Variance Misspecification and the One-Dimensional Gaussian Mixture NPMLE

Abstract

Can a one-atom Gaussian mixture NPMLE serve as evidence that a sample comes from a single population? Without regularization the answer is no: even under one Gaussian component, the fitted support diverges. We study whether a small variance inflation prevents this spurious support growth. The observations are i.i.d. $\mathcal{N}(0, 1-\varepsilon_n)$, while the fitted location mixture uses unit-variance Gaussian components.

If $\varepsilon_n = n^{-r_n}$ and $r_n \to r$, the one-atom probability converges to a phase curve $p(r)$. The curve is continuous, equals one at $r = 0$, is strictly decreasing on $(0, 1/4)$, and vanishes for $r \geq 1/4$. We identify its behavior at both endpoints; the upper endpoint is governed by the persistence exponent of a stationary Gaussian process. The limiting curve is defined through two independent compensated Poisson edge fields, which arise from the upper and lower sample extremes. More generally, for $r < 1/4$ the entire fitted atom count converges to an integer-valued law that depends only on r and varies continuously with r. These limiting counts diverge in probability as $r \uparrow 1/4$, whereas the finite-sample counts are uniformly tight whenever $\limsup r_n < 1/4$.

Note. This draft was produced through an iterative process in which Mark Sellke repeatedly asked GPT-5.6 in OpenAI Codex to prove, refine, and write up further extensions.

Contents

1 Introduction

Can one detect from a finite sample whether the data come from a single Gaussian by fitting a Gaussian mixture model? For the nonparametric maximum likelihood estimator, this becomes a question about the fitted mixing distribution: does maximum likelihood choose one component mean, or does it introduce additional atoms? Polyanskiy and Wu asked how many atoms the NPMLE selects even under one Gaussian component [PW20, Section 5.6]. Our companion paper shows that, when the fitted component variance matches the data-generating variance, the support in fact diverges in probability [Ano26b]. Thus an unregularized atom count is not by itself a calibrated diagnostic of latent heterogeneity.

Here we ask whether fitting slightly wider Gaussian components prevents this spurious growth of the fitted support. Let $Z_1, \ldots, Z_n$ be i.i.d. $\mathcal{N}(0,1)$, and observe

$$X_i = \sqrt{1-\varepsilon_n}\, Z_i, \qquad 0 \le \varepsilon_n < 1. \tag{1.1}$$

We fit location mixtures of unit-variance Gaussians. Thus the fitted kernels are slightly wider than the data-generating component, and spreading the mixing distribution incurs a deterministic variance penalty. Our main results determine the scale at which this penalty first makes a one-atom fit optimal.

For a probability measure G on $\mathbb{R}$, write

$$q_G(x) = \int_{\mathbb{R}} \phi(x-u)\, dG(u), \qquad \ell_n(G) = \sum_{i=1}^n \log q_G(X_i),$$

where ϕ and Φ are the standard normal density and distribution function, respectively. Let $\widehat{G}_n$ maximize ℓ_n over all probability measures on $\mathbb{R}$. The one-dimensional Gaussian-mixture NPMLE is almost surely unique [LR93], so its support size is unambiguous.

The geometry behind the transition is simple to state. Put $Y_i = X_i - \bar{X}_n$ and

$$D_n(a) = \frac{1}{n} \sum_{i=1}^n \exp\left(aY_i - \frac{a^2}{2} \right). \tag{1.2}$$

The measure $\delta_{\bar{X}_n}$ is an NPMLE if and only if $\sup_a D_n(a) \le 1$. For the sample with variance defect ε, denote this field by $D_{n,\varepsilon}$. After writing $t = a\sqrt{1-\varepsilon}$,

$$D_{n,\varepsilon}(a) = e^{-\kappa(\varepsilon)t^2} K_n(t), \qquad \kappa(\varepsilon) = \frac{\varepsilon}{2(1-\varepsilon)},$$

where K_n depends only on the standardized sample $Z_{1:n}$. Increasing ε therefore suppresses every nonzero perturbation and can only favor a one-atom fit. Lemma 1.7 proves that the one-atom set is an upper interval in ε. We may therefore define the samplewise critical inflation

$$\varepsilon_{*,n} = \inf\left\{ \varepsilon \in [0,1) : \sqrt{1-\varepsilon}\, Z_1, \ldots, \sqrt{1-\varepsilon}\, Z_n \text{ have a one-atom NPMLE} \right\},$$

and its exponent

$$R_{*,n} = -\frac{\log \varepsilon_{*,n}}{\log n}, \qquad -\log 0 := \infty.$$

This monotonicity is the deterministic foundation of the paper, and it is specific to one atom. Proposition 1.8 shows how completely monotonicity fails for larger support counts: for any $m \ge 2$ and any component variances $S < T$, a finite sample can have m atoms at variance T and two at variance S. The value two is optimal, since a path that is one-atomic at S remains one-atomic at every larger variance. Thus one-atomicity has a well-defined critical variance, whereas no larger support threshold is monotone.

The nontrivial scale is polynomial. Choose u_n so that $n\mathbb{P}\{Z > u_n\} = 1$. The rescaled upper-edge points $u_n(Z_i - u_n)$ converge to a Poisson point process on $\mathbb{R}$ with intensity $e^{-x}\,dx$. For score locations

$t = \theta u_n$, $1/2 < \theta < 1$, the centered directional field converges to a compensated Poisson integral. The lower sample edge gives an independent copy.

Let N be a Poisson point process on $\mathbb{R}$ with intensity $\mu(dx) = e^{-x}\,dx$. For $1/2 < \theta < 1$, define

$$S(\theta) = \int_{\mathbb{R}} e^{\theta x}\{N - \mu\}(dx). \tag{1.3}$$

The integral on $(-\infty, 0)$ is a compensated L^2 limit. On $[0, \infty)$, the process has only finitely many points almost surely, so the integral is the ordinary Poisson sum minus its compensator. This split is important: the positive-tail contribution need not be square-integrable.

Let S_+ and S_- be independent copies of S. For $0 < r < 1/4$, set $\theta_r = 1 - \sqrt{r}$ and

$$p(r) = \mathbb{P}\left\{ \max_{\xi \in \{+,-\}} \sup_{\theta_r \leq \theta < 1} S_\xi(\theta) \leq 0 \right\}. \tag{1.4}$$

Extend p to $[0, \infty]$ by $p(0) = 1$ and $p(r) = 0$ for $r \geq 1/4$.

The first theorem identifies the limiting one-atom probability at every polynomial exponent and, equivalently, the law of the samplewise critical inflation.

Theorem 1.1. *Assume $0 < \varepsilon_n < 1$, and write $\varepsilon_n = n^{-r_n}$. If $r_n \to r \in [0, \infty]$, then*

$$\mathbb{P}\{|\operatorname{supp}(\widehat{G}_n)| = 1\} \longrightarrow p(r).$$

More generally, if $\liminf_{n\to\infty} r_n \geq 1/4$, then $\mathbb{P}\{|\operatorname{supp}(\widehat{G}_n)| = 1\} \to 0$.

Together with the continuity in Proposition 1.2, this is equivalent to $R_{*,n} \Rightarrow R_*$, where

$$\mathbb{P}\{R_* \geq r\} = p(r), \qquad 0 < r < 1/4.$$

The limiting curve has no flat pieces in its nontrivial interval. Its two endpoints are governed by different mechanisms: one large Poisson point near $r = 0$, and Gaussian persistence near $r = 1/4$.

Proposition 1.2. *The function p is continuous and nonincreasing on $[0, \infty]$, strictly decreasing on $(0, 1/4)$, and satisfies $0 < p(r) < 1$ there. Moreover,*

$$1 - p(r) \sim 2\sqrt{r} \qquad (r \downarrow 0),$$

while

$$p(r) = \left(\frac{1}{4} - r\right)^{3/8 + o(1)} \qquad (r \uparrow 1/4).$$

Near the upper endpoint, the small jumps of the Poisson field are coupled to a stationary Gaussian process with sech covariance. Its one-sided persistence exponent is $3/16$ [DPSZ02, FTZ22]; the two independent sample edges double this to $3/8$. The transfer from the Poisson field uses a Komlós–Major–Tusnády (KMT) strong approximation and the small-level persistence estimates of Feldheim–Feldheim–Mukherjee [KMT75, KMT76, FFM25].

Figure 1 illustrates the phase curve and the two endpoint asymptotics.

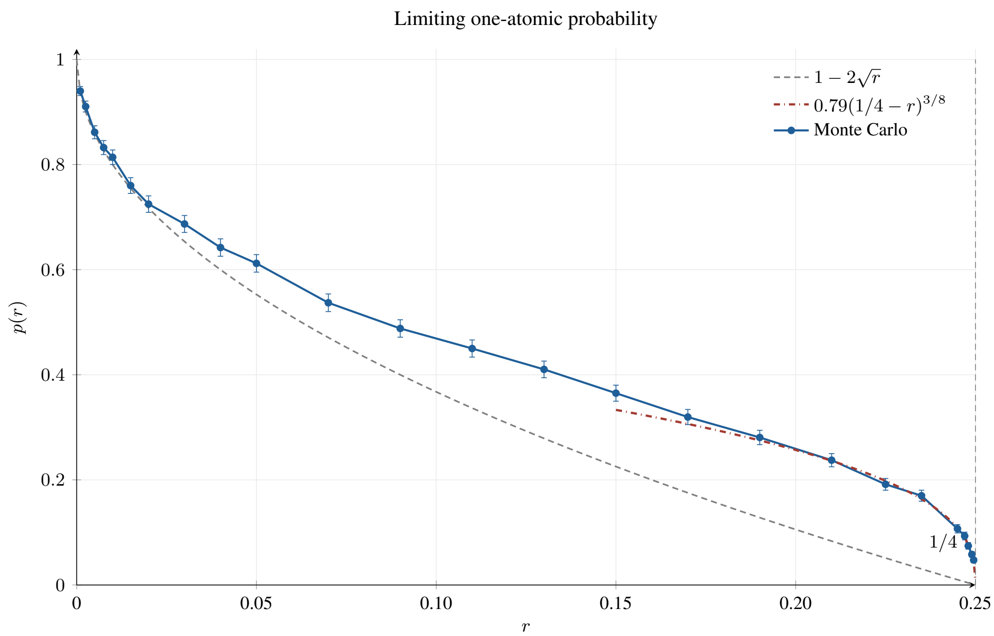

Figure 1: Numerical illustration of the limiting one-atom phase curve. The blue points are Monte Carlo estimates of $p(r)$, with two-standard-error bars, and the gray dashed line is the lower-endpoint asymptotic $1-2\sqrt{r}$. Proposition 1.2 proves that $p(r)=(1/4-r)^{3/8+o(1)}$ as $r\uparrow 1/4$; numerically, $0.79(1/4-r)^{3/8}$ appears to match near the endpoint, as shown by the red dot-dashed line.

The Poisson limit also determines the entire support size. For $b\in(1/2,1)$, a finite positive measure ν compactly supported in $[b,1)$, and $h(y)=\log(1+y)-y$, put

$$F_\nu(x)=\int e^{\theta x}\,\nu(d\theta),\qquad \mathcal{J}_{N,b}(\nu)=\int S(\theta)\,\nu(d\theta)+\int_{\mathbb{R}} h\{F_\nu(x)\}\,N(dx). \tag{1.5}$$

Proposition 5.10 proves that $\mathcal{J}_{N,b}$ has an almost surely unique optimizer ν_b^*. Let $\nu_b^{*,+}$ and $\nu_b^{*,-}$ be the optimizers constructed from two independent copies of N. Define $\mathsf{K}_0=1$ and, for $0<r<1/4$,

$$\mathsf{K}_r=1+|\operatorname{supp}(\nu^{*,+}_{1-\sqrt{r}})|+|\operatorname{supp}(\nu^{*,-}_{1-\sqrt{r}})|. \tag{1.6}$$

The next theorem shows that this explicit Poisson functional gives the unique limit of the fitted support size.

Theorem 1.3. *If $\varepsilon_n=n^{-r_n}$ and $r_n\to r<1/4$, then*

$$|\operatorname{supp}(\widehat{G}_n)|\Rightarrow \mathsf{K}_r.$$

For $0<r<1/4$,

$$\mathbb{P}\{\mathsf{K}_r=1\}=p(r).$$

The map $r\mapsto\mathcal{L}(\mathsf{K}_r)$ is continuous in total variation on $[0,1/4)$.

Although each interior limiting law is finite, the Poisson optimizer becomes increasingly complex as its hard wall approaches $1/2$. The endpoint likelihood comparison gives a lower bound for its atom count at the scale $\log(1/\delta)/\log\log(1/\delta)$.

Proposition 1.4. *For $0 < \delta < 1/4$, set $r_\delta = 1/4 - \delta$. Then*

$$\lim_{\eta \downarrow 0} \limsup_{\delta \downarrow 0} \mathbb{P}\left\{ \mathsf{K}_{r_\delta} \le \eta \frac{\log(1/\delta)}{\log\log(1/\delta)} \right\} = 0.$$

In particular, $\mathsf{K}_r \to \infty$ in probability as $r \uparrow 1/4$.

Proposition 1.4 concerns the limiting variables K_r: one first takes the sample-size limit with r fixed and then lets $r \uparrow 1/4$. It does not give a uniform finite-sample approximation for a triangular sequence $r_n \uparrow 1/4$.

The convergence theorem also gives uniform tightness away from the critical endpoint: if $\limsup_n r_n < 1/4$, then $|\operatorname{supp}(\widehat{G}_n)| = O_{\mathbb{P}}(1)$. Indeed, failure along some sequence would persist along a subsequence on which r_n converges to a point of $[0, 1/4)$, contradicting Theorem 1.3.

The one-dimensional mechanism is exceptional. In dimensions at least two, the angular geometry produces logarithmic radial transitions instead of the two-sided polynomial persistence curve above; that problem, including growing dimension, is treated in a companion paper [Ano26a].

1.1 Preliminaries

We first settle the deterministic geometry of the variance path. The directional-derivative criterion reduces one-atom optimality to a scalar supremum, and increasing the variance defect can only lower this supremum. This makes the samplewise critical inflation well defined. We then show that no analogous monotonicity survives for any larger support threshold. The one-atom monotonicity and the counterexamples for larger support thresholds form the deterministic foundation for the scaling limits proved later.

1.1.1 Notation and convergence

All asymptotic notation is as $n \to \infty$, unless another limit is specified. For a deterministic positive sequence a_n, the relation $X_n = O_{\mathbb{P}}(a_n)$ means that $|X_n|/a_n$ is tight, while $X_n = o_{\mathbb{P}}(a_n)$ means that $X_n/a_n \to 0$ in probability. Deterministic $O(\cdot)$ and $o(\cdot)$ have their usual meanings. We write $a_n \sim b_n$ if $a_n/b_n \to 1$, $a_n \ll b_n$ if $a_n/b_n \to 0$, and $a_n \lesssim b_n$ if $a_n \le C b_n$ for a constant independent of n. The reverse relations are $\gg$ and $\gtrsim$, and $a_n \asymp b_n$ means that both one-sided bounds hold. Constants $c, C, c_0, C_0, \dots$ may change from line to line.

We write $X_n \Rightarrow X$ for weak convergence. Random functions on a compact set K are viewed in $C(K)$ with the supremum norm. Point processes carry the vague topology. For a measure G, $\operatorname{supp}(G)$ is its support and δ_x is the unit point mass at x. Finally, $J \Subset I$ means that the interval J is compactly contained in I. For a measurable function f, we use the empirical-measure notation $P_n f = n^{-1} \sum_{i=1}^n f(Z_i)$ and $Pf = \mathbb{E} f(Z)$, where $Z \sim \mathcal{N}(0,1)$.

We use bracketing only in the support-tightness argument. If $\mathcal{F}$ is a class of measurable functions and P is a probability measure, an $L^1(P)$-bracket of width η is a pair $[l, u]$ with $l \le u$, $\|u - l\|_{L^1(P)} \le \eta$, and all functions between l and u. The bracketing number $N_{[]}\{\eta, \mathcal{F}, L^1(P)\}$ is the least number of such brackets needed to cover $\mathcal{F}$.

1.1.2 One-atom optimality and uniqueness

The following criterion turns the infinite-dimensional likelihood problem into a scalar supremum. Its monotonicity outside the sample range will later exclude score locations beyond the extreme observations.

Lemma 1.5. *For arbitrary observations $X_1, \dots, X_n \in \mathbb{R}$, the measure $\delta_{\bar{X}_n}$ is an NPMLE if and only if*

$$\sup_{a \in \mathbb{R}} D_n(a) \le 1,$$

where D_n is defined in (1.2). Moreover, $D_n(0) = 1$, D_n is strictly decreasing on $(\max_i Y_i, \infty)$, and it is strictly increasing on $(-\infty, \min_i Y_i)$.

Proof. For $a \in \mathbb{R}$ and $0 \le t \le 1$, perturb the one-atom measure in the direction of a second atom:

$$G_t = (1-t)\delta_{\bar{X}_n} + t\delta_{\bar{X}_n + a}.$$

Since

$$\frac{q_{G_t}(X_i)}{q_{\delta_{\bar{X}_n}}(X_i)} = 1 - t + t \exp\left(aY_i - \frac{a^2}{2}\right),$$

the right derivative at $t = 0$ is

$$\frac{d}{dt}\ell_n(G_t)\Big|_{t=0+} = n\{D_n(a) - 1\}.$$

Among point masses, $\delta_{\bar{X}_n}$ uniquely maximizes the likelihood. Concavity of $G \mapsto \ell_n(G)$, together with Lindsay's directional-derivative criterion [Lin83], therefore shows that it is globally optimal exactly when $D_n(a) \le 1$ for every a.

The identity $D_n(0) = 1$ is immediate. Differentiating gives

$$D_n'(a) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - a)\exp\left(aY_i - \frac{a^2}{2}\right).$$

Every summand is negative when $a > \max_i Y_i$, and every summand is positive when $a < \min_i Y_i$, proving the final assertions. □

The NPMLE is known to be unique in one dimension. For the one-atom event we will use the following narrower consequence.

Proposition 1.6. *Let $n \ge 2$ and let $X_1, \dots, X_n$ be i.i.d. from a nondegenerate Gaussian distribution. Almost surely, if a one-atom NPMLE exists, then it is the unique NPMLE.*

Proof. The only possible one-atom maximizer is $G_0 = \delta_{\bar{X}_n}$. Suppose that G_0 and a different mixing measure G both maximize the likelihood. Strict concavity of $(q_1, \dots, q_n) \mapsto \sum_i \log q_i$ implies

$$q_G(X_i) = q_{G_0}(X_i), \qquad 1 \le i \le n;$$

otherwise their midpoint would have larger likelihood. Lemma 1.5 gives $D_n(a) \le 1$ for every a, while

$$\int D_n(u - \bar{X}_n)\, dG(u) = \frac{1}{n}\sum_{i=1}^{n}\frac{q_G(X_i)}{q_{G_0}(X_i)} = 1.$$

Thus $D_n(u - \bar{X}_n) = 1$ for G-almost every u. Since $G \ne G_0$, there is a nonzero a with $D_n(a) = 1$.

It remains to show that such a contact has probability zero on the one-atom event. Regard the centered sample $Y = (Y_1, \dots, Y_n)$ as a Gaussian vector in the hyperplane $\mathcal{H} = \{y \in \mathbb{R}^n : \sum_i y_i = 0\}$. Its law is spherical and nondegenerate on $\mathcal{H}$, so $Y = RU$, where $R > 0$ is independent of U and has a density. For a centered sample rU, set

$$\mathcal{D}_{r,U}(t) = D_{r,U}(t/r) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(tU_i - \frac{t^2}{2r^2}\right).$$

For fixed U and $t \neq 0$, this quantity is strictly increasing in r. If two radii $r_1 < r_2$ both admitted a second maximizer tied with the one-atom fit, the KKT equality for that second maximizer would give $t_1 \neq 0$ with $\mathcal{D}_{r_1,U}(t_1) = 1$. Then

$$\mathcal{D}_{r_2,U}(t_1) = \exp\left\{\frac{t_1^2}{2}\left(\frac{1}{r_1^2} - \frac{1}{r_2^2}\right)\right\} > 1,$$

contradicting Lemma 1.5 at radius r_2. Hence for each fixed U, at most one radius can produce a tie. Conditional on U, the radius R has a density, so it hits this at-most-singleton set with probability zero. Integrating over U proves the claim. □

1.1.3 Deterministic variance paths

We now determine which support properties are monotone along a fixed-sample variance path. A larger variance defect damps every nonzero likelihood perturbation, so the one-atom set is an upper interval in ε.

Lemma 1.7. *Fix $z_1, \dots, z_n \in \mathbb{R}$, and for $0 \le \varepsilon < 1$ set $x_i^{(\varepsilon)} = \sqrt{1-\varepsilon}\, z_i$. Let $D_{n,\varepsilon}$ be the field (1.2) formed from this sample. If $0 \le \varepsilon_1 \le \varepsilon_2 < 1$, then*

$$\sup_{a\in\mathbb{R}} D_{n,\varepsilon_2}(a) \le \sup_{a\in\mathbb{R}} D_{n,\varepsilon_1}(a).$$

Proof. Let $\bar{z} = n^{-1}\sum_i z_i$ and define

$$K_n(t) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(t(z_i - \bar{z}) - \frac{t^2}{2}\right).$$

For $x_i^{(\varepsilon)} = \sqrt{1-\varepsilon}\, z_i$, the change of variables $t = a\sqrt{1-\varepsilon}$ gives

$$D_{n,\varepsilon}(a) = e^{-\kappa(\varepsilon)t^2} K_n(t), \qquad \kappa(\varepsilon) = \frac{\varepsilon}{2(1-\varepsilon)}.$$

The map $\varepsilon \mapsto \kappa(\varepsilon)$ is increasing, and t ranges over all of $\mathbb{R}$ for every $\varepsilon < 1$. Taking suprema proves the displayed inequality; Lemma 1.5 gives the corresponding monotonicity of the one-atom event. □

This monotonicity is special to one atom. To state the sharp obstruction, fix a finite sample X, let $\widehat{G}_{X,t}$ denote its Gaussian location-mixture NPMLE with component variance t. The NPMLE is unique in one dimension, so we may write $K_X(t)$ for its number of atoms. After rescaling to unit component variance, Lemma 1.7 shows that

$$K_X(S) = 1 \quad\Longrightarrow\quad K_X(t) = 1 \quad \text{for every } t > S.$$

Thus two is the smallest possible value at S in any example with $K_X(T) > 1$ for some $T > S$. The next proposition reaches this minimum while making $K_X(T)$ arbitrarily large. The construction, proved in Appendix B, crucially uses that the Gaussian kernel decays faster than exponentially; it does not directly extend to Laplace mixtures.

Proposition 1.8. *Fix $0 < S < T$ and an integer $m \geq 2$. There is a finite sample $X \subset \mathbb{R}$ for which the NPMLE is unique at variances S and T, and*

$$K_X(T) = m, \qquad K_X(S) = 2.$$

Finally, the sample X may be chosen to consist of exactly m distinct values.

1.2 Related literature

The mixture NPMLE originates in Robbins and Kiefer–Wolfowitz [Rob50, KW56]. In the normal-means empirical Bayes model, the fitted mixing distribution estimates the latent prior and the induced Bayes rule is used for inference; see, for example, Jiang–Zhang [JZ09], Koenker–Mizera [KM14], and Dicker–Zhao [DZ16]. A point-mass prior describes a homogeneous population of signals, so whether the fitted prior has one atom is a simple, deliberately stringent structural diagnostic.

The geometry of exact mixture likelihood maximizers was developed by Lindsay [Lin83]. Koenker and Gu documented sparse NPMLE fits in *Minimalist g-Modeling* [KG19]; Polyanskiy and Wu proved an $O_{\mathbb{P}}(\log n)$ support upper bound under subgaussian sampling [PW20]. Our support-divergence companion shows that the support nevertheless diverges under any fixed finitely supported mixing distribution [Ano26b]. These results concern a well-specified component variance. Here the fitted variance is deliberately inflated, and the object is the threshold at which this perturbation collapses the optimizer to one atom.

Mixture homogeneity tests provide the closest inferential comparison. Gu–Koenker–Volgushev compare the point-mass likelihood with the likelihood at the Kiefer–Wolfowitz NPMLE [GKV18]; related likelihood-ratio statistics have nonstandard null behavior [Har85, AGM09, JZ19]. Our statistic is different: it records whether the unrestricted maximizer itself has one atom after a prescribed variance inflation. A useful analogue is Silverman's critical bandwidth, the smallest amount of kernel smoothing needed to reduce a density estimate to a prescribed number of modes [Sil81, HY01].

The fixed-sample path $t \mapsto \widehat{G}_{X,t}$ also appears in scalar quadratic rate–distortion theory, with reciprocal variance acting as inverse temperature. In his analysis of deterministic annealing, Rose conjectured that the optimizer's support size grows monotonically as inverse temperature increases [Ros94, Section IV-C, p. 1946]. While Lemma 1.7 confirms the corresponding monotonicity for one-atomicity, Proposition 1.8 disproves every higher-support analogue in a strong form.

A complementary line of NPMLE theory studies density or decision risk [Zha09, SG20, DWYZ23]. Kim and Sen's smooth NPMLE is particularly close in spirit: their Gaussian smoothing layer also inflates the marginal component variance, but their goal is estimation and inference in a smoothed mixing model [KS26]. We instead let the inflation vanish and study the exact support selected by maximum likelihood.

2 The One-Dimensional Edge Field

This section supplies the probabilistic input shared by Theorems 1.1 and 1.3. We resolve the Karush–Kuhn–Tucker (KKT) field at score locations $t = \theta u_n$: each sample tail produces a Poisson cloud of extreme observations, and the centered exponential sum converges on compact subintervals of $(1/2, 1)$ to the compensated field in (1.3). The remaining estimates restore empirical centering, combine the two tails, and exclude score ranges outside this compact limit.

We use the unit-intensity extreme-value normalization: for $Z \sim \mathcal{N}(0,1)$, choose u_n by $n\mathbb{P}\{Z > u_n\} = 1$. Throughout this section, assume $0 \leq \varepsilon_n < 1$ and write

$$X_i = \sigma_n Z_i, \qquad \sigma_n = \sqrt{1-\varepsilon_n}, \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1).$$

For a candidate shift a, put $t = a\sigma_n$. The exact KKT field contains an empirical-centering factor, which we restore below. Without that factor, the field is

$$e^{-\lambda_n(t/u_n)^2}\frac{1}{n}\sum_{i=1}^{n}\exp\Big(tZ_i - \frac{t^2}{2}\Big), \qquad \lambda_n = \frac{\varepsilon_n u_n^2}{2(1-\varepsilon_n)}.$$

The exponential sum is the correctly specified field; the factor $e^{-\lambda_n(t/u_n)^2}$ is the deterministic penalty caused by variance misspecification. Since $u_n^2/2 = \log n + o(\log n)$, λ_n and ε_n have the same polynomial exponent whenever $\varepsilon_n \to 0$. The one-atom question is whether the penalized field can exceed the KKT level one.

We write $t = \theta u_n$. After division by the scale $c_n(\theta)$ defined below, only observations within distance $O(1/u_n)$ of the sample maximum contribute at order one. These observations need not lie near the score location θu_n. Writing an extreme observation as $Z_i = u_n + x/u_n$, its contribution is

$$\frac{1}{n}\exp\Big(\theta u_n^2 - \frac{\theta^2 u_n^2}{2}\Big)e^{\theta x} = c_n(\theta)e^{\theta x}$$

to the exponential sum, where

$$c_n(\theta) := \frac{1}{n}\exp\Big(\theta u_n^2 - \frac{\theta^2 u_n^2}{2}\Big) = n^{-(1-\theta)^2+o(1)}.$$

For $\theta > 1/2$, the centered fluctuation of the correctly specified field is of order $c_n(\theta)$ and is governed by the Poisson process of extreme points. Proposition 2.1 proves this convergence uniformly for θ in any fixed compact interval $J \Subset (1/2, 1)$.

If $\lambda_n = n^{-\alpha+o(1)}$, balancing the damping against $c_n(\theta)$ gives $\theta_\alpha = 1 - \sqrt{\alpha}$. To pass from compact convergence to the phase theorem, we must also combine the two sample tails and exclude crossings below $1/2$ and near 1. The estimates below do so and establish the needed regularity of the limiting supremum.

2.1 Compact edge process

The first input to the polynomial sandwich in Theorem 3.1 is compact convergence of the normalized empirical exponential sum to the compensated Poisson field.

Proposition 2.1. *Let $Z_1, \dots, Z_n$ be i.i.d. $\mathcal{N}(0,1)$, and let u_n be defined by*

$$n\mathbb{P}\{Z_1 > u_n\} = 1.$$

For $\theta \in (0,1)$ set

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\exp\Big(\theta u_n Z_i - \frac{\theta^2 u_n^2}{2}\Big), \qquad c_n(\theta) = \frac{1}{n}\exp\Big(\theta u_n^2 - \frac{\theta^2 u_n^2}{2}\Big).$$

If $J = [\theta_0, \theta_1] \Subset (1/2, 1)$, then

$$\left\{\frac{H_n(\theta) - 1}{c_n(\theta)} : \theta \in J\right\} \Rightarrow \{S(\theta) : \theta \in J\} \qquad \text{in } C(J),$$

where N is a Poisson point process on $\mathbb{R}$ with intensity $\mu(dx) = e^{-x}\,dx$, and

$$\begin{aligned} S(\theta) &= S^{\mathrm{neg}}(\theta) + S^{\mathrm{pos}}(\theta), \\ S^{\mathrm{neg}}(\theta) &= \lim_{M\to\infty}\int_{-M}^{0} e^{\theta x}\{N - \mu\}(dx), \\ S^{\mathrm{pos}}(\theta) &= \sum_{X \in N \cap [0,\infty)} e^{\theta X} - \int_0^\infty e^{\theta x}\mu(dx). \end{aligned}$$

The first limit holds in $L^2(\Omega; C(J))$. The second sum is almost surely finite because $\mu([0,\infty)) = 1$, and its compensator equals $(1-\theta)^{-1}$. Thus $S^{\mathrm{pos}}(\theta)$ is integrable, though not square-integrable when $\theta \geq 1/2$. The process has an almost surely real-analytic version on $(1/2, 1)$. Moreover

$$c_n(\theta) = n^{-(1-\theta)^2 + o(1)}$$

uniformly for θ in compact subintervals of $(0,1)$.

Let $X_i = \sigma_n Z_i$, where $\sigma_n = \sqrt{1-\varepsilon_n}$, and let D_n be the centered directional derivative from (1.2). Put

$$a_n(\theta) = \frac{\theta u_n}{\sigma_n}, \qquad \lambda_n = \frac{\varepsilon_n u_n^2}{2(1-\varepsilon_n)}.$$

Then, uniformly on every $J \Subset (1/2, 1)$,

$$\frac{D_n(a_n(\theta)) - 1}{c_n(\theta)} = e^{-\lambda_n \theta^2} \frac{H_n(\theta) - 1}{c_n(\theta)} + \frac{e^{-\lambda_n \theta^2} - 1}{c_n(\theta)} + o_{\mathbb{P}}(1).$$

Proof. Define the edge point process

$$N_n = \sum_{i=1}^n \delta_{u_n(Z_i - u_n)}, \qquad \mu_n = \mathbb{E} N_n.$$

Its intensity has density

$$\mu_n(dx) = n \frac{\phi(u_n + x/u_n)}{u_n}\, dx = \frac{n\phi(u_n)}{u_n} \exp\Bigl(-x - \frac{x^2}{2u_n^2}\Bigr)\, dx.$$

Since $n\phi(u_n)/u_n \to 1$, the standard Poisson convergence theorem for extremes [Kal17] gives vague convergence of N_n on compact intervals to a Poisson point process N with intensity $e^{-x}dx$. Also

$$\frac{H_n(\theta) - 1}{c_n(\theta)} = \int_{\mathbb{R}} e^{\theta x} \{N_n - \mu_n\}(dx).$$

Fix $M < \infty$. Because the limiting intensity is diffuse, the Poisson process almost surely has no atom at either endpoint of $[-M, M]$. On this interval, vague convergence, the continuous mapping theorem for the equicontinuous family $\{e^{\theta x} : \theta \in J\}$, and the uniform convergence $\mu_n \to \mu$ imply convergence in $C(J)$ of the truncated centered integrals. We now show that the portions with $x < -M$ and $x > M$ vanish uniformly in θ as $M \to \infty$.

For the upper tail, monotonicity in θ gives

$$\mathbb{E} \sup_{\theta \in J} \int_M^\infty e^{\theta x} \{N_n + \mu_n\}(dx) \leq 2 \int_M^\infty e^{\theta_1 x} \mu_n(dx) \leq C_J e^{-(1-\theta_1)M}.$$

Replacing N_n, μ_n by their limits N, μ gives the identical upper-tail bound. For the lower tail, write

$$R_n(\theta) = \int_{-\infty}^{-M} e^{\theta x} \{N_n - \mu_n\}(dx).$$

Since $2\theta_0 > 1$,

$$\mathbb{E}|R_n(\theta_0)|^2 \leq \int_{-\infty}^{-M} e^{2\theta_0 x} \mu_n(dx) \leq C_J e^{-(2\theta_0 - 1)M}.$$

Differentiating in θ inserts a factor x; the corresponding second-moment estimate is

$$\sup_{\theta\in J}\mathbb{E}|R_n'(\theta)|^2 \le C_J(1+M^2)e^{-(2\theta_0-1)M}.$$

The squared Sobolev bound

$$\|f\|_\infty^2 \le C_J\{|f(\theta_0)|^2 + \|f'\|^2_{L^2(J)}\}$$

on the interval J therefore gives

$$\mathbb{E}\sup_{\theta\in J}|R_n(\theta)|^2 \le C_J(1+M^2)e^{-c_J M}\to 0 \qquad (M\to\infty),$$

uniformly in n. The identical estimate with N_n, μ_n replaced by N, μ proves the asserted $L^2(\Omega; C(J))$ construction of S^{neg}. Together with the upper-tail L^1 bound, this proves the functional convergence. Repeating the negative-tail estimate after any fixed number of θ-derivatives gives locally uniform convergence of the derivative series. The positive part is a finite exponential sum minus $(1-\theta)^{-1}$, so the resulting version of S is real analytic on $(1/2, 1)$.

Mills' ratio gives

$$n = u_n\sqrt{2\pi}\, e^{u_n^2/2}(1+o(1)).$$

Hence

$$c_n(\theta) = (u_n\sqrt{2\pi})^{-1}\exp\Big(-\frac{(1-\theta)^2u_n^2}{2}\Big)(1+o(1)) = n^{-(1-\theta)^2+o(1)}$$

uniformly on compact θ-intervals.

It remains to restore the damping and empirical centering in the KKT field. For $a_n(\theta) = \theta u_n/\sigma_n$,

$$D_n(a_n(\theta)) = e^{-\theta u_n\bar{Z}_n}e^{-\lambda_n\theta^2}H_n(\theta), \qquad \bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i.$$

On $J = [\theta_0, \theta_1] \Subset (1/2, 1)$,

$$\inf_{\theta\in J} c_n(\theta) \ge n^{-(1-\theta_0)^2+o(1)} \gg n^{-1/2}\sqrt{\log n},$$

because $(1-\theta_0)^2 < 1/4$. Since $u_n\bar{Z}_n = O_{\mathbb{P}}(\sqrt{\log n/n})$ and

$$\sup_{\theta\in J}\frac{|H_n(\theta)-1|}{c_n(\theta)} = O_{\mathbb{P}}(1),$$

the empirical-centering factor contributes $o_{\mathbb{P}}(c_n(\theta))$ uniformly. Dividing by $c_n(\theta)$ gives the stated expansion. □

2.2 Two tails and empirical centering

Proposition 2.1 treats positive candidate shifts and therefore uses the right tail of the sample. Applying it to the reflected sample $-Z_1, \dots, -Z_n$ treats negative shifts and the left tail. We now introduce both signs explicitly. For $\xi \in \{-1, 1\}$ and $\theta \ge 0$, set

$$a_n(\theta) = \frac{\theta u_n}{\sqrt{1-\varepsilon_n}}, \qquad \lambda_n = \frac{\varepsilon_n u_n^2}{2(1-\varepsilon_n)},$$

$$H_{n,\xi}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \exp\Big(\theta u_n \xi Z_i - \frac{\theta^2 u_n^2}{2}\Big), \qquad \overline{D}_{n,\xi}(\theta) = e^{-\lambda_n \theta^2} H_{n,\xi}(\theta),$$

and

$$D_{n,\xi}(\theta) = D_n\big(\xi a_n(\theta)\big).$$

Thus

$$D_{n,\xi}(\theta) = e^{-\xi\theta u_n \bar{Z}_n}\overline{D}_{n,\xi}(\theta), \qquad \bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i.$$

The exact KKT field $D_{n,\xi}$ differs from $\overline{D}_{n,\xi}$ only by the sample-mean factor. Uniformly over both signs and $0 \le \theta \le 1$, this factor changes the field by $o_{\mathbb{P}}(\lambda_n)$.

Lemma 2.2. *Assume $\lambda_n = n^{-\alpha+o(1)}$ for some $\alpha \in (0, 1/4)$. Then*

$$\sup_{\xi=\pm 1}\sup_{0\le\theta\le 1}\big|D_{n,\xi}(\theta) - \overline{D}_{n,\xi}(\theta)\big| = o_{\mathbb{P}}(\lambda_n).$$

Proof. From

$$D_{n,\xi}(\theta) = e^{-\xi\theta u_n \bar{Z}_n}\overline{D}_{n,\xi}(\theta)$$

and $u_n|\bar{Z}_n| = O_{\mathbb{P}}(\sqrt{\log n/n}) = o_{\mathbb{P}}(1)$, we have

$$\sup_{\xi,\theta}\big|D_{n,\xi}(\theta) - \overline{D}_{n,\xi}(\theta)\big| \le O_{\mathbb{P}}(u_n|\bar{Z}_n|)\sup_{\xi,\theta}\overline{D}_{n,\xi}(\theta).$$

Thus it is enough to prove

$$\sup_{\xi=\pm 1}\sup_{0\le\theta\le 1}\overline{D}_{n,\xi}(\theta) = O_{\mathbb{P}}(\sqrt{\log n}).$$

For fixed data, each summand $\exp(\theta u_n \xi Z_i - \theta^2 u_n^2/2)$ is maximized over $0 \le \theta \le 1$ either at $\theta = 0$, at $\theta = 1$, or at $\theta = \xi Z_i/u_n$ when $0 \le \xi Z_i \le u_n$. Hence

$$\sup_{0\le\theta\le 1} H_{n,\xi}(\theta) \le 1 + \frac{1}{n}\sum_{i=1}^{n}\exp\Big(\frac{(\xi Z_i)_+^2}{2}\Big)\mathbf{1}_{\{0\le \xi Z_i\le u_n\}} + H_{n,\xi}(1).$$

The middle term has expectation $\mathbb{E}[e^{Z^2/2}\mathbf{1}_{\{0\le Z\le u_n\}}] = u_n/\sqrt{2\pi}$, so Markov's inequality gives an $O_{\mathbb{P}}(u_n)$ bound for the average. The endpoint term $H_{n,\xi}(1)$ is $O_{\mathbb{P}}(1)$. Indeed, for

$$A_n = \frac{1}{n}\sum_{i=1}^{n} e^{u_n\xi Z_i - u_n^2/2}\mathbf{1}_{\{\xi Z_i \le u_n\}},$$

exponential tilting gives $\mathbb{E}A_n = 1/2$ and

$$\mathrm{Var}(A_n) \le n^{-1}e^{u_n^2}\Phi(-u_n) = O(u_n^{-2}).$$

If $u_n < \xi Z_i \le u_n + L/u_n$, then the corresponding term in the average, including the factor $1/n$, is at most C_L/u_n. The number of observations in this window is tight. The probability of any observation above $u_n + L/u_n$ is $O(e^{-L})$. Letting $L \to \infty$ gives $H_{n,\xi}(1) = O_{\mathbb{P}}(1)$. Since $e^{-\lambda_n\theta^2} \le 1$, the desired $O_{\mathbb{P}}(\sqrt{\log n})$ bound follows. Multiplying by $u_n|\bar{Z}_n|$ gives $O_{\mathbb{P}}(\log n/\sqrt{n}) = o_{\mathbb{P}}(\lambda_n)$ because $\alpha < 1/4$. $\square$

2.3 Uniform exclusion and regularity

Compact convergence does not by itself control the full KKT supremum. The lemmas in this subsection supply the exclusion and regularity estimates used to pass from Proposition 2.1 to the polynomial sandwich in Theorem 3.1. The Poisson normalization is sharp only for $\theta > 1/2$; below this point, ordinary fluctuations of the full sample dominate the extreme-tail scale. A direct second-moment argument controls the larger range $\theta \le \theta_* < 1/\sqrt{2}$ at order $n^{-1/2+\theta_*^2+o(1)}$. The two methods therefore overlap on $(1/2, 1/\sqrt{2})$: the Poisson limit is sharper there, while the cruder second-moment bound is sufficient for excluding subthreshold crossings.

Lemma 2.3. *Fix $0 < \theta_* < 1/\sqrt{2}$ and define*

$$K_{n,\xi}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \exp\Big(\theta u_n \xi (Z_i - \bar{Z}_n) - \frac{\theta^2 u_n^2}{2}\Big), \qquad 0 \le \theta \le \theta_*.$$

Then

$$\sup_{\xi=\pm 1}\sup_{0<\theta\le\theta_*} \frac{|K_{n,\xi}(\theta)-1|}{\theta^2} = O_{\mathbb{P}}\big(n^{-1/2+\theta_*^2+o(1)}\big).$$

Proof. We give the proof for $\xi = 1$; the other sign is identical. Let

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \exp\Big(\theta u_n Z_i - \frac{\theta^2 u_n^2}{2}\Big).$$

For $r = 1, 2, 3$, Gaussian integration gives, uniformly for $0 \le \theta \le \theta_*$,

$$\mathbb{E}\left[\left|\partial_\theta^r \exp\Big(\theta u_n Z - \frac{\theta^2 u_n^2}{2}\Big)\right|^2\right] \le u_n^{C_r} e^{\theta^2 u_n^2}.$$

Since $\mathbb{E}H_n^{(r)}(\theta) = 0$ for $r \ge 1$, independence and the displayed derivative bound imply

$$\sup_{0\le\theta\le\theta_*} \|H_n^{(r)}(\theta)\|_2 \le u_n^{C_r} n^{-1/2+\theta_*^2+o(1)}.$$

Applying the elementary Sobolev bound

$$\sup_{0\le\theta\le\theta_*} |f(\theta)| \le |f(0)| + \int_0^{\theta_*} |f'(\theta)|\, d\theta$$

to H_n' and H_n'' gives

$$\sup_{0\le\theta\le\theta_*} \big(|H_n'(\theta)| + |H_n''(\theta)|\big) = O_{\mathbb{P}}\big(n^{-1/2+\theta_*^2+o(1)}\big).$$

Applying the derivative argument once more to $H_n - 1$, and using $H_n(0) = 1$, gives

$$\sup_{0\le\theta\le\theta_*} |H_n(\theta) - 1| = O_{\mathbb{P}}\big(n^{-1/2+\theta_*^2+o(1)}\big).$$

Write $b_n = u_n \bar{Z}_n = O_{\mathbb{P}}(\sqrt{\log n/n})$. Then

$$K_{n,1}(\theta) = e^{-b_n\theta} H_n(\theta),$$

so

$$K''_{n,1}(\theta) = e^{-b_n\theta}\{H''_n(\theta) - 2b_nH'_n(\theta) + b_n^2H_n(\theta)\}.$$

The displayed bounds and $b_n = o_{\mathbb{P}}(1)$ give

$$\sup_{0\le\theta\le\theta_*} |K''_{n,1}(\theta)| = O_{\mathbb{P}}\big(n^{-1/2+\theta_*^2+o(1)}\big).$$

Finally $K_{n,1}(0) = 1$ and $K'_{n,1}(0) = 0$, hence

$$K_{n,1}(\theta) - 1 = \int_0^\theta (\theta - s)K''_{n,1}(s)\, ds,$$

and the asserted bound follows. □

Below θ_α by a fixed margin, the deterministic damping $\lambda_n\theta^2$ dominates the fluctuations. Gaussian estimates handle the range up to a fixed $\theta_* > 1/2$, and the compact Poisson estimate handles the remainder.

Lemma 2.4. *Fix $\alpha \in (0, 1/4)$, let*

$$\theta_\alpha = 1 - \sqrt{\alpha},$$

and assume $\lambda_n = n^{-\alpha+o(1)}$. For every

$$0 < \eta < \theta_\alpha - \frac{1}{2}$$

we have

$$\mathbb{P}\Big\{\sup_{\xi=\pm1}\ \sup_{0\le\theta\le\theta_\alpha-\eta} D_{n,\xi}(\theta) > 1\Big\} \to 0.$$

Proof. Put $\theta_0 = \theta_\alpha - \eta$. Choose

$$\theta_* \in \Big(\frac{1}{2}, \theta_0\Big) \quad \text{so that} \quad \theta_*^2 < \frac{1}{2} - \alpha;$$

this is possible because $\theta_0 > 1/2$ and $\sqrt{1/2-\alpha} > 1/2$.

On $[0, \theta_*]$ we use the Gaussian-regime bound for the centered null field. Write

$$D_{n,\xi}(\theta) = e^{-\lambda_n\theta^2}K_{n,\xi}(\theta),$$

where $K_{n,\xi}$ is the centered null field from Lemma 2.3. Since $\theta_*^2 < 1/2 - \alpha$,

$$\sup_{\xi=\pm1}\ \sup_{0<\theta\le\theta_*} \frac{|K_{n,\xi}(\theta) - 1|}{\theta^2} = o_{\mathbb{P}}(\lambda_n).$$

Consequently,

$$\sup_{\xi=\pm1}\ \sup_{0<\theta\le\theta_*} \frac{D_{n,\xi}(\theta) - 1}{\theta^2} \le -\lambda_n + o_{\mathbb{P}}(\lambda_n),$$

because $(1-e^{-\lambda_n\theta^2})/\theta^2 = \lambda_n + O(\lambda_n^2)$ uniformly for $0 \le \theta \le 1$. Therefore $\sup_{\xi=\pm1}\sup_{0\le\theta\le\theta_*} D_{n,\xi}(\theta) \le 1$ with probability tending to one.

On the compact interval $[\theta_*, \theta_0] \Subset (1/2, 1)$, Lemma 2.2 lets us replace $D_{n,\xi}$ by $\overline{D}_{n,\xi}$ at an $o_{\mathbb{P}}(\lambda_n)$ cost. Proposition 2.1 controls the null edge field uniformly on this compact interval: $H_{n,\xi}(\theta) - 1 = O_{\mathbb{P}}(c_n(\theta))$, uniformly in θ and ξ. Moreover,

$$\sup_{\theta\le\theta_0} c_n(\theta) = n^{-(1-\theta_0)^2+o(1)} = o(\lambda_n), \qquad 1 - e^{-\lambda_n\theta^2} \ge c_{\theta_*}\lambda_n$$

for $\theta_* \le \theta \le \theta_0$. We therefore also have

$$\sup_{\xi=\pm 1} \sup_{\theta_* \le \theta \le \theta_0} \overline{D}_{n,\xi}(\theta) \le 1 - c\lambda_n + o_{\mathbb{P}}(\lambda_n) \qquad \text{with probability tending to one.}$$

Lemma 2.2 shows that replacing $\overline{D}_{n,\xi}$ by the empirically centered field $D_{n,\xi}$ changes the supremum over $\xi = \pm 1$ and $\theta_* \le \theta \le \theta_0$ by $o_{\mathbb{P}}(\lambda_n)$. Thus the supremum remains strictly below one with probability tending to one. □

We also need to exclude crossings near $\theta = 1$, where compact convergence does not apply and the limiting compensator diverges. The next lemma controls this endpoint uniformly for the finite-sample field. It also covers the score window of width $O(u_n^{-2})$ immediately to the right of $\theta = 1$, beyond the edge score $t = u_n$.

Lemma 2.5. *Assume $\lambda_n = n^{-\alpha+o(1)}$ for some $\alpha \in (0, 1/4)$. For every fixed $R \in [0, \infty)$,*

$$\lim_{\rho \downarrow 0} \limsup_{n\to\infty} \mathbb{P}\Big\{ \sup_{\xi=\pm 1} \sup_{1-\rho \le \theta \le 1+R/u_n^2} D_{n,\xi}(\theta) > 1 \Big\} = 0.$$

Proof. We give the argument for $\xi = 1$; the other tail is identical.

We first record the endpoint behavior of the limiting field. Its compensated negative-half-line part extends continuously to $\theta = 1$. Indeed, for any $b > 1/2$, the stochastic integrals

$$T(\theta) = \int_{-\infty}^{0} e^{\theta x}\{N - \mu\}(dx), \qquad T'(\theta) = \int_{-\infty}^{0} x e^{\theta x}\{N - \mu\}(dx)$$

have uniformly bounded second moments on $[b, 1]$. Stochastic Fubini and the one-dimensional Sobolev inequality therefore give a continuous version of T on $[b, 1]$. The positive half-line contains only finitely many Poisson points, so its random sum has a finite limit as $\theta \uparrow 1$, whereas its compensator is $(1-\theta)^{-1}$. Consequently,

$$S(\theta) \longrightarrow -\infty \qquad \text{almost surely as } \theta \uparrow 1. \tag{2.1}$$

We now treat $1 - \rho \le \theta < 1$, taking $\rho < 1/4$ and $\rho^2 < \alpha$. The Mills-ratio formula in the proof of Proposition 2.1 is uniform for $1 - \rho \le \theta \le 1$ and gives

$$c_n(\theta) = (u_n\sqrt{2\pi})^{-1} \exp\Big\{ -\frac{(1-\theta)^2 u_n^2}{2} \Big\} (1 + o(1)).$$

Thus

$$\inf_{1-\rho \le \theta < 1} c_n(\theta) = n^{-\rho^2 + o(1)} \gg \lambda_n.$$

Since $c_n(\theta) \ge n^{-\rho^2+o(1)}$ and $\rho^2 < \alpha$, $\lambda_n = o(c_n(\theta))$ uniformly on this interval. Lemma 2.2 and the fact that the damping factor is at most one show that, for every fixed $\eta > 0$, a crossing of $D_{n,1}$ implies, with probability tending to one, a crossing above level $-\eta$ by

$$X_n(\theta) := \frac{H_{n,1}(\theta) - 1}{c_n(\theta)} = \int_{\mathbb{R}} e^{\theta x}\{N_n - \mu_n\}(dx).$$

It remains to control X_n on a noncompact parameter interval. Split $[1 - \rho, 1)$ into

$$[1 - \rho, 1 - \rho/2] \quad \text{and} \quad [1 - \rho/2, 1).$$

The first interval is compactly contained in $(1/2, 1)$, so Proposition 2.1 and the portmanteau theorem give

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{1-\rho\le\theta\le 1-\rho/2} X_n(\theta) > -\eta \right\} \le \mathbb{P}\left\{ \sup_{1-\rho\le\theta\le 1-\rho/2} S(\theta) \ge -\eta \right\}.$$

By (2.1), the probability on the right tends to zero as $\rho \downarrow 0$.

For the remaining interval, fix $M > 0$ and write

$$A_{n,M}(\theta) = \int_{-\infty}^{M} e^{\theta x}\{N_n - \mu_n\}(dx), \qquad C_{n,M}(\theta) = \int_{M}^{\infty} e^{\theta x}\mu_n(dx).$$

For $3/4 \le \theta \le 1$, the Poisson isometry and the intensity formula in Proposition 2.1 give

$$\sup_{n,\theta} \mathbb{E}\{|A_{n,M}(\theta)|^2 + |A'_{n,M}(\theta)|^2\} \le C_M.$$

The Sobolev inequality therefore makes $\{\|A_{n,M}\|_\infty : n \ge 1\}$ tight. Given $\beta > 0$, choose $K = K(M, \beta)$ so that

$$\sup_n \mathbb{P}\{\|A_{n,M}\|_\infty > K\} < \beta.$$

On $\{N_n(M, \infty) = 0\}$, we have $X_n = A_{n,M} - C_{n,M}$. Since $C_{n,M}$ is increasing in θ,

$$\inf_{1-\rho/2\le\theta<1} C_{n,M}(\theta) = C_{n,M}(1-\rho/2) \longrightarrow \frac{2e^{-\rho M/2}}{\rho}$$

for each fixed M, ρ. This limit exceeds $K + \eta$ once ρ is small enough. Moreover, $\limsup_n \mathbb{P}\{N_n(M, \infty) > 0\} \le e^{-M}$. Hence

$$\lim_{\rho\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{1-\rho/2\le\theta<1} X_n(\theta) > -\eta \right\} \le \beta + e^{-M}.$$

Send first $\beta \downarrow 0$ and then $M \to \infty$. Together with the compact interval, this proves the exclusion on $1 - \rho \le \theta < 1$.

We now treat the window immediately to the right of one. For each fixed R, we claim that

$$\sup_{0\le s\le R} H_{n,1}\left(1 + \frac{s}{u_n^2}\right) \le \frac{3}{4} \qquad \text{with probability tending to one.}$$

Put $t_n(s) = u_n + s/u_n$, and split the average defining $H_{n,1}(1 + s/u_n^2)$ at u_n. For the bulk term

$$B_n(s) = \frac{1}{n}\sum_{i=1}^{n} e^{t_n(s)Z_i - t_n(s)^2/2}\mathbf{1}_{\{Z_i\le u_n\}},$$

exponential tilting gives, uniformly for $0 \le s \le R$,

$$\mathbb{E}B_n(s) = \mathbb{P}\{Z \le u_n - t_n(s)\} = \Phi(-s/u_n) \le \frac{1}{2} + o(1),$$

and

$$\operatorname{Var}(B_n(s)) \le \frac{1}{n} e^{t_n(s)^2}\Phi(u_n - 2t_n(s)) \le \frac{C_R}{u_n^2}.$$

Differentiating the summand in s gives a factor $(Z_i - t_n(s))/u_n$, and the same tilted calculation yields

$$\sup_{0\le s\le R} \mathrm{Var}(B_n'(s)) \le \frac{C_R}{u_n^2}.$$

Since both $\sup_s \mathrm{Var}(B_n(s))$ and $\sup_s \mathrm{Var}(B_n'(s))$ are $O(u_n^{-2})$, the Sobolev inequality on $[0,R]$ gives $\sup_{0\le s\le R} |B_n(s) - \mathbb{E}B_n(s)| \to 0$ in probability.

For the edge part, first restrict to $u_n < Z_i \le u_n + L/u_n$. On this event each averaged summand is at most $C_{R,L}/u_n$, using $ne^{-u_n^2/2} \asymp u_n$, while the number of such observations is tight; hence this contribution is $o_{\mathbb{P}}(1)$ for fixed L. The probability of any observation above $u_n + L/u_n$ has limit at most e^{-L}, so

$$\limsup_{n\to\infty} \mathbb{P}\left\{ \sup_{0\le s\le R} \text{edge part} > \varepsilon \right\} \le e^{-L}$$

for every $\varepsilon > 0$. Letting $L \to \infty$ proves the claim.

The damping factor is at most one, and $e^{-\theta u_n \bar{Z}_n} = 1 + o_{\mathbb{P}}(1)$ uniformly on this window. Thus the bound for $H_{n,1}$ implies, for example, $\sup_{1\le\theta\le 1+R/u_n^2} D_{n,1}(\theta) \le 0.9$ with probability tending to one. This proves the right-endpoint exclusion. Combining the two endpoint ranges and then sending $\rho \downarrow 0$ completes the proof. □

The phase theorem compares crossing probabilities with events determined by $\sup S$. To pass between strict and weak inequalities, we need this supremum to have no atom at zero. We prove that fact by conditioning on all but one Poisson point: moving the remaining point changes the supremum strictly monotonically.

Lemma 2.6. *For every $b \in (1/2, 1)$, define*

$$M_b = \sup_{b\le\theta<1} S(\theta).$$

Then

$$\mathbb{P}\{M_b = 0\} = 0.$$

The same holds for the maximum of two independent copies of M_b.

Proof. Use the version of S that is continuous on compact subintervals of $[b, 1)$ and satisfies $S(\theta) \to -\infty$ as $\theta \uparrow 1$, as constructed in the proof of Lemma 2.5. For each bounded interval I, this version can be chosen jointly for every possible location of a labeled Poisson point in I. Hence M_b is finite, and its supremum is attained on a compact subinterval $[b, 1-\delta]$.

For $m \ge 1$, let

$$I_m = [-m-1, -m].$$

Condition on the Poisson configuration outside I_m and on the number of points inside I_m. On the event $N(I_m) = k \ge 1$, use the standard labeled representation of the k points in I_m [Kal17]: after conditioning on $k-1$ labels, the remaining location, call it Y, has a continuous density on I_m proportional to e^{-y}.

For any conditioned configuration for which the supremum is finite and attained, write the resulting supremum as a function of Y:

$$\Psi(Y) = \sup_{b\le\theta<1} \{R(\theta) + e^{\theta Y}\},$$

where R includes all contributions independent of Y, including the compensator. If $Y_2 > Y_1$ and θ_1 attains the supremum for Y_1, then $\theta_1 \ge b > 0$ and

$$\Psi(Y_2) \ge R(\theta_1) + e^{\theta_1 Y_2} > R(\theta_1) + e^{\theta_1 Y_1} = \Psi(Y_1).$$

Thus Ψ is strictly increasing, so the equation $\Psi(Y) = 0$ has at most one solution. Since Y has a continuous conditional density,

$$\mathbb{P}\{M_b = 0 \mid N(I_m) = k, \text{ the conditioned data}\} = 0 \qquad (k \geq 1).$$

After averaging over the conditioning, $\mathbb{P}\{M_b = 0,\ N(I_m) \geq 1\} = 0$. Therefore

$$\mathbb{P}\{M_b = 0\} \leq \mathbb{P}\{N(I_m) = 0\} = \exp\{-\mu(I_m)\}.$$

But $\mu(I_m) = \int_{-m-1}^{-m} e^{-x}\, dx \to \infty$, so $\mathbb{P}\{M_b = 0\} = 0$.

If M_b^+ and M_b^- are independent copies, then $\mathbb{P}\{\max(M_b^+, M_b^-) = 0\} \leq \mathbb{P}\{M_b^+ = 0\} + \mathbb{P}\{M_b^- = 0\} = 0$. □

It remains to control $M_b = \sup_{b\leq\theta<1} S(\theta)$ as $b \downarrow 1/2$. The compensated integral over the negative half-line then accumulates variance of order $(\theta - 1/2)^{-1}$. On a sparse geometric grid, its normalized values approach nearly independent centered Gaussians; consequently, the probability that every grid value is nonpositive tends to zero.

Lemma 2.7. *For*

$$M_b = \sup_{b\leq\theta<1} S(\theta), \qquad b \in (1/2, 1),$$

we have

$$\mathbb{P}\{M_b > 0\} \to 1 \qquad \textit{as } b \downarrow \frac{1}{2}.$$

Proof. Split $S(\theta)$ into its negative and positive half-line contributions. On the negative half-line put $r = e^{-x}$. Then the intensity $e^{-x}dx$ becomes Lebesgue measure dr on $[1, \infty)$, and the negative contribution is

$$T(\theta) = \int_1^\infty r^{-\theta}\{N(dr) - dr\}, \qquad \theta > \frac{1}{2},$$

where N is a unit-rate Poisson process on $[1, \infty)$. The positive half-line contribution is finite and continuous on compact subintervals of $[1/2, 1)$, and hence is negligible after multiplication by $\sqrt{\theta - 1/2}$. For the finite grid of θ-values introduced below, this positive half-line contribution is almost surely bounded and vanishes after multiplication by $\sqrt{\theta - 1/2}$.

We first record the finite-dimensional limit of T near $1/2$. Fix $m \geq 1$ and positive constants $a_1, \ldots, a_m$. As $\beta \downarrow 0$,

$$\left\{\sqrt{2a_j\beta}\, T\Big(\frac{1}{2} + a_j\beta\Big)\right\}_{j=1}^m \Rightarrow (G_1, \ldots, G_m),$$

where $(G_1, \ldots, G_m)$ is centered Gaussian with covariance

$$\mathbb{E}G_iG_j = \frac{2\sqrt{a_ia_j}}{a_i + a_j}.$$

Indeed, for any $c_1, \ldots, c_m$, write

$$g_\beta(r) = \sum_{j=1}^m c_j \sqrt{2a_j\beta}\, r^{-1/2-a_j\beta}.$$

Then $\sup_{r\geq 1} |g_\beta(r)| = O(\sqrt{\beta})$, while

$$\int_1^\infty g_\beta(r)^2\, dr$$

converges to the quadratic form determined by the displayed covariance. Expanding the compensated Poisson characteristic exponent gives

$$\int_1^\infty \{e^{itg_\beta(r)} - 1 - itg_\beta(r)\}\, dr = -\frac{t^2}{2}\int_1^\infty g_\beta(r)^2\, dr + o(1),$$

because the remainder is bounded by $C_t\|g_\beta\|_\infty \int g_\beta^2 = o(1)$. This proves the stated Gaussian finite-dimensional limit.

Fix m and $q \in (0,1)$. For $b > 1/2$ set

$$\beta_b = b - \frac{1}{2}, \qquad \theta_j(b) = \frac{1}{2} + \beta_b q^{-j}, \qquad 0 \le j \le m-1.$$

For b close enough to $1/2$, all these points lie in $[b,1)$. Applying the Gaussian finite-dimensional limit just proved with $a_j = q^{-j}$ gives

$$\left\{\sqrt{2\beta_b q^{-j}}\, S(\theta_j(b))\right\}_{j=0}^{m-1} \Rightarrow (G_0, \ldots, G_{m-1}),$$

where

$$\mathbb{E}G_iG_j = \frac{2q^{|i-j|/2}}{1+q^{|i-j|}}.$$

Thus

$$\limsup_{b\downarrow 1/2} \mathbb{P}\{M_b \le 0\} \le \mathbb{P}\{G_0 \le 0, \ldots, G_{m-1} \le 0\}.$$

For fixed m, the covariance matrix above converges to the identity as $q \downarrow 0$. Hence the right-hand side tends to 2^{-m} as $q \downarrow 0$. Letting first $q \downarrow 0$ and then $m \to \infty$ proves the lemma. $\square$

3 The Phase Diagram

This section proves Theorem 1.1 and the interior continuity and strict monotonicity assertions in Proposition 1.2. The compact process limit and the exclusion estimates first give an η-sandwich between two Poisson no-crossing events; the no-atom theorem for the limiting supremum then closes the sandwich. Section 4 proves the two remaining assertions of Proposition 1.2.

3.1 The polynomial sandwich

The compact limit and the two exclusion estimates sandwich the full KKT event between Poisson events whose score intervals differ by 2η.

Theorem 3.1. *Fix $\alpha \in (0, 1/4)$ and assume*

$$\varepsilon_n = n^{-\alpha+o(1)}.$$

Let

$$\theta_\alpha = 1 - \sqrt{\alpha}.$$

Let S_+ and S_- be independent copies of the compensated Poisson process

$$S(\theta) = \int_{\mathbb{R}} e^{\theta x}\{N - \mu\}(dx), \qquad \mu(dx) = e^{-x}\, dx.$$

Here S is interpreted using the half-line construction in Proposition 2.1. For

$$0 < \eta < \min\Big\{\theta_\alpha - \frac{1}{2}, \sqrt{\alpha}\Big\},$$

define

$$p_-(\alpha, \eta) = \mathbb{P}\Big\{\max_{\xi\in\{+,-\}} \sup_{\theta_\alpha-\eta\le\theta<1} S_\xi(\theta) < 0\Big\},$$

and

$$p_+(\alpha, \eta) = \mathbb{P}\Big\{\max_{\xi\in\{+,-\}} \sup_{\theta_\alpha+\eta\le\theta<1} S_\xi(\theta) \le 0\Big\}.$$

Then

$$p_-(\alpha,\eta) \le \liminf_{n\to\infty} \mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n)| = 1\big\} \le \limsup_{n\to\infty} \mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n)| = 1\big\} \le p_+(\alpha,\eta).$$

Moreover

$$p_-(\alpha,\eta) > 0, \qquad p_+(\alpha,\eta) < 1.$$

Proof. Put

$$\lambda_n = \frac{\varepsilon_n u_n^2}{2(1-\varepsilon_n)}.$$

Then $\lambda_n = n^{-\alpha+o(1)}$, since $u_n^2/2 = \log n + o(\log n)$ and $\varepsilon_n \to 0$.

The Poisson argument in Proposition 2.1 applies jointly to both tails. After truncation to compact edge windows, the multinomial rare-event formula gives independent Poisson limits; the positive- and negative-edge truncation bounds in that proposition remove both cutoffs. Thus the two compact fields converge jointly to (S_+, S_-).

Let

$$E_n = \{\sup_a D_n(a) > 1\}.$$

By Lemma 1.5, the one-atom event is E_n^c.

For the upper bound on the one-atom probability, fix $\rho > 0$ so small that

$$\theta_\alpha + \eta < 1 - \rho.$$

On the compact interval $[\theta_\alpha + \eta, 1-\rho]$ we have

$$(1-\theta)^2 < \alpha$$

uniformly. Hence

$$\lambda_n \theta^2 / c_n(\theta) \to 0$$

uniformly there, and Proposition 2.1 plus Lemma 2.2 gives

$$\left\{\frac{D_{n,\xi}(\theta) - 1}{c_n(\theta)} : \xi = \pm 1,\ \theta \in [\theta_\alpha+\eta, 1-\rho]\right\} \Rightarrow \{S_\xi(\theta) : \xi = \pm1,\ \theta\in[\theta_\alpha+\eta,1-\rho]\}.$$

Therefore

$$\liminf_n \mathbb{P}(E_n) \ge \mathbb{P}\Big\{\max_{\xi\in\{+,-\}} \sup_{\theta_\alpha+\eta\le\theta\le1-\rho} S_\xi(\theta) > 0\Big\}.$$

Letting $\rho \downarrow 0$ and using $S_\xi(\theta) \to -\infty$ as $\theta \uparrow 1$ gives

$$\limsup_n \mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n)| = 1\big\} \le p_+(\alpha,\eta).$$

For the lower bound, Lemma 2.4 excludes crossings $D_{n,\xi}(\theta) > 1$ for

$$0 \le \theta \le \theta_\alpha - \eta$$

with high probability. Fix $R < \infty$. Lemma 2.5 lets us discard

$$1 - \rho \le \theta \le 1 + R/u_n^2$$

at a cost that vanishes as $\rho \downarrow 0$. For the part to the right of this window, use the event

$$\max_i Z_i \le u_n + \frac{R}{u_n}, \qquad \max_i(-Z_i) \le u_n + \frac{R}{u_n}.$$

Its complement has limiting probability $O(e^{-R})$, by Mills' ratio. Intersect it with

$$|\bar{Z}_n| \le \frac{\varepsilon_n}{2(1-\varepsilon_n)}\left(u_n + \frac{R}{u_n}\right),$$

an event whose probability tends to one because $\sqrt{n}\,\varepsilon_n u_n \to \infty$. On this intersection, if $\theta > 1+R/u_n^2$, then $\xi a_n(\theta)$ lies strictly beyond the corresponding centered sample extreme. Lemma 1.5 then makes $D_{n,\xi}(\theta)$ decreasing for larger θ, so any crossing there would already force a crossing in the discarded endpoint window. We send $R \to \infty$ at the end. On the remaining compact interval

$$[\theta_\alpha - \eta, 1 - \rho],$$

write

$$W_{n,\xi}(\theta) = \frac{D_{n,\xi}(\theta) - 1}{c_n(\theta)}.$$

The deterministic damping term in Proposition 2.1 is nonpositive, and the empirical-centering error is $o_{\mathbb{P}}(1)$ after division by $c_n(\theta)$ on this compact interval. Thus, for every fixed $\delta > 0$, with probability tending to one, a crossing in the compact interval implies

$$\max_{\xi\in\{+,-\}} \sup_{\theta_\alpha-\eta\le\theta\le1-\rho} \frac{H_{n,\xi}(\theta) - 1}{c_n(\theta)} \ge -\delta.$$

The compact functional convergence and the portmanteau theorem for the closed set $\{f : \sup f \ge -\delta\}$ bound the limsup of the compact crossing probability by

$$\mathbb{P}\Big\{\max_{\xi\in\{+,-\}} \sup_{\theta_\alpha-\eta\le\theta\le1-\rho} S_\xi(\theta) \ge -\delta\Big\}.$$

Letting $\delta \downarrow 0$ decreases these events to the event in which both Poisson suprema are at most zero. Hence

$$\limsup_n \mathbb{P}(E_n) \le \mathbb{P}\Big\{\max_{\xi\in\{+,-\}} \sup_{\theta_\alpha-\eta\le\theta\le1-\rho} S_\xi(\theta) \ge 0\Big\} + o_\rho(1) + O(e^{-R}).$$

Letting first $\rho \downarrow 0$ and then $R \to \infty$ gives

$$\liminf_n \mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n)| = 1\big\} \ge p_-(\alpha,\eta).$$

We finish by proving that neither sandwich bound is trivial. For $p_- > 0$, put $b = \theta_\alpha - \eta > 1/2$. We show that one copy of S is strictly negative on $[b, 1)$ with positive probability. Choose M large. On the event that N has no points in $[-M, \infty)$, the contribution of this interval is the deterministic negative function

$$-\int_{-M}^{\infty} e^{\theta x} e^{-x}\, dx = -\frac{e^{(1-\theta)M}}{1-\theta}, \qquad b \le \theta < 1.$$

If $M > 1/(1-b)$, the positive compensator $e^{(1-\theta)M}/(1-\theta)$ has minimum eM over $[b, 1)$. Let T_M be the compensated contribution from $(-\infty, -M)$. By the L^2 isometry, applied also to the θ-derivative, $\sup_{b\le\theta\le 1} |T_M(\theta)| \to 0$ in probability. Thus, for all large M, the event

$$\sup_{b\le\theta\le 1} T_M(\theta) \le e\,M/2$$

has positive probability. This event is independent of $\{N([-M, \infty)) = 0\}$, and their intersection has positive probability. On that intersection $S(\theta) < 0$ for all $b \le \theta < 1$. Since S_+ and S_- are independent, both copies are negative on $[b, 1)$ with positive probability. Hence $p_-(\alpha, \eta) > 0$.

For $p_+ < 1$, choose any $\theta_0 \in (\theta_\alpha + \eta, 1)$. The random variable $S_+(\theta_0)$ is centered, integrable, and nondegenerate. Hence $\mathbb{P}\{S_+(\theta_0) > 0\} > 0$, which gives $p_+(\alpha, \eta) < 1$. □

3.2 Continuity and strict monotonicity

The sandwich identifies the interior limit, but it does not by itself show that changing the exponent changes the limiting probability. We now prove the interior regularity assertions in Proposition 1.2.

Proof of the interior assertions in Proposition 1.2. With

$$\mathcal{M}_b = \max_{\xi\in\{+,-\}} \sup_{b\le\theta<1} S_\xi(\theta),$$

fix $0 < r < 1/4$ and write $\theta_r = 1 - \sqrt{r}$. The proof of $p_- > 0$ in Theorem 3.1, applied with $b = \theta_r$, gives $\mathbb{P}\{\mathcal{M}_{\theta_r} < 0\} > 0$, and hence $p(r) > 0$. The strict inequality $p(r) < 1$ follows by evaluating one copy at any fixed $\theta_0 \in (\theta_r, 1)$: the centered variable $S_+(\theta_0)$ is nondegenerate, so it is positive with positive probability.

For monotonicity and continuity, put

$$F(b) = \mathbb{P}\{\mathcal{M}_b \le 0\}.$$

The random variables $\mathcal{M}_b$ decrease as b increases, while $\theta_r = 1 - \sqrt{r}$ decreases as r increases. Hence $p(r) = F(\theta_r)$ is nonincreasing in r.

To prove strictness, for one copy of the field put

$$\pi(b) = \mathbb{P}\Big\{\sup_{b\le\theta<1} S(\theta) \le 0\Big\}, \qquad b \in (1/2, 1),$$

so that $F(b) = \pi(b)^2$. Fix $1/2 < b_0 < b_1 < 1$. The event defining $\pi(b_0)$ is contained in the event defining $\pi(b_1)$; to make the containment strict, it suffices to find a positive-probability event on which S is positive somewhere in $[b_0, b_1)$ but negative on $[b_1, 1)$.

Under the change of variables $y = e^{-x}$, the field has the representation

$$S(\theta) = \int_0^\infty y^{-\theta}\{\Pi(dy) - dy\},$$

where Π is a unit-rate Poisson process. Choose a constant c strictly between $1/(1-b_0)$ and $1/(1-b_1)$. For large Y, set $k = \lfloor cY \rfloor$. Choose $\delta > 0$ small and put $L = (1+\delta)Y$. On the event E_Y that Π has no points in $(0, Y)$ and exactly k points in $[Y, L]$, the contribution from $(0, L]$, including its compensator, is bounded below at b_0 by

$$kL^{-b_0} - \frac{L^{1-b_0}}{1-b_0} \geq mY^{1-b_0}$$

and bounded above, uniformly for $\theta \geq b_1$, by

$$kY^{-\theta} - \frac{L^{1-\theta}}{1-\theta} \leq -mY^{1-\theta},$$

for some $m > 0$ and all large Y. These inequalities follow by first setting $\delta = 0$, where they are precisely the two strict inequalities defining c, and then taking δ small. Notice that E_Y has positive probability for every fixed Y.

It remains to control the independent compensated tail

$$R_L(\theta) = \int_L^\infty y^{-\theta}\{\Pi(dy) - dy\}.$$

For $Q_Y(\theta) = Y^{\theta-1}R_L(\theta)$, the L^2 isometry gives, uniformly on $[b_0, 1]$,

$$\mathbb{E}|Q_Y(\theta)|^2 + \mathbb{E}|Q_Y'(\theta)|^2 \leq \frac{C}{Y}.$$

The derivative estimate follows by inserting $\log(Y/y)$ in the integrand. The Sobolev inequality therefore yields $\sup_{b_0 \leq \theta \leq 1} |Q_Y(\theta)| \to_{\mathbb{P}} 0$. For all large Y, the tail preserves both strict signs above with positive probability. By independence from E_Y, the separating event has positive probability, so $\pi(b_0) < \pi(b_1)$, hence $F(b_0) < F(b_1)$. Because $r \mapsto \theta_r$ is strictly decreasing, $p(r) = F(\theta_r)$ is strictly decreasing on $(0, 1/4)$.

If $b_m \downarrow b$, then the domains $[b_m, 1)$ increase and $\mathcal{M}_{b_m} \uparrow \mathcal{M}_b$ almost surely. Therefore $\mathbf{1}_{\{\mathcal{M}_{b_m} \leq 0\}} \downarrow \mathbf{1}_{\{\mathcal{M}_b \leq 0\}}$, and

$$\lim_m F(b_m) = F(b).$$

If $b_m \uparrow b$, then the domains decrease and $\mathcal{M}_{b_m} \downarrow \mathcal{M}_b$ almost surely; the endpoint fact $S_\xi(\theta) \to -\infty$ as $\theta \uparrow 1$ reduces the supremum to compact subintervals, where the paths are continuous. Hence $\mathbf{1}_{\{\mathcal{M}_{b_m} \leq 0\}} \uparrow \mathbf{1}_{\{\mathcal{M}_b < 0\}}$. Lemma 2.6 gives $\mathbb{P}\{\mathcal{M}_b = 0\} = 0$, so $F(b_m) \to F(b)$. This proves continuity of F, and hence of p, on the interior interval. □

Because the limiting supremum has no atom at zero, the strict and weak no-crossing events have the same probability. The sandwich therefore yields the interior case of Theorem 1.1.

Proposition 3.2. *Fix* $\alpha \in (0, 1/4)$ *and assume*

$$\varepsilon_n = n^{-\alpha + o(1)}.$$

Then

$$\mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n)| = 1\big\} \to p(\alpha).$$

Proof. Write

$$\mathcal{M}_b = \max_{\xi \in \{+,-\}} \sup_{b \le \theta < 1} S_\xi(\theta).$$

Lemma 2.6 gives $\mathbb{P}\{\mathcal{M}_{\theta_\alpha} = 0\} = 0$. As $\eta \downarrow 0$,

$$p_-(\alpha, \eta) = \mathbb{P}\{\mathcal{M}_{\theta_\alpha - \eta} < 0\} \uparrow \mathbb{P}\{\mathcal{M}_{\theta_\alpha} < 0\}.$$

Here the increasing convergence follows from continuity on compact subintervals and the endpoint fact $S_\xi(\theta) \to -\infty$ as $\theta \uparrow 1$. Similarly,

$$p_+(\alpha, \eta) = \mathbb{P}\{\mathcal{M}_{\theta_\alpha + \eta} \le 0\} \downarrow \mathbb{P}\{\mathcal{M}_{\theta_\alpha} \le 0\} = \mathbb{P}\{\mathcal{M}_{\theta_\alpha} < 0\}.$$

The η-sandwich in Theorem 3.1 now forces the asserted limit. $\square$

Proof of Theorem 1.1. For $0 < r < 1/4$, the assumption $r_n \to r$ is exactly $\varepsilon_n = n^{-r+o(1)}$, so the result is Proposition 3.2.

Suppose $r = 0$. Fix $\alpha \in (0, 1/4)$ and let $\varepsilon_n^{(\alpha)}$ be chosen so that $\varepsilon_n^{(\alpha)} = n^{-\alpha}$. Since $r_n \to 0$, we have $\varepsilon_n \ge \varepsilon_n^{(\alpha)}$ for all large n. Coupling both samples through the same standardized variables Z_i, Lemma 1.7 gives

$$\big\{|\operatorname{supp}(\widehat{G}_n^{(\varepsilon_n^{(\alpha)})})| = 1\big\} \subseteq \big\{|\operatorname{supp}(\widehat{G}_n^{(\varepsilon_n)})| = 1\big\}$$

for all large n, where $\widehat{G}_n^{(\varepsilon)}$ denotes the NPMLE computed from $\sqrt{1-\varepsilon}\, Z_1, \ldots, \sqrt{1-\varepsilon}\, Z_n$. Therefore

$$\liminf_{n\to\infty} \mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n^{(\varepsilon_n)})| = 1\big\} \ge p(\alpha).$$

Letting $\alpha \downarrow 0$ and using $p(\alpha) \to 1$ gives the limit one.

Finally suppose either $r_n \to r \ge 1/4$, with $r = \infty$ allowed, or more generally $\liminf_n r_n \ge 1/4$. Fix $\alpha \in (0, 1/4)$. Then $\varepsilon_n \le \varepsilon_n^{(\alpha)}$ for all large n. Monotonicity gives

$$\big\{|\operatorname{supp}(\widehat{G}_n^{(\varepsilon_n)})| = 1\big\} \subseteq \big\{|\operatorname{supp}(\widehat{G}_n^{(\varepsilon_n^{(\alpha)})})| = 1\big\}$$

for all large n. Hence

$$\limsup_{n\to\infty} \mathbb{P}\big\{|\operatorname{supp}(\widehat{G}_n^{(\varepsilon_n)})| = 1\big\} \le p(\alpha).$$

Letting $\alpha \uparrow 1/4$ and using $p(\alpha) \to 0$ gives the limit zero. $\square$

4 Endpoint Behavior

This section proves the two endpoint asymptotics in Proposition 1.2. They have different probabilistic origins. Near $r = 0$, a crossing is caused by one unusually large positive Poisson point. Near $r = 1/4$, many small jumps accumulate and, after logarithmic reparametrization, behave like a stationary Gaussian process.

4.1 Lower endpoint and continuity

This subsection proves the $r \downarrow 0$ assertion of Proposition 1.2. A single large positive Poisson point governs the one-sided crossing probability.

Lemma 4.1. *For one copy of the limiting Poisson field,*

$$R_{1-\varepsilon} := \mathbb{P}\Big\{ \sup_{1-\varepsilon \le \theta < 1} S(\theta) > 0 \Big\} \sim \varepsilon \qquad (\varepsilon \downarrow 0).$$

Proof. Write $h = 1 - \theta$. Split $S(1-h) = A_h + B_h$, where

$$A_h = \int_0^\infty e^{(1-h)x} \{N - \mu\}(dx), \qquad B_h = \int_{-\infty}^0 e^{(1-h)x} \{N - \mu\}(dx).$$

Enumerate the points of $N \cap [0, \infty)$ as X_j, and put $W_j = e^{X_j}$. Then the W_j form a Poisson point process on $[1, \infty)$ with intensity $w^{-2}\, dw$, and hence

$$A_h = \sum_j W_j^{1-h} - \frac{1}{h}.$$

The negative half-line is negligible at the scale $1/h$. Indeed, for every fixed $q \ge 2$, the compensated Poisson integrals defining B_h and $\partial_h B_h$ have uniformly bounded qth moments for $0 \le h \le 1/4$. The fundamental theorem of calculus therefore gives $\mathbb{E} \sup_{0 \le h \le 1/4} |B_h|^q < \infty$. Consequently, with $K_\varepsilon = \varepsilon^{-1/3}$ and any $q > 3$,

$$\mathbb{P}\Big\{ \sup_{0<h\le\varepsilon} |B_h| > K_\varepsilon \Big\} = o(\varepsilon).$$

It therefore suffices to estimate the positive-half-line crossing probability with a perturbation of size K_ε. For each fixed sign $\sigma \in \{-1, 1\}$, consider

$$\mathbb{P}\Big\{ \sup_{0<h\le\varepsilon} \Big(\sum_j W_j^{1-h} - \frac{1}{h} \Big) > \sigma K_\varepsilon \Big\}.$$

The sign does not affect the first-order answer, since $K_\varepsilon = o(\varepsilon^{-1})$.

For fixed h, put $Y_j = W_j^{1-h}$. Then the Y_j form a Poisson cloud on $[1, \infty)$ with tail intensity

$$\nu_h(y, \infty) = y^{-1/(1-h)}.$$

This regularly varying tail is subexponential; we need its one-large-jump estimate uniformly over $0 < h \le \varepsilon$ [FKZ13, Chapter 4]. Write

$$\beta = (1-h)^{-1}, \qquad x = x_h = h^{-1} + \sigma K_\varepsilon, \qquad \delta = 1/\log x.$$

The total number of points is $\mathrm{Pois}(1)$, and

$$\nu_h(y, \infty) = y^{-\beta}, \qquad \int_1^{\delta x} y\, \nu_h(dy) = \beta \int_1^{\delta x} y^{-\beta}\, dy \le C \log x$$

uniformly for $0 < h \le \varepsilon$, once $\varepsilon < 1/3$. If $\sum_j Y_j > x$ but $\max_j Y_j \le x$, then at least one of the following events occurs:

1. there is a point in $((1-\delta)x, x]$;
2. there are at least two points above δx;
3. every point is at most δx, and the total number of points exceeds δ^{-1};
4. there is exactly one point in $(\delta x, (1-\delta)x]$, and the sum of the points below δx exceeds δx.

The Poisson formula, followed by Markov's inequality in the last case, bounds the four probabilities by

$$C\delta x^{-\beta}, \qquad C(\delta x)^{-2\beta}, \qquad \mathbb{P}\{\mathrm{Pois}(1) > \delta^{-1}\}, \qquad C(\delta x)^{-\beta}\frac{\log x}{\delta x},$$

respectively. Since $1/2 \le hx \le 2$, their sum is $a_\varepsilon h$ for a deterministic modulus $a_\varepsilon \downarrow 0$, uniformly over $0 < h \le \varepsilon$. Thus

$$\mathbb{P}\left\{\sum_j Y_j > x_h,\ \max_j Y_j \le x_h\right\} \le a_\varepsilon h.$$

To pass from fixed h to the supremum, choose $\delta_\varepsilon \downarrow 0$ so slowly that $a_\varepsilon/\delta_\varepsilon \to 0$ and $K_\varepsilon \varepsilon = o(\delta_\varepsilon)$, and use the mesh $h_k = \varepsilon(1+\delta_\varepsilon)^{-k}$. Fix a large numerical constant C and define the enlarged one-large-jump event

$$\mathcal{J}_\varepsilon^{\mathrm{enl}} = \left\{\max_j W_j > \inf_{0<h\le\varepsilon}\left[(1-C\delta_\varepsilon)x_h\right]^{1/(1-h)}\right\}.$$

The infimum is attained at $h = \varepsilon$ for all small ε. Indeed, the derivative of $\log\{[(1-C\delta_\varepsilon)x_h]^{1/(1-h)}\}$ has the sign of

$$\log\{(1-C\delta_\varepsilon)x_h\} - \frac{1-h}{h(1+\sigma K_\varepsilon h)},$$

which is negative uniformly on $0 < h \le \varepsilon$, because $K_\varepsilon\varepsilon = o(\delta_\varepsilon)$ and $h^{-1} \gg \log(1/h)$. The intensity of the enlarged event is therefore

$$\left[(1-C\delta_\varepsilon)x_\varepsilon\right]^{-1/(1-\varepsilon)} = (1 + O(\delta_\varepsilon) + o(1))\varepsilon.$$

Thus the enlarged event differs from the exact one-large-jump event by probability $o(\varepsilon)$.

On the complement of $\mathcal{J}_\varepsilon^{\mathrm{enl}}$, every atom lies below the enlarged threshold. If $h_{k+1} < h \le h_k$, then $W^{1-h_{k+1}} \ge W^{1-h}$ for every $W \ge 1$, while $x_{h_{k+1}}/x_h = 1 + O(\delta_\varepsilon)$ uniformly in k. Thus a crossing somewhere in the kth cell gives, at h_{k+1}, a subcritical-sum event at threshold $(1-C\delta_\varepsilon)x_{h_{k+1}}$, after increasing C if necessary. The fixed-h calculation above applies uniformly to this threshold as well, because $h_{k+1}(1-C\delta_\varepsilon)x_{h_{k+1}}$ stays between two fixed positive constants; enlarge the modulus $a_\varepsilon \downarrow 0$ if necessary to cover both thresholds. On the complement of $\mathcal{J}_\varepsilon^{\mathrm{enl}}$, its largest atom is below this same threshold. The probability of the mesh event is therefore at most $Ca_\varepsilon h_k$. Since $\sum_k h_k = O(\varepsilon/\delta_\varepsilon)$, the total contribution is $o(\varepsilon)$.

Thus, to first order, a crossing is caused by one atom. For either sign σ, differentiation shows that the required atom threshold is decreasing on $0 < h \le \varepsilon$, for all small ε, and is therefore minimized at $h = \varepsilon$:

$$\inf_{0<h\le\varepsilon}(h^{-1} + \sigma K_\varepsilon)^{1/(1-h)} = (\varepsilon^{-1} + \sigma K_\varepsilon)^{1/(1-\varepsilon)} = \varepsilon^{-1/(1-\varepsilon)}(1 + o(1)).$$

The intensity of atoms above this threshold is

$$\int_{(\varepsilon^{-1}+\sigma K_\varepsilon)^{1/(1-\varepsilon)}}^{\infty} w^{-2}\,dw = (\varepsilon^{-1} + \sigma K_\varepsilon)^{-1/(1-\varepsilon)} \sim \varepsilon.$$

Since this intensity tends to zero, the probability of at least one such atom is also asymptotic to ε. Combining this with the $o(\varepsilon)$ contribution of the negative half-line and of the subcritical positive cloud proves $R_{1-\varepsilon} \sim \varepsilon$. $\square$

Taking $\varepsilon = \sqrt{r}$ in Lemma 4.1 and using the independence of S_+ and S_- gives

$$p(r) = \left(1 - R_{1-\sqrt{r}}\right)^2, \qquad 1 - p(r) \sim 2\sqrt{r} \quad (r \downarrow 0).$$

In particular, $p(r) \to 1 = p(0)$ at the lower endpoint.

4.2 Upper-endpoint persistence exponent

This subsection proves the $r \uparrow 1/4$ assertion of Proposition 1.2. Lemma 2.7 gives convergence to zero but not its rate; the sharper estimate comes from the persistence exponent of a stationary Gaussian process with sech covariance.

Lemma 4.2. *Let $G = (G(t))_{t\in\mathbb{R}}$ be the centered stationary Gaussian process with covariance*

$$\mathbb{E}G(s)G(t) = \operatorname{sech}\left(\frac{t-s}{2}\right).$$

For every deterministic $a_T \to 0$,

$$\mathbb{P}\{G(t) \le a_T \text{ for all } 0 \le t \le T\} = \exp\left\{-\frac{3}{16}T + o(T)\right\}. \tag{4.1}$$

Proof. To keep track of the normalization, we derive the stated exponent directly from the cited results. Dembo–Poonen–Shao–Zeitouni introduce, in [DPSZ02, equation (1.3)], the constant

$$b = -4 \lim_{T\to\infty} \frac{1}{T} \log \mathbb{P}\{G(t) \le 0 \text{ for all } 0 \le t \le T\}.$$

Here their process Y is exactly the centered stationary Gaussian process with correlation $\operatorname{sech}(t/2)$; see [DPSZ02, equation (1.4) and the sentence following it]. Their Theorem 1.1 is the no-real-zero statement, and Theorems 1.3 and 1.4 identify the same constant b as the persistence exponent obtained after the standard reduction from the polynomial near ± 1 to the stationary sech process.

FitzGerald–Tribe–Zaboronski rigorously compute this constant. In [FTZ22, Proposition 5] they prove that the Kac-polynomial persistence probability satisfies

$$\lim_{n\to\infty} \frac{\log p_{2n+1}}{\log n} = -\frac{3}{4}.$$

Their proof explicitly invokes the Dembo–Poonen–Shao–Zeitouni characterization: see [FTZ22, equations (19)–(20)]. Their Pfaffian formula for the sech-kernel gap probability then gives [FTZ22, equation (21)]

$$\log \mathbb{P}\{G(t) \neq 0 \text{ for all } 0 \le t \le T\} = -\frac{3}{16}T + o(T). \tag{*}$$

Because G is symmetric and its zeros are almost surely simple, the no-zero event on $[0, T]$ is the disjoint union of the positive- and negative-persistence events, which have equal probability. Simplicity follows, for example, from Bulinskaya's lemma, since the sech process is smooth and $(G(t), G'(t))$ is nondegenerate for each fixed t. Multiplication by the constant factor 2 does not affect the exponential rate, so $(*)$ gives

$$\mathbb{P}\{G(t) \le 0 \text{ for all } 0 \le t \le T\} = \exp\left\{-\frac{3}{16}T + o(T)\right\}.$$

The same value was identified earlier in the physics literature by Poplavskyi–Schehr [PS18].

We also need persistence at a vanishing deterministic level $a_T = o(1)$, rather than exactly at zero. Feldheim–Feldheim–Mukherjee prove existence of fixed-level persistence exponents in [FFM25, Theorem 1] and continuity of the exponent in the level in [FFM25, Theorem 4(II)]. The sech process is covered by their hypotheses: its spectral density is proportional to $\operatorname{sech}(\pi\lambda)$, hence is absolutely continuous, positive at the origin, and has exponential tails. Their statements use strict inequalities, but continuity in the level lets us squeeze the weak event needed here between the two fixed levels $\pm\delta$. For every fixed $\delta > 0$ and all large T,

$$\{G(t) \le -\delta \text{ for all } 0 \le t \le T\} \subseteq \{G(t) \le a_T \text{ for all } 0 \le t \le T\} \subseteq \{G(t) \le \delta \text{ for all } 0 \le t \le T\}.$$

Taking logarithms, dividing by T, and then letting $\delta \downarrow 0$ gives (4.1). □

The factor $3/16$ is the persistence exponent in logarithmic time T. In Proposition 4.3, T will be comparable to $\log(1/\beta)$, so the one-sided no-crossing probability is $\beta^{3/16+o(1)}$; the two-sided one-atom probability is its square, hence has exponent $3/8$. This endpoint refinement is not needed for the phase diagram in Theorem 1.1, but it identifies the rate at which the phase curve reaches zero.

Proposition 4.3. *As $\alpha \uparrow 1/4$,*

$$p(\alpha) = \left(\frac{1}{4} - \alpha\right)^{3/8+o(1)}.$$

Equivalently, if

$$q(\alpha) := \mathbb{P}\left\{\sup_{\theta_\alpha \le \theta < 1} S(\theta) \le 0\right\}, \qquad \theta_\alpha = 1 - \sqrt{\alpha},$$

then $q(\alpha) = (1/4 - \alpha)^{3/16+o(1)}$.

Proof. The proof has two main inputs. Near $\theta = 1/2$, the contribution of the negative half-line has a Gaussian tangent process, and in logarithmic distance from the endpoint this tangent process has the sech covariance of Lemma 4.2. Away from $\theta = 1/2$, the remaining interval has only a subexponential persistence cost.

Put $\beta = \theta_\alpha - 1/2 = 1/2 - \sqrt{\alpha}$. Then $\beta = (1/4 - \alpha)/(1/2 + \sqrt{\alpha}) \sim 1/4 - \alpha$ as $\alpha \uparrow 1/4$. Define the one-sided no-crossing probability

$$F(\beta) = \mathbb{P}\left\{S\left(\frac{1}{2} + h\right) \le 0 \text{ for every } \beta \le h < \frac{1}{2}\right\}.$$

By definition, $q(\alpha) = F(\beta)$ and $p(\alpha) = F(\beta)^2$. It is therefore enough to prove

$$F(\beta) = \beta^{3/16+o(1)} \qquad (\beta \downarrow 0). \tag{4.2}$$

Split the Poisson field as $S(1/2 + h) = T_h + U_h$, where

$$T_h = \int_{-\infty}^{0} e^{(1/2+h)x}\{N - \mu\}(dx), \qquad U_h = \int_{0}^{\infty} e^{(1/2+h)x}\{N - \mu\}(dx).$$

The negative half-line is the singular part as $h \downarrow 0$: its variance is of order $1/h$. The positive half-line stays tight on the short window $0 < h \le h_0$ and will only move the persistence threshold by $o(1)$ after normalization by $\sqrt{h}$.

After the change of variables $r = e^{-x}$ on $x < 0$,

$$T_h = \int_1^\infty r^{-1/2-h}\{\Pi(dr) - dr\},$$

where Π is a unit-rate Poisson process on $[1,\infty)$. Let $M(r) = \Pi([1,r]) - (r-1)$. Integration by parts gives

$$T_h = \left(\frac{1}{2} + h\right)\int_1^\infty r^{-3/2-h} M(r)\, dr.$$

We use the KMT approximation for a Poisson counting process. Corollary 5.3 of Chapter 7 in Ethier–Kurtz [EK86, pp. 358–359] gives a coupling of M to a Brownian motion B on $[1,\infty)$, with $B(1) = 0$ and covariance $\mathbb{E}B(r)B(s) = (r \wedge s) - 1$, for which

$$\Gamma := \sup_{r\ge 1} \frac{|M(r) - B(r)|}{C_0(1+\log r)}$$

has a finite exponential moment. Consequently, for constants $C, c > 0$ and every $y \ge 1$,

$$\mathbb{P}\left\{\sup_{r\ge 1} \frac{|M(r) - B(r)|}{C_0(1+\log r)} > y\right\} \le Ce^{-cy}. \tag{4.3}$$

This is the Poisson-process form of the Komlós–Major–Tusnády strong approximation [KMT75, KMT76]. For completeness, the weighted infinite-horizon form also follows from the finite-horizon estimate by applying it at horizons e^{m+1}: on the block $e^m \le r \le e^{m+1}$, a violation at level y requires an error of order $y(m+1)$, and summing the resulting $Ce^{-cy(m+1)}$ bounds over $m \ge 0$ gives (4.3). Define the Gaussian comparison field

$$\widetilde{T}_h = \int_1^\infty r^{-1/2-h}\, dB(r) = \left(\frac{1}{2} + h\right)\int_1^\infty r^{-3/2-h} B(r)\, dr.$$

On the complement of the event in (4.3),

$$\sqrt{2h}\,|T_h - \widetilde{T}_h| \le Cy\sqrt{h}\int_1^\infty (1+\log r) r^{-3/2-h}\, dr \le Cy\sqrt{h}.$$

Let $h_0 = (\log(1/\beta))^{-4}$ and $T_\beta = \log(h_0/\beta)$. Taking $y = AT_\beta$ with A fixed and large gives, outside an event of probability $O(e^{-cAT_\beta})$,

$$\sup_{\beta\le h\le h_0} \sqrt{2h}\,|T_h - \widetilde{T}_h| \le \Delta_\beta, \tag{4.4}$$

where $\Delta_\beta = C_A T_\beta \sqrt{h_0} = o(1)$.

Now parametrize $h = \beta e^t$, $0 \le t \le T_\beta$, and set $G_\beta(t) = \sqrt{2\beta e^t}\,\widetilde{T}_{\beta e^t}$. This process has exactly the law of G restricted to $[0, T_\beta]$, since

$$\mathbb{E}G_\beta(s)G_\beta(t) = 2\sqrt{\beta e^s\,\beta e^t}\int_1^\infty r^{-1-\beta e^s - \beta e^t}\, dr = \frac{2e^{(s+t)/2}}{e^s + e^t} = \operatorname{sech}\left(\frac{t-s}{2}\right).$$

Thus the logarithmic distance from the endpoint, rather than the original parameter h, is the natural time variable. The covariance computation is exact; the only approximation in the finite window is the KMT replacement of the Poisson counting process by Brownian motion.

We first estimate the finite window

$$E_\beta(h_0) = \left\{ S\left(\frac{1}{2}+h\right) \le 0 \text{ for every } \beta \le h \le h_0 \right\}.$$

For $0 < h \le h_0 < 1/4$,

$$U_h = \sum_{x_j>0} e^{(1/2+h)x_j} - \frac{1}{1/2-h} \ge -4.$$

Thus $E_\beta(h_0)$ implies $T_h \le 4$ for every $\beta \le h \le h_0$, hence $\sqrt{2h}T_h \le 4\sqrt{2h_0}$ on the same window. Combining this implication with (4.4), and choosing A in (4.3) large enough that the coupling error is negligible on the $\exp\{-(3/16)T_\beta\}$ scale, gives

$$\mathbb{P}(E_\beta(h_0)) \le \mathbb{P}\{G(t) \le 4\sqrt{2h_0} + \Delta_\beta \text{ for all } 0 \le t \le T_\beta\} + o\Big(e^{-3T_\beta/16}\Big).$$

By (4.1), this upper bound is $\exp\{-(3/16)T_\beta + o(T_\beta)\}$.

For the lower bound, let

$$V = \sum_{x_j>0} e^{3x_j/4}.$$

Since $\mathbb{E}V = \int_0^\infty e^{-x/4}\,dx = 4$, choose $K < \infty$ such that $c_K := \mathbb{P}\{V \le K\} > 0$. For all small β, $h_0 < 1/4$, so $U_h \le V$ for every $0 < h \le h_0$. On $\{V \le K\}$, the event $\{T_h \le -K \text{ for all } \beta \le h \le h_0\}$ implies $E_\beta(h_0)$. The positive and negative half-line Poisson processes are independent, so

$$\mathbb{P}(E_\beta(h_0)) \ge c_K\,\mathbb{P}\{T_h \le -K \text{ for all } \beta \le h \le h_0\}.$$

On the coupling event (4.4), the condition

$$G_\beta(t) \le -K\sqrt{2h_0} - \Delta_\beta \qquad (0 \le t \le T_\beta)$$

implies $T_h \le -K$ throughout $\beta \le h \le h_0$. The same choice of A makes the complement of the coupling event negligible on this scale. Another application of (4.1), with the fixed factor c_K absorbed into the $o(T_\beta)$ term, therefore gives

$$\mathbb{P}(E_\beta(h_0)) \ge \exp\left\{-\frac{3}{16}T_\beta + o(T_\beta)\right\}.$$

Together with the upper bound,

$$\mathbb{P}(E_\beta(h_0)) = \exp\left\{-\frac{3}{16}T_\beta + o(T_\beta)\right\} = \beta^{3/16+o(1)}, \tag{4.5}$$

because $T_\beta = \log(1/\beta) + o(\log(1/\beta))$.

The upper bound in (4.2) follows immediately from $F(\beta) \le \mathbb{P}(E_\beta(h_0))$. For the lower bound, use the Harris–FKG inequality for Poisson processes: two decreasing events of a Poisson configuration are positively correlated; see, for example, [Kal17, Theorem 20.4]. For an interval $I \subset (0, 1/2)$, the event

$$\left\{ S\left(\frac{1}{2}+h\right) \le 0 \text{ for every } h \in I \right\}$$

is decreasing in the Poisson configuration, because adding a point increases $S(\theta)$ for every θ. Hence

$$F(\beta) \ge \mathbb{P}(E_\beta(h_0))\,\mathbb{P}\left\{ S\left(\frac{1}{2}+h\right) \le 0 \text{ for every } h_0 \le h < \frac{1}{2} \right\}.$$

We now show that the second factor is subexponential on the $\log(1/\beta)$ scale.

The interval $[h_0, 1/2)$, which lies away from the singular endpoint $h = 0$, contributes only a subexponential factor. We cover its initial portion by a slowly expanding sequence of windows. Each window has the same sech-persistence estimate with a new left endpoint, and the sum of their logarithmic costs is only $O(\log(1/h_0))$.

Choose $h_* > 0$ so small that $(\log(1/h))^{-4} > h$ for $0 < h \le h_*$. Starting from the present h_0, define $h_{j+1} = (\log(1/h_j))^{-4}$ and stop when h_j first exceeds h_*. The finite-window lower bound just proved applies uniformly with β replaced by any sufficiently small left endpoint a and h_0 replaced by $b = (\log(1/a))^{-4}$: the KMT error is still $o(1)$ after normalization, the positive-half-line event $\{V \le K\}$ has the same fixed probability, and the level shift in Lemma 4.2 remains $o(1)$. Thus

$$\mathbb{P}\left\{S\left(\frac{1}{2}+h\right) \le 0 \text{ for every } a \le h \le b\right\} \ge \exp\left\{-O\left(1+\log\frac{b}{a}\right)\right\}, \qquad b = (\log(1/a))^{-4},$$

successively on the intervals $[h_j, h_{j+1}]$. Using FKG again gives

$$\mathbb{P}\left\{S\left(\frac{1}{2}+h\right) \le 0 \text{ for every } h_0 \le h \le h_*\right\} \ge \exp\{-O(\log(1/h_0))\}.$$

Since $\log(1/h_0) = O(\log\log(1/\beta)) = o(\log(1/\beta))$, this factor is $\exp\{-o(\log(1/\beta))\}$.

Finally, $\mathbb{P}\{S(1/2+h) \le 0 \text{ for every } h_* \le h < 1/2\} > 0$ by the positive-probability construction used for $p_-(\alpha, \eta) > 0$ in Theorem 3.1, with $b = 1/2 + h_*$. Multiplying the three lower bounds gives $F(\beta) \ge \beta^{3/16+o(1)}$. Together with the matching upper bound, this proves (4.2); the identities $q(\alpha) = F(\beta)$ and $p(\alpha) = F(\beta)^2$ finish the proof. □

Proposition 4.3 implies $p(r) \to 0 = p(1/4)$ as $r \uparrow 1/4$. Together with the interior continuity proved in Section 3, this completes the proof of Proposition 1.2.

5 Support Size and Its Limiting Law

This section proves Theorem 1.3 and Proposition 1.4. The phase diagram only decides whether the central atom is globally optimal; controlling the entire support requires excluding cascades of additional KKT contacts. We first show that the small fitted mass near each sample edge cannot create such a cascade. This localization reduces the two edges to compact random-analytic optimization problems. We then identify their unique optimizers, transfer the active sets back to the finite-sample likelihood, and finally let $r \uparrow 1/4$ to prove divergence of the limiting count.

5.1 Likelihood localization

The first step toward Theorem 1.3 is a deterministic comparison around a fit that already carries a small amount of mass away from its central atom. The comparison shows that such edge mass cannot create KKT contacts where the central directional field has a quantitative gap below one.

It is convenient to work in score coordinates. Replace a component location u by $t = \sqrt{1-\varepsilon_n}\, u$; this bijection preserves support size. We reuse G for the transformed mixing measure and set

$$\kappa_n = \frac{\varepsilon_n}{2(1-\varepsilon_n)}, \qquad \gamma_n = (1-\varepsilon_n)^{-1} = 1 + 2\kappa_n.$$

If the central atom is at τ, the likelihood ratio of the atom at $\tau + a$ to the atom at τ, evaluated at Z_i, is

$$q_{n,\tau,a}(i) = \exp\left\{a(Z_i - \gamma_n \tau) - \frac{\gamma_n a^2}{2}\right\}.$$

We also write

$$q_t(z) = \exp\left\{tz - \frac{\gamma_n t^2}{2}\right\}$$

for the likelihood ratio of the score atom t to the atom at zero. For a perturbation $G = (1 - W)\delta_\tau + G^{\mathrm{out}}$, write

$$m_G(i) = 1 - W + \int q_{n,\tau,u-\tau}(i)\, G^{\mathrm{out}}(du)$$

for its likelihood ratio to δ_τ, and define

$$D_{n,\tau}(a) = \frac{1}{n}\sum_{i=1}^{n} q_{n,\tau,a}(i), \qquad \Psi_G(\tau + a) = \frac{1}{n}\sum_{i=1}^{n} \frac{q_{n,\tau,a}(i)}{m_G(i)}.$$

The KKT conditions are $\Psi_G \leq 1$, with equality G-almost everywhere; continuity of Ψ_G upgrades this to equality at every support point.

The deterministic no-propagation estimate below separates the total mass of the edge atoms from the geometry of their locations.

Lemma 5.1. *For every measure $G = (1 - W)\delta_\tau + G^{\mathrm{out}}$ defined above,*

$$\Psi_G(\tau + a) \leq \frac{D_{n,\tau}(a)}{1 - W}, \qquad a \in \mathbb{R}.$$

Consequently, if $D_{n,\tau} \leq 1 - \eta$ on a set A and $W \leq \eta/2$, then G has no support point in $\tau + A$.

Proof. Every likelihood ratio is nonnegative, so $m_G(i) \geq 1 - W$. Substitution in the definition of Ψ_G gives the first inequality, and $(1 - \eta)/(1 - \eta/2) < 1$ gives the second. □

An NPMLE places almost all its mass near the true score origin. The following coarse estimate isolates a central component before the edge analysis. For a mixing measure G in score coordinates, put

$$m_G(z) = \int \exp\left\{tz - \frac{\gamma_n t^2}{2}\right\} G(dt), \qquad \Delta_n(G) = \int \{1 - e^{-\kappa_n |t|^2}\}\, G(dt).$$

If P is standard Gaussian measure, then $P m_G = 1 - \Delta_n(G)$, and hence

$$P \log m_G \leq \log P m_G = \log\{1 - \Delta_n(G)\} \leq -\Delta_n(G). \tag{5.1}$$

The population loss yields a coarse uniform empirical bound. Its polylogarithmic precision is sufficient to extract a central atom before the critical-edge analysis begins.

Lemma 5.2. *There is a constant $C < \infty$ such that, uniformly over $0 \leq \varepsilon_n < 1$,*

$$\Delta_n(\widehat{G}_n) = O_{\mathbb{P}}\left\{\frac{(\log n)^C}{n}\right\}$$

for the NPMLE $\widehat{G}_n$.

Proof. It suffices to optimize over mixing measures supported in $\mathrm{conv}\{Z_1/\gamma_n, \dots, Z_n/\gamma_n\}$. Indeed, if p is the Euclidean projection of t onto this convex hull, then

$$(t-p)Z_i - \frac{\gamma_n}{2}(t^2 - p^2) = (t-p)(Z_i - \gamma_n p) - \frac{\gamma_n}{2}(t-p)^2 \leq 0,$$

so moving t to p weakly increases every fitted density. With probability tending to one this interval lies in $[-C\sqrt{\log n}, C\sqrt{\log n}]$.

For $\mathcal{M}(R) = \{m_G : \mathrm{supp}(G) \subset [-R, R]\}$, we use the compact Gaussian-mixture entropy bound from [Ano26b, Section 2]; see also [GVDV01, Section 3]:

$$\log N_{[]}\{\eta, \mathcal{M}(R), L^1(P)\} \leq C(1+R)\{1 + \log(1/\eta)\}^2, \qquad 0 < \eta \leq \frac{1}{2}. \tag{5.2}$$

The constant is uniform in κ_n, since

$$m_G(z)\phi(z) = \int e^{-\kappa_n t^2}\phi(z-t)\, G(dt),$$

and an $L^1(P)$ bracket for m_G is an ordinary L^1 bracket for a Gaussian location mixture of total mass at most one. To pass from the cited probability-mixture bound to subprobabilities, write the mixing measure as $s\widetilde{G}$. Replace the endpoints of the probability-mixture brackets by their positive parts, which does not increase their width. Bracket $s \in [0,1]$ on an $\eta/4$-grid and use probability-mixture brackets of width $\eta/4$. If $s_j \leq s \leq s_j + \eta/4$ and $l \leq m_{\widetilde{G}} \leq u$, then $[s_j l, (s_j + \eta/4)u]$ has width at most $\eta/2 + \eta^2/16 < \eta$. The additional $O\{\log(1/\eta)\}$ entropy is absorbed by (5.2). Bracket inequalities are understood P-almost everywhere; the displayed classes are separable, so the suprema below are measurable.

Fix a dyadic shell $\delta < \Delta_n(G) \leq 2\delta$, and cover it by upper brackets u of $L^1(P)$-width $\delta/4$. If $m_G \leq u$, then $Pu \leq 1 - 3\delta/4$, and Markov's inequality gives

$$\mathbb{P}\{P_n \log m_G \geq -\delta/4,\ m_G \leq u\} \leq e^{n\delta/4}(Pu)^n \leq e^{-n\delta/2}.$$

For $R = O(\{\log n\}^{1/2})$, the logarithm of the number of brackets is at most $(\log n)^C$ whenever $\delta \geq n^{-2}$. A union bound over brackets and dyadic shells therefore shows that every measure with $\Delta_n(G) > C(\log n)^C/n$ has negative likelihood gain relative to δ_0, with probability tending to one. Such a measure cannot be an NPMLE, proving the claim. $\square$

The defect bound does not rule out several nearby atoms. The local collapse lemma shows that the NPMLE carries all central mass on one atom, with the remaining edge mass smaller than the KKT gaps used later.

Lemma 5.3. *Assume $\kappa_n = n^{-\alpha+o(1)}$ for a fixed $0 < \alpha < 1/4$. For a sufficiently small fixed $r_0 > 0$, the NPMLE can, with probability tending to one, be written*

$$\widehat{G}_n = (1 - W_n)\delta_{\tau_n} + G_n^{\mathrm{out}}, \qquad \mathrm{supp}(G_n^{\mathrm{out}}) \subset [-r_0, r_0]^c, \tag{5.3}$$

where

$$W_n + |\tau_n|^2 = O_{\mathbb{P}}\left\{\frac{(\log n)^C}{n\kappa_n}\right\}.$$

Moreover, $W_n = o_{\mathbb{P}}(\kappa_n)$.

Proof. Lemma 5.2 and $1 - e^{-\kappa_n |t|^2} \geq c\kappa_n \min\{|t|^2, 1\}$ imply the displayed bounds for the mass outside $[-r_0, r_0]$ and for the second moment of the restriction to that interval. Normalize the central restriction to a probability measure G^{cen}, and let τ and V be its mean and variance. Cauchy–Schwarz gives the asserted bound for $|\tau|^2$.

We show that replacing G^{cen} by δ_τ strictly increases the likelihood unless $V = 0$. Keep the outside part fixed and interpolate

$$G_u = (1 - W)\{(1 - u)\delta_\tau + uG^{\text{cen}}\} + G^{\text{out}}, \qquad 0 \leq u \leq 1.$$

By concavity it is enough to show that the derivative at $u = 0$ is negative. If m_i^0 is the likelihood ratio under G_0, put $\omega_i = (1 - W)q_\tau(Z_i)/m_i^0$. With $h = t - \tau$, the derivative is

$$\sum_{i=1}^{n} \omega_i \int \left[\exp\left\{h(Z_i - \gamma_n\tau) - \frac{\gamma_n h^2}{2}\right\} - 1\right] G^{\text{cen}}(dt). \tag{5.4}$$

First replace ω_i by one. The bound on τ allows us to choose a deterministic $\delta_n \downarrow 0$ such that $|\tau| \leq \delta_n\sqrt{\kappa_n}$ with probability tending to one. On this event, a compact empirical-process bound gives

$$\sup_{\substack{|\tau| \leq \delta_n\sqrt{\kappa_n} \\ |h| \leq 2r_0}} \left|\partial_h^2 P_n e^{h(Z-\gamma_n\tau)-\gamma_n h^2/2} - \partial_h^2 P e^{h(Z-\gamma_n\tau)-\gamma_n h^2/2}\right| = O_{\mathbb{P}}\{n^{-1/2}(\log n)^C\} = o_{\mathbb{P}}(\kappa_n).$$

Indeed, after two differentiations the envelope on the fixed parameter rectangle is a fixed polynomial in $|Z|$ times $e^{2r_0|Z|}$, and hence has finite moments of every order. The population average is

$$F_\tau(h) = \exp\{-\kappa_n h^2 - \gamma_n\tau h\}.$$

Its second derivative is at most $-\kappa_n$ on $[-2r_0, 2r_0]$, for small enough r_0, with probability tending to one. Since $\int h\, G^{\text{cen}}(dt) = 0$, the unweighted version of (5.4) is at most $-cn\kappa_n V$.

It remains to restore the weights. Write

$$A_i = \int \left[e^{h(Z_i-\gamma_n\tau)-\gamma_n h^2/2} - 1\right] G^{\text{cen}}(dt), \qquad r_i^0 = 1 - \omega_i.$$

Thus the likelihood ratio of the NPMLE differs from m_i^0 by the factor $1 + \omega_i A_i$. Taylor's theorem and $\int h\, G^{\text{cen}}(dt) = 0$ give

$$\left|\int \left[e^{h(Z_i-\gamma_n\tau)-\gamma_n h^2/2} - 1\right] G^{\text{cen}}(dt)\right| \leq C(1 + Z_i^2)e^{2r_0|Z_i|+C}V.$$

On an event of probability tending to one, $\max_i |Z_i| \leq C_0\sqrt{\log n}$ for a fixed C_0. Hence

$$\max_i |A_i| \leq C(1 + \log n)e^{2C_0 r_0\sqrt{\log n}}V = o_{\mathbb{P}}(1).$$

To bound $\sum_i r_i^0$, let $\widehat{r}_i$ be the posterior responsibility of G^{out} under the original NPMLE. Integrating its KKT equality over G^{out} gives the exact identity $\sum_i \widehat{r}_i = nW$. The denominator comparison yields, uniformly in i,

$$r_i^0 = \widehat{r}_i(1 + \omega_i A_i) \leq \{1 + o_{\mathbb{P}}(1)\}\widehat{r}_i.$$

Consequently $\sum_i r_i^0 = O_{\mathbb{P}}(nW)$. Replacing ω_i by one in (5.4) therefore changes it by at most

$$O_{\mathbb{P}}\left\{nW(1 + \log n)e^{2C_0 r_0\sqrt{\log n}}V\right\} = o_{\mathbb{P}}\{n\kappa_n V\}.$$

Indeed, the ratio of the first quantity to $n\kappa_n V$ is $O_{\mathbb{P}}(n^{-1+2\alpha+o(1)})$. The derivative is therefore negative whenever $V \neq 0$. Optimality forces $V = 0$, which proves (5.3); the final assertion follows from $\alpha < 1/4$. $\square$

We now control the total number of KKT contacts. On a fixed score interval, the likelihood converges to a Poisson optimization with an analytic KKT function and hence finitely many zeros. At the moving lower endpoint, where the variance penalty changes on the shorter scale u_n^{-2}, a separate zero-counting argument controls the crossover.

Fix for now $0 < \alpha < 1/4$, assume $\varepsilon_n = n^{-\alpha+o(1)}$, and put $\theta_\alpha = 1 - \sqrt{\alpha}$. Let N be the Poisson process from Proposition 2.1, with intensity $\mu(dx) = e^{-x}\,dx$, and let S be its compensated edge field. For a compact interval $K = [b, c] \Subset (1/2, 1)$ and a finite positive measure ν on K, define

$$F_\nu(x) = \int_K e^{\theta x}\,\nu(d\theta), \qquad h(y) = \log(1+y) - y,$$

and

$$\mathcal{J}_N^K(\nu) = \int_K S(\theta)\,\nu(d\theta) + \int_{\mathbb{R}} h(F_\nu(x))\,N(dx). \tag{5.5}$$

The last integral is almost surely finite. For $x < 0$, use $|h(y)| \le y^2/2$ and $F_\nu(x) \le \nu(K)e^{bx}$. The resulting envelope is integrable against N because $2b > 1$. On $[0, \infty)$, the Poisson process has only finitely many points.

To obtain (5.5), place masses $w_j = s_j/\{nc_n(\theta_j)\}$ at finitely many score locations $\theta_j u_n$. If v_i is the resulting likelihood increment at the ith observation, then

$$\sum_i \log(1+v_i) = \sum_i v_i + \sum_i \{\log(1+v_i) - v_i\}.$$

The first sum converges to $\sum_j s_j S(\theta_j)$. At an edge observation $Z_i = u_n + x/u_n$, the second sum sees $v_i \to \sum_j s_j e^{\theta_j x}$, while ordinary observations contribute only a negligible quadratic remainder. This gives the two terms in (5.5).

5.2 Compact Poisson likelihoods

After localization, tightness of the atom count reduces to finiteness of the active set in a compact limiting Poisson optimization. For a fixed realization of the Poisson process, the next proposition gives this finiteness uniformly over all maximizers.

Proposition 5.4. *The functional $\mathcal{J}_N^K$ is concave and coercive on finite positive measures on K, and it has at least one maximizer. For every maximizer ν, its first-variation field*

$$\Gamma_\nu(\theta) = S(\theta) - \int_{\mathbb{R}} e^{\theta x}\frac{F_\nu(x)}{1+F_\nu(x)}N(dx) \tag{5.6}$$

is real analytic on $(1/2, 1)$ and satisfies

$$\Gamma_\nu \le 0 \quad \text{on } K, \qquad \Gamma_\nu = 0 \quad \nu\text{-almost everywhere}.$$

In particular, $\operatorname{supp}(\nu)$ is finite. Almost surely, the support sizes are bounded uniformly over all maximizers of $\mathcal{J}_N^K$. More generally, for every $M < \infty$, the numbers of zeros of Γ_ν on K are bounded uniformly over all positive measures ν with $\nu(K) \le M$.

Proof. Concavity follows from concavity of $\log(1+y)$. To prove coercivity, let $m = \nu(K) \to \infty$, choose k with $2^k \le m < 2^{k+1}$, and put $L_k = k \log 2/(2c)$. On $I_k = [-L_k - 1, -L_k]$,

$$F_\nu(x) \ge m e^{cx} \ge e^{-c} 2^{k/2}.$$

The count $N(I_k)$ has mean $\asymp 2^{k/(2c)}$. Chernoff's bound and Borel–Cantelli imply that it is at least half its mean for every large k, almost surely. Since $h(y) \le -y/2$ for large y, the contribution from I_k is at most $-C2^{k(1/2+1/(2c))}$. This dominates the linear term, which is $O(m) = O(2^k)$, because $c < 1$. Thus the functional is coercive. On sets of bounded mass it is continuous for weak convergence: the negative Poisson tail is uniform by $|h(F_\nu(x))| \le C_M e^{2bx}$, and the positive tail is finite. Compactness of bounded sets of positive measures on K then gives existence.

Differentiation in the direction of a point mass gives (5.6) and the KKT inequalities. Both terms extend analytically to a complex neighborhood of each compact subinterval of $(1/2, 1)$; on the negative half-line the integrand in the second term is bounded by $Ce^{(b+\operatorname{Re}\theta)x}$, which is integrable when $b + \operatorname{Re}\theta > 1$. The function Γ_ν is not identically zero, since $S(\theta) \to -\infty$ as $\theta \uparrow 1$ and the subtracted integral is nonnegative. Its zeros on K are therefore finite, and the KKT conditions place $\operatorname{supp}(\nu)$ among them.

For uniformity, let $\nu_j(K) \le M$. Compactness of bounded sets of positive measures on K gives a weakly convergent subsequence, along which the functions Γ_{ν_j} converge locally uniformly on a complex neighborhood of K. The limit is Γ_ν for the limiting measure and is not identically zero: $S(\theta) \to -\infty$ as $\theta \uparrow 1$, whereas the integral subtracted in (5.6) is nonnegative. Choose a bounded domain U containing K, with closure in this neighborhood, such that Γ_ν has no zeros on ∂U. Rouché's theorem then bounds the number of zeros in K for all sufficiently large j. This proves the uniform zero bound, which in turn bounds the support of every maximizer. $\square$

To transfer the limiting problem to the finite-sample likelihood, Lemma 5.3 supplies a central atom τ_n. Put $Y_{n,i} = Z_i - \gamma_n \tau_n$. For a sign $\xi \in \{-1, 1\}$ and $\theta \in K$, the likelihood ratio relative to that central atom is

$$q_{n,\xi,\theta}(Z_i) = \exp\left\{\xi\theta u_n Y_{n,i} - \frac{\theta^2 u_n^2}{2} - \lambda_n \theta^2\right\}, \qquad \lambda_n = \kappa_n u_n^2.$$

Given positive measures ν_+, ν_- on K, place physical mass $\nu_\xi(d\theta)/\{nc_n(\theta)\}$ at $\tau_n + \xi\theta u_n$, removing the same total mass from the central atom. Denote the resulting likelihood gain by $\mathcal{L}_{n,K}(\nu_+, \nu_-)$. The deterministic part of a normalized first variation is

$$\mathcal{P}_n(\theta) = \frac{1 - e^{-\lambda_n \theta^2}}{c_n(\theta)}. \tag{5.7}$$

Uniform convergence and a scaled-mass bound allow the optimizing measure, rather than only a fixed perturbation, to pass to the limit.

Proposition 5.5. *If $K \Subset (\theta_\alpha, 1)$, then on every set $\nu_+(K) + \nu_-(K) \le M$,*

$$\mathcal{L}_{n,K}(\nu_+, \nu_-) \Rightarrow \mathcal{J}^K_{N_+}(\nu_+) + \mathcal{J}^K_{N_-}(\nu_-)$$

in the space of continuous functions on the compact set of such measure pairs, where N_+ and N_- are independent copies of N. The first variations converge jointly, locally uniformly in a fixed complex neighborhood of every smaller interval $K' \Subset K$.

If $K \Subset (1/2, 1)$ is allowed to contain θ_α, an NPMLE whose noncentral support lies in the two signed copies of K has total scaled edge mass $O_{\mathbb{P}}(1)$.

More precisely, let $K = [b, c] \Subset (1/2, 1)$ contain θ_α and assume

$$\frac{\alpha}{2} + (1-b)^2 < \frac{1}{2}.$$

For NPMLEs whose noncentral support lies in the signed copies of K, let $\nu_{n,+}, \nu_{n,-}$ be their scaled edge measures. The normalized first-variation fields, after adding $\mathcal{P}_n$, are jointly tight for locally uniform convergence on a fixed complex neighborhood of every $K' \Subset K$. Along any subsequence on which $\nu_{n,\xi} \Rightarrow \nu_\xi$, their limits are the analytic fields Γ_{ν_ξ} in (5.6).

Proof. For finitely atomic measures, write

$$v_i = \sum_{\xi=\pm 1} \int_K \frac{q_{n,\xi,\theta}(Z_i) - 1}{nc_n(\theta)} \, \nu_\xi(d\theta).$$

The identity $\sum_i \log(1+v_i) = \sum_i v_i + \sum_i \{\log(1+v_i) - v_i\}$ separates the two limits. Proposition 2.1 sends the linear term to $\int S_+ d\nu_+ + \int S_- d\nu_-$. The deterministic damping vanishes uniformly on $K \Subset (\theta_\alpha, 1)$. Translation by the central atom is also negligible after normalization by $c_n(\theta)$, because, for $b = \min K$,

$$\frac{u_n |\tau_n|}{c_n(b)} = O_{\mathbb{P}}\Big\{(\log n)^C n^{-1/2+\alpha/2+(1-b)^2+o(1)}\Big\} = o_{\mathbb{P}}(1).$$

At a right-edge observation $Z_i = u_n + x/u_n$, the positive part of v_i converges to $F_{\nu_+}(x)$, while the negative part is exponentially small; the left edge is analogous. This yields the nonlinear Poisson integrals.

The convergence is uniform over bounded measure sets. Indeed, the physical edge mass is at most $2M/\{nc_n(b)\}$, so its square contributes $o(1)$. On a bounded edge window, the kernels and their first few derivatives are uniformly bounded and equicontinuous. Below $-A$, the likelihood remainder is bounded by $C_M e^{2bx}$, and its first variation by $C_M e^{(b+\operatorname{Re}\theta)x}$. The finite-sample analogues follow from the exponential-moment bounds used in Proposition 2.1. A finite weak net of the compact measure set, followed by $A \to \infty$, proves the asserted functional and analytic convergence.

The bounded-window equicontinuity and the negative-tail envelopes $C_M e^{2bx}$ and $C_M e^{(b+\operatorname{Re}\theta)x}$ also prove the analytic tightness assertion for an interval that crosses θ_α. The random linear term still converges to S, while its deterministic mean is exactly $-\mathcal{P}_n$; adding $\mathcal{P}_n$ removes the only term that is not tight across the crossover. In the nonlinear term, $e^{-\lambda_n \theta^2} = 1 + o(1)$ uniformly on the fixed complex neighborhood, so the envelopes $C_M e^{2bx}$ and $C_M e^{(b+\operatorname{Re}\theta)x}$ on the negative tail remain valid. The translation by τ_n is negligible, together with all derivatives, by the displayed condition on b. Uniformity over bounded measure sets allows the fields to be evaluated at the sample-dependent $\nu_{n,\xi}$; after a weakly convergent subsequence, their limits are precisely Γ_{ν_ξ}.

For the mass bound, put $m = \nu_+(K) + \nu_-(K)$ and choose a sign carrying at least $m/2$. Uniformly on K, the linear term is at most $B_n m$, where $B_n = O_{\mathbb{P}}(1)$; this remains true when K slightly crosses θ_α, because the normalized variance penalty is nonpositive. Let $N_n(I)$ denote the number of observations whose edge coordinate lies in I. For a suitable constant $c_1 > 0$,

$$\lim_{L\to\infty} \liminf_{n\to\infty} \mathbb{P}\big\{N_n([-L-1, -L]) \ge c_1 e^L\big\} = 1,$$

by the edge-process convergence. Each such observation has likelihood increment at least $(m/3)e^{-c(L+1)}$, where $c = \max K < 1$. Since $h(y) \le -c_2 y$ for large y, this block contributes at most $-c_3 m e^{(1-c)L}$; all other nonlinear terms are nonpositive. First restrict to $B_n \le B$, then choose L so that this negative coefficient exceeds $2B$, and finally take m large. The resulting likelihood gain is negative, while zero edge mass has gain zero. Hence the maximizing scaled mass is tight. $\square$

5.3 The moving lower edge

The compact Poisson limit leaves one gap in the proof of support tightness: a narrow crossover where the normalized variance penalty changes from dominant to negligible. There the penalty varies exponentially faster than the random analytic remainder, which yields the deterministic zero bound needed to control KKT contacts at the moving lower edge.

Lemma 5.6. *Let I be a compact interval, and suppose R_n and b_n are analytic on a fixed complex neighborhood of I and real-valued on I. Assume that every subsequence has a further subsequence converging locally uniformly to a nonzero analytic function. Let*

$$\mathcal{P}_n(\theta) = \exp\{a_n\varphi(\theta) + b_n(\theta)\}, \qquad a_n \to \infty,$$

where φ is analytic on the same neighborhood, real-valued on I, with $\varphi' < 0$ there, and $\{b_n - b_n(\theta_0)\}$ is locally uniformly bounded there for one $\theta_0 \in I$. Along every subsequence on which R_n converges, the number of real zeros of $R_n - \mathcal{P}_n$ on I is bounded.

Proof. Pass to a subsequence with $R_n \to R \not\equiv 0$. Away from small disjoint complex disks around the finitely many zeros of R, both R_n'/R_n and b_n' are bounded. Hence, on each component where $R_n > 0$,

$$\frac{d}{d\theta}\log\frac{R_n(\theta)}{\mathcal{P}_n(\theta)} = \frac{R_n'(\theta)}{R_n(\theta)} - a_n\varphi'(\theta) - b_n'(\theta) > 0$$

for large n. There is at most one intersection on that component and none where $R_n < 0$.

It remains to count intersections near a zero θ_0 of R, of multiplicity m. Analytic factorization and Rouché's theorem give

$$R_n(z) = g_n(z)\prod_{j=1}^{m}(z - z_{n,j}),$$

where $z_{n,j} \to \theta_0$, while g_n, $1/g_n$, and g_n'/g_n are uniformly bounded. On a real interval where $R_n > 0$, set $H_n = \log R_n - a_n\varphi - b_n$. Its critical points satisfy

$$0 = H_n'(\theta) = A_n(\theta) + \sum_{j=1}^{m}\frac{1}{\theta - z_{n,j}}, \qquad A_n = \frac{g_n'}{g_n} - a_n\varphi' - b_n'. \tag{5.8}$$

If $c_0 = \min_I(-\varphi')$, then $A_n \geq c_0 a_n/2$ on the real axis. Thus every critical point lies within $O(a_n^{-1})$ of one of the roots $z_{n,j}$.

In each cluster of such roots, rescale $y = a_n(\theta - x_n)$. After a subsequence, the rescaled roots either converge or leave every compact set. Multiplying $a_n^{-1}H_n'$ by the polynomial Q_n formed from the convergent roots gives

$$F_n(y) = C_n(y)Q_n(y) + Q_n'(y), \qquad C_n \to C = -\varphi'(\theta_0) > 0.$$

The limit $CQ + Q'$ is not identically zero: at a root of Q of multiplicity q, the derivative has order $q-1$, whereas CQ has order at least q. Rouché's theorem therefore bounds the number of critical points in each cluster. Rolle's theorem then bounds the zeros of H_n, equivalently the intersections $R_n = \mathcal{P}_n$. Summing over the finitely many zero disks and complementary components proves the claim. □

We apply this lemma to the finite-sample KKT certificate. For the penalty $\mathcal{P}_n$ in (5.7), on a fixed neighborhood of θ_α,

$$\frac{d}{d\theta}\log\mathcal{P}_n(\theta) = -(1-\theta)u_n^2 + O(1),$$

so its transition has width $O(u_n^{-2})$. Choose $I = [b, c] \Subset (1/2, 1)$ around θ_α with

$$\frac{\alpha}{2} + (1-b)^2 < \frac{1}{2}. \tag{5.9}$$

For the fitted measure, normalize the KKT certificate by

$$D_{\tau_n,\xi}(\theta) = \frac{1}{n}\sum_{i=1}^{n} q_{n,\xi,\theta}(Z_i), \qquad \Gamma_{n,\xi}(\theta) = \frac{\Psi_{\widehat{G}_n}(\tau_n + \xi\theta u_n) - 1}{c_n(\theta)}.$$

Write the fitted likelihood ratio relative to its central atom as $1 + v_i$. Since that atom has positive mass, its KKT equality gives $n^{-1}\sum_i (1+v_i)^{-1} = 1$. Consequently,

$$\begin{aligned}\Gamma_{n,\xi}(\theta) &= \frac{1}{nc_n(\theta)}\sum_{i=1}^{n}\frac{q_{n,\xi,\theta}(Z_i) - 1}{1+v_i} \\ &= \frac{D_{\tau_n,\xi}(\theta) - 1}{c_n(\theta)} - \frac{1}{nc_n(\theta)}\sum_{i=1}^{n}\{q_{n,\xi,\theta}(Z_i) - 1\}\frac{v_i}{1+v_i}. \end{aligned}\tag{5.10}$$

There is no residual contribution from the -1 in the last sum, because

$$\sum_i \frac{v_i}{1+v_i} = \sum_i \left(1 - \frac{1}{1+v_i}\right) = 0.$$

Proposition 2.1 controls the first term. In the second, an observation with right-edge coordinate x converges to $-e^{\theta x}F_{\nu_+}(x)/\{1 + F_{\nu_+}(x)\}$, and the left edge is analogous. The negative-tail envelope is $C_M e^{(b+\operatorname{Re}\theta)x}$, which is integrable because $b + \operatorname{Re}\theta > 1$. Once Lemma 5.7 confines the noncentral support to the signed copies of I, the crossover assertion of Proposition 5.5 gives

$$\Gamma_{n,\xi}(\theta) = R_{n,\xi}(\theta) - \mathcal{P}_n(\theta), \tag{5.11}$$

where $\{R_{n,\xi}\}$ is tight for locally uniform convergence on the complex neighborhood, and every subsequential limit is the nonzero function Γ_ν in (5.6). The only extra term is the translation by τ_n, which is negligible with all derivatives by (5.9) and

$$\sup_{\theta\in I}\frac{u_n|\tau_n|}{c_n(\theta)} = O_{\mathbb{P}}\Big\{(\log n)^C n^{-1/2+\alpha/2+(1-b)^2+o(1)}\Big\} = o_{\mathbb{P}}(1).$$

Finally,

$$\mathcal{P}_n(\theta) = \exp\left\{u_n^2(-\theta + \theta^2/2) + \log n + \log(1 - e^{-\lambda_n\theta^2})\right\}.$$

On a fixed simply connected complex neighborhood of I, bounded away from zero, $1 - e^{-\lambda_n z^2}$ has no zeros for all large n. We choose the analytic logarithm there that is real on I. After subtraction at one fixed point, this logarithm and all its derivatives are locally uniformly bounded. Lemma 5.6 therefore bounds the number of contacts in the moving crossover.

To exclude contacts outside a compact edge interval, we collect the quantitative gaps already proved in the phase-diagram analysis. Put

$$\tau_{0,n} = \frac{\bar{Z}_n}{\gamma_n}, \qquad D_{0,\xi}(\theta) = D_{n,\tau_{0,n}}(\xi\theta u_n),$$

so $D_{0,\xi}$ is the KKT field relative to the best one-atom fit.

Lemma 5.7. *Fix $r_0 > 0$ and $0 < \eta < \theta_\alpha - 1/2$. There is $c > 0$ such that, with probability tending to one,*

$$\sup_{\xi=\pm 1} \sup_{r_0/u_n \le \theta \le \theta_\alpha - \eta} D_{0,\xi}(\theta) \le 1 - c\kappa_n.$$

Moreover, for every $\delta > 0$, there are $\rho > 0$ and $a > 0$, with $\rho^2 < \alpha$, such that, with probability at least $1 - \delta + o(1)$,

$$\sup_{\xi=\pm 1} \sup_{1-\rho \le \theta < 1} \frac{D_{0,\xi}(\theta) - 1}{c_n(\theta)} \le -a, \qquad \sup_{\xi=\pm 1} \sup_{\theta \ge 1} D_{0,\xi}(\theta) \le 0.9.$$

Proof. For the lower range, choose $\theta_* \in (1/2, \theta_\alpha - \eta)$ with $\theta_*^2 < 1/2 - \alpha$. The Gaussian-range estimate used in Lemma 2.4 gives

$$\sup_{\xi=\pm 1} \sup_{0 < \theta \le \theta_*} \frac{|K_{n,\xi}(\theta) - 1|}{\theta^2} = O_{\mathbb{P}}\{n^{-1/2+\theta_*^2+o(1)}\} = o_{\mathbb{P}}(\lambda_n).$$

Since $D_{0,\xi} = e^{-\lambda_n \theta^2} K_{n,\xi}$, this leaves a deficit at least $c\kappa_n$ for $r_0/u_n \le \theta \le \theta_*$. On $[\theta_*, \theta_\alpha - \eta]$, Proposition 2.1 and Lemma 2.2 make the random error $o_{\mathbb{P}}(\lambda_n)$, while the damping loss is at least $c\lambda_n$. This proves the first assertion.

For the upper endpoint, use the decomposition in the proof of Lemma 2.5. Fix M. After points above the edge level M are removed, the compensated field and its derivative are tight on $[3/4, 1]$. Thus, for a suitable $K = K(M, \delta)$, both truncated fields have supremum at most K with probability at least $1 - \delta/4 + o(1)$. The probability that either edge process has a point above M is at most $2e^{-M} + o(1)$. On the complementary event, the omitted compensator is at least

$$\frac{e^{-(1-\theta)M}}{1 - \theta}, \qquad 1 - \rho \le \theta < 1.$$

Choose M so that $2e^{-M} < \delta/4$, and then choose $\rho > 0$ so small that this lower bound exceeds $K + 2$ throughout the interval. The normalized null field is then at most -2. The damping is nonpositive, and $\rho^2 < \alpha$ makes the empirical-centering error $o_{\mathbb{P}}(c_n)$; after decreasing the margin, this proves the first upper-endpoint inequality with $a = 1$.

For $\theta \ge 1$, choose R so large that both centered sample extremes are below the score $1 + R/u_n^2$, except on an event of probability at most $\delta/4 + o(1)$. The right-of-one estimate in the proof of Lemma 2.5, applied up to this score, gives $D_{0,\xi} \le 0.9$. Beyond it, Lemma 1.5 makes the field monotone away from the corresponding sample extreme. This proves the second inequality, simultaneously for both signs, with total failure probability at most $\delta + o(1)$. □

The limiting-law argument requires compactness of the finite-sample active sets before the limiting optimizer can be identified. The central-atom localization, the exclusion margins, and the analytic zero bound combine to give this fixed-exponent tightness.

Lemma 5.8. *If $r_n \to \alpha \in (0, 1/4)$, then*

$$|\operatorname{supp}(\widehat{G}_n)| = O_{\mathbb{P}}(1).$$

Proof. Lemma 5.3 gives one central atom and outside mass

$$W_n = O_{\mathbb{P}}\{(\log n)^C/(n\kappa_n)\} = o_{\mathbb{P}}(\kappa_n).$$

Relative to the best one-atom center, write $D_{\tau_n,\xi}(\theta) = D_{n,\tau_n}(\xi\theta u_n)$. The central atom introduces the exact translation

$$D_{\tau_n,\xi}(\theta) = e^{-\xi\theta u_n \delta_n} D_{0,\xi}(\theta), \qquad \delta_n = \gamma_n \tau_n - \bar{Z}_n,$$

where

$$|\delta_n| = O_{\mathbb{P}}\Big\{(\log n)^C n^{-1/2+\alpha/2+o(1)}\Big\}.$$

Choose $\eta > 0$ so small that $\alpha/2 + (\sqrt{\alpha} + 2\eta)^2 < 1/2$. The first bound in Lemma 5.7, the translation estimate, and Lemma 5.1 exclude contacts between the fixed core and $(\theta_\alpha - \eta)u_n$. For the upper endpoint choose ρ from the second bound, with $1 - \rho > \theta_\alpha - \eta$. There the normalized gap is at least $ac_n(1-\rho) = n^{-\rho^2+o(1)}$, whereas

$$W_n = O_{\mathbb{P}}\Big\{(\log n)^C n^{-1+\alpha+o(1)}\Big\}, \qquad u_n|\delta_n| = O_{\mathbb{P}}\Big\{(\log n)^C n^{-1/2+\alpha/2+o(1)}\Big\}.$$

Both are negligible relative to the gap because $\rho^2 < \alpha < 1/4$. It remains only to note that a support point cannot have an arbitrarily large translated score. The convex-hull localization in the proof of Lemma 5.2 places every score atom of $\widehat{G}_n$ in $[\min_i Z_i/\gamma_n, \max_i Z_i/\gamma_n]$. After increasing the fixed R in Lemma 5.7, this implies that every displacement from τ_n has

$$\theta \le 1 + \frac{R+1}{u_n^2}$$

outside an event of probability $\delta + o(1)$; here $u_n|\tau_n| = o_{\mathbb{P}}(1)$. On this bounded endpoint window the translation factor $e^{-\xi\theta u_n \delta_n}$ is uniformly $1 + o_{\mathbb{P}}(1)$. The normalized gap for $1 - \rho \le \theta < 1$, the fixed (0.1) gap for $1 \le \theta \le 1 + (R+1)/u_n^2$, and Lemma 5.1 therefore exclude all contacts beyond $(1-\rho)u_n$, apart from an event of probability $O(\delta) + o(1)$.

Every noncentral support point now lies in a signed copy of $[\theta_\alpha - \eta, 1-\rho]$, which is compactly contained in

$$K = [\theta_\alpha - 2\eta, 1 - \rho/2] \Subset (1/2, 1).$$

Proposition 5.5 makes the scaled mass on K tight and gives joint analytic convergence on a complex neighborhood of the smaller support interval. Prokhorov's theorem and a Skorokhod representation therefore allow us, from every subsequence, to extract a further subsequence on which the edge measures converge weakly and the analytic remainders in (5.11) converge locally uniformly, almost surely on a common probability space. Away from a fixed neighborhood of θ_α, the uniform analytic zero bound in Proposition 5.4 applies to the limiting remainders. Inside that neighborhood, Lemma 5.6 controls the intersections with the rapidly varying penalty. Rouché's theorem transfers both bounds to the exact KKT certificates along the subsequence. Since every support point is a zero, the support counts are tight. Otherwise one could choose a subsequence on which they escape every deterministic bound with fixed positive probability. Repeating the Prokhorov–Skorokhod subsequence extraction just described would then give a contradiction.

□

5.4 The limiting Poisson optimizer

To prove Theorem 1.3, we must identify a canonical limit for the compact likelihoods used above. Replacing the moving variance penalty by its hard wall produces the optimizer ν_b^* from (1.5); its uniqueness turns support tightness into convergence to a law depending only on r.

The functional $\mathcal{J}_{N,b}$ agrees with $\mathcal{J}_N^K$ in (5.5) whenever ν is supported on the compact interval $K \subset [b, 1)$. Its first variation is

$$\Gamma_\nu(\theta) = S(\theta) - \int_{\mathbb{R}} e^{\theta x} \frac{F_\nu(x)}{1 + F_\nu(x)}\, N(dx). \tag{5.12}$$

Thus a maximizing measure satisfies $\Gamma_\nu \le 0$ on $[b, 1)$, with equality on its support.

The negative Poisson points form a uniqueness set for the Laplace transforms appearing in (1.5). This is much stronger than uniqueness on any fixed finite set of observations.

Lemma 5.9. *Write the negative points of N as $-T_j$. Almost surely, if a finite signed measure δ is supported on a compact interval $[b, c] \subset (0, \infty)$ and*

$$\int_b^c e^{-T_j\theta}\,\delta(d\theta) = 0 \qquad \textit{for every } j,$$

then $\delta = 0$.

Proof. For $\operatorname{Re} z > 0$, set

$$H(z) = e^{bz}\int_b^c e^{-z\theta}\,\delta(d\theta).$$

This is a bounded analytic function on the right half-plane. Unless it vanishes identically, its zeros satisfy the half-plane Blaschke condition [Gar07, Chapter II],

$$\sum_{H(z)=0} \frac{\operatorname{Re} z}{1 + |z|^2} < \infty,$$

with multiplicity.

The points T_j form a Poisson process on $(0, \infty)$ with intensity $e^t\,dt$. If Q_m counts its points in $[m, m+1]$, then the Q_m's are independent Poisson variables with means $(e-1)e^m$. A Chernoff bound and Borel–Cantelli give $Q_m \ge (e-1)e^m/2$ for all sufficiently large m, almost surely. Consequently,

$$\sum_j \frac{T_j}{1 + T_j^2} \ge \sum_{m \text{ large}} \frac{mQ_m}{1 + (m+1)^2} = \infty.$$

Thus $H(T_j) = 0$ for every j forces $H \equiv 0$. Uniqueness of the Laplace transform then gives $\delta = 0$. □

We can now attach a canonical optimizer to each hard-wall location. Its support size will be the one-sided contribution to the limiting atom count.

Proposition 5.10. *For every fixed $b \in (1/2, 1)$, the functional $\mathcal{J}_{N,b}$ has an almost surely unique maximizer ν_b^*. This measure has finite support in a random compact subinterval of $[b, 1)$.*

Proof. Since $S(\theta) \to -\infty$ as $\theta \uparrow 1$, there is a random $c_N \in (b, 1)$ such that $S(\theta) < 0$ on $[c_N, 1)$. The second term in (5.12) is nonnegative, so deleting any mass in this interval strictly increases the objective. Proposition 5.4, applied on a compact interval containing $[b, c_N]$, gives existence and finite support.

The first term in (1.5) is linear, while h is strictly concave. If two measures ν_0, ν_1 were maximizers, equality in the concavity inequality for their midpoint would require $F_{\nu_0}(X) = F_{\nu_1}(X)$ at every point X of N. In particular, the two transforms agree at every negative point. Lemma 5.9, applied to $\nu_1 - \nu_0$, gives $\nu_0 = \nu_1$. □

For later use, write

$$K_b = |\operatorname{supp}(\nu_b^*)|.$$

The value zero is allowed: $K_b = 0$ exactly when the one-sided Poisson field does not cross zero on $[b, 1)$.

5.5 The moving penalty and the hard wall

The next step in Theorem 1.3 shows why no subpolynomial feature of $\varepsilon_n = n^{-r_n}$ survives in the limiting optimizer. The deterministic penalty changes abruptly at $b_r = 1 - \sqrt{r}$, but its value exactly at that point need not converge. A variational argument moves any boundary mass an asymptotically negligible distance to the vanishing-penalty side.

Recall $\lambda_n = \varepsilon_n u_n^2/\{2(1-\varepsilon_n)\}$, the scale $c_n(\theta)$ from Section 2, and the penalty

$$\mathcal{P}_n(\theta) = \frac{1 - e^{-\lambda_n \theta^2}}{c_n(\theta)}$$

from (5.7). Assume $r_n \to r \in (0, 1/4)$. Since $u_n^2/2 = \log n + O(\log\log n)$ and $\lambda_n \to 0$, a direct expansion gives, uniformly on compact subintervals of $(1/2, 1)$,

$$\log \mathcal{P}_n(\theta) = \{(1-\theta)^2 - r_n\} \log n + O(\log\log n). \tag{5.13}$$

Consequently, for every fixed $\eta > 0$,

$$\sup_{\theta \geq b_r + \eta} \mathcal{P}_n(\theta) \longrightarrow 0, \qquad \inf_{\theta \leq b_r - \eta} \mathcal{P}_n(\theta) \longrightarrow \infty. \tag{5.14}$$

The following deterministic lemma records the variational consequence. It is stated on a compact interval because all random likelihoods have already been localized there.

Lemma 5.11. *Let $K = [a, c] \Subset (1/2, 1)$, with $a < b_r < c$, and suppose concave functionals $\widetilde{L}_n$ converge locally uniformly on every bounded-mass set of finite positive measures on K to $\mathcal{J}_N^K$. Define*

$$L_n(\nu) = \widetilde{L}_n(\nu) - \int_K \mathcal{P}_n(\theta)\, \nu(d\theta).$$

If the masses of the maximizing measures are tight, every weak limit of maximizers of L_n is supported on $[b_r, c]$ and maximizes $\mathcal{J}_N^K$ there. Conversely, every measure on $[b_r, c]$ has a recovery sequence with the same limiting value.

Proof. The two bounds in (5.14) make the penalty vanish uniformly on compact subsets of $(b_r, c]$ and force every bounded-objective sequence to put asymptotically no mass on $[a, b_r - \eta]$, for each $\eta > 0$. Tightness of the masses gives the asserted support and the usual limsup inequality.

For recovery, put

$$\delta_n = |r_n - r|^{1/2} + \left(\frac{\log\log n}{\log n}\right)^{1/2}.$$

Then $\delta_n \downarrow 0$ and $\mathcal{P}_n(b_r + \delta_n) \to 0$ by (5.13). Leave mass above $b_r + \delta_n$ fixed and move all mass in $[b_r, b_r + \delta_n]$ to $b_r + \delta_n$. The penalty is decreasing on a fixed neighborhood of b_r for all large n, as follows by differentiating its definition, so the recovery penalties vanish. Local uniform convergence of $\widetilde{L}_n$ completes the proof. $\square$

The compact expansion already used for support tightness becomes uniform across the moving wall after the deterministic penalty is added back. We record the strengthened form needed to identify the optimizer itself.

Lemma 5.12. *Let $K = [a, c] \Subset (1/2, 1)$ contain b_r, and assume*

$$\frac{r}{2} + (1-a)^2 < \frac{1}{2}.$$

On every set of edge-measure pairs with total mass at most M, the functionals

$$\mathcal{L}_{n,K}(\nu_+, \nu_-) + \int_K \mathcal{P}_n \, d\nu_+ + \int_K \mathcal{P}_n \, d\nu_-$$

converge uniformly to $\mathcal{J}^K_{N_+}(\nu_+) + \mathcal{J}^K_{N_-}(\nu_-)$. The first variations converge jointly and locally uniformly, together with their first two derivatives, on every smaller complex neighborhood. Here N_+ and N_- are independent copies of N.

Proof. Truncate the two Gaussian edge processes to a bounded edge window. On this window the kernel and its first two parameter derivatives are uniformly bounded and equicontinuous over measures of mass at most M, so compact point-process convergence gives the assertion. On the negative tail, the nonlinear likelihood remainder has envelope $C_M e^{2ax}$; its differentiated first variations are bounded by a polynomial in $|x|$ times $C_M e^{(a+\operatorname{Re}\theta)x}$. These envelopes are integrable because $2a > 1$. The positive tail contains only finitely many limiting points. A finite weak net of the compact measure set makes the truncation uniform.

The deterministic part of the normalized first variation is exactly $-\mathcal{P}_n$. Once it is added back, $e^{-\lambda_n \theta^2} = 1 + o(1)$ uniformly with two derivatives on K, so the nonlinear limit is unchanged. Translation by the central atom is negligible with two derivatives under the displayed condition on a. Finally, integrate the uniform first-variation convergence along $s \mapsto (s\nu_+, s\nu_-)$, $0 \le s \le 1$, to obtain convergence of the functionals. Joint convergence of the two extreme point processes gives independent N_+ and N_-. $\square$

An atom of physical mass w at signed score displacement θu_n from the central atom receives scaled mass $nc_n(\theta)w$. The next proposition identifies the weak limit of these two scaled edge measures before addressing their exact support sizes.

Proposition 5.13. *Assume $r_n \to r \in (0, 1/4)$. If $\nu_{n,+}$ and $\nu_{n,-}$ are the scaled edge measures of the NPMLE, then*

$$(\nu_{n,+}, \nu_{n,-}) \Rightarrow (\nu^{*,+}_{b_r}, \nu^{*,-}_{b_r}),$$

where the two limits are independent copies of the optimizer in Proposition 5.10. In particular, their joint law depends on the variance sequence only through r.

Proof. Lemma 5.3 isolates one central atom and shows that the total physical edge mass tends to zero. The subthreshold margin confines all remaining support from below. For every $\delta > 0$, Lemma 5.7 supplies a deterministic $c < 1$ such that, except on an event of probability at most $\delta + o(1)$, both edge measures are supported on a fixed interval $[a, c]$ containing b_r. Proposition 5.5 makes their total scaled masses tight.

On a Skorokhod coupling, apply Lemmas 5.12 and 5.11. Every subsequential limit maximizes the two hard-wall functionals. Proposition 5.10 identifies these maximizers uniquely. Letting first $n \to \infty$ and then $\delta \downarrow 0$ proves the joint convergence. The wall is b_r by (5.13). $\square$

Weak convergence alone does not preserve an exact atom count: atoms with vanishing mass could disappear, and several atoms could merge. The missing input is nondegeneracy of every limiting KKT contact.

Proposition 5.14. *For every fixed $b \in (1/2, 1)$, the following statements hold almost surely for the optimizer ν^*_b and its KKT field $\Gamma_{\nu^*_b}$.*

1. *The zero set of* $\Gamma_{\nu_b^*}$ *on* $[b,1)$ *is exactly* $\operatorname{supp}(\nu_b^*)$.

2. *Every interior contact* θ *is quadratic:* $\Gamma''_{\nu_b^*}(\theta) < 0$.

3. *At the lower boundary, either* $\Gamma_{\nu_b^*}(b) < 0$*, or* $b \in \operatorname{supp}(\nu_b^*)$ *and* $\Gamma'_{\nu_b^*}(b) < 0$.

The strict inequalities have the natural constrained interpretation. An interior contact is a local maximum of a nonpositive function, while a contact at the left endpoint has a nonpositive inward derivative. Appendix A uses the continuously distributed negative Poisson points to rule out equality and inactive contacts.

5.6 Convergence of the atom count

This subsection completes the proof of Theorem 1.3 by passing from weak convergence of the edge measures to convergence of their active sets. Recall from (1.6) that $\mathsf{K}_r = 1 + K^+_{b_r} + K^-_{b_r}$, where $b_r = 1 - \sqrt{r}$. The central one is the atom isolated by Lemma 5.3.

The nondegenerate contacts of Proposition 5.14 persist one-for-one under the finite-sample approximation, including a possible contact at the moving hard wall.

Proposition 5.15. *If* $r_n \to r \in (0, 1/4)$*, then*

$$|\operatorname{supp}(\widehat{G}_n)| \Rightarrow \mathsf{K}_r.$$

Moreover, $\mathbb{P}\{\mathsf{K}_r = 1\} = p(r)$.

Proof. Proposition 5.13 and positivity of every limiting mass force at least one finite-sample atom into each small, disjoint neighborhood of a limiting atom. Lemma 5.8 gives the compactness needed to extract the corresponding KKT fields. For the reverse bound, work on the event from Proposition 5.13 on which both signed edge measures are supported in a fixed compact interval and have bounded scaled mass. Its probability is at least $1 - \delta - o(1)$.

Away from the finite limiting contact set, Proposition 5.14 gives a strict negative KKT margin and excludes support points. Near an interior active contact, the limiting second derivative is strictly negative. Two-derivative convergence therefore makes the finite-sample KKT field strictly concave there. A strictly concave, nonpositive function has at most one zero, while weak convergence of the positive limiting mass supplies at least one support point. Hence exactly one finite-sample atom lies near each interior limiting atom.

It remains to treat the hard wall. If the boundary is inactive, its strict KKT margin and the nonnegative penalty exclude a crossover contact. Suppose instead that $\nu^*_{b_r}$ has an atom at b_r. On a fixed two-sided neighborhood of b_r, write the normalized finite-sample KKT field as

$$\Gamma_n(\theta) = R_n(\theta) - \mathcal{P}_n(\theta).$$

The penalty-free remainder satisfies $R'_n < -a < 0$, while direct differentiation gives

$$\frac{\mathcal{P}'_n}{\mathcal{P}_n} = -(1-\theta)u_n^2 + O(1), \qquad \frac{\mathcal{P}''_n}{\mathcal{P}_n} = (1-\theta)^2 u_n^4 + O(u_n^2).$$

Wherever $\Gamma'_n \geq 0$, the positive term $-\mathcal{P}'_n$ is at least a; the first display then gives $\mathcal{P}_n \gtrsim u_n^{-2}$, and the second gives $\mathcal{P}''_n \gtrsim u_n^2$. Since $R''_n = O_{\mathbb{P}}(1)$, we have $\Gamma''_n < 0$ there. Thus Γ'_n cannot cross from negative to positive. Two distinct zeros of the nonpositive function Γ_n would force such a crossing, so there is at most one crossover contact. Weak convergence of the boundary mass supplies at least one.

The argument applies independently at the two sample edges. Letting $\delta \downarrow 0$ proves count convergence. Finally, $\nu_b^* = 0$ exactly when $\sup_{b \le \theta < 1} S(\theta) \le 0$. The two edges are independent, so (1.4) gives $\mathbb{P}\{\mathsf{K}_r = 1\} = p(r)$. $\square$

Proposition 5.14 and the implicit-function theorem also control changes in the hard-wall location. Under a common Poisson realization, the atom count is locally constant at each fixed b, almost surely; this yields total-variation rather than only weak continuity.

Proposition 5.16. *The map $b \mapsto \mathcal{L}(K_b)$ is continuous in total variation on $(1/2, 1)$. Consequently, $r \mapsto \mathcal{L}(\mathsf{K}_r)$ is continuous in total variation on $[0, 1/4)$.*

Proof. Couple all optimizers with the same Poisson process and fix deterministic b. If $b \notin \operatorname{supp}(\nu_b^*)$, Proposition 5.14 gives a strict boundary inequality. The support lies a positive distance to the right of b, and the KKT field is strictly negative on a neighborhood of b. For every b' sufficiently close to b, the same measure is feasible and satisfies the KKT conditions on $[b', 1)$. Uniqueness gives $\nu_{b'}^* = \nu_b^*$.

Suppose instead that

$$\nu_b^* = s_1 \delta_b + \sum_{j=2}^{k} s_j \delta_{\theta_j}, \qquad b < \theta_2 < \cdots < \theta_k.$$

Parameterize this boundary stratum by its k masses and $k-1$ interior locations. The objective is analytic in these variables and in b. Its Hessian in the atomic variables is negative definite. Indeed, for a tangent vector v, put

$$g_v(x) = \sum_{j=1}^{k} e^{\theta_j x} \{\dot{s}_j + s_j x \dot{\theta}_j\}, \qquad \dot{\theta}_1 = 0.$$

At the critical point,

$$v^\top H v = -\int_{\mathbb{R}} \frac{g_v(x)^2}{\{1 + F_{\nu_b^*}(x)\}^2} N(dx) + \sum_{j=2}^{k} s_j \Gamma''_{\nu_b^*}(\theta_j) \dot{\theta}_j^2.$$

The second term is nonpositive. If the first vanishes, g_v vanishes at every negative Poisson point. Lemma A.1 and linear independence of the confluent exponentials give $v = 0$. Thus the Hessian is nonsingular.

The implicit-function theorem supplies a nearby critical k-atomic measure on the corresponding boundary stratum, with positive masses. Proposition 5.14 gives a strict negative KKT margin away from the contacts, strict quadratic maxima at the interior contacts, and a strictly negative inward derivative at the boundary contact. These inequalities persist as b varies, so the critical branch satisfies the global KKT conditions and is the unique optimizer. Hence $K_{b'} = K_b$ for all sufficiently close b'.

We have shown that if $b_m \to b$, then $\mathbf{1}_{\{K_{b_m} \neq K_b\}} \to 0$ almost surely. Dominated convergence therefore gives

$$d_{\mathrm{TV}}\{\mathcal{L}(K_{b_m}), \mathcal{L}(K_b)\} \le \mathbb{P}\{K_{b_m} \neq K_b\} \longrightarrow 0.$$

The same coupling handles the sum of two independent copies, and $r \mapsto b_r$ is continuous on $(0, 1/4)$. At zero, $\mathbb{P}\{\mathsf{K}_r \neq 1\} = 1 - p(r) \to 0$ by Proposition 1.2, which proves total-variation continuity on $[0, 1/4)$. $\square$

Proof of Theorem 1.3. For $r \in (0, 1/4)$, the theorem is Proposition 5.15 together with Proposition 5.16. If $r_n \to 0$, Theorem 1.1 gives $\mathbb{P}\{|\operatorname{supp}(\widehat{G}_n)| = 1\} \to 1$, so the count converges to $\mathsf{K}_0 = 1$. $\square$

5.7 Divergence at the critical endpoint

This subsection proves Proposition 1.4. In one-sided coordinates, write $b = 1/2 + \beta$. The logarithmic distance from the wall to the endpoint $1/2$ then grows like $\log(1/\beta)$. A positive density of separated score locations can each contribute order-one likelihood gain, whereas a single atom can exploit only the largest score on this interval.

For $h > 0$, define the normalized endpoint score and its negative-tail quadratic variation by

$$X(h) = \sqrt{2h}\, S(1/2 + h), \qquad Q_h = \int_{-\infty}^{0} e^{(1+2h)x}\, N(dx).$$

The KMT construction in the proof of Proposition 4.3 supplies all the uniform information needed below. We collect those consequences together with the Poisson strong law.

Lemma 5.17. *On an extension of the probability space, for every $0 < h_0 < 1/4$ there is a centered stationary Gaussian process G_{h_0} with covariance*

$$\mathbb{E} G_{h_0}(s) G_{h_0}(t) = \operatorname{sech}\left(\frac{t-s}{2}\right)$$

such that

$$\sup_{t\geq 0} |X(h_0 e^{-t}) - G_{h_0}(t)| \leq R\sqrt{h_0},$$

where $R < \infty$ almost surely and can be chosen simultaneously for all $h_0 < 1/4$. Moreover,

$$\sup_{0\leq t\leq L} G_{h_0}(t) = O_{\mathbb{P}}\{\sqrt{\log(L+2)}\}$$

uniformly in h_0, and each fixed-spacing sequence $(G_{h_0}(j\Delta))_{j\geq 0}$ is stationary and ergodic. Finally,

$$2hQ_h \longrightarrow 1 \quad \textit{almost surely as } h \downarrow 0, \qquad \sup_{0<h\leq h_0} \sqrt{h}\, X(h)^+ \longrightarrow 0 \quad \textit{in probability as } h_0 \downarrow 0.$$

Proof. Under $t = e^{-x}$, the negative Poisson process becomes a unit-rate Poisson process Π on $[1, \infty)$, and

$$\int_{-\infty}^{0} e^{(1/2+h)x} \{N - \mu\}(dx) = \int_{1}^{\infty} t^{-1/2-h} \{\Pi(dt) - dt\}.$$

The infinite-horizon KMT coupling in (4.3) replaces the centered counting process by Brownian motion with an error bounded by an almost surely finite random multiple of $1 + \log t$. Integration by parts makes the normalized error $O(R\sqrt{h_0})$, uniformly for $0 < h \leq h_0$. The positive half-line contains only finitely many Poisson points almost surely, and its contribution to the normalized process is another almost surely finite random multiple of $\sqrt{h_0}$. The normalized Brownian integrals have covariance

$$\frac{2\sqrt{hk}}{h+k};$$

putting $h = h_0 e^{-t}$ gives the stated sech covariance.

A unit-interval Gaussian maximal inequality and a union bound give the maximum estimate. Since the covariance tends to zero at infinity, every fixed-spacing sampled sequence is ergodic. The same maximum estimate and Borel–Cantelli show that $\sup_{t\geq 0} e^{-t/2} G_{h_0}(t)^+ < \infty$ almost surely, with a law independent of h_0; the coupling then gives the second limiting display.

Finally, if $A(t) = \Pi([1,t])$, integration by parts gives

$$Q_h = (1+2h) \int_1^\infty A(t) t^{-2-2h} \, dt.$$

The Poisson strong law $A(t)/t \to 1$, followed by the usual Abelian splitting at a fixed large cutoff, yields $2hQ_h \to 1$. □

Let

$$V_b = \mathcal{J}_{N,b}(\nu_b^*)$$

be the unrestricted one-sided optimum. A logarithmically spaced collection of small atoms obtains likelihood proportional to the length of the endpoint window.

Proposition 5.18. *There is a deterministic $c > 0$ such that*

$$\mathbb{P}\left\{V_{1/2+\beta} \geq c \log(1/\beta)\right\} \longrightarrow 1 \qquad (\beta \downarrow 0).$$

Proof. Put $L_0 = \log(1/\beta)$, $H_\beta = L_0^{-4}$, and $L_\beta = \log(H_\beta/\beta) \sim L_0$. Fix $\Delta > 0$, and take

$$h_j = H_\beta e^{-j\Delta}, \qquad 0 \leq j < J_\beta, \qquad J_\beta = \lfloor L_\beta/\Delta \rfloor.$$

By Lemma 5.17, this score vector is uniformly $o_{\mathbb{P}}(1)$ from a sampled stationary sech process. Choose $q > 0$. Ergodicity gives constants $p, q > 0$ such that, with probability tending to one, at least pJ_β indices satisfy $X(h_j) \geq q$.

For a small constant $a > 0$, place mass $a\sqrt{2h_j}$ at $1/2 + h_j$ for each favorable index. The linear gain is at least $apqJ_\beta$. Since $h(y) \geq -y^2/2$, its nonlinear cost is at most one half of $\int F^2 \, dN$. Dominate this integral by the competitor using all grid points. On the negative half-line its expectation is

$$a^2 \sum_{j,l<J_\beta} \frac{2\sqrt{h_j h_l}}{h_j + h_l} = a^2 \sum_{j,l<J_\beta} \operatorname{sech}\left(\frac{(j-l)\Delta}{2}\right) \leq C_\Delta a^2 J_\beta.$$

Because the full-grid transform at zero is $O(a\sqrt{H_\beta})$, the variance of this Poisson integral is at most $C_\Delta a^4 H_\beta J_\beta$. Its positive-half-line contribution is $o_{\mathbb{P}}(1)$, since that half-line contains finitely many points. Consequently,

$$\mathcal{J}_{N,1/2+\beta}(\nu_\beta) \geq \{apq - C_\Delta a^2 + o_{\mathbb{P}}(1)\} J_\beta.$$

Choose a so that the deterministic coefficient is positive. □

The sparse upper bound rests on an exact subadditivity identity. It avoids having to cover the class of all sparse measures.

Lemma 5.19. *If $\nu = \sum_{j=1}^D s_j \delta_{\theta_j}$, then*

$$\mathcal{J}_{N,b}(\nu) \leq \sum_{j=1}^D \mathcal{J}_{N,b}(s_j \delta_{\theta_j}).$$

Proof. For $y, z \geq 0$,

$$h(y+z) - h(y) - h(z) = \log \frac{1+y+z}{(1+y)(1+z)} \leq 0.$$

Induction proves subadditivity of the nonlinear term, and the score term is linear. □

Optimizing a single atom costs only the square of the largest normalized score. The Gaussian maximum over logarithmic time therefore contributes the smaller factor $\log\log(1/\beta)$.

Proposition 5.20. *Let*

$$M_\beta = \sup_{\beta \le h < 1/2} \sup_{s \ge 0} \{\mathcal{J}_{N,1/2+\beta}(s\delta_{1/2+h})\}_+.$$

Then

$$M_\beta = O_{\mathbb{P}}\{\log\log(1/\beta)\}.$$

Proof. For $x < 0$, the inequality $h(y) \le -y^2/\{2(1+y)\}$ gives

$$\mathcal{J}_{N,1/2+\beta}(s\delta_{1/2+h}) \le sS(1/2+h) - \frac{s^2 Q_h}{2(1+s)}. \tag{5.15}$$

Fix $\delta > 0$. Lemma 5.17 allows us to choose deterministic $h_0 > 0$ such that, with probability at least $1-\delta$,

$$Q_h \ge \frac{1}{4h}, \qquad \sqrt{h}\,X(h)^+ \le \frac{1}{16} \qquad (0 < h \le h_0).$$

On this event $S(1/2+h) \le Q_h/4$. The right side of (5.15) is nonpositive for $s \ge 1$; for $0 \le s \le 1$, quadratic optimization gives

$$\mathcal{J}_{N,1/2+\beta}(s\delta_{1/2+h}) \le \frac{S(1/2+h)_+^2}{Q_h} \le 2X(h)_+^2.$$

Lemma 5.17 bounds the Gaussian maximum over this $O\{\log(1/\beta)\}$-length interval by $O_{\mathbb{P}}(\sqrt{\log\log(1/\beta)})$.

On $h_0 \le h < 1/2$, the best one-atom gain is almost surely finite and independent of β. Indeed, $S(\theta) \to -\infty$ as $\theta \uparrow 1$, so only a random compact interval matters. Uniform coercivity in s follows from the negative-point block argument in the proof of Proposition 5.4. Since δ was arbitrary, the result follows. □

We now compare the unrestricted logarithmic gain with the best gain available to a sparse optimizer.

Proof of Proposition 1.4. Put $r = r_\delta = 1/4 - \delta$ and $\beta = b_r - 1/2 = 1/2 - \sqrt{r}$, so $\beta \sim \delta$. Write $a_\delta = \log(1/\delta)/\log\log(1/\delta)$. Fix $\zeta > 0$. Proposition 5.20 allows us to choose $C < \infty$ so that

$$\limsup_{\beta \downarrow 0} \mathbb{P}\{M_\beta > C\log\log(1/\beta)\} \le \zeta.$$

On the complementary event, if $K_{b_r} \le \eta a_\delta$, then Lemma 5.19 gives

$$V_{b_r} \le K_{b_r} M_\beta \le \{C\eta + o(1)\}\log(1/\beta).$$

Proposition 5.18 supplies a deterministic $c > 0$ for which $V_{b_r} \ge c\log(1/\beta)$ with probability tending to one. Thus, for every $\eta < c/(2C)$,

$$\limsup_{\delta \downarrow 0} \mathbb{P}\{K_{b_r} \le \eta a_\delta\} \le \zeta.$$

Because ζ was arbitrary, the nested limit is zero for the one-sided count. Finally, $\mathsf{K}_r = 1 + K_{b_r}^+ + K_{b_r}^-$, and the event in the proposition implies $K_{b_r}^+ \le \eta a_\delta$, proving the stated result. □

A Transversality for the Limiting Poisson Likelihood

This appendix proves Proposition 5.14, the transversality input used to transfer limiting active sets in Theorem 1.3. The argument shows that every active KKT contact of the limiting Poisson likelihood is nondegenerate and that no inactive contact occurs. We condition on finitely many negative Poisson points and apply parametric transversality; the elementary zero bound below supplies the required Jacobian rank.

Lemma A.1. *Let $\alpha_1, \dots, \alpha_m$ be distinct real numbers, and let p_j be real polynomials of degrees at most d_j. Unless*

$$A(x) = \sum_{j=1}^{m} p_j(x) e^{\alpha_j x}$$

vanishes identically, it has at most $\sum_j (d_j + 1) - 1$ real zeros, counted with multiplicity.

Proof. The functions

$$e^{\alpha_j x}, x e^{\alpha_j x}, \dots, x^{d_j} e^{\alpha_j x}, \qquad 1 \le j \le m,$$

form an extended complete Chebyshev system on $\mathbb{R}$. After their common exponential factors are removed, the successive Wronskians are nonzero confluent Vandermonde determinants. The repeated-Rolle argument for such systems gives the stated zero bound; see [KS66, Chapter I]. □

We now prove the contact statement used in the main text. The continuously distributed Poisson locations supply more perturbation parameters than any bad-contact stratum has equations.

Proof of Proposition 5.14. We show that each bad-contact event is null. Suppose first that the optimizer has k distinct interior atoms with positive masses. Its parameters range over the open stratum

$$s_j > 0, \qquad b < \theta_1 < \cdots < \theta_k < 1.$$

Choose $q = 6(k+1)^2$ negative Poisson points and isolate them in disjoint intervals with rational endpoints, each containing exactly one point. Conditional on the process outside these intervals and on the count events, their locations $x_1, \dots, x_q$ have a joint density positive on the product of the intervals. The rest of the process contributes a fixed real-analytic background. The selected points contribute

$$\sum_{\ell=1}^{q} \log\{1 + F_\nu(x_\ell)\}$$

to the objective: inserting a point at x adds $F_\nu(x)$ to the linear field and $h\{F_\nu(x)\}$ to the nonlinear term.

We work first on rational compact substrata on which masses are bounded away from zero, atom locations are separated, and all free locations remain away from their interval endpoints. These substrata countably cover every finite atomic measure with distinct positive atoms.

The active parameters obey the $2k$ analytic stationarity equations

$$\Gamma_\nu(\theta_j) = 0, \qquad \Gamma'_\nu(\theta_j) = 0, \qquad 1 \le j \le k.$$

An inactive interior contact θ_0 adds the two equations $\Gamma_\nu(\theta_0) = \Gamma'_\nu(\theta_0) = 0$ and the one variable θ_0. A flat active contact adds $\Gamma''_\nu(\theta_j) = 0$ and no variable. Let Φ collect the active equations and the appropriate bad-contact equations.

We claim that Φ, viewed as a function of the atomic parameters, the possible location θ_0, and $x_1, \dots, x_q$, is a submersion. Every row of Φ is a first variation in one of the generalized directions

$$\delta_{\theta_j}, \quad \delta'_{\theta_j}, \quad \delta_{\theta_0}, \quad \delta'_{\theta_0}, \quad \delta''_{\theta_j}.$$

Take a linear combination of these rows and denote the corresponding generalized signed measure by η. The contribution of a selected point at x to this row combination is

$$\frac{F_\eta(x)}{1+F_\nu(x)}.$$

If the combination annihilates every Jacobian column corresponding to a selected location, then

$$\frac{d}{dx}\left\{\frac{F_\eta(x)}{1+F_\nu(x)}\right\}=0 \qquad \text{at } x=x_1,\dots,x_q. \tag{A.1}$$

After multiplication by $\{1+F_\nu(x)\}^2$, the numerator is the exponential polynomial

$$A_\eta(x)=F_\eta'(x)\{1+F_\nu(x)\}-F_\eta(x)F_\nu'(x).$$

It uses at most $(k+1)^2$ distinct exponents, with polynomial coefficients of degree at most two. Lemma A.1 bounds the number of zeros of a nonzero A_η by less than $3(k+1)^2$. Equation (A.1) supplies twice that many, so $A_\eta\equiv 0$.

It follows that $F_\eta/(1+F_\nu)$ is constant. Both numerator and denominator tend respectively to zero and one as $x\to-\infty$, so the constant is zero and $F_\eta\equiv 0$. Linear independence of the generalized exponentials gives $\eta=0$. The generalized point masses listed above are linearly independent because the active locations are distinct and an inactive location is not active. Hence the Jacobian columns in $x_1,\dots,x_q$ already have full row rank.

The parametric transversality theorem [GP10, Chapter 2] now applies. For almost every selected-point configuration, the map obtained by fixing those locations is transverse to zero. For an inactive interior contact, the domain has dimension $2k+1$ and the target has dimension $2k+2$; for a flat active contact, the corresponding dimensions are $2k$ and $2k+1$. In either case the zero set is empty.

The boundary strata are analogous. If b is inactive, the bad equation $\Gamma_\nu(b)=0$ adds one equation and no variable. If b is active, its location is fixed, so the active stratum has $2k-1$ parameters, and the bad equality $\Gamma_\nu'(b)=0$ adds one equation. The same generalized-direction argument, now using δ_b or δ_b', makes both systems submersions with targets larger than their domains.

Finally remove the conditioning. Every locally finite simple Poisson configuration contains arbitrarily many negative points that can be isolated in disjoint rational intervals. There are only countably many choices of k, compact parameter substrata, contact windows, and rational interval tuples, and every conditional bad event above has probability zero. Colliding locations and zero masses require no extra stratum because the optimizer is represented by its distinct support points with positive masses.

Since $\Gamma_\nu\le 0$, an active interior zero has $\Gamma_\nu''\le 0$, and an active boundary zero has $\Gamma_\nu'(b)\le 0$. The null events just excluded are exactly the equality cases. This proves all three assertions. $\square$

B Nonmonotonicity of the Full Support Count

We prove Proposition 1.8. Throughout this appendix,

$$k_t(x,y)=\exp\left\{-\frac{(x-y)^2}{2t}\right\};$$

the omitted Gaussian normalizing constant is independent of the mixing distribution. For a weighted empirical measure

$$P=\sum_{i=1}^q w_i\delta_{x_i}, \qquad w_i>0, \quad \sum_i w_i=1,$$

write $q_{t,G}(x_i) = \int k_t(x_i, y)\, dG(y)$ and

$$D_{t,P,G}(y) = \sum_i w_i \frac{k_t(x_i, y)}{q_{t,G}(x_i)}. \tag{B.1}$$

Lindsay's directional-derivative criterion [Lin83] says that G is an NPMLE exactly when

$$D_{t,P,G}(y) \leq 1 \quad (y \in \mathbb{R}). \tag{B.2}$$

The identity

$$\int D_{t,P,G}(y)\, dG(y) = 1. \tag{B.3}$$

then makes equality automatic G-almost everywhere. Strict concavity of the log likelihood in the fitted vector implies that all maximizers have the same fitted vector. Thus, if the equality set in (B.2) is finite and its kernel columns are linearly independent, the NPMLE is unique.

We also use the standard stability of nondegenerate zeros: if $F_N \to F$ in C^1 near z_0, $F(z_0) = 0$, and $DF(z_0)$ is invertible, then F_N has a zero converging to z_0.

The first lemma produces the high-variance dual certificate. Its first r data locations remain fixed, while the last location and its contact move to $+\infty$.

Lemma B.1. *Fix $r \geq 1$ and $t > 0$. There are fixed points*

$$a_1 < \cdots < a_r = 0$$

such that, for every sufficiently large B, there are positive coefficients $c_1(B), \ldots, c_{r+1}(B)$ for which

$$C_B(y) = \sum_{i=1}^{r} c_i(B) k_t(a_i, y) + c_{r+1}(B) k_t(B, y) \leq 1. \tag{B.4}$$

Equality holds at exactly $r+1$ nondegenerate maxima $\theta_1(B), \ldots, \theta_{r+1}(B)$. Moreover,

$$c_i(B) \to c_i^0 > 0, \quad \theta_i(B) \to \alpha_i \quad (i \leq r), \qquad c_{r+1}(B) \to 1, \quad \theta_{r+1}(B) - B \to 0,$$

and the square contact matrix

$$K_B = \big(k_t(x_i, \theta_j(B))\big)_{i,j=1}^{r+1}, \qquad (x_1, \ldots, x_{r+1}) = (a_1, \ldots, a_r, B), \tag{B.5}$$

is invertible.

Proof. We first build the r fixed peaks. Choose a large separation L and put $a_i = (i - r)L$. Let the coefficients $c = (c_1, \ldots, c_r)$ range over a fixed neighborhood of $\mathbf{1}$. For fixed $R > 0$, local coordinates $y = a_i + s$ give, uniformly over i, c and $|s| \leq R$,

$$\sum_{j=1}^{r} c_j k_t(a_j, a_i + s) = c_i e^{-s^2/(2t)} + E_{i,L,c}(s),$$

where

$$\|E_{i,L,c}\|_{C^2[-R,R]} \leq C_{r,R,t}(1 + L^2) \exp\left\{-\frac{(L-R)^2}{2t}\right\} \longrightarrow 0. \tag{B.6}$$

Choose $0 < \delta < \min\{R, \sqrt{t}\}$. The derivative of the limiting single peak is bounded away from zero on $\{\delta \le |s| \le R\}$, while its second derivative is negative on $[-\delta, \delta]$. Hence each interval $[a_i - R, a_i + R]$ contains exactly one critical point

$$\alpha_i(c, L) = a_i + s_i(c, L), \qquad s_i(c, L) \to 0,$$

which is a nondegenerate maximum.

Let $M_i(c, L)$ be the value at this maximum. Since the spatial derivative vanishes there,

$$\frac{\partial M_i}{\partial c_j}(c, L) = k_t(a_j, \alpha_i(c, L)).$$

Consequently $M(c, L) - \mathbf{1} \to c - \mathbf{1}$ in C^1 near $\mathbf{1}$. Stability of nondegenerate zeros gives coefficients $c(L) \to \mathbf{1}$ for which all r maximum heights equal one. Choose R so large that $2re^{-R^2/(2t)} < 1$, and then choose $L > 2R$ large enough that the coefficients $c(L)$ and critical points $\alpha_i(c(L), L)$ constructed above exist. Outside the union of the r intervals, every center is at distance at least R, so the Gaussian sum is strictly below one. Inside each interval its unique critical point is its maximum, of height one. We have therefore constructed

$$H_0(y) = \sum_{i=1}^{r} c_i^0 k_t(a_i, y) \le 1$$

with exactly r nondegenerate contacts α_i. The matrix

$$A = \big(k_t(a_i, \alpha_j)\big)_{i,j=1}^{r}$$

is invertible because it converges to the identity as $L \to \infty$.

We now add the remote peak. For coefficients near $(c_1^0, \ldots, c_r^0, 1)$, the old maxima persist near the α_i, and a new nondegenerate maximum appears near B. The map from the coefficients to these $r+1$ maximum heights converges in C^1, as $B \to \infty$, to the decoupled height map. Its derivative at the limiting coefficient vector is

$$\begin{pmatrix} A^{\mathsf{T}} & 0 \\ 0 & 1 \end{pmatrix}.$$

Stability of nondegenerate zeros gives positive coefficients for which all $r + 1$ heights are one, with the asserted limits.

It remains to check the global inequality. On small fixed neighborhoods of the old contacts, the perturbed old cluster has a unique strictly concave maximum of height one, and the remote term is negligible. On the complement of these neighborhoods in a large fixed interval, H_0 has a fixed gap below one. An identical local argument controls the new maximum in a fixed neighborhood of B. Finally, enlarge the old and remote regions until the entire old cluster is below $1/4$ outside the old region and the remote term is below $1/4$ outside the region around B. This proves (B.4) globally and identifies every contact. The matrix K_B converges blockwise to $\mathrm{diag}(A, 1)$, so it is invertible for large B. □

An equal-height certificate can be inverted into empirical weights. This lets us prescribe positive masses at all high-variance contacts before choosing the sample multiplicities.

Lemma B.2. *Suppose*

$$C(y) = \sum_{i=1}^{q} c_i k_t(x_i, y) \le 1$$

has equality exactly at $\theta_1, \ldots, \theta_q$, where every $c_i > 0$, and suppose $K = (k_t(x_i, \theta_j))_{i,j=1}^q$ is invertible. For $p \in \Delta_q^\circ$, define

$$M = \operatorname{diag}(c_1, \ldots, c_q)K, \qquad w = Mp, \qquad G = \sum_{j=1}^q p_j \delta_{\theta_j}.$$

Then $w \in \Delta_q^\circ$, and G is the unique NPMLE for the weighted empirical distribution $\sum_i w_i \delta_{x_i}$.

Proof. Every column of M sums to one because

$$\sum_i M_{ij} = \sum_i c_i k_t(x_i, \theta_j) = C(\theta_j) = 1.$$

Thus w is a strictly positive probability vector. Moreover,

$$q_{t,G}(x_i) = (Kp)_i, \qquad \frac{w_i}{q_{t,G}(x_i)} = c_i.$$

The dual certificate of G is exactly C, so (B.2) proves optimality. Any other maximizer is supported on the same contact set and has a mass vector p' satisfying $Kp' = Kp$. Invertibility gives $p' = p$. □

We next show that a suitable two-cluster empirical distribution has exactly two atoms at the smaller variance. The exponent condition below says that the smaller cluster has much more mass than the Gaussian overlap between the two clusters.

Lemma B.3. *Fix $a_1 < \cdots < a_r = 0$. For $B \to \infty$, let*

$$P_B = \sum_{i=1}^r w_{i,B}\delta_{a_i} + w_{\star,B}\delta_B$$

be a probability measure with all weights positive, and put $\varepsilon_B = w_{r,B}$. Suppose

$$\varepsilon_B \to 0, \tag{B.7}$$

$$\log \varepsilon_B = -\alpha B^2 + o(B^2) \quad \textit{for some } 0 < \alpha < \frac{1}{2}, \tag{B.8}$$

$$\frac{w_{i,B}}{\varepsilon_B k_1(a_i, 0)} \to 0 \qquad (i < r). \tag{B.9}$$

Then, for all sufficiently large B, the variance-one NPMLE is unique and has exactly two atoms.

Proof. Condition (B.9) gives $\sum_{i<r} w_{i,B} = o(\varepsilon_B)$, and hence

$$w_{\star,B} = 1 - \varepsilon_B\{1 + o(1)\}. \tag{B.10}$$

We seek an NPMLE of the form

$$G_{\rho,u,z} = \varepsilon_B\rho\,\delta_u + (1 - \varepsilon_B\rho)\delta_{B+z},$$

where (ρ, u, z) is near $(1, 0, 0)$. Let

$$\begin{aligned} q_i(\rho, u, z) &= \varepsilon_B\rho k_1(a_i, u) + (1 - \varepsilon_B\rho)k_1(a_i, B + z), \\ q_\star(\rho, u, z) &= \varepsilon_B\rho k_1(B, u) + (1 - \varepsilon_B\rho)k_1(B, B + z), \end{aligned}$$

and denote the corresponding dual function by $D_{\rho,u,z}$.

For every fixed $C > 0$ and integer $j \geq 0$, (B.8) implies

$$B^j \frac{\exp\{-(B-C)^2/2\}}{\varepsilon_B} \longrightarrow 0. \tag{B.11}$$

Indeed, the logarithm of the left side is $-(\frac{1}{2}-\alpha)B^2 + O(B) + o(B^2)$. Since every derivative of a Gaussian kernel is the kernel times a polynomial, (B.11) controls all parameter and spatial derivatives of the cross-cluster terms below.

Set $\lambda_i = w_{i,B}/q_i$ and $\lambda_\star = w_{\star,B}/q_\star$. Uniformly for (ρ, u, z) near $(1, 0, 0)$,

$$\lambda_i \to 0 \qquad \text{in } C^1,\ i < r, \tag{B.12}$$

$$\lambda_r \to \frac{1}{\rho k_1(0,u)} \qquad \text{in } C^1, \tag{B.13}$$

$$\lambda_\star \to \frac{1}{k_1(0,z)} \qquad \text{in } C^1. \tag{B.14}$$

For example, $q_i \geq \varepsilon_B \rho k_1(a_i, u)$, so

$$0 \leq \lambda_i \leq C \frac{w_{i,B}}{\varepsilon_B k_1(a_i, 0)} \longrightarrow 0.$$

Differentiation introduces either a bounded local derivative or a cross-cluster derivative divided by ε_B, which is controlled by (B.11). Differentiating the identity $w_{r,B} = \varepsilon_B$ in the same way proves (B.13). Finally, (B.14) follows from (B.10) and

$$q_\star = (1 - \varepsilon_B \rho) k_1(0, z) + \varepsilon_B \rho k_1(B, u).$$

Consider the three equations

$$F_B(\rho, u, z) = \begin{pmatrix} D_{\rho,u,z}(u) - 1 \\ D'_{\rho,u,z}(u) \\ D'_{\rho,u,z}(B+z) \end{pmatrix} = 0.$$

Equations (B.11)–(B.14) imply convergence in C^1 near $(1, 0, 0)$ to

$$F_\infty(\rho, u, z) = \begin{pmatrix} \rho^{-1} - 1 \\ -u/\rho \\ -z \end{pmatrix}.$$

Since $F_\infty(1, 0, 0) = 0$ and $DF_\infty(1, 0, 0) = -I_3$, there is a solution

$$\rho_B \to 1, \qquad u_B \to 0, \qquad z_B \to 0.$$

Put

$$\pi_B = \varepsilon_B \rho_B, \qquad v_B = B + z_B, \qquad G_B = \pi_B \delta_{u_B} + (1 - \pi_B)\delta_{v_B}.$$

For all large B, $0 < \pi_B < 1$, and construction gives

$$D_B(u_B) = 1, \qquad D'_B(u_B) = D'_B(v_B) = 0.$$

The second contact height is automatic: by (B.3),

$$1 = \pi_B D_B(u_B) + (1 - \pi_B) D_B(v_B),$$

so $D_B(v_B) = 1$.

It remains to exclude other contacts. At the solution,

$$\lambda_r \to 1, \qquad \sum_{i<r} \lambda_i \to 0, \qquad \lambda_\star \to 1.$$

Consequently,

$$D_B(y) \to k_1(0, y) \quad \text{in } C^2 \text{ on fixed compact sets,}$$

and

$$D_B(B + s) \to k_1(0, s) \quad \text{in } C^2 \text{ on fixed compact } s\text{-sets.}$$

Choose $0 < \delta < 1$. The limiting Gaussian is strictly concave on $[-\delta, \delta]$, so u_B is the unique contact there. Uniform convergence gives a strict gap below one on $[-R, R] \setminus (-\delta, \delta)$. The translated argument shows that v_B is the only contact in $[B - R, B + R]$. Finally choose R so large that $3e^{-R^2/2} < 1$. Outside these intervals,

$$D_B(y) \le (1 + o(1))k_1(0, y) + (1 + o(1))k_1(B, y) + o(1) \le 2e^{-R^2/2} + o(1) < 1.$$

Thus G_B satisfies (B.2), with equality exactly at its two atoms.

The two kernel columns are linearly independent because the rows corresponding to 0 and B have determinant

$$\begin{aligned} &\det \begin{pmatrix} k_1(0, u_B) & k_1(0, v_B) \\ k_1(B, u_B) & k_1(B, v_B) \end{pmatrix} \\ &\quad = k_1(0, v_B) k_1(B, u_B) \left\{ e^{B(v_B - u_B)} - 1 \right\} > 0. \end{aligned}$$

The NPMLE is therefore unique. □

We now choose the high-variance fitted masses so that the induced empirical weights satisfy the two-cluster lemma at variance one.

Proof of Proposition 1.8. Rescaling the observations and mixing locations by $S^{-1/2}$ reduces the claim to $S = 1$ and $T = \tau > 1$. Put $r = m - 1$, and apply Lemma B.1 at variance τ. Thus

$$x_1 = a_1 < \cdots < x_r = a_r = 0 < x_m = B$$

support a certificate

$$C_B(y) = \sum_{i=1}^{m} c_i(B) k_\tau(x_i, y) \le 1$$

with exactly m contacts, the last of which satisfies $\theta_m(B) = B + o(1)$. Let K_B be the contact matrix (B.5), and set

$$M_B = \operatorname{diag}\{c_1(B), \ldots, c_m(B)\} K_B.$$

Choose the high-variance NPMLE masses

$$\eta_B = e^{-B^2/\tau}, \qquad p_j^{(B)} = \frac{\eta_B}{r} \quad (j \le r), \qquad p_m^{(B)} = 1 - \eta_B,$$

and define the empirical weights $w^{(B)} = M_B p^{(B)}$. Lemma B.2 gives a unique m-atom NPMLE at variance τ.

For $i \le r$, write

$$q_i^{(B)} = \sum_{j=1}^{m} p_j^{(B)} k_\tau(a_i, \theta_j(B)).$$

Since the first r masses sum to η_B,

$$q_i^{(B)} = (1 - \eta_B) k_\tau(a_i, \theta_m(B)) + O(\eta_B).$$

Because a_i is fixed and $\theta_m(B) = B + o(1)$,

$$\frac{\eta_B}{k_\tau(a_i, \theta_m(B))} = \exp\left\{-\frac{B^2}{2\tau} + O(B) + o(B)\right\} \longrightarrow 0.$$

It follows that

$$q_i^{(B)} = \{1 + o(1)\} k_\tau(a_i, \theta_m(B)). \tag{B.15}$$

Put $\varepsilon_B = w_r^{(B)} = c_r(B) q_r^{(B)}$. Since $c_r(B) \to c_r^0 > 0$,

$$\log \varepsilon_B = -\frac{B^2}{2\tau} + o(B^2).$$

Thus (B.8) holds with $\alpha = 1/(2\tau) < 1/2$. For $i < r$,

$$\begin{aligned} \frac{w_i^{(B)}}{\varepsilon_B k_1(a_i, 0)} &= \{1 + o(1)\} \frac{c_i^0}{c_r^0} \frac{k_\tau(a_i, \theta_m(B))}{k_\tau(0, \theta_m(B)) k_1(a_i, 0)} \\ &= \{1 + o(1)\} \frac{c_i^0}{c_r^0} \exp\left\{\frac{a_i \theta_m(B)}{\tau} + \frac{a_i^2}{2}\left(1 - \frac{1}{\tau}\right)\right\} \longrightarrow 0, \end{aligned}$$

because $a_i < 0$ and $\theta_m(B) \to \infty$. Lemma B.3 now gives a unique two-atom NPMLE at variance one for all sufficiently large B.

It remains to replace the weights by integer multiplicities. The matrix M_B is invertible and satisfies $\mathbf{1}^\mathsf{T} M_B = \mathbf{1}^\mathsf{T}$. Hence $M_B \Delta_m^\circ$ is relatively open in the affine probability hyperplane and contains $w^{(B)}$. Along a sequence $B_N \to \infty$, choose a rational vector

$$\widetilde{w}^{(N)} \in M_{B_N} \Delta_m^\circ \cap \mathbb{Q}^m$$

so close to $w^{(B_N)}$ that

$$\max_i \left| \frac{\widetilde{w}_i^{(N)}}{w_i^{(B_N)}} - 1 \right| \longrightarrow 0.$$

Then $\widetilde{p}^{(N)} = M_{B_N}^{-1} \widetilde{w}^{(N)}$ lies in Δ_m°, so the same certificate gives a unique m-atom NPMLE at variance τ. The relative approximation preserves (B.8) and (B.9), so the variance-one NPMLE still has two atoms. Expressing the coordinates of one such rational vector over a common denominator gives an ordinary sample with exactly m distinct values and positive integer multiplicities. Rescaling by $\sqrt{S}$ completes the construction. □

References

[AGM09] Jean-Marc Azaïs, Élisabeth Gassiat, and Cécile Mercadier. The likelihood ratio test for general mixture models with or without structural parameter. *ESAIM Probab. Stat.*, 13:301–327, 2009.

[Ano26a] Radial phase transitions for Gaussian mixture NPMLEs in fixed and growing dimension, 2026. In preparation.

[Ano26b] Support divergence of mixture NPMLEs under finite mixtures, 2026. In preparation.

[DPSZ02] Amir Dembo, Bjorn Poonen, Qi-Man Shao, and Ofer Zeitouni. Random polynomials having few or no real zeros. *J. Amer. Math. Soc.*, 15(4):857–892, 2002.

[DWYZ23] Natalie Doss, Yihong Wu, Pengkun Yang, and Harrison H Zhou. Optimal estimation of high-dimensional Gaussian location mixtures. *Ann. Stat.*, 51(1):62–95, 2023.

[DZ16] Lee H Dicker and Sihai D Zhao. High-dimensional classification via nonparametric empirical Bayes and maximum likelihood inference. *Biometrika*, 103(1):21–34, 2016.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 1986.

[FFM25] Naomi D. Feldheim, Ohad N. Feldheim, and Sumit Mukherjee. Persistence and ball exponents for Gaussian stationary processes. *Comm. Pure Appl. Math.*, 78(10):1949–2000, 2025.

[FKZ13] Sergey Foss, Dmitry Korshunov, and Stan Zachary. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer, 2013.

[FTZ22] Will FitzGerald, Roger Tribe, and Oleg Zaboronski. Asymptotic expansions for a class of Fredholm Pfaffians and interacting particle systems. *Ann. Probab.*, 50(6):2409–2474, 2022.

[Gar07] John B. Garnett. *Bounded Analytic Functions*, volume 236 of *Graduate Texts in Mathematics*. Springer, revised first edition, 2007.

[GKV18] Jiaying Gu, Roger Koenker, and Stanislav Volgushev. Testing for homogeneity in mixture models. *Econometric Theory*, 34(4):850–895, 2018.

[GP10] Victor Guillemin and Alan Pollack. *Differential topology*, volume 370. American Mathematical Soc., 2010.

[GVDV01] Subhashis Ghosal and Aad W Van Der Vaart. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Stat.*, 29(5):1233–1263, 2001.

[Har85] John A. Hartigan. A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, volume 2, pages 807–810. Wadsworth, Belmont, CA, 1985.

[HY01] Peter Hall and Matthew York. On the calibration of Silverman's test for multimodality. *Statist. Sinica*, 11(2):515–536, 2001.

[JZ09] Wenhua Jiang and Cun-Hui Zhang. General maximum likelihood empirical Bayes estimation of normal means. *Ann. Stat.*, 37(4):1647–1684, 2009.

[JZ19] Wenhua Jiang and Cun-Hui Zhang. Rate of divergence of the nonparametric likelihood ratio test for Gaussian mixtures. *Bernoulli*, 25(4B), 2019.

[Kal17] Olav Kallenberg. *Random Measures, Theory and Applications*. Springer, 2017.

[KG19] Roger Koenker and Jiaying Gu. Minimalist g-Modeling. *Statistical Science*, 34(2):209 – 213, 2019.

[KM14] Roger Koenker and Ivan Mizera. Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *Journal of the American Statistical Association*, 109(506):674–685, 2014.

[KMT75] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent RV's, and the sample DF. I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 32:111–131, 1975.

[KMT76] János Komlós, Péter Major, and Gábor Tusnády. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 34:33–58, 1976.

[KS66] Samuel Karlin and William J. Studden. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, 1966.

[KS26] Taehyun Kim and Bodhisattva Sen. Empirical Bayes estimation and inference via smooth nonparametric maximum likelihood. *arXiv preprint arXiv:2603.27843*, 2026.

[KW56] Jack Kiefer and Jacob Wolfowitz. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, pages 887–906, 1956.

[Lin83] Bruce G Lindsay. The Geometry of Mixture Likelihoods: A General Theory. *Ann. Stat.*, 11(1):86–94, 1983.

[LR93] Bruce G Lindsay and Kathryn Roeder. Uniqueness of estimation and identifiability in mixture models. *Canadian J. Stat.*, 21(2):139–147, 1993.

[PS18] Mihail Poplavskyi and Grégory Schehr. Exact persistence exponent for the 2D-diffusion equation and related Kac polynomials. *Phys. Rev. Lett.*, 121(15):150601, 2018.

[PW20] Yury Polyanskiy and Yihong Wu. Self-Regularizing Property of Nonparametric Maximum Likelihood Estimator in Mixture Models. *arXiv preprint arXiv:2008.08244*, 2020.

[Rob50] Herbert Robbins. A generalization of the method of maximum likelihood: Estimating a mixing distribution (Abstract). *Annals of Mathematical Statistics*, 21(2):314–315, 1950.

[Ros94] Kenneth Rose. A mapping approach to rate-distortion computation and analysis. *IEEE Transactions on Information Theory*, 40(6):1939–1952, 1994.

[SG20] Sujayam Saha and Adityanand Guntuboyina. On the nonparametric maximum likelihood estimator for Gaussian location mixture densities with application to Gaussian denoising. *Ann. Stat.*, 48(2):738–762, 2020.

[Sil81] B. W. Silverman. Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B*, 43(1):97–99, 1981.

[Zha09] Cun-Hui Zhang. Generalized maximum likelihood estimation of normal mixture densities. *Statistica Sinica*, pages 1297–1318, 2009.