log-likelihood ratio), so the test rejecting $U_n/\|\psi\| > z_\alpha$ has local power $\Phi(\langle \psi, S\rangle h/\|\psi\| - z_\alpha)$ and efficacy $\langle \psi, S\rangle^2/\|\psi\|^2$. For $\psi \in V_m$, $\langle \psi, S\rangle = \langle \psi, P_{V_m}S\rangle$, so Cauchy–Schwarz maximizes the efficacy at $\psi \propto P_{V_m}S$ with value $\|P_{V_m}S\|^2 = gI$; $\psi = S$ gives I; the ratio is g. *(i)* The estimating-function (Godambe) information is the same functional $\langle \psi, S\rangle^2/\|\psi\|^2$ via the Bartlett identity $\mathbb{E}_\theta[\partial_\theta \psi] = -\langle \psi, S\rangle$, so the same maximizer gives the efficiency statement—Proposition 1 read as a relative efficiency. □

Part (i) restates Proposition 1 as an efficiency; the new content is the testing efficacy (ii) and that both branches reduce to the *same* optimization $\max_{\psi \in V_m}\langle \psi, S\rangle^2/\|\psi\|^2$. This is the local face of Proposition 2: the "single

g for both branches" is a contiguous-limit statement (both read the score there) and does not contradict the off-limit fact that the branch readings are in general different numbers (§4).

Corollary 3 (Heavy-tailed power the orthogonal basis cannot state)**.** *For a symmetric heavy-tailed location family of tail index α (Student-t_ν, $\alpha = \nu$; Cauchy, $\alpha = 1$), the admissible fractional test of Proposition 3 has efficacy $g(m)\, I$ with $g(m) \uparrow 1$ as the exponents densify in $(0, \alpha/2)$ (Theorem 3) and local power $\Phi(\sqrt{g(m)I}\, h - z_\alpha)$. The Hermite/polynomial smooth test of the same order has* undefined *Pitman efficacy on these laws: its coefficients require $\mathbb{E}[x^{2k}]$, which diverge for $2k \geq \alpha$, so the orthogonal system is not in $L^2(\nu)$. The fractional efficacy already reaches $g \approx 0.96$ (t_3) and $g \approx 0.75$ (Cauchy) at low order (§6), where the Hermite Gram entry has diverged; these are low-order values, not ceilings (Corollary 4).*

Remark (error exponent). For simple $H_0 = p_0$ versus $H_1 = p_\delta$ the optimal exponent is the Chernoff information $C = -\min_{0 \leq s \leq 1} \log \int p_0^{1-s} p_\delta^s$, with local expansion $C(p_0, p_\delta) = (\delta^2/8)I + o(\delta^2)$ [8, 9]. A linear test $\sum_i \psi(X_i)$, $\psi \in V_m$, has local error exponent $(\delta^2/8)\langle \psi, S\rangle^2 / \|\psi\|^2$, maximized over V_m at $(\delta^2/8)\, gI$; hence the *local* error-exponent ratio is again $g = \kappa_m(S)$. For fixed alternatives C is a cumulant-generating (non-L^2) functional of Λ_δ, so the captured fraction $\kappa_m(\Lambda_\delta)$ does not determine it; a sharp fixed-alternative exponent is left to future work.

5.5 Completeness: the basis switch

Whether the truncated series can reach *any* target as $m \to \infty$—the question left open in §3—has a clean answer that depends on the tail, and that dependence is the precise reason for the non-orthogonal basis.

Proposition 4 (Completeness under a light tail)**.** *If ν has a finite exponential moment, $\int e^{c|x|}\, d\nu < \infty$ for some $c > 0$, then polynomials are dense in $L^2_0(\nu)$; since the integer-power subsystem spans them ($|x|^{2k} = x^{2k}$, $\operatorname{sgn}(x)|x|^{2k+1} = x^{2k+1}$, with the constant), $\kappa_m(z; \nu) \uparrow 1$ as $m \to \infty$ for every $z \in L^2_0(\nu)$.*

A finite exponential moment forces Carleman's condition and polynomial density [1, 5]; this is the regime (Gaussian, Laplace, sub-exponential) in which the classical orthogonal systems already apply.

Theorem 3 (Completeness of the admissible fractional system)**.** *Let ν be symmetric on $\mathbb{R}$ with moments finite up to order $< 2s_0$ for some $s_0 > 0$*

(a regularly varying tail of index α gives $s_0 = \alpha/2$). Let $A \subset (0, s_0)$ index the admissible elements ($|x|^a, \operatorname{sgn}(x)|x|^a \in L^2(\nu) \iff a < s_0$). If A has an accumulation point in the open *interval $(0, s_0)$—in particular if A is dense in $(0, s_0)$—then $\operatorname{span}\{|x|^a, \operatorname{sgn}(x)|x|^a, 1 : a \in A\}$ is dense in $L^2(\nu)$, so $\kappa_m(z; \nu) \uparrow 1$ as $m \to \infty$ for every $z \in L^2_0(\nu)$. Completeness can fail only when A accumulates solely at the boundary s_0.*

Proof. By symmetry the even and odd subspaces of $L^2(\nu)$ are orthogonal; treat each. Let $g \in L^2(\nu)$ be even with $\langle g, |x|^a \rangle_\nu = 0$ for all $a \in A$. Write $t = |x|$, let μ be the law of $|X|$ and $G(t) = g$, and set $M(a) = \int_0^\infty G(t)\, t^a \, d\mu(t)$. For $\sigma = \operatorname{Re} a \in (0, s_0)$, Cauchy–Schwarz gives $\int |G| t^\sigma \, d\mu \le \|G\|_{L^2(\mu)} (\mathbb{E}|X|^{2\sigma})^{1/2} < \infty$, and the $|\log t|$-dominated derivative is integrable on compact substrips, so M is holomorphic on the strip $\Sigma = \{0 < \operatorname{Re} a < s_0\}$. Since $M \equiv 0$ on A, which has an accumulation point in Σ, the identity theorem gives $M \equiv 0$ on Σ. Fix $\sigma_0 \in (0, s_0)$: then $\tau \mapsto M(\sigma_0 + i\tau) = \int_0^\infty G(t) t^{\sigma_0} e^{i\tau \log t} \, d\mu(t)$ is the Fourier transform of the finite complex measure obtained by pushing $G(t) t^{\sigma_0} d\mu(t)$ forward under $t \mapsto \log t$; its vanishing forces, by Fourier uniqueness, $G(t) t^{\sigma_0} = 0$ μ-almost everywhere, hence $g = 0$. The odd part is identical with $\operatorname{sgn}(x)|x|^a$. □

The hypothesis is mild: any symmetric law with some finite moments and exponents accumulating inside $(0, s_0)$—for instance the design "a_i dense in $(0, \hat{\alpha}/2)$" with $\hat{\alpha}$ an estimated tail index (§6). For non-symmetric ν the same Mellin argument runs on each half-line using $|x|^a$ and $\operatorname{sgn}(x)|x|^a$ together; Müntz–Szász approximation [6] is the bounded-interval shadow of this statement. The result is not a corollary of the classical weighted Müntz–Szász theorems: those characterize density of $\{|x|^a\}_{a \in A}$ in L^p of a *fixed* weight on a bounded interval or half-line under a Blaschke/gap condition on A [6], whereas here the weight is the unknown law ν itself, the admissible range $(0, s_0)$ is fixed by its tail index $\alpha = 2s_0$, and what is characterized is closure of the *centered* system in $L^2_0(\nu)$. The statistical content—that the basis regains the L^2 closure precisely on the heavy tails where the Hermite/polynomial system leaves the space—is what the bounded-interval theorems do not address.

Corollary 4 (Completeness by basis switch; score capture)**.** *Completeness holds in both tail regimes, by a change of basis: integer powers under a light tail (Proposition 4), bounded accumulating fractional powers under a heavy tail (Theorem 3)—exactly where the polynomial/Hermite system leaves L^2. In particular, since $S, \Lambda \in L^2_0(\nu)$, $\kappa_m(S) \uparrow 1$ and $\kappa_m(\Lambda) \uparrow 1$, so the low-order heavy-tailed values (Corollary 3) are not ceilings. By Proposition 3*

the operationally decisive quantity is the low-order $g = \kappa_m(S)$*, already high at orders where the orthogonal Gram entries diverge.*

5.6 Decomposition

Lemma 2 (Decomposition leg)**.** *Define the decomposition reading consistently with Definition 2 as* $R^2 = \kappa_m(\log p; \nu) = 1 - \mathrm{MSED}_m / \mathrm{Var}_\nu(\log p)$*, where* MSED_m *is the affine (intercept-included) order-m reconstruction error of the centered log-density. For the Gaussian-in-basis model the between-class reconstruction-energy difference equals the log-ratio up to the log-determinant offset,* $\mathrm{logMSED}_0 - \mathrm{logMSED}_1 = \frac{1}{2}D_0 - \frac{1}{2}D_1 = \Lambda + \frac{1}{2}(\log|F_1| - \log|F_0|)$*, hence* $= \Lambda$ *when* $|F_0| = |F_1|$*.*

Here $\mathrm{logMSED}_c$ is the unnormalized reconstruction energy $\frac{1}{2}D_c$ up to a class-independent constant (the per-class normalizer $\frac{1}{2}\log|F_c|$ is not part of it). Part two is quadratic discriminant analysis [3]; the decomposition target is the log-density $\log p$, distinct from the log-ratio Λ (which is the detection object), so R^2 does not collapse to c.

6 Numerical experiments

The framework predicts: on a family whose carrier is captured, all readings equal one at finite order; off it, all are below one and rise with the basis; the estimation and detection readings coincide in the small-gap limit; and the fractional element captures information on heavy tails where a polynomial/Hermite basis is not square-integrable. We test these in a controlled study with all thresholds fixed before computation (global seed 2026; reproduction code in the supplement, §10). Fifteen checks all hold: the ten unification checks (A1, A2, T1–T4, A3, D1, H1, L1), collected in Table 2; and five operating-characteristic checks (U1, P1, P2, S1, C1), reported with Monte-Carlo standard errors in Table 3 and Figure 3.

Termination on the Gaussian. The score $S = x$ is linear, so the order-one estimation efficiency is $g = 1.000$ (A1); the detection branch is exact, $c = 1.000$, with the QDA discriminant equal to the true log-ratio to machine precision (A2, the empirical face of Theorem 2). The decomposition R^2 of the log-density is 0.000 at order two ($\{1, x\}$, which misses the quadratic) and 1.000 at order three ($\{1, x, x^2\}$): the decomposition terminates exactly when x^2 enters (D1), as Theorem 1(c) requires.

Joint rise on the Laplace. With the fractional basis $\mathrm{sgn}(x)|x|^a$ the estimation efficiency rises $0.50 \to 0.92 \to 0.96 \to 0.98$, and the model-based and empirical efficiencies agree to 0.014 (T1, T2). The detection captured-ratio rises $0.71 \to 0.94$ (T3). The decomposition R^2 of the log-density rises $0.00 \to 0.80 \to 1.00$, terminating once $|x|$ enters the basis—because the Laplace log-density is exactly affine in $|x|$ (D1). The estimation jump from 0.50 to 0.92 occurs precisely when $|x|^{0.5}$ is added (the Laplace score has a kink a monotone polynomial cannot capture), and detection stays at 0.71 while the basis is smooth and jumps only when $|x|$ enters: what matters is the *choice* of non-orthogonal element, not the degree.

The covariance-shift anchor. On $\mathcal{N}(0,1)$ vs $\mathcal{N}(0,2.25)$ the order-one projection captures essentially nothing ($c = 10^{-6}$: an even signal is invisible to an odd basis, the blindness predicted by Corollary 2) and the order-two projection is exact ($c = 1.000$). The Mahalanobis-affine head equals the log-ratio to 2.7×10^{-15}, the projected statistic's sign agrees with Bayes on all 4×10^6 samples, and the change-point normal system yields a statistic with correlation 1.000000000 to the log-ratio (A3). The sequential information ratio $J(s)/\|z\|^2$ equals c to 1.2×10^{-13} across all eight family×order cells, saturating at one from order three on the Gaussian (T4).

The local equality. On the Laplace, the detection reading converges to the estimation reading as the hypothesis gap vanishes: with $g = \kappa_1(S; p_0) = 0.4995$, the detection $c(\delta) = \kappa_2(\Lambda_\delta; p_0)$ runs $0.588, 0.541, 0.522, 0.511, 0.504$ for $\delta = 0.5, 0.2, 0.1, 0.05, 0.02$, i.e. $|c - g|$ shrinks from 0.088 to 0.005 (L1)—Proposition 2 numerically.

Heavy tails. On Student-t_3, where $\mathbb{E}[x^4] = \infty$ and the monomial Gram entry $\mathbb{E}[x^4]$ is empirically 5,152 (divergent in population, so $\{x, x^3\}$ is not L^2), the admissible fractional basis ($2 \max_i a_i < 3$) yields estimation efficiency rising $0.88 \to 0.96$; on the Cauchy law, where no polynomial moment exists, the fractional basis with $2 \max_i a_i < 1$ still captures $0.60 \to 0.75$ of the Fisher information (H1). This is the regime the fractional element is designed for, and the one a Hermite expansion cannot enter.

Operating characteristics and completeness. Five further checks, thresholds fixed in advance, quantify the new theory (Figure 3, Table 3). *Uncertainty (U1):* over $R = 30$ independent seeds the headline Laplace readings are $g = 0.977 \pm 0.0001$, $c = 0.940 \pm 0.001$, $R^2 = 1.000 \pm 0.000$

Table 2: Controlled study; the ten unification checks, all passing. On the Gaussian the series terminates (all readings reach one); on the Laplace they rise together; the sequential information ratio tracks c; the detection reading converges to the estimation reading as the gap vanishes; the fractional element captures Fisher information on heavy tails where a polynomial/Hermite basis is not square-integrable.

Family	Check	Quantity	Value
Gaussian	A1	g (order-1 $\{x\}$)	1.000
Gaussian	A2	$\lvert\text{QDA} - \Lambda\rvert_{\max}$	0.0
Gaussian	D1	$R^2(\log p)$, order $2 \to 3$	$0.00 \to 1.00$
Laplace	T1	$g(m)$ (fractional)	$0.50 \to 0.98$
Laplace	T2	$\lvert g_{\text{model}} - g_{\text{emp}}\rvert$	0.014
Laplace	T3	$c(m)$ (detection)	$0.71 \to 0.94$
Laplace	D1	$R^2(\log p)$ $(m = 2, 3, 4)$	$0.00, 0.80, 1.00$
Laplace	L1	$c(\delta) \to g$ as $\delta \downarrow 0$	$\lvert c - g\rvert : 0.088 \to 0.005$
Gauss. cov-shift	A3	c, order $1 \to 2$	$10^{-6} \to 1.00$
Gauss. cov-shift	A3	$\lvert\text{Ghosh} - \Lambda\rvert_{\max}$	2.7×10^{-15}
both	T4	$\max\lvert c - J/\lVert z\rVert^2\rvert$	1.2×10^{-13}
t_3 / Cauchy	H1	fractional g rises	$0.88 \to 0.96$ / $0.60 \to 0.75$
t_3	H1	monomial $\mathbb{E}[x^4]$ (diverges)	$\approx 5{,}152$

(mean $\pm$ standard error, $n = 2 \times 10^5$ each)—the point estimates are tight. *Power (P1):* the order-m projected-score test, size calibrated to 0.05, has power rising $0.59 \to 0.80$ on the Laplace at a fixed local shift, matching the prediction $\Phi(\sqrt{gI}\,h - z_\alpha)$ of Proposition 3 to within 0.03; the Student-t_3 test behaves likewise, in a regime where the Hermite test has no defined efficacy (Corollary 3). *Efficiency (P2):* the one-step estimator's RMSE falls and its empirical efficiency rises $0.49 \to 0.96$, tracking the analytic g to within 0.07. *Selection (S1):* the held-out captured fraction selects the order ($m^\star = 6$ here), and a Hill tail-index estimate ($\hat{\alpha} \approx 2.5$ on t_3) sets the admissible exponent ceiling $\hat{\alpha}/2 \approx 1.23$, giving a data-driven basis with $g = 0.99$. *Completeness (C1):* for a generic smooth target the captured fraction $\kappa_m(z)$ climbs to 0.999 (Laplace) and 0.980 (t_3, exponents confined to $(0, 3/2)$) as the fractional exponents densify—the falsifiable signature of Theorem 3.

7 The invariance corollary

By Definition 1 the series depends on x only through the basis evaluation $\varphi(x) = T(x)$. Hence any transformation $x \mapsto gx$ with $\varphi(gx) = \varphi(x)$—any g

Table 3: Operating-characteristic and completeness checks (means with Monte-Carlo standard errors where applicable). Power and efficiency rise with order and track the captured fraction; the held-out and tail-index rules select the basis; $\kappa_m(z) \to 1$ as the fractional exponents densify.

Check	Quantity	Result
U1	Laplace $g\pm$SE (R=30, n=2×10^5)	0.977 ± 0.0001
U1	Laplace $c\pm$SE	0.940 ± 0.001
P1	Laplace power, order $1 \to 4$ (vs. theory)	$0.59 \to 0.80$ ($\lvert\Delta\rvert < 0.03$)
P2	Laplace efficiency, order $1 \to 4$	$0.49 \to 0.96$
S1	held-out κ selects order	$m^\star = 6$
S1	t_3 Hill $\hat{\alpha} \to$ ceiling $\hat{\alpha}/2$	$2.5 \to 1.23$
C1	$\kappa_m(z)$ densified (Laplace / t_3)	0.999 / 0.980

in the stabilizer Stab(T)—leaves the score, the log-ratio, the change-point statistic, and the reconstruction error simultaneously unchanged. Precisely: if $\varphi(gx) = \varphi(x)$ for all x, all branch readings are invariant under g; equivalently, all branches are blind to Stab(T). This is the *reduction principle*, here read generatively: the invariant behavior is a consequence of the single series factoring through T. That generative reason is the present contribution, distinct from—and complementary to—a characterizational treatment of the same invariance, which we do not develop here. When the blindness is to a direction rather than a group action, the statement weakens from exact invariance to projection blindness (§9).

A nontrivial instance. The Gaussian covariance-shift anchor (Corollary 2) is a case where the reduction is *not* recoverable by applying sufficiency to the mean. The first-order rung $\{x\}$—the natural location statistic—is provably blind to a pure scale change ($c \to 0$ as the means coincide), while the order-two rung $\{x, x^2\}$ captures the exact log-ratio ($c = 1$). The captured-fraction criterion therefore selects the two-dimensional carrier $\{x, x^2\}$ as the minimal sufficient reduction, a choice a naive mean-based sufficiency argument misses entirely; §6 (check A3) exhibits the $c(S{=}1) \approx 0 \to c(S{=}2) = 1$ jump. This answers the natural objection that the corollary is a restatement of feature invariance: here the invariance is to a scale group the first moment cannot see.

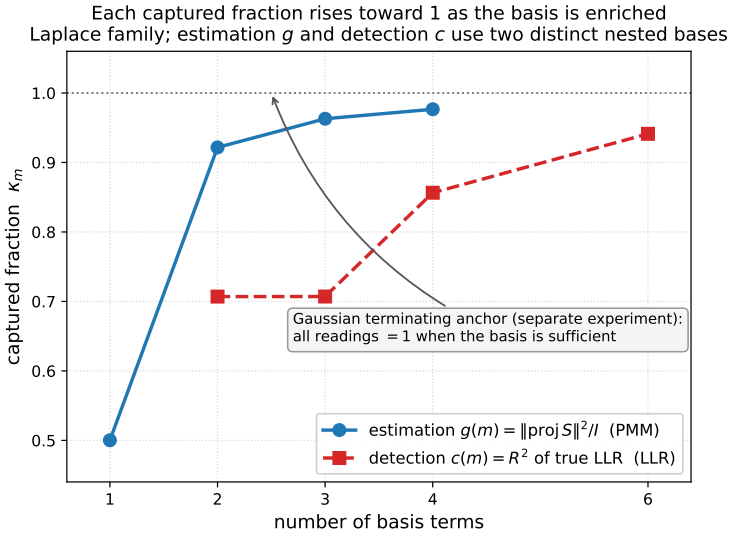


Figure 1: Laplace family: the estimation efficiency g (captured Fisher fraction, fractional basis) and the detection captured-ratio c each rise toward one as the basis is enriched (two distinct nested bases; the abscissa is the number of basis terms, not a shared index). The dotted line marks the Gaussian terminating anchor, where both equal one (a separate experiment).

8 Consequences

Organizing the tasks around one series turns previously documented results into consequences.

When decomposition cannot beat moment statistics. On a profile where the chosen basis is already a sufficient statistic, Theorem 1 makes the series terminate and the second-order reconstruction exact; there is then no residual for a learned decomposition to exploit, so it cannot beat closed-form moment statistics at an equal feature budget. In this framework that equal-budget parity is a theorem about a terminated series, not a surprise.

The crossover as truncation error. The sample size at which a decomposition overtakes a moment baseline is the finite-sample bias–variance profile of the truncation residual R_m: at small n the variance of the estimated weights dominates and decomposition loses; at large n the truncation

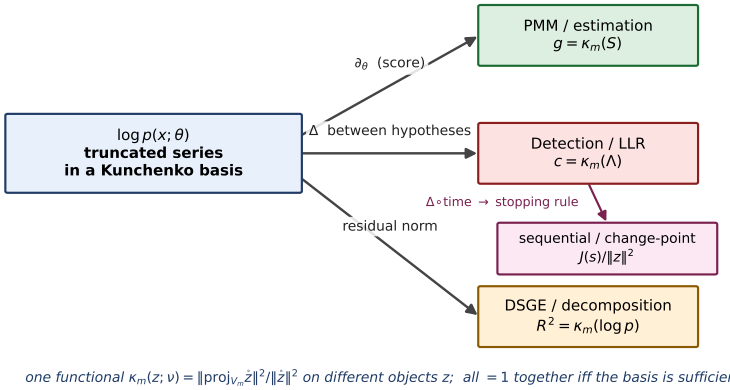


Figure 2: The master object: one truncated log-likelihood series, read by the operations ∂_θ (estimation, reading g), Δ between hypotheses (detection, c; its temporal composition $\Delta \circ \mathrm{time}$ feeding a stopping rule is the change-point line, $J(s)/\|z\|^2$), and the residual norm (decomposition, R^2). Each reading is the same functional κ_m on a different object.

bias dominates and it wins. These are retrospective consistencies, not independent predictions; we state them as such.

9 Limitations

Definitional core vs. falsifiable content. That the score is $\partial_\theta \log p$, the log-ratio is $\Delta \log p$, and the reconstruction error is a residual is definitional. The falsifiable content is that the *same* basis at the *same* order makes all branches exact *simultaneously* on sufficiency (Theorem 1) and rise together off it (Lemma 1, §6), and that the estimation and detection readings coincide in the small-gap limit (Proposition 2). We claim no numerical equality of the readings off the anchor.

Scope of the machine-checked core. The Lean core certifies the algebraic identities on the Gaussian terminating anchor only (§5); the off-family, non-orthogonal, and general-basis results rest on the pen-and-paper proofs and the numerical study, not on Lean.

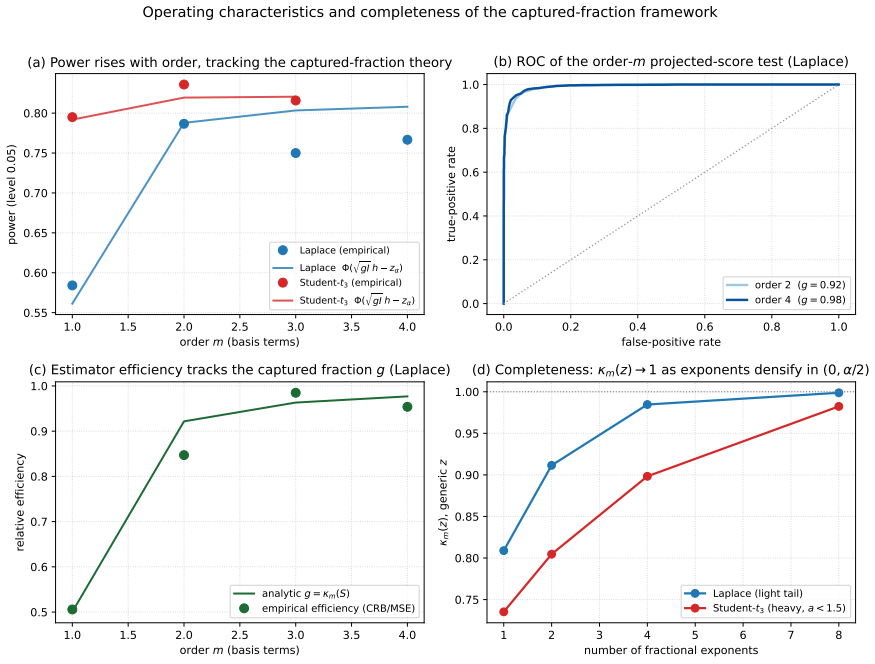


Figure 3: Operating characteristics and completeness. (a) Power of the order-m projected-score test (size 0.05) rises with order and tracks $\Phi(\sqrt{gI}\,h - z_\alpha)$ (Proposition 3); (b) its ROC improves with order; (c) the estimator's empirical efficiency tracks the analytic $g = \kappa_m(S)$; (d) for a generic target $\kappa_m(z) \to 1$ as the fractional exponents densify in $(0, \alpha/2)$, on both a light (Laplace) and a heavy (Student-t_3) tail—the signature of Theorem 3.

Convergence of the non-orthogonal series. We work at finite-m L^2 truncation, with the explicit admissibility $2\max_i a_i < \alpha$. Completeness as $m \to \infty$ is settled for symmetric laws by Theorem 3: $\overline{\bigcup_m V_m} = L^2_0(\nu)$, hence $\kappa_m \to 1$, whenever the exponents accumulate inside $(0, \alpha/2)$. What remains open is the *rate* of $\kappa_m \to 1$, the non-symmetric case (the half-line argument needs the two parities jointly), and the degenerate boundary-accumulation case that the theorem excludes.

Synthetic-only verification; subtype without a group. The numerical evidence is a controlled synthetic study plus the machine-checked anchor; a real-data illustration is optional and not included, and the theoretical claim does not depend on it. The invariance corollary is group-exact only when $\mathrm{Stab}(T)$ is a group; otherwise it weakens to projection blindness. Off

the terminating anchor the sequential statistic is a quasi-log-ratio whose run-length and detection-delay theory under truncation, and the effect of plug-in moments, are inherited open questions of the change-point line.

10 Conclusion

Reading estimation, detection, and decomposition as ∂/Δ/residual of one truncated likelihood series turns three separate quality measures into one captured-fraction functional on three objects. The picture is deliberately modest: off the terminating anchor the readings stay distinct numbers and agree only there, or—for estimation and detection—in the small-gap limit. What carries weight beyond this bookkeeping is the non-orthogonal fractional basis, which restores L^2 completeness on heavy tails where the Hermite system leaves the space (Theorem 3); the Gaussian covariance-shift anchor, exact at order two and machine-checked, on which the classical change-point optimality transfers verbatim (Corollary 2); and the reduction principle, recovered as an invariance corollary. A controlled synthetic study confirms the same branch-wise picture across the Gaussian, Laplace, Student-t, and Cauchy examples (§6). The existence of likelihood series themselves remains classical; the contribution is their conjunction under one functional.

Code and data availability

All data are synthetic Monte-Carlo samples generated in code (global seed 2026); no external dataset is used. The full reproduction package—the verification script (fifteen pre-specified checks against recorded output), the figure script, and a standalone Lean 4/Mathlib project for the machine-checked core—is openly available at `https://github.com/SZabolotnii/Ku-LSU-code-supplement`.

A Machine-checked lemmas and reproducibility

The Lean core (`LikelihoodSeries.lean`, Mathlib v4.26.0; a clean `lake build` with zero `sorry`) mechanizes, for Theorem 2 and Corollary 2, the following lemmas:

- `llr_eq_score_difference` (and its equal-log-determinant corollary) — the bridge $\Lambda = \Delta(\frac{1}{2}D^2)$;

- `argmax_logDensity_iff_argmin_penalty` — the decision equivalence;
- `gauss_llr_eq_quadratic` and `gauss_llr_mem_span_quadratics` — the degree-two structure of the Gaussian log-ratio;
- `gauss_llr_eq_mahala_affine` — the Mahalanobis-head (Ghosh) form;
- `gauss_bayes_iff_min_penalty` and `llr_nonneg_iff_penalty_le` — the Bayes-decision sign;
- `sum_inner_sq_eq_norm_sq_of_mem_span` and `J_saturates` — Parseval saturation on a finite span;
- `gaussLogDensity_eq_logDensity` — the concrete scalar Gaussian instantiates the abstract model, so the hypotheses are satisfied rather than assumed.

These cover the Gaussian anchor only. The figures recompute every plotted number from the verification script; no result is hardcoded.

Declarations

Funding. No funding was received for this work.
Competing interests. The author declares that he has no competing interests.
Data availability. This is a theoretical study; the accompanying controlled study uses only synthetic data generated by the publicly available code (see Code availability).
Code availability. Code reproducing all reported checks and figures is openly available at `https://github.com/SZabolotnii/Ku-LSU-code-supplement`.
Author contributions. Serhii Zabolotnii is the sole author.

References

[1] N. I. Akhiezer. *The Classical Moment Problem and Some Related Questions in Analysis*. Oliver and Boyd, Edinburgh, 1965. Translated by N. Kemmer; Dover reprint 2020, ISBN 978-0-486-84555-5.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society and Oxford University Press, 2000.

[3] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, Hoboken, NJ, 3 edition, 2003.

[4] A. R. Barron and C.-H. Sheu. Approximation of density functions by sequences of exponential families. *The Annals of Statistics*, 19(3):1347–1369, 1991. doi:10.1214/aos/1176348252.

[5] Christian Berg and J. P. Reus Christensen. Density questions in the classical theory of moments. *Annales de l'Institut Fourier*, 31(3):99–114, 1981. doi:10.5802/aif.840.

[6] Peter Borwein and Tamás Erdélyi. *Polynomials and Polynomial Inequalities*, volume 161 of *Graduate Texts in Mathematics*. Springer, New York, 1995. ISBN 978-0-387-94509-5. doi:10.1007/978-1-4612-0793-1.

[7] L. D. Brown. *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, volume 9 of *IMS Lecture Notes — Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1986. doi:10.1214/lnms/1215466757.

[8] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952. doi:10.1214/aoms/1177729330.

[9] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, Hoboken, NJ, 2 edition, 2006. ISBN 978-0-471-24195-9. doi:10.1002/047174882X.

[10] I. Csiszár. i-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975. doi:10.1214/aop/1176996454.

[11] Fabing Duan, François Chapeau-Blondeau, and Derek Abbott. Fisher information as a metric of locally optimal processing and stochastic resonance. *PLoS ONE*, 7(4):e34282, 2012. doi:10.1371/journal.pone.0034282.

[12] A. Feuerverger and P. McDunnough. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B*, 43(1):20–27, 1981. doi:10.1111/j.2517-6161.1981.tb01142.x.

[13] Annesha Ghosh, Anil K. Ghosh, Rita SahaRay, and Soham Sarkar. Classification using global and local Mahalanobis dis-

tances. *Journal of Multivariate Analysis*, 207:105417, 2025. doi:10.1016/j.jmva.2025.105417.

[14] V. P. Godambe. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4):1208–1211, 1960. doi:10.1214/aoms/1177705693.

[15] V. P. Godambe, editor. *Estimating Functions*. Oxford University Press, Oxford, 1991. ISBN 9780198522287. doi:10.1093/oso/9780198522287.001.0001.

[16] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, 1986. doi:10.1002/9781118186435.

[17] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982. doi:10.2307/1912775.

[18] Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi:10.1214/aoms/1177703732.

[19] A. Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6:695–709, 2005. URL `https://jmlr.org/papers/v6/hyvarinen05a.html`.

[20] Yu. P. Kunchenko. *Polynomial Parameter Estimations of Close to Gaussian Random Variables*. Shaker Verlag, Aachen, 2002.

[21] Lucien Le Cam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3:37–98, 1960. No DOI (pre-DOI university serial).

[22] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, New York, 2 edition, 1998.

[23] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971. doi:10.1214/aoms/1177693055.

[24] D. G. Luenberger. *Optimization by Vector Space Methods*. Wiley, New York, 1969.

[25] P. McCullagh. *Tensor Methods in Statistics*. Number 29 in Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1987.

[26] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387, 1986. doi:10.1214/aos/1176350164.

[27] J. Neyman. "smooth test" for goodness of fit. *Scandinavian Actuarial Journal*, 1937(3–4):149–199, 1937. doi:10.1080/03461238.1937.10404821.

[28] Yakov Nikitin. *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press, Cambridge, 1995. ISBN 978-0-521-47029-2. doi:10.1017/CBO9780511530081.

[29] Gottfried E. Noether. On a theorem of pitman. *The Annals of Mathematical Statistics*, 26(1):64–68, 1955. doi:10.1214/aoms/1177728593.

[30] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954. doi:10.1093/biomet/41.1-2.100.

[31] Moshe Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, 13(1):206–227, 1985. doi:10.1214/aos/1176346587.

[32] J. C. W. Rayner and D. J. Best. *Smooth Tests of Goodness of Fit*. Oxford University Press, New York, 1989.

[33] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998. doi:10.1017/CBO9780511802256.

[34] Serhii Zabolotnii. Generalized stochastic approximation of the log-likelihood ratio for robust sequential change-point detection, 2026. Information functional $J(s)$; Lean `InfoFunctional.lean`. doi:10.48550/arXiv.2605.23419.

[35] Serhii Zabolotnii. Efficient frequentist fractional polynomials for skewed dose–response and survival data: a variance-reducing alternative to OLS-FP, 2026. doi:10.48550/arXiv.2605.16846.

[36] Serhii Zabolotnii. Parametrically adaptive transition polynomial: a signed-parity continuous-α extension of Kunchenko stochastic polynomials, 2026. doi:10.48550/arXiv.2605.14610.

[37] Serhii Zabolotnii. From Volterra series to Kunchenko stochastic polynomials: Half a century of non-Gaussian estimation methodology, 2026. Survey of the Kunchenko school. doi:10.48550/arXiv.2605.22354.