Pseudovariance via Forma Potentials, Laplace–Parametric Multimodal Densities, and Polynomial Universality over Gaussian Mixtures

Christopher Lee Burgess

Abstract

Classical variance is an expected quadratic energy and therefore privileges unimodal central tendency. We define *pseudovariance* as the expected value of a nonnegative *forma* potential, and we introduce a companion moment calculus based on iterated antiderivatives that recovers classical central moments as a special case. We then construct a Laplace–parametric family of analytic densities of Gibbs type $p(\tau) \propto \exp(-\phi(\tau, c)/(2\Sigma))$ and derive a closed-form normalization via Laplace approximation about a dominant global mode. Finally, we prove a non-tautological universality theorem: restricting ϕ to the class of *coercive polynomials* (degree allowed to grow), the induced Gibbs densities are dense (in L^1 / total variation) in the class of finite weighted Gaussian mixtures. The proof uses compact approximation of log-densities by polynomials plus tail control via high-degree coercive terms.

1 Axioms and motivating viewpoint

Axiom 1 (Dispersion as expected energy). Any "dispersion-like" functional can be viewed as an expectation of a nonnegative energy/loss.

Axiom 2 (Multimodality is geometric). Multimodality is most naturally represented by an energy landscape with multiple wells (local minima).

Axiom 3 (Publishable generalization requires constraints). If the forma class is unrestricted, universality becomes tautological (e.g., by setting $\phi = -2\Sigma \log p$). A meaningful theory therefore restricts ϕ to a structured family and proves approximation/stability under that restriction.

2 Pseudovariance and pseudocovariance

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ measurable.

Definition 1 (Forma)**.** *A* forma *is a measurable function* $\phi : \mathbb{R} \to [0, \infty]$.

Definition 2 (Pseudovariance)**.** *If* $\mathbb{E}[\phi(X)] < \infty$*, define the* pseudovariance *of* X *w.r.t.* ϕ *by*

$$\Sigma_\phi(X) \;:=\; \mathrm{VAR}_\phi(X) \;:=\; \mathbb{E}[\phi(X)].$$

Interpretation. Σ_ϕ is the expected "shape energy" under the potential ϕ. Choosing ϕ with multiple wells yields a dispersion functional sensitive to modality and geometry, not merely spread around one mean.

Definition 3 (Pseudocovariance)**.** *Let* (X, Y) *be a pair of real random variables and let* $\phi : \mathbb{R}^2 \to [0, \infty]$ *be measurable with* $\mathbb{E}[\phi(X, Y)] < \infty$*. Define*

$$\mathrm{COV}_\phi(X, Y) \;:=\; \mathbb{E}[\phi(X, Y)].$$

3 Pseudomeans and a moment calculus

3.1 Pseudomeans

To represent "modal centers" induced by a shape potential, we use minima of ϕ.

Assumption 1 (Smoothness for pseudomeans). *In this section assume $\phi \in C^2(\mathbb{R})$.*

Definition 4 (Local minima set). *Define*

$$C(\phi) \ := \ \{\tau \in \mathbb{R} : \ \phi'(\tau) = 0, \ \phi''(\tau) > 0\}.$$

Definition 5 (Pseudomean). *Any $\mu \in C(\phi)$ is called a* pseudomean *(modal center) induced by ϕ.*

Remark 1. *This is a* shape-first *notion: pseudomeans depend on ϕ. In data-driven use one typically introduces a parameterized family $\phi(\cdot; c)$ and fits c from data; then pseudomeans of $\phi(\cdot; c)$ become distribution-adaptive.*

3.2 Pseudomoments and recovery of classical central moments

Assumption 2 (Antiderivative convention). *Assume ϕ is smooth enough that iterated antiderivatives exist. For an integer $m \geq 0$, write $\phi^{(-m)}$ for the m-fold antiderivative with all integration constants set to 0 (e.g., by requiring $\phi^{(-m)}(0) = \phi^{(-m)\prime}(0) = \cdots = \phi^{(-m)(m-1)}(0) = 0$).*

Definition 6 (Pseudomoments). *For integer $k \geq 1$ such that $\mathbb{E}|\phi^{(2-k)}(X)| < \infty$, define*

$$\mu_k^{(\phi)} \ := \ \frac{k!}{2}\, \mathbb{E}\big[\phi^{(2-k)}(X)\big].$$

Lemma 1 (Classical central moments are recovered by the quadratic forma). *Let $\phi(x) = (x-\mu)^2$ with $\mu = \mathbb{E}[X]$ and $\mathbb{E}[|X|^k] < \infty$. Then for $k = 3$,*

$$\mu_3^{(\phi)} = \mathbb{E}[(X-\mu)^3].$$

Proof. With the stated convention, $\phi^{(-1)}(x) = \frac{1}{3}(x-\mu)^3$. Thus $\mu_3^{(\phi)} = \frac{3!}{2}\mathbb{E}[\phi^{(-1)}(X)] = \frac{6}{2}\mathbb{E}[\frac{1}{3}(X-\mu)^3] = \mathbb{E}[(X-\mu)^3]$. □

Remark 2. *Analogous computations recover higher classical central moments (subject to the antiderivative convention). For nonquadratic ϕ, the resulting pseudomoments define shape-sensitive analogues.*

4 Laplace–parametric analytic densities from forma

4.1 Energy-based (Gibbs) family

Fix parameters c and scale $\Sigma > 0$. Consider the Gibbs/Laplace template

$$p(\tau \mid c, \Sigma) \ := \ \frac{1}{Z(c,\Sigma)} \exp\Big(-\frac{\phi(\tau,c)}{2\Sigma}\Big), \qquad Z(c,\Sigma) = \int_{\mathbb{R}} \exp\Big(-\frac{\phi(\tau,c)}{2\Sigma}\Big)\, d\tau.$$

This is analytic in τ when $\phi(\cdot, c)$ is analytic and coercive enough that $Z < \infty$. Multimodality is induced by multiple wells of $\phi(\cdot, c)$.

4.2 Laplace approximation about a dominant global mode

Assumption 3 (Laplace regime)**.** *Fix c and assume: (i) $\phi(\cdot, c) \in C^2(\mathbb{R})$ and $Z(c,\Sigma) < \infty$; (ii) $\phi(\cdot, c)$ has a* unique global minimizer *τ_0 with $\phi''(\tau_0, c) > 0$.*

Theorem 1 (Laplace normalization)**.** *Under the Laplace regime, as $\Sigma \downarrow 0$,*

$$Z(c, \Sigma) \;=\; \exp\Big(-\frac{\phi(\tau_0, c)}{2\Sigma}\Big)\sqrt{\frac{4\pi\Sigma}{\phi''(\tau_0, c)}}\,(1 + o(1)).$$

Consequently,

$$p(\tau \mid c, \Sigma) \;=\; \sqrt{\frac{\phi''(\tau_0, c)}{4\pi\Sigma}}\exp\Big(-\frac{\phi(\tau, c) - \phi(\tau_0, c)}{2\Sigma}\Big)\,(1 + o(1)),$$

uniformly on compact sets.

Proof. Standard Laplace method: Taylor expand $\phi(\tau, c)$ to second order at τ_0 and integrate the resulting Gaussian. □

Remark 3 ("Analytic multimodal approximation")**.** *The density above is* not *in general a Gaussian mixture. It is a single analytic exponential-of-potential whose multimodality is governed by the landscape of ϕ. Gaussian behavior appears only as the local quadratic surrogate near the dominant mode.*

5 A non-tautological density theorem: coercive polynomial forma

We now restrict the forma family in a meaningful way.

Definition 7 (Coercive polynomial forma class)**.** *Let Φ_{poly} be the class of functions $\phi : \mathbb{R} \to \mathbb{R}$ of the form*

$$\phi(\tau) = P(\tau) + \alpha\tau^{2m},$$

where P is a real polynomial, $m \in \mathbb{N}$ (degree $2m$), and $\alpha > 0$. Then $\phi(\tau) \to +\infty$ as $|\tau| \to \infty$, so $e^{-\phi(\tau)}$ is integrable.

Given $\phi \in \Phi_{\mathrm{poly}}$, define the induced density (set $\Sigma = 1/2$ for notational convenience)

$$q_\phi(\tau) \;:=\; \frac{e^{-\phi(\tau)}}{\int_{\mathbb{R}} e^{-\phi(t)}\,dt}.$$

(Any fixed $\Sigma > 0$ can be absorbed into ϕ by scaling.)

5.1 Target class: finite Gaussian mixtures

Let

$$g(\tau) = \sum_{j=1}^{K} w_j\,\mathcal{N}(\tau; \mu_j, \sigma_j^2), \qquad w_j > 0,\ \sum_j w_j = 1,\ \sigma_j > 0.$$

Such g is C^∞ and strictly positive on $\mathbb{R}$.

5.2 Two technical lemmas: tail domination and tail mass bounds

Lemma 2 (Polynomial lower bound)**.** *Let P be a real polynomial of degree $d \geq 0$. Then there exists a constant $C_P \geq 0$ such that*

$$P(t) \geq -C_P\,(1+|t|^d) \qquad \text{for all } t \in \mathbb{R}.$$

Proof. Write $P(t) = \sum_{k=0}^{d} a_k t^k$. Then $|P(t)| \leq \sum_{k=0}^{d} |a_k|\,|t|^k \leq \Big(\sum_{k=0}^{d} |a_k|\Big)(1+|t|^d)$ for all t (since $|t|^k \leq 1+|t|^d$ when $k \leq d$). Thus $P(t) \geq -C_P(1+|t|^d)$ with $C_P = \sum_{k=0}^{d} |a_k|$. □

Lemma 3 (Explicit tail bound for coercive polynomial Gibbs weights)**.** *Fix $m \in \mathbb{N}$, $\alpha > 0$, and a polynomial P of degree $d < 2m$. Let*

$$\phi(t) = P(t) + \alpha t^{2m}, \qquad W(t) = e^{-\phi(t)}.$$

Let C_P be as in the previous lemma. Define the tail threshold

$$R_0 := \max\left\{1,\ \left(\frac{2C_P}{\alpha}\right)^{\frac{1}{2m-d}}\right\}.$$

Then for all $R \geq R_0$ and all $t \geq R$,

$$W(t) \leq \exp(C_P)\,\exp\Big(-\frac{\alpha}{2}\,t^{2m}\Big).$$

Consequently, for all $R \geq R_0$,

$$\int_{|t|>R} W(t)\,dt \ \leq\ 2\,e^{C_P}\,\frac{1}{2m}\left(\frac{2}{\alpha}\right)^{\frac{1}{2m}}\Gamma\Big(\frac{1}{2m},\ \frac{\alpha}{2}R^{2m}\Big),$$

where $\Gamma(a,x) = \int_x^\infty u^{a-1}e^{-u}\,du$ is the upper incomplete gamma function.

Proof. By the polynomial lower bound, $P(t) \geq -C_P(1+t^d)$ for $t \geq 0$, so

$$-\phi(t) = -P(t) - \alpha t^{2m} \leq C_P(1+t^d) - \alpha t^{2m}.$$

For $t \geq R \geq R_0$, we have $\alpha t^{2m} \geq 2C_P t^d$ (by the definition of R_0), hence

$$C_P t^d - \alpha t^{2m} \leq -\frac{\alpha}{2}t^{2m}.$$

Therefore $-\phi(t) \leq C_P - \frac{\alpha}{2}t^{2m}$ and so $W(t) \leq e^{C_P}\exp(-\frac{\alpha}{2}t^{2m})$ for $t \geq R$; symmetry gives the same for $t \leq -R$. Integrating and substituting $u = \frac{\alpha}{2}t^{2m}$ yields

$$\int_R^\infty e^{-\frac{\alpha}{2}t^{2m}}\,dt = \frac{1}{2m}\left(\frac{2}{\alpha}\right)^{\frac{1}{2m}}\Gamma\Big(\frac{1}{2m},\ \frac{\alpha}{2}R^{2m}\Big),$$

and multiplying by $2e^{C_P}$ gives the stated bound. □

5.3 Main universality theorem

Theorem 2 (Polynomial Gibbs densities are dense in finite Gaussian mixtures)**.** *Let g be any finite weighted Gaussian mixture density on $\mathbb{R}$. For every $\varepsilon > 0$, there exists a coercive polynomial forma $\phi \in \Phi_{\text{poly}}$ such that the induced Gibbs density q_ϕ satisfies*

$$\|q_\phi - g\|_{L^1(\mathbb{R})} \;<\; \varepsilon.$$

Equivalently, $\{q_\phi : \phi \in \Phi_{\text{poly}}\}$ is dense in the class of finite Gaussian mixtures in total variation.

Proof. Fix $\varepsilon > 0$. We construct ϕ in three steps: (1) truncate tails, (2) approximate the log-density on a compact interval by a polynomial, (3) enforce tail control by a high-degree coercive term that is negligible on the compact interval.

Step 1: Choose a truncation radius. Since g has Gaussian tails, choose $R > 0$ so that

$$\int_{|\tau|>R} g(\tau)\, d\tau \;<\; \varepsilon/6.$$

Step 2: Approximate the energy on $[-R, R]$ by a polynomial. Define the (bounded, continuous) energy on $[-R, R]$:

$$E(\tau) := -\log g(\tau).$$

Because g is continuous and strictly positive on the compact interval, E is continuous on $[-R, R]$. By the Weierstrass approximation theorem, there exists a polynomial P such that

$$\sup_{|\tau|\le R} |P(\tau) - E(\tau)| \;<\; \delta,$$

where $\delta > 0$ will be chosen later.

Step 3: Add a coercive high-degree tail term with negligible effect on $[-R, R]$. Choose an integer m large and set $\alpha := \delta/R^{2m}$ so that for $|\tau| \le R$,

$$0 \le \alpha\tau^{2m} \le \delta.$$

Define the coercive polynomial forma

$$\phi(\tau) := P(\tau) + \alpha\tau^{2m}.$$

Then on $[-R, R]$ we have

$$|\phi(\tau) - E(\tau)| \le |P(\tau) - E(\tau)| + \alpha\tau^{2m} \;<\; 2\delta.$$

Hence for $|\tau| \le R$,

$$e^{-2\delta}\, g(\tau) \;\le\; e^{-\phi(\tau)} \;\le\; e^{2\delta}\, g(\tau). \tag{$\star$}$$

Step 4: Control normalization and compare densities on $[-R, R]$. Let $Z_\phi = \int e^{-\phi}$. Integrating ($\star$) over $[-R, R]$ gives

$$e^{-2\delta} \int_{|\tau|\le R} g(\tau)\, d\tau \;\le\; \int_{|\tau|\le R} e^{-\phi(\tau)}\, d\tau \;\le\; e^{2\delta} \int_{|\tau|\le R} g(\tau)\, d\tau.$$

Since $\int_{|\tau|\le R} g = 1 - \int_{|\tau|>R} g \ge 1 - \varepsilon/6$, we obtain a two-sided estimate for the bulk contribution to Z_ϕ.

Step 5 (Explicit tail control). Let $\phi(t) = P(t) + \alpha t^{2m}$ with $d = \deg P < 2m$. By Lemma 5.?, for all $R \geq R_0(P, \alpha, m)$,

$$\int_{|t|>R} e^{-\phi(t)}\, dt \;\leq\; B_{\text{tail}}(R) := 2\, e^{C_P}\, \frac{1}{2m} \left(\frac{2}{\alpha}\right)^{\frac{1}{2m}} \Gamma\!\left(\frac{1}{2m},\; \frac{\alpha}{2} R^{2m}\right).$$

Since $\Gamma(a, x) \to 0$ as $x \to \infty$ for fixed $a > 0$, we can choose R (or, equivalently, increase m) so that $B_{\text{tail}}(R) < \frac{\varepsilon}{6}\, B_{\text{bulk}}$ where

$$B_{\text{bulk}} := \int_{|t|\leq R} e^{-\phi(t)}\, dt.$$

This implies $\int_{|t|>R} q_\phi(t)\, dt < \varepsilon/6$.

Step 6: Assemble the L^1 bound. Split

$$\|q_\phi - g\|_1 \leq \int_{|\tau|\leq R} |q_\phi - g|\, d\tau + \int_{|\tau|>R} q_\phi\, d\tau + \int_{|\tau|>R} g\, d\tau.$$

The last two terms are each $< \varepsilon/6$ by construction, so it remains to bound the bulk term.

On $[-R, R]$, $(\star)$ gives multiplicative closeness between unnormalized $e^{-\phi}$ and g. Since $q_\phi = e^{-\phi}/Z_\phi$, the bulk discrepancy can be bounded by choosing δ small enough to make $e^{\pm 2\delta}$ close to 1 and to control Z_ϕ relative to 1 (using the bulk/tail bounds from Steps 4–5). Concretely, choose δ such that $e^{4\delta} - 1 < \varepsilon/3$. Then one obtains

$$\int_{|\tau|\leq R} |q_\phi(\tau) - g(\tau)|\, d\tau \;<\; \varepsilon/3.$$

Hence $\|q_\phi - g\|_1 < \varepsilon/3 + \varepsilon/6 + \varepsilon/6 = \varepsilon$. □

Remark 4 (Why this is non-tautological)**.** *We did* not *set $\phi = -\log g$. We restricted ϕ to coercive polynomials and proved approximation by: (i) approximating $-\log g$ on a compact set by a polynomial, and (ii) enforcing integrable tails via an added high-degree coercive term that is negligible on the compact set. The theorem is therefore a genuine universality result for a structured forma family.*

Corollary 1 (Density over finite Gaussian mixtures)**.** *Let $\mathcal{G}$ be the set of finite weighted Gaussian mixtures on $\mathbb{R}$. Then $\{q_\phi : \phi \in \Phi_{\text{poly}}\}$ is dense in $\mathcal{G}$ in total variation.*

6 Connecting back to pseudovariance and Laplace approximation

Given a fitted forma $\phi(\cdot, c)$, pseudovariance $\Sigma_\phi = \mathbb{E}[\phi(X, c)]$ provides a scale/temperature that can be used (or estimated) in the Gibbs model $p(\tau) \propto \exp(-\phi(\tau, c)/(2\Sigma_\phi))$. Under a dominant global mode regime, Laplace normalization yields the analytic approximation in §4.

Remark 5 (What I think remains for a "full" paper)**.** *To reach journal-level strength, I should add at least one of: (a) rates (how degree grows with ε for mixtures), (b) a data-fitting procedure for (P, α, m, c) with consistency guarantees, (c) a modality theorem tying wells of $\phi(\cdot, c)$ to modes/components under identifiable conditions, (d) experiments comparing the induced estimators against Gaussian mixture EM in defined regimes.*