Location estimation at variance-gamma poles and cusps

Scott Schubr
Luminary Oddities

November 26, 2026 (edition 4)

Abstract

Nitithumbundit and Chan asked for the optimal rate of convergence and the asymptotic distribution of the maximum leave-one-out likelihood location estimator of Podgórski and Wallin for variance-gamma laws, studied by simulation for shapes ν in $(0, 1]$. Podgórski and Wallin asked whether this estimator reaches the optimal rate $n^{-1/(1+\alpha)}$ at a power pole $|x|^\alpha$. We answer the first question for the symmetric univariate model with known scale at every shape, and the second for even densities. At a power pole, $0 < \nu < 1/2$, the global maximizer converges at rate $n^{-1/(2\nu)}$, with all moments, to the maximizer of an explicit functional of a Poisson process with intensity $|t|^{2\nu-1}$: the midpoint of the consecutive pair that best balances the closeness of the pair against the logarithmic potential of the other points. This holds for every even density decreasing in $|x|$ with a power pole, exponential moments and mild score bounds, so the optimal rate is attained there. At the logarithmic pole $\nu = 1/2$ the rate is $(\log n)/n$, with a drifted Brownian limit. At cusps, $1/2 < \nu < 3/4$, the rate is $n^{-1/(4\nu-1)}$, slower than the exponent $1/(2\nu)$ used in the simulations, with a fractional Brownian limit; at $\nu = 3/4$ it is $(n \log n)^{-1/2}$, and above it the root-n Fisher limit holds. The proofs use capped likelihoods and Hellinger bounds on dyadic shells. The article also treats the minimax constant at the logarithmic pole and dimension at least two.

Keywords. Variance-gamma distribution; location estimation; leave-one-out likelihood; unbounded density; super-efficiency; Poisson point process; argmax; cusp; fractional Brownian motion; minimax risk.
MSC 2020. Primary 62F12; Secondary 62E20, 60G55, 62C20.

Contents

1 Introduction **2**
1.1 The questions . . . 2
1.2 Main results . . . 3
1.3 Further results . . . 3

2 The leave-one-out location estimator at the variance-gamma logarithmic pole **4**
2.1 Estimator and results . . . 4
2.2 Density estimates and a coarse bound . . . 5
2.3 Fine approximation and increment bounds . . . 6
2.4 Process convergence and maximization . . . 7
2.5 The limiting absolute and squared risks . . . 8

3 The leave-one-out location estimator at power poles and cusps 9
3.1 The questions and the results . . . 9
3.2 Variance-gamma densities . . . 12
3.3 The Poisson limit criterion . . . 13
3.4 A capped bound for the leave-one-out criterion . . . 14
3.5 Localization . . . 15
3.6 Local convergence and proof of Theorem 3.2 . . . 16
3.7 Cusps: proof of Theorem 3.4 . . . 17
3.8 Numerical illustration . . . 19

4 Sharp location rates for multivariate variance-gamma laws 20
4.1 Model and phase diagram . . . 20
4.2 Density and weighted differences . . . 21
4.3 The subcritical Poisson phase . . . 22
4.4 Critical experiments . . . 23
4.5 Attaining the critical information . . . 24
4.6 Compact parameter sets and the minimax constant . . . 25

5 Critical variance-gamma location with unknown shape, skew and scatter 25
5.1 Model and results . . . 25
5.2 Density bounds near the pole . . . 26
5.3 A moment pilot for the full parameter . . . 27
5.4 The joint local experiment . . . 28
5.5 Estimating location with unknown nuisance . . . 29

6 The sharp minimax risk at the variance-gamma logarithmic pole 30
6.1 Model and results . . . 30
6.2 Capped comparison . . . 31
6.3 Fine likelihood integrals and moments . . . 32
6.4 Minimax identity and constant . . . 34

7 Location inference as variance-gamma shape approaches one 34
7.1 Results . . . 34
7.2 Annular information . . . 35
7.3 The local experiments . . . 37
7.4 An attaining estimator . . . 39

1 Introduction

1.1 The questions

In the gamma-mixing parameterization used throughout, with mixing law Gamma$(\nu, 1)$, a variance-gamma density in dimension d has a pole at its centre exactly when $\nu \le d/2$: the power pole $|x|^{2\nu-d}$ for $\nu < d/2$ and a logarithmic pole at $\nu = d/2$. In one dimension the density has the power pole $|x|^{2\nu-1}$ for $\nu < 1/2$, a logarithmic pole at $\nu = 1/2$, and a cusp $f(0) - b|x|^{2\nu-1}$ for $1/2 < \nu < 1$. At a pole the likelihood of a location parameter is infinite at every observation, so the maximum likelihood estimator does not exist. Podgórski and Wallin [1] proposed to maximize instead the leave-one-out likelihood, the product of the densities at all observations except the one closest to

the candidate location. For densities $p(x)|x|^\alpha$ with $-1 < \alpha < 0$ and a factor p with bounded derivative they proved consistency and $n^\delta(\widehat{\theta}_n - \theta) \to 0$ in probability for every $\delta < 1/(1+\alpha)$ [1, Theorem 1], a lower bound on the rate, which their proof gives up to a power-of-logarithm factor. They call $n^{-1/(1+\alpha)}$ the optimal rate and record, in their concluding section [1, Section 6], the open question whether the estimator reaches it.

Question 1.1 (Podgórski and Wallin)**.** *Does the maximum leave-one-out likelihood estimator of a location parameter attain the rate $n^{-1/(1+\alpha)}$ at a power pole $|x|^\alpha$, $-1 < \alpha < 0$?*

Nitithumbundit and Chan [2] computed this estimator for the symmetric univariate variance-gamma law at fifty shapes $\nu \in (0, 1]$. They took $1/(1+\alpha) = 1/(2\nu)$ as the "proposed optimal rate" [2, Section 3], fitted rates and limit laws by simulation [2, Section 5], and state that proving the optimal rate of convergence and the asymptotic distribution analytically "is still an open question" [2, Abstract and Section 6]; the journal version is [3].

Question 1.2 (Nitithumbundit and Chan)**.** *For the symmetric univariate variance-gamma law of shape ν, what are the optimal rate of convergence and the asymptotic distribution of the maximum leave-one-out likelihood location estimator?*

1.2 Main results

We answer Question 1.2 at every shape $\nu > 0$, for known scale and the exact global maximizer of the leave-one-out likelihood, and Question 1.1 for every even (symmetric) density satisfying condition (P) of Definition 3.1.

- *Power poles,* $0 < \nu < 1/2$ (Theorem 3.2 and Corollary 3.3). With $r_n = \sigma(nA_{\nu,1})^{-1/(2\nu)}$, $(\widehat{\theta}_n - \theta)/r_n$ converges in law and in every moment to the maximizer $T_{2\nu-1}$ of the Poisson functional (4), which is the midpoint of the consecutive pair of Poisson points maximizing (5). The general theorem gives the rate $n^{-1/(1+\alpha)}$ with an explicit limit for every density with condition (P), and this rate is the minimax rate, as Polfeldt showed [4, 5]. Condition (P) concerns even densities; the class of Podgórski and Wallin also contains asymmetric and one-sided poles. Condition (P) also covers the variance-gamma factor p, whose derivative is of order $|x|^{-2\nu}$ at the origin.

- *Logarithmic pole,* $\nu = 1/2$ (Theorem 2.1). The rate is $(\log n)/n$ and the limit is a drifted Brownian argmax, with limiting absolute and squared risks $3/I$ and $26/I^2$, $I = \pi/\sigma$.

- *Cusps and regular shapes,* $\nu > 1/2$ (Theorem 3.4). The rate is $n^{-1/(4\nu-1)}$ with a fractional Brownian argmax limit for $1/2 < \nu < 3/4$, $(n \log n)^{-1/2}$ with a Gaussian limit at $\nu = 3/4$, and $n^{-1/2}$ with the Fisher limit for $\nu > 3/4$.

Thus the rate exponent is $1/(2\nu)$ on $(0, 1/2)$, $1/(4\nu-1)$ on $(1/2, 3/4)$ and $1/2$ above, with logarithmic corrections at $\nu = 1/2$ and $\nu = 3/4$. The exponent $1/(2\nu)$ of the simulations is exact on the pole range and is replaced by the smaller exponent $1/(4\nu - 1)$ on the cusp range. Section 3.8 compares these rates with exact computations of the estimator and with the exponents fitted in [2, Tables 1 and 2].

1.3 Further results

Section 4 determines the sharp rates for every shape and the minimax constants at the critical shape, in every dimension at least two. Location estimation is critical at shape $\nu = 1$ there, where

the pole is logarithmic in the plane and is the power pole $|x|^{2-d}$, with logarithmically divergent location Fisher information, in dimension $d \geq 3$; it falls between the Poisson regime of a jump and the regular Gaussian regime. Section 5 shows that the critical constant is unaffected when skew and scatter must also be estimated on compact brackets and the shape is known to within a root-n neighbourhood of one. The Brownian argmax law and its second moment 26 are due to Terent'yev [6]; see also [7].

Section 6 identifies the sharp minimax constant in one dimension at the logarithmic pole. The flat-prior posterior and the normalized leave-one-out likelihood converge jointly, as random probability measures, to the same Brownian likelihood measure. Their means therefore share the limiting risk $16\zeta(3)/I^2$ of the canonical Pitman estimator, computed by Rubin and Song [8] for a discontinuous signal in white noise. The posterior mean is Pitman's best equivariant estimator [9], and it is exactly minimax at each sample size, as Girshick and Savage [10] proved for Pitman estimators; Section 6.4 gives a direct proof in this model. The general theorems of Ibragimov and Has'minskii on Bayes estimators [11, Chapter I, Section 10] derive the limit law and moment convergence of posterior means from finite-dimensional convergence of the likelihood ratio together with Hellinger bounds of the kind proved in Section 2; the new step in Section 6 is the leave-one-out average, whose criterion is not a likelihood ratio. The rate $(\log n)/n$ is minimax for critical univariate variance-gamma location, also with unknown gamma rates, by [12, Theorem 6.3]. Since the leave-one-out maximizer has risk $26/I^2$, averaging the leave-one-out likelihood improves it by the classical Rubin–Song efficiency factor $8\zeta(3)/13 \approx 0.740$, here transferred to the variance-gamma likelihood. Section 7 lets the shape ν_n move with the sample size and derives the exact information in the windows $(\nu_n - 1)\log n \to \tau \in \mathbb{R}$ in dimension at least three, and in the plane for every sequence $\nu_n \downarrow 1$ and for shape deficits $1 - \nu_n$ up to the triple-logarithmic threshold; one estimator attains the sharp constant in each of these windows.

This is the fourth edition of this article; it adds Section 3, which answers the two questions above at power poles and cusps, and revises the title, abstract and introduction accordingly. The third edition corrected the description of where the density has a logarithmic pole, stated the scope of the crossover results precisely and added references to classical results on Bayes and Pitman estimators; the second edition added Sections 6 and 7.

2 The leave-one-out location estimator at the variance-gamma logarithmic pole

For a symmetric univariate variance-gamma density at shape 1/2, the maximum leave-one-out likelihood location estimator converges at rate $(\log n)/n$ to the maximizer of a two-sided Brownian motion with negative drift proportional to distance from zero. All fixed absolute moments converge after normalization, with explicit limiting absolute and squared risks. This gives an exact limit for the critical case simulated by Nitithumbundit and Chan. Two capped likelihoods, used only in the proof, connect the actual maximizer to the local likelihood process.

2.1 Estimator and results

Let $X_i = \theta_0 + Z_i$, $i = 1, \ldots, n$, be independent, with $n \geq 2$, and let the density of Z_i be

$$f_\sigma(x) = \frac{1}{\pi\sigma} K_0(|x|/\sigma), \qquad \sigma > 0.$$

The scale and shape are fixed and known. Write $j(t)$ for a sample index closest to t and let

$$Q_n(t) = \sum_{i \neq j(t)} \log f_\sigma(X_i - t).$$

The data are distinct almost surely. Ties at a midpoint give the same criterion. Define $\widehat{\theta}_n$ to be any measurable global maximizer. It can be obtained by checking the adjacent sample midpoints. Indeed, the integral representation of $K_0(x)$ as a positive mixture of e^{-xu} for $u \geq 1$ gives strict radial decrease and strict log convexity for $x > 0$. On each Voronoi interval of a data point, that point is omitted and the remaining log densities are convex in t. The two unbounded intervals have their maxima at their finite ends by monotonicity. This also proves existence and bounds the maximizer by the sample range.

Let $\mathcal{B}$ be standard two-sided Brownian motion and write $d_n = (\log n)/n$.

Theorem 2.1. *For every θ_0, with $I = \pi/\sigma$,*

$$\frac{\widehat{\theta}_n - \theta_0}{d_n} \implies U_\sigma = \operatorname*{argmax}_{u \in \mathbb{R}} \left\{ \sqrt{I}\, \mathcal{B}(u) - \frac{I}{2}|u| \right\}.$$

For every fixed $a > 0$, $\mathbb{E}|(\widehat{\theta}_n - \theta_0)/d_n|^a \to \mathbb{E}|U_\sigma|^a$. In particular the limiting first absolute moment is $3/I$ and the limiting second moment is $26/I^2$. On every fixed compact interval in u, the centered process $Q_n(\theta_0 + ud_n) - Q_n(\theta_0)$ converges in the space of continuous functions to the Brownian process in the display.

In the notation of Nitithumbundit and Chan [3], this is the model with dimension one, $\nu = 1/2$, $\gamma = 0$ and known $\Sigma = \sigma^2$, with the other parameters fixed. The theorem concerns the exact global maximizer of the leave-one-out likelihood of Podgórski and Wallin, as used by Nitithumbundit and Chan, which the midpoint check above computes, rather than the output of a particular numerical search. Their simulated power-law fit does not determine the logarithmic factor. The theorem also supplies the limiting distribution they examined numerically. Podgórski and Wallin [1] prove super-efficiency for power-law poles under their assumption $-1 < \alpha < 0$. The logarithmic pole has a different normalization. The fixed-parameter logarithmic contact experiment in [12] gives the finite-dimensional likelihood limit here. We prove uniform process control and localization of the unmodified leave-one-out maximizer.

2.2 Density estimates and a coarse bound

Translation and dilation reduce the proof to $\theta_0 = 0$, $\sigma = 1$. Let $f = f_1$, $A = 1/\pi$, and $L = \log n$, so $d_n = L/n$. There is a fixed $0 < x_0 < 1/10$ such that

$$f(x) = A \log(1/|x|) + c_f + o(1), \qquad |f'(x)| \leq C/|x| \quad (0 < |x| < x_0).$$

These and the exponential tails follow from the Bessel representations and expansions in [13]. The logarithmic derivative is bounded on $|x| \geq x_0$. Also $|\log f(x)| \leq C(1 + |x|)$ there, using the large-argument expansion. In particular $f \in L^2(\mathbb{R})$.

For $0 < r < x_0$, put $f_r(x) = f(|x| \vee r)$, with $g_r = f_r / \int f_r$. This description uses symmetry of f. By integrating the derivative from $|x|$ to r,

$$\|f - g_r\|_1 \leq Cr, \qquad \|f^{\otimes n} - g_r^{\otimes n}\|_{\mathrm{TV}} \leq Cnr.$$

Writing $\phi_r = \log f_r$, use the centered capped process

$$C_{n,r}(t) = \sum_{i=1}^{n} [\phi_r(X_i - t) - \phi_r(X_i)].$$

At any t with at most one observation within r, the difference between the centered leave-one-out and capped criteria is exactly the change in the capped omitted term, provided the same holds at zero. This identity includes $t = X_i$, where the leave-one-out criterion omits that observation.

We first show that for every fixed $b_0 > 0$ one can choose B such that all maximizers lie within

$$W = BL^2/n \tag{1}$$

with probability $1 - O(n^{-b_0})$. Pick $D > b_0 + 5$ and use a coarse cap $r_g = n^{-D}$. The probability of a pair within $2r_g$ is $O(n^2 r_g)$ by Cauchy–Schwarz and $f \in L^2$. All data are in $[-C_0 L, C_0 L]$ except with probability $O(n^{-b_0})$ for C_0 large. On these two events the deleted capped terms change by at most $C_1 L$ between zero and any t in that deterministic range. Indeed the nearest distance is at most $2C_0 L$, and the largest capped log density is $O(\log L)$. A maximizer there must therefore satisfy $C_{n,r_g}(t) \geq -C_1 L$.

Under $g_{r_g}^{\otimes n}$, the exponential half moment of $C_{n,r_g}(t)$ is the n-th power of the affinity of g_{r_g} and its translate. With $H^2(g,h) = \int(\sqrt{g} - \sqrt{h})^2$, it is bounded by $\exp(-nH^2(g_{r_g}, g_{r_g}(\cdot - t))/2)$. For small nonzero t, the squared distance is at least $c|t|/\log(1/|t|)$. By symmetry take $t > 0$ and restrict the integral to $x = tz$, $2 \leq z \leq 3$. Here $f(tz) - f(t(z-1))$ tends uniformly to $A\log((z-1)/z)$ while the squared sum of the square roots is of order $\log(1/t)$. This proves the bound. For $|t|$ bounded away from zero, this distance is bounded away from zero by identifiability, continuity of L^2 translations, and decay of the L^2 correlation of $\sqrt{f}$ at infinity. The triangle inequality for H transfers these lower bounds to g_{r_g}, since $H(f, g_{r_g}) \leq C\sqrt{r_g}$. In particular for $|t| \geq W/2$, $nH^2 \geq cBL$ for sufficiently large n. Markov's inequality on a deterministic mesh of width $n^{-(D+3)}$ in the data range rules out $C_{n,r_g} \geq -C_1 L - 1$ at mesh points there except with probability $O(n^{-b_0})$, upon enlarging B. The capped log derivative is bounded globally by C/r_g, so the mesh error for the sum is $o(1)$. Product total variation transfers the union bound from g_{r_g} to f; this costs $O(nr_g)$. Together with the pair and sample range bounds, this proves (1).

2.3 Fine approximation and increment bounds

Fix $b_1 > 0$, choose $K > b_1 + 5$, and use the fine cap $r = n^{-1}L^{-K}$ only on $[-W, W]$. The probability of a pair within $2r$ with both observations in $[-W - 2r, W + 2r]$ is at most

$$Cn^2 r \int_{-W-4r}^{W+4r} f(x)^2\,dx \leq CL^{4-K}.$$

For the first bound apply Cauchy–Schwarz to the density product at each pair displacement. With probability $1 - O(\exp(-cL^3))$ there is an observation in $[-W, W]$, which gives nearest distances at most $2W$ on the entire interval. Hence uniformly on this interval the largest variation of the capped omitted term is at most $\phi_r(0) - \phi_r(2W) = O(\log L/L)$. Consequently $Q_n(t) - Q_n(0)$ and $C_{n,r}(t)$ differ uniformly by $O(\log L/L)$ except with probability $O(L^{-b_1})$. All these observations at true density f can be transferred to g_r at cost $O(L^{-K})$ in probability, and conversely.

Now work under g_r. For $|u| \leq BL$ define

$$Y_u(x) = \phi_r(x - ud_n) - \phi_r(x), \qquad Z_n(u) = \sum_i [Y_u(X_i) - \mathbb{E}_{g_r} Y_u(X_i)].$$

For $|u|, |v| \leq BL$ the following estimates hold with $\epsilon_n = C\log L/L$:

$$\|Y_u - Y_v\|_\infty \leq \epsilon_n, \qquad n\mathbb{E}_{g_r}(Y_u - Y_v)^2 \leq C|u - v|. \tag{2}$$

For $|x| \le 4W$, use $|\frac{d}{dt} \log f(t)| \le C/(Lt)$ for $r \le t \le 5W$. Since $\log(5W/r) = O(\log L)$ this gives the first bound there. For the second, composition with the lower cap $\log(|x| \vee r) = \max(\log|x|, \log r)$ is a contraction. The squared L^2 shift difference of the uncapped $\log|x|$ is π^2 times the shift length, by the transform of the logarithm, $-\pi/|\xi|$ away from zero, and Plancherel. The reference density g_r is globally at most CL, so the contribution of this region is at most $Cn\eta/L$, where $\eta = d_n|u-v|$.

For $|x| > 4W$ the shifted arguments are comparable to x and outside the cap. The mean value theorem gives the supremum bound $C\eta/(W\log(1/W)) + C\eta$, at most C/L. For the squared integral use the same derivative bound under $g_r(x) \le C\log(1/|x|)$ near zero. Its value times n is bounded by

$$Cn\eta^2\left[1 + \int_{4W}^{x_0} \frac{dx}{x^2 \log(1/x)}\right] \le Cn\eta^2/(W\log(1/W)) \le Cn\eta/L.$$

The integral bound follows by splitting at $\sqrt{W}$. This proves (2).

We will truncate the dyadic grids at a polynomially fine power of L. To bound the interpolation error, the number of observations in $[-4W, 4W]$ is $O(L^{b_1+5})$ except with probability $O(L^{-b_1})$, by the expected count $O(L^3)$ and Markov. The individual Lipschitz bound in u there is Cd_n/r. For the rest of the observations the expected sum of their supremum Lipschitz constants is $O(L^2)$: integrate the log derivative times g_r, bounded near zero by $C/|x|$ on $|x| > 4W$, and multiply by $nd_n = L$. Markov gives a bound $O(L^{b_1+4})$ with the same failure probability. The expectation term subtracted in Z_n has deterministic Lipschitz bound of the same polynomial order, using the expected near count. Thus there is a fixed D_1 for which the Lipschitz constant of Z_n on the entire interval is at most L^{D_1} except with probability $O(L^{-b_1})$. Take $J_n = \lceil (D_1+4)\log_2 L \rceil$. Every dyadic grid of level J_n on an interval of length at most $2BL$ then has interpolation error $o(1)$ on this one common event.

2.4 Process convergence and maximization

We give the stochastic bounds with the grid error understood on the common event just described. On an interval of length R, anchored at an endpoint, use dyadic parent and child edges of length at most $R2^{-j+1}$. Bernstein's inequality and (2) show that, except with probability Ce^{-s}, the magnitude of all the level j increments is at most

$$C\big(\sqrt{R2^{-j}(s+2j+1)} + \epsilon_n(s+2j+1)\big).$$

Here the probability bound after summing over all levels uses at most 2^{j+1} edges at a level and includes the top endpoint increment. Summing the thresholds through level J_n bounds the supremum from the anchor by

$$C\sqrt{R}(1+\sqrt{s}) + C\epsilon_n[J_n(1+s) + J_n^2] + o(1). \tag{3}$$

The result holds also with a specified interior anchor, by applying it on the two intervals to its left and right.

On a fixed compact interval, cover by intervals of length at most 2δ, including enough overlap so any pair at distance at most δ is covered. Apply (3) with $s = c + \log(C/\delta)$. After a union bound the failure probability is $O(e^{-c})$ plus the common interpolation failure probability. For fixed δ, c the term with ϵ_n tends to zero. Sending δ to zero gives stochastic equicontinuity of Z_n. The deterministic drift is equicontinuous asymptotically too. Indeed e^{Y_u} is a true one-observation likelihood ratio under g_r. On fixed compacts $\|Y_u\|_\infty = o(1)$, so uniformly there

$$n\mathbb{E}_{g_r} Y_u = -\tfrac{1}{2} n\mathbb{E}_{g_r} Y_u^2 + o(1)$$

by Taylor expansion with (2). The quadratic term is equicontinuous by Cauchy–Schwarz and (2).

For the finite-dimensional limit, put $x = d_n y$. At fixed $y \notin \{0, u\}$, the small-argument expansion gives

$$LY_u(d_n y) \;\longrightarrow\; D_u(y) := \log|y| - \log|y-u|, \qquad \frac{g_r(d_n y)}{L} \;\longrightarrow\; A.$$

On $|x| \le 4W$, the logarithmic contraction used above bounds $|LY_u(d_n y)|$ by $C|D_u(y)|$, and g_r/L is bounded. Thus dominated convergence applies on the whole mark line since $D_u \in L^2$ and $4W/d_n \to \infty$. For $|x| > 4W$ the earlier squared increment bound yields $O(nd_n^2/(W\log(1/W))) = o(1)$ for fixed u. Using $nd_n = L$ and polarization,

$$n\mathbb{E}_{g_r}(Y_u Y_v) \;\longrightarrow\; A\int_{\mathbb{R}} D_u(y)D_v(y)\,dy = \frac{\pi}{2}(|u| + |v| - |u-v|).$$

The product of means times n tends to zero, by the Taylor drift bound. The uniform single-observation envelope tends to zero, so the triangular central limit theorem gives the centered Gaussian vector, and the Taylor drift has limit $-\pi|u|/2$. Combined with tightness and the fine comparison, this proves compact process convergence of the leave-one-out criterion under the true density.

It remains to confine the maximizer to a bounded range in $u = t/d_n$. For $1 \le |u| \le BL$, the same Hellinger comparison as in the coarse bound gives $nH^2(g_r, g_r(\cdot - ud_n)) \ge c|u|$ uniformly. Indeed $nH^2(f, f(\cdot - ud_n)) \ge c'|u|$ by the pole lower bound, and $nH^2(f, g_r) = o(1)$. The log likelihood drift has upper bound $-nH^2$, by $\log x \le 2(\sqrt{x} - 1)$. Thus the drift is at most $-c|u|$ across the range. Apply (3) on the interval from zero to $\min(2R, BL)$, and on its negative, with $s = \gamma R$, where $\gamma > 0$ is sufficiently small. For all sufficiently large R and n the right-hand side is at most $cR/2$: the square-root term has coefficient $O(\sqrt{\gamma})$ at order R, and $\epsilon_n(J_n + J_n^2) = o(1)$ and $\epsilon_n J_n \gamma R = o(R)$. Except with probability $O(e^{-\gamma R})$, the centered capped process therefore cannot overcome the drift on the shell $R \le |u| \le \min(2R, BL)$. The fine comparison changes the criterion by $o(1)$. Sum these bounds over dyadic shells. The single common interpolation event and the fine comparison are paid once. Consequently, for R above a fixed constant and $R \le BL$, the likelihood maximizer has tail bounded by $Ce^{-c_2 R} + O(L^{-b_1}) + O(n^{-b_0})$. The coarse bound already controls the event $|u| > BL$. The probability transfer between the true and capped reference is $O(L^{-K})$, absorbed in these estimates.

The limiting Brownian path with negative linear radial drift tends to minus infinity at both ends and has an almost surely unique global maximizer. On either half-line uniqueness on compact intervals follows from the Brownian maximum property and absolute continuity under addition of linear drift. The two independent half-line suprema have continuous exponential distributions by the Brownian hitting probability with drift. The compact uniform process limit, tightness of all maximizers and the argmax mapping now give the stated limit in law.

Finally the coarse and fine analysis used arbitrary b_0, b_1, while the estimator stayed the same. On the coarse event the scaled maximizer has magnitude at most BL. Integration of the preceding shell tail bound makes any fixed power uniformly integrable there upon choosing b_1 larger than that power. Off the coarse event use the sample range bound $|\widehat{\theta}_n| \le \max_i |X_i|$; exponential tails give $\mathbb{E}(\max_i |X_i|)^{2a} = O(L^{2a})$. Cauchy–Schwarz and the arbitrary n^{-b_0} bound make its contribution to the a-th moment of the scaled maximizer vanish. This proves convergence of all fixed absolute moments. Rescaling the data restores σ, with Brownian information $I = \pi/\sigma$ by Brownian scaling.

2.5 The limiting absolute and squared risks

The canonical Brownian argmax law and the moment 26 go back to Terent'yev [6], as recorded in [7]. We give a short derivation for both risk constants.

By Brownian scaling, $|IU_\sigma|$ is distributed as the absolute maximizing time of $B(t) - |t|/2$ for standard two-sided Brownian motion B. For a one-sided Brownian motion with drift $-a$, $a > 0$, write M, T for the eventual maximum and its time. The first-passage transform to level $m > 0$ is

$$\mathbb{E}[e^{-s\tau_m}; \tau_m < \infty] = e^{-m(a+\sqrt{a^2+2s})}, \qquad s \geq 0.$$

This follows from the exponential Brownian martingale stopped at τ_m and a lower level tending to minus infinity. The case $s = 0$ gives the exponential tail of M. For a grid of levels $m = j\delta$, the strong Markov property at τ_m gives

$$\mathbb{E}[e^{-s\tau_m};\, m \leq M < m + \delta] = e^{-m(a+\sqrt{a^2+2s})}(1 - e^{-2a\delta}).$$

The first-passage times at lower grid approximations to M converge almost surely to T, by continuity, uniqueness of the maximum and the fact that the path tends to minus infinity. Riemann sums against bounded continuous functions therefore identify the joint transform

$$\mathbb{E}[e^{-sT};\, M \in dm] = 2a\, e^{-m(a+\sqrt{a^2+2s})}\, dm.$$

For the two-sided maximum take $a = 1/2$, multiply by the probability $1 - e^{-m}$ that the independent opposite maximum is smaller, and double. The Laplace transform of $|IU_\sigma|$ is thus

$$\frac{2}{(1/2 + \sqrt{1/4 + 2s})(3/2 + \sqrt{1/4 + 2s})}.$$

Its first two moments are 3 and 26, as follows by differentiation at zero. The moment convergence in Theorem 2.1 proves the corresponding absolute and squared risk constants for the leave-one-out estimator.

3 The leave-one-out location estimator at power poles and cusps

For a symmetric univariate variance-gamma density of shape $0 < \nu < 1/2$, the density has the power pole $|x|^{2\nu-1}$ and the maximum leave-one-out likelihood estimator converges at rate $n^{-1/(2\nu)}$. Its limit is the argmax of an explicit functional of a Poisson process: the midpoint of the consecutive pair that best balances the closeness of the pair against the logarithmic potential of the other points. All moments converge. The same theorem holds for every even density decreasing in $|x|$ with a power pole, exponential moments and mild score bounds, so the optimal rate is attained. For shapes $1/2 < \nu < 3/4$ the density has a cusp and the rate is $n^{-1/(4\nu-1)}$, with a fractional Brownian argmax limit; at $\nu = 3/4$ the rate is $(n \log n)^{-1/2}$, and above it the root-n Gaussian limit holds.

3.1 The questions and the results

Let $X_i = \theta_0 + Z_i$, $i = 1, \ldots, n$, be independent, where Z_i has an even density f. As in Section 2.1, write $j(t)$ for a sample index closest to t and

$$Q_n(t) = \sum_{i \neq j(t)} \log f(X_i - t),$$

the leave-one-out likelihood of Podgórski and Wallin [1, Section 3.2], and let $\widehat{\theta}_n$ be the smallest global maximizer of Q_n; Lemma 3.6 shows that it exists and is measurable. In the words of Podgórski and Wallin, "an open question is if this rate is actually reached by the estimator" [1, Section 6], the rate being $n^{-1/(1+\alpha)}$ at a pole $|x|^\alpha$ (Question 1.1). Section 2 answers Question 1.2 of Nitithumbundit and Chan at the logarithmic pole $\nu = 1/2$. This section answers it at every other shape, and answers Question 1.1 for the class of densities described next.

Definition 3.1. An even probability density f on $\mathbb{R}$ satisfies *condition* (P) *with exponent* $\alpha \in (-1,0)$ *and constant* $A>0$ if the following hold, with a constant C_0:

(P1) f is positive, twice continuously differentiable and strictly decreasing on $(0,\infty)$;

(P2) $f(x) = A|x|^{\alpha}(1+\rho(x))$ with $|\rho(x)| \le C_0|x|^{\epsilon}$ for $0<|x|\le 1$ and some $\epsilon \in (0,1]$;

(P3) the score $\psi = (\log f)'$ satisfies $|\psi(x)| \le C_0(1+|x|^{-1})$ and $|\psi'(x)| \le C_0(1+x^{-2})$ for $x \neq 0$;

(P4) $\int e^{c_0|x|} f(x)\,dx < \infty$ for some $c_0 > 0$.

The limit is described by a Poisson process. Let $-1<\alpha<0$ and let $\mathcal{N}$ be a Poisson process on $\mathbb{R}$ with intensity $|t|^{\alpha}\,dt$. The intensity is locally integrable and has infinite mass on each half-line, and $\mathbb{E}\sum_{t\in\mathcal{N},|t|>1}|t|^{-1} = 2/|\alpha|$. Hence almost surely $\mathcal{N}$ is locally finite, has infinitely many points on each side of the origin, has no point at the origin, and satisfies $\sum_{t\in\mathcal{N}}\min(1,|t|^{-1})<\infty$. For such a configuration write $\tau(u)$ for a point of $\mathcal{N}$ nearest to u, and define

$$\Psi_{\alpha}(u) = \alpha\Big[\sum_{t\in\mathcal{N}\setminus\{\tau(u),\tau(0)\}} \log\Big|1-\frac{u}{t}\Big| + \mathbf{1}_{\{\tau(u)\neq\tau(0)\}}\big(\log|\tau(0)-u| - \log|\tau(u)|\big)\Big]. \tag{4}$$

The series converges absolutely, since $|\log|1-u/t|| \le 2|u|/|t|$ for $|t|\ge 2|u|$. Formally, $\Psi_{\alpha}(u)$ is $\alpha[\sum_{t\neq\tau(u)}\log|t-u| - \sum_{t\neq\tau(0)}\log|t|]$, the limit of the centered leave-one-out log-likelihood; at a point u with two nearest points both choices give the same value. Lemma 3.7 shows that Ψ_{α} has almost surely a unique maximizer T_{α}, and that T_{α} is the midpoint $m_k = (s_k+s_{k+1})/2$ of two consecutive points $s_k < s_{k+1}$ of $\mathcal{N}$. Equivalently, with $g_k = s_{k+1}-s_k$, the index of T_{α} maximizes

$$\log\frac{g_k}{2} - \sum_{t\in\mathcal{N}}\log\Big|1-\frac{m_k}{t}\Big|, \tag{5}$$

because $\Psi_{\alpha}(m_k)$ equals $|\alpha|$ times (5) plus a quantity that does not depend on k. Thus the limit estimator chooses the consecutive pair whose closeness, after the two endpoint terms cancel against $\log(g_k/2)$, best balances the logarithmic potential of the other points at its midpoint. The law of T_{α} is symmetric and depends only on α.

Theorem 3.2. *Let f satisfy condition* (P)*, with exponent α and constant A. Put $r_n = (nA)^{-1/(1+\alpha)}$. For every θ_0 the following hold.*

1. *for every $M>0$ the process $u \mapsto Q_n(\theta_0+ur_n) - Q_n(\theta_0)$ converges in law in $C[-M,M]$ to Ψ_{α};*
2. *$(\widehat{\theta}_n-\theta_0)/r_n \Longrightarrow T_{\alpha}$, and $\mathbb{E}|(\widehat{\theta}_n-\theta_0)/r_n|^p \to \mathbb{E}|T_{\alpha}|^p < \infty$ for every $p>0$;*
3. *for every $b>0$ there are C_b and n_b such that $\Pr(|\widehat{\theta}_n-\theta_0| > Mr_n) \le C_bM^{-b}$ for all $M\ge 1$ and $n\ge n_b$;*
4. *the rate is optimal: there is $c>0$ such that every sequence of estimators $\widetilde{\theta}_n$ satisfies*

$$\liminf_{n\to\infty}\ \max_{\theta\in\{\theta_0,\,\theta_0+cr_n\}} \Pr_{\theta}\big(|\widetilde{\theta}_n-\theta| \ge cr_n/2\big) \ge \frac{1}{4}.$$

Corollary 3.3. *Let $0<\nu<1/2$ and $\sigma>0$, and let Z_i have the variance-gamma density $f_{\nu,\sigma}(x) = \sigma^{-1}f_{\nu}(x/\sigma)$,*

$$f_{\nu}(x) = \frac{(|x|/2)^{\nu-1/2}K_{\nu-1/2}(|x|)}{\sqrt{\pi}\,\Gamma(\nu)},$$

of a centered normal law with variance $2\sigma^2 G$, $G \sim \mathrm{Gamma}(\nu, 1)$. Then the conclusions of Theorem 3.2 hold with $\alpha = 2\nu - 1$ and

$$r_n = \sigma\,(nA_{\nu,1})^{-1/(2\nu)}, \qquad A_{\nu,1} = \frac{\Gamma(1/2-\nu)}{4^\nu\sqrt{\pi}\,\Gamma(\nu)},$$

the constant $A_{\nu,d}$ of Theorem 4.1 evaluated at $d = 1$. In particular

$$n^{1/(2\nu)}(\widehat{\theta}_n - \theta_0) \Longrightarrow \sigma A_{\nu,1}^{-1/(2\nu)}\, T_{2\nu-1},$$

with convergence of all moments, and $\widehat{\theta}_n$ is a midpoint of two adjacent order statistics. In the parameterization of Nitithumbundit and Chan with scale $\Sigma = s^2$, $\sigma = s/\sqrt{2\nu}$.

Theorem 3.2 answers the question of Podgórski and Wallin for every even density satisfying condition (P): the estimator attains the optimal rate $n^{-1/(1+\alpha)}$, with an explicit limit law. Their class also contains asymmetric and one-sided poles, which condition (P) does not include. The rate $n^{-1/(1+\alpha)}$ is the minimax rate (Polfeldt [4, 5]), and Ibragimov and Khas'minskii [14] proved limit laws at this rate for Bayesian estimators; part 4 is the standard two-point bound, included for completeness. Corollary 3.3 answers the rate and limit-distribution questions of Nitithumbundit and Chan at every shape $0 < \nu < 1/2$. Their proposed exponent $1/(2\nu)$ is exact on this range. The variance-gamma factor $p(x) = |x|^{1-2\nu} f_\nu(x)$ has the derivative of order $|x|^{-2\nu}$ at the origin (Lemma 3.5), so condition (P) also covers densities whose factor p has an unbounded derivative. Another example is the symmetric gamma density $|x|^\alpha e^{-|x|}/(2\Gamma(1+\alpha))$, which satisfies condition (P) with $\epsilon = 1$, $\psi(x) = \alpha/x - \operatorname{sgn} x$ and $\psi'(x) = -\alpha/x^2$.

For $\nu > 1/2$ the density is bounded and the leave-one-out likelihood differs from the full likelihood by one bounded term. Write $\beta = 2\nu - 1$ and, for $1/2 < \nu < 1$,

$$f_\nu(0) = \frac{\Gamma(\nu - 1/2)}{2\sqrt{\pi}\,\Gamma(\nu)}, \qquad b_\nu = \frac{2\,\Gamma(3/2-\nu)}{(2\nu-1)\,4^\nu\sqrt{\pi}\,\Gamma(\nu)}, \tag{6}$$

so that $f_\nu(x) = f_\nu(0) - b_\nu|x|^\beta + O(x^2)$ (Lemma 3.5). Let W_H be a standard two-sided fractional Brownian motion, with $\mathbb{E}W_H(u)^2 = |u|^{2H}$.

Theorem 3.4. *Let $\nu > 1/2$, $\sigma > 0$, and let Z_i have the variance-gamma density $f_{\nu,\sigma}$ of Corollary 3.3.*

1. *If $1/2 < \nu < 3/4$, put $H = 2\nu - \frac{1}{2}$,*

$$\gamma_\nu^2 = \frac{b_\nu^2}{f_\nu(0)}\int_{\mathbb{R}} \big(|y-1|^\beta - |y|^\beta\big)^2 dy = \frac{8\,b_\nu^2}{\pi f_\nu(0)}\,\Gamma(1+\beta)^2\sin^2\frac{\pi\beta}{2}\,\sin(\pi\beta)\,\Gamma(-1-2\beta),$$

   and $s_n = \sigma(n\gamma_\nu^2)^{-1/(4\nu-1)}$. Then

$$\frac{\widehat{\theta}_n - \theta_0}{s_n} \;\Longrightarrow\; \operatorname*{argmax}_{u\in\mathbb{R}}\Big\{W_H(u) - \frac{1}{2}|u|^{2H}\Big\},$$

   and this maximizer is almost surely unique.

2. *If $\nu = 3/4$, then*

$$\sqrt{n\log n}\,(\widehat{\theta}_n - \theta_0) \Longrightarrow N(0, \sigma^2/I_*), \qquad I_* = \frac{b_{3/4}^2}{4f_{3/4}(0)} = \frac{\Gamma(3/4)}{\sqrt{\pi}\,\Gamma(1/4)}.$$

3. *If $\nu > 3/4$, then $\sqrt{n}\,(\widehat{\theta}_n - \theta_0) \Longrightarrow N(0, \sigma^2/I_\nu)$, where $I_\nu = \int \psi_\nu^2 f_\nu$ is the location Fisher information of f_ν and $\psi_\nu = (\log f_\nu)'$.*

Together with Theorem 2.1, Corollary 3.3 and Theorem 3.4 give the exact rate and the limit law of the leave-one-out estimator at every shape $\nu > 0$ of the symmetric univariate variance-gamma law with known scale. The rate exponent is

$$\frac{1}{2\nu}\ (0 < \nu < \tfrac{1}{2}), \qquad \frac{1}{4\nu - 1}\ (\tfrac{1}{2} < \nu < \tfrac{3}{4}), \qquad \frac{1}{2}\ (\nu > \tfrac{3}{4}),$$

with the logarithmic corrections $(\log n)/n$ at $\nu = 1/2$ and $(n \log n)^{-1/2}$ at $\nu = 3/4$. On the cusp range the exponent is $1/(4\nu - 1)$, smaller than the proposed exponent $1/(2\nu)$; for example it is $5/7$ at $\nu = 0.6$, where $1/(2\nu) = 5/6$. The cusp limits for the full maximum likelihood estimator go back to Prakasa Rao [15] and to the classification of singularities of Ibragimov and Has'minskii [11, Chapter VI]; the part of Theorem 3.4 specific to the leave-one-out likelihood is that removing the nearest observation leaves the limit unchanged, and we give a complete proof for the variance-gamma cusp. The Poisson experiment that underlies Theorem 3.2 is the one-dimensional analogue of the Poisson phase of Theorem 4.1; for unbounded densities Ibragimov and Khas'minskii [14] proved limit laws at rate $n^{-1/(1+\alpha)}$ for Bayesian estimators of the shift. The leave-one-out criterion is not a likelihood ratio, and the new steps are the capped bound of Lemma 3.8, the shell estimate of Proposition 3.10, which uses only exponential moments of capped likelihood ratios, and the description of the limit through nearest points.

3.2 Variance-gamma densities

Lemma 3.5. *Let $\nu > 0$ and $f = f_\nu$, $\psi = (\log f)'$.*

1. *f is even, positive and analytic on $(0, \infty)$, strictly decreasing there, and $\int e^{s|x|} f(x)\,dx < \infty$ for $|s| < 1$. For $x > 0$, $\psi(x) = -K_{|\nu - 3/2|}(x)/K_{|\nu - 1/2|}(x)$.*

2. *If $0 < \nu < 1/2$, then f satisfies condition* (P) *with $\alpha = 2\nu - 1$, $A = A_{\nu,1}$ and $\epsilon = 1 - 2\nu$; more precisely $f(x) = A_{\nu,1}|x|^{2\nu - 1} - B_\nu + O(|x|^{2\nu+1})$ with $B_\nu = \Gamma(1/2 + \nu)/((1 - 2\nu)\sqrt{\pi}\,\Gamma(\nu))$, and $\log f$ is strictly convex on $(0, \infty)$. The factor $p(x) = |x|^{1-2\nu} f(x)$ satisfies $|p'(x)| \sim (1 - 2\nu) B_\nu |x|^{-2\nu}$ as $x \to 0$.*

3. *If $1/2 < \nu < 1$, then $f(x) = f(0) - b_\nu |x|^\beta + O(x^2)$ with $\beta = 2\nu - 1$ and the constants* (6)*, and $|\psi(x)| \le C(1 + |x|^{\beta - 1})$. If $\nu \ge 1$, then $|\psi| \le 1$.*

Proof. For $x > 0$ write $\mu = \nu - 1/2$, so that $f(x) = (x/2)^\mu K_\mu(x)/(\sqrt{\pi}\Gamma(\nu))$. The normal mixture representation shows that f is even and strictly decreasing on $(0, \infty)$, and $\mathbb{E}e^{sZ} = \mathbb{E}e^{s^2 G} = (1 - s^2)^{-\nu}$ for $|s| < 1$. By [13, 10.29.4], $(x^\mu K_\mu)' = -x^\mu K_{\mu - 1}$, so $\psi = -K_{\mu-1}/K_\mu$, and $K_{-s} = K_s$ gives part 1.

Let $0 < \nu < 1/2$ and $m = 1/2 - \nu \in (0, 1/2)$. By [13, 10.27.4, 10.25.2],

$$(x/2)^{-m} K_m(x) = \frac{\pi}{2\sin(m\pi)} \Big[(x/2)^{-2m} \sum_{k \ge 0} \frac{(x^2/4)^k}{k!\,\Gamma(k + 1 - m)} - \sum_{k \ge 0} \frac{(x^2/4)^k}{k!\,\Gamma(k + 1 + m)} \Big],$$

with entire series. The reflection formula gives the leading coefficients $\Gamma(m)/2$ and $\Gamma(1 - m)/(2m)$, hence the stated expansion of f, and $\rho(x) = -(B_\nu/A_{\nu,1})|x|^{1-2\nu} + O(x^2)$, which is (P2) with $\epsilon = 1 - 2\nu$. The same expansion differentiated termwise gives the stated asymptotic of p'. For (P3) put $R = -\psi = K_{m+1}/K_m$. The two forms of [13, 10.29.2] give the Riccati identity

$$\psi' = 1 + \frac{2m + 1}{x} R - R^2.$$

By [13, 10.30.2], $R(x) \sim 2m/x$ as $x \to 0$, and by [13, 10.40.2], $R(x) \to 1$ as $x \to \infty$; since R is continuous on $(0,\infty)$, $R \le C(1+x^{-1})$, and the identity gives $|\psi'| \le C(1+x^{-2})$. For convexity, K_m is a positive mixture of the functions e^{-xc}, $c \ge 1$, by [13, 10.32.9], hence log-convex by the Cauchy–Schwarz inequality, and x^{-m} is strictly log-convex; their product is f up to a constant.

Let $1/2 < \nu < 1$, so $\mu \in (0, 1/2)$. The same series with m replaced by μ give

$$(x/2)^{\mu} K_{\mu}(x) = \frac{\Gamma(\mu)}{2} - \frac{\Gamma(1-\mu)}{2\mu}(x/2)^{2\mu} + O(x^2),$$

which is the stated cusp expansion. Here $\psi = -K_{1-\mu}/K_{\mu} \sim -c\, x^{2\mu-1}$ as $x \to 0$ by [13, 10.30.2], and $\psi \to -1$ at infinity, which gives the bound on ψ. If $\nu \ge 1$, then $|\mu - 1| \le \mu$, and $K_s(x)$ is nondecreasing in $s \ge 0$ by [13, 10.32.9], so $|\psi| \le 1$. □

Monotonicity in $|x|$ and evenness give two facts used below. For a density satisfying (P1), $2|x|f(x) \le \int_{-|x|}^{|x|} f \le 1$, so with (P2) there are constants $0 < c_1 \le C_1$ with

$$c_1|x|^{\alpha} \le f(x) \le C_1|x|^{\alpha} \quad (0 < |x| \le 1), \qquad f(x) \le C_1|x|^{\alpha} \quad (x \ne 0). \tag{7}$$

Consequently, for every interval I of length $\ell \le 1$,

$$\Pr(Z \in I) \le \frac{2C_1}{1+\alpha}\Big(\frac{\ell}{2}\Big)^{1+\alpha}, \qquad \Pr(Z \in I) \le C_1 \ell\, \operatorname{dist}(0, I)^{\alpha}, \tag{8}$$

the first because $|x|^{\alpha}$ is symmetric and decreasing in $|x|$.

Lemma 3.6. *Let f be even, positive and strictly decreasing on $(0,\infty)$, and let the data be distinct. Then Q_n is continuous on $\mathbb{R}$, strictly decreasing on $[X_{(n)}, \infty)$ and strictly increasing on $(-\infty, X_{(1)}]$, and its maximizers form a nonempty compact subset of $[X_{(1)}, X_{(n)}]$; the smallest one is a measurable function of the data and is translation equivariant. If $\log f$ is strictly convex on $(0,\infty)$, every maximizer is a midpoint of two adjacent order statistics.*

Proof. On the open Voronoi cell of an observation, that observation is omitted and the remaining terms are continuous. At the midpoint m of adjacent order statistics $X_{(k)} < X_{(k+1)}$, the two one-sided expressions differ by $\log f(X_{(k+1)} - m) - \log f(X_{(k)} - m) = 0$, by evenness. For $t > X_{(n)}$ the omitted point is $X_{(n)}$ and every remaining distance $t - X_i$ increases with t, so Q_n decreases strictly; the left end is symmetric. Hence the maximum is attained on the compact interval $[X_{(1)}, X_{(n)}]$, the maximizers form a closed set, and its smallest element is measurable. Equivariance holds because Q_n for the shifted data is the shifted Q_n. Under strict convexity, on the closed Voronoi cell of $X_{(k)}$ the function Q_n is the sum of the functions $t \mapsto \log f(X_i - t)$, $i \ne k$, each strictly convex on that cell, which contains no X_i with $i \ne k$ in its interior. A sum of convex functions with a strictly convex term attains its maximum over a closed interval only at the endpoints, which are midpoints or, for the two unbounded cells, excluded by monotonicity. □

3.3 The Poisson limit criterion

Lemma 3.7. *Almost surely, Ψ_{α} is continuous on $\mathbb{R}$, vanishes at zero, and is strictly convex on the closed Voronoi cell of each point of $\mathcal{N}$. The maximizers of Ψ_{α} on $\mathbb{R}$ are midpoints of consecutive points, and if $\sup_{|u| \ge M} \Psi_{\alpha}(u) < 0$ for some M, then Ψ_{α} has exactly one maximizer T_{α}. The law of T_{α} is symmetric. If $\mathcal{N}$ is replaced by a Poisson process of intensity $a|t|^{\alpha}\,dt$, $a > 0$, the maximizer has the law of $a^{-1/(1+\alpha)} T_{\alpha}$.*

Proof. Fix a configuration with the almost sure properties listed before (4). On the cell of a point s, with $\tau(u) = s$, the bracket in (4) is $\sum_{t\neq s}(\log|t-u| - \log|t|)$ up to a constant, a series of functions each concave in u on the cell, converging uniformly on compact sets. Multiplied by $\alpha < 0$ it is convex, and strictly so, since a neighbouring point contributes the strictly convex term $\alpha \log|t-u|$. Continuity inside cells is clear, and at a midpoint the two nearest points are equidistant, so the two expressions agree. Hence a maximizer lies at an endpoint of a cell. At a midpoint m_k the bracket equals (5) with the sign changed, plus $\sum_t \log|t| - \sum_{t\neq\tau(0)} \log|t| = \log|\tau(0)|$ after regrouping the absolutely convergent series; this proves the identity used for (5).

If $\sup_{|u|\geq M} \Psi_\alpha < 0 = \Psi_\alpha(0)$, the maximum over $[-M, M]$ is a global maximum. Uniqueness follows from the Mecke formula [16, Theorem 4.1]: for measurable $h \geq 0$, $\mathbb{E}\sum_{s\in\mathcal{N}} h(s, \mathcal{N}\setminus\{s\}) = \int \mathbb{E}h(s,\mathcal{N})|s|^\alpha\, ds$. Let $h(s,\mu) = 1$ if $\mu\cup\{s\}$ has two distinct midpoints of consecutive pairs, none of whose four endpoints is s, with equal values of Ψ_α. If two maximizers exist, some point s of $\mathcal{N}$ is not an endpoint of the two pairs, so the event is bounded by $\sum_s h(s, \mathcal{N}\setminus\{s\})$. Fix a typical configuration μ and a pair of midpoints $m \neq m'$ of μ. Adding a point s that is not adjacent to them keeps the nearest points of m and m', and changes $\Psi(m) - \Psi(m')$ by $\alpha\log|(s-m)/(s-m')|$. This is a nonconstant real-analytic function of s on each component of $\mathbb{R}\setminus\{m, m'\}$, so it takes any given value at no more than two points. Hence $\{s : h(s,\mu) = 1\}$ is countable, the integral vanishes, and two maximizers occur with probability zero. Symmetry follows from the reflection invariance of $\mathcal{N}$, since the functional of $-\mathcal{N}$ at $-u$ equals the functional of $\mathcal{N}$ at u. The image of $\mathcal{N}$ under $t \mapsto ct$ is a Poisson process of intensity $c^{-1-\alpha}|t|^\alpha\, dt$, and (4) is unchanged when u and all points are multiplied by c, since its terms are logarithms of ratios; take $c = a^{-1/(1+\alpha)}$. □

3.4 A capped bound for the leave-one-out criterion

From now on f satisfies condition (P), $\theta_0 = 0$ by equivariance, and $r_n = (nA)^{-1/(1+\alpha)}$. Note $nr_n^{1+\alpha} = 1/A$ and $nr_n \to 0$, because $1+\alpha < 1$. For $0 < r \leq 1$, as in Section 2.2, put $f_r(x) = f(|x|\vee r)$, $\phi_r = \log f_r$, and let $C_{n,r}(t) = \sum_i[\phi_r(X_i - t) - \phi_r(X_i)]$. The excess of the uncapped log-density is $\chi_r(x) = \log f(x) - \phi_r(x) = \log(f(x)/f(r))\,\mathbf{1}_{\{|x|<r\}} \geq 0$. Write $d(t) = \min_i|X_i - t|$.

Lemma 3.8. *For every $t \in \mathbb{R}$ and $0 < r \leq 1$,*

$$Q_n(t) - Q_n(0) \leq C_{n,r}(t) + \log\frac{f(r)}{f(d(t)\vee r)} + \sum_{i\neq j(t)} \chi_r(X_i - t).$$

For every $h \in \mathbb{R}$, $\mathbb{E}\exp\{[\phi_r(Z-h) - \phi_r(Z)]/2\} \leq 1 + 2rf(r) - \frac{1}{2}H^2(f_r, f_r(\cdot - h))$.

Proof. Since $\log f = \phi_r + \chi_r$, $Q_n(t) = \sum_i \phi_r(X_i - t) - \phi_r(X_{j(t)} - t) + \sum_{i\neq j(t)} \chi_r(X_i - t)$. At $t = 0$ drop the nonnegative χ_r terms and use $\phi_r \leq \log f(r)$, which holds because f decreases in $|x|$. Also $\phi_r(X_{j(t)} - t) = \log f(d(t)\vee r)$. Subtracting proves the first claim. For the second, the integrand $f(x)e^{[\phi_r(x-h)-\phi_r(x)]/2}$ equals $\sqrt{f_r(x)f_r(x-h)}$ for $|x| \geq r$ and is at most $f(x)$ for $|x| < r$, since $\phi_r(x) = \log f(r) \geq \phi_r(x-h)$ there. Hence the expectation is at most $\int\sqrt{f_r f_r(\cdot - h)} + \Pr(|Z| < r) = \int f_r - \frac{1}{2}H^2(f_r, f_r(\cdot - h)) + \Pr(|Z| < r)$, and $\int f_r = 1 - \Pr(|Z| < r) + 2rf(r)$. □

Lemma 3.9. *There are $h_0 \in (0, 1/4]$, $r_0 > 0$ and $c_2, c_4 > 0$ such that $H^2(f_r, f_r(\cdot - h)) \geq c_2|h|^{1+\alpha}$ for $2r \leq |h| \leq h_0$, and $H^2(f_r, f_r(\cdot - h)) \geq c_4$ for $|h| \geq h_0/2$ and $r \leq r_0$. Moreover $H^2(f, f(\cdot - h)) \leq C|h|^{1+\alpha}$ for $|h| \leq 1$.*

Proof. Let $0 < h$ and $x \in [2h, 3h]$, so that $x - h \geq h \geq 2r$ and both values are uncapped. By (P2), for $3h_0$ small the ratio $f(x-h)/f(x)$ is at least $((x-h)/x)^\alpha(1-\eta)/(1+\eta) \geq (3/2)^{|\alpha|}(1-\eta)/(1+\eta)$, with $\eta = C_0(3h_0)^\epsilon$. Choosing h_0 with this bound at least $1+c$ gives $\sqrt{f(x-h)} - \sqrt{f(x)} \geq c'\sqrt{f(x)} \geq$

$c''h^{\alpha/2}$, and integration over $[2h,3h]$ gives the first bound. For the second, $h\mapsto\int(\sqrt{f(\cdot-h)}-\sqrt{f})^2$ is continuous, positive for $h\neq 0$ (otherwise f would be periodic) and tends to two at infinity, so its infimum over $|h|\geq h_0/2$ is positive, and $\|\sqrt{f_r}-\sqrt{f}\|_2^2\leq \Pr(|Z|<r)\to 0$ together with the triangle inequality gives the claim for small r. For the upper bound, the part $|x|\leq 2|h|$ costs at most $2\int_{|x|\leq 3|h|} f\leq C|h|^{1+\alpha}$ by (8). For $|x|>2|h|$ use $|(\sqrt{f})'|=|\psi|\sqrt{f}/2$, the bound (P3) and (7) at points of modulus at least $|x|/2$: the part $2|h|<|x|\leq 2$ costs at most $Ch^2\int_{2|h|}^2 x^{\alpha-2}dx\leq C|h|^{1+\alpha}$, and the part $|x|>2$ costs at most Ch^2. □

3.5 Localization

Write $t_i=X_i/r_n$ and $N_n=\sum_i\delta_{t_i}$. By (7) the mean measure of N_n satisfies $\mathbb{E}N_n(ds)=nr_nf(r_ns)\,ds\leq(C_1/A)|s|^\alpha ds$, and the factorial moment measures of order k are bounded by its k-fold product. We use C for constants depending only on f and on the exponents b,p below.

Proposition 3.10. *For every $b>0$ there are U_0, n_b and C_b such that for $n\geq n_b$:*

1. *for $U_0\leq U\leq h_0/(2r_n)$, $\Pr(\sup_{U\leq|u|\leq 2U}[Q_n(ur_n)-Q_n(0)]\geq -1)\leq C_bU^{-b}$;*
2. *$\Pr(\sup_{|t|\geq h_0/2}[Q_n(t)-Q_n(0)]\geq -1)\leq C_bn^{-b}$.*

Consequently $\Pr(|\widehat{\theta}_n|>Mr_n)\leq C_bM^{-b}$ for all $M\geq 1$ and $n\geq n_b$.

Proof. Apply Lemma 3.8 with $r=r_n$. By Lemma 3.8 and (7), $2r_nf(r_n)\leq 2C_1/(nA)$, so independence gives

$$\mathbb{E}e^{C_{n,r_n}(t)/2}\leq\exp\{2C_1/A-\tfrac{n}{2}H^2(f_{r_n},f_{r_n}(\cdot-t))\}. \tag{9}$$

Part 1. Fix a shell $S_U=\{U\leq|u|\leq 2U\}$ and let $t=ur_n$, $u\in S_U$. Then $2r_n\leq|t|\leq h_0$, so Lemma 3.9 gives $nH^2\geq(c_2/A)|u|^{1+\alpha}$. We bound the three terms of Lemma 3.8 on five events.

(E1) The function $u\mapsto C_{n,r_n}(ur_n)$ is Lipschitz, with almost everywhere derivative bounded by $r_n\sum_i|\psi(X_i-ur_n)|\mathbf{1}_{\{|X_i-ur_n|>r_n\}}\leq C_0\sum_i(r_n+\min(1,|t_i-u|^{-1}))$ by (P3). For $u\in S_U$, $\min(1,|t_i-u|^{-1})$ is at most one if $|t_i|\leq 4U$ and at most $2/|t_i|$ otherwise. The mean measure bound gives $\mathbb{E}\sup_{S_U}\leq C(nr_n+U^{1+\alpha}+U^\alpha)\leq CU^{1+\alpha}$. Let E_1 be the event that the Lipschitz constant is at most $U^{1+\alpha+b}$; then $\Pr(E_1^c)\leq CU^{-b}$ by Markov's inequality.

(E2) Take a grid in S_U of mesh $U^{-1-\alpha-b}$, with at most $4U^{2+\alpha+b}$ points u_k. By Markov's inequality and (9), $\Pr(C_{n,r_n}(u_kr_n)\geq-\frac{c_2}{4A}|u_k|^{1+\alpha})\leq C\exp\{-\frac{c_2}{8A}U^{1+\alpha}\}$. Let E_2 be the event that none of these occurs. On $E_1\cap E_2$, every $u\in S_U$ has a grid point u_k with $|C_{n,r_n}(ur_n)-C_{n,r_n}(u_kr_n)|\leq U^{1+\alpha+b}\cdot U^{-1-\alpha-b}=1$, so $\sup_{S_U}C_{n,r_n}(ur_n)\leq-\frac{c_2}{4A}U^{1+\alpha}+1$.

(E3) Let $E_3=\{d(0)\leq Ur_n\}$. Since $\Pr(|Z|\leq s)\geq(2c_1/(1+\alpha))s^{1+\alpha}$ for $s\leq 1$, $\Pr(E_3^c)\leq(1-\Pr(|Z|\leq Ur_n))^n\leq\exp\{-cU^{1+\alpha}\}$. On E_3, $d(ur_n)\leq|u|r_n+d(0)\leq 3Ur_n\leq 1$ on S_U, so (7) gives $\log[f(r_n)/f(d\vee r_n)]\leq\log(C_1/c_1)+|\alpha|\log(3U)$.

(E4) Let K be an integer with $1+K\alpha\leq -b$ and let E_4 be the event that no ball of radius one centred in S_U contains K of the points t_i. If K points lie in such a ball, K points lie within distance two of the leftmost of them, which lies in $U-1\leq|s|\leq 2U+1$. By the factorial moment bound and the mean measure bound, which gives $\mathbb{E}N_n(I)\leq C\,\ell\,U^\alpha$ for intervals I of length $\ell\leq 4$ inside $|s|\geq U/2$, $\Pr(E_4^c)\leq C\,U\,U^\alpha(U^\alpha)^{K-1}=CU^{1+K\alpha}\leq CU^{-b}$.

(E5) Let $\eta_U=\exp(-U^{(1+\alpha)/2})$ and let E_5 be the event that no two points $t_i\neq t_j$ with $U-1\leq|t_i|\leq 2U+1$ satisfy $|t_i-t_j|<\eta_U$. In the same way $\Pr(E_5^c)\leq C\eta_UU^{1+2\alpha}$.

On $E_4 \cap E_5$, every $i \neq j(ur_n)$ with $\chi_{r_n}(X_i - ur_n) \neq 0$ has $|t_i - u| < 1$, there are at most $K-1$ of them, and $|t_i - u| \geq |t_i - t_{j(ur_n)}|/2 \geq \eta_U/2$; by (7) each term is at most $\log(C_1/c_1) + |\alpha| \log(2/\eta_U)$. On $E_1 \cap \cdots \cap E_5$, for $u \in S_U$,

$$Q_n(ur_n) - Q_n(0) \leq -\frac{c_2}{4A} U^{1+\alpha} + 1 + C \log(3U) + K(C + |\alpha| U^{(1+\alpha)/2}),$$

which is less than -1 for $U \geq U_0$. The failure probabilities add up to at most $C_b U^{-b}$.

Part 2. By Lemma 3.6, maximizers lie in $[X_{(1)}, X_{(n)}]$, and $\Pr(\max_i |X_i| > K_0 \log n) \leq C n^{1-c_0 K_0}$ by (P4); take K_0 with $c_0 K_0 > b+1$. For $h_0/2 \leq |t| \leq K_0 \log n$, Lemma 3.9 and (9) give $\mathbb{E} e^{C_{n,r_n}(t)/2} \leq C e^{-c_4 n/2}$. The derivative of $t \mapsto C_{n,r_n}(t)$ is at most $2C_0 n/r_n$, so a grid of mesh r_n/n^2, with $O(n^2 r_n^{-1} \log n)$ points, controls $C_{n,r_n} \leq -c_4 n/4$ on the range except with superpolynomially small probability. On the range, $d(t) \leq 2K_0 \log n$ when $d(0) \leq 1$, and (P3) and (7) give $\log[f(r_n)/f(d(t) \vee r_n)] \leq C \log n$. The excess terms concern points within r_n of t, hence with $|X_i| \geq h_0/4$, where $f \leq f(h_0/8)$; the probability that K observations with $|X| \geq h_0/4$ lie within $2r_n$ of one of them is at most $C n^K r_n^{K-1} = C n^{1-(K-1)(1/(1+\alpha)-1)}$, which is at most n^{-b} for K large. The probability that two observations lie within $r_n e^{-\sqrt{n}}$ is at most $C n^2 (r_n e^{-\sqrt{n}})^{1+\alpha}$ by (8). Off these events the excess terms total at most $K(C + |\alpha|(\sqrt{n} + \log 2))$, and $Q_n(t) - Q_n(0) \leq -c_4 n/4 + 1 + C \log n + CK\sqrt{n} < -1$.

For the last statement, parts 1 and 2 hold for every exponent; apply them with an exponent $b' \geq b + b/(1+\alpha)$. For $M \leq U_0$ the bound is trivial. For $U_0 \leq M \leq h_0/(2r_n)$, sum part 1 over the dyadic shells $U = 2^k M$ with $2^{k+1} M \leq h_0/r_n$, which reach $|u| \geq h_0/(2r_n)$, and add part 2; since $M^{-b} \geq (2r_n/h_0)^b \geq c n^{-b'}$, the total is at most $C_b M^{-b}$. For $M > h_0/(2r_n)$ the probability is at most $\min\{C_b n^{-b'}, \Pr(\max_i |X_i| > M r_n)\} \leq \min\{C_b n^{-b'}, C n e^{-c_0 M r_n}\}$ by (P4); the first term is at most $C M^{-b}$ when $M r_n \leq n$, and the second when $M r_n > n$. □

3.6 Local convergence and proof of Theorem 3.2

Lemma 3.11. *For every $R > 0$, the restriction of N_n to $[-R, R]$ converges in total variation to the restriction of $\mathcal{N}$.*

Proof. Let $p_n = \Pr(|Z| \leq R r_n) \leq C(R r_n)^{1+\alpha}$. Given its number of points, the restricted binomial process consists of independent points with the conditional law, and so does the Poisson process with mean measure $\mu_n(ds) = n r_n f(r_n s)\, ds$ on $[-R, R]$; hence their total variation distance is that of the counts, at most $n p_n^2 \to 0$ by Le Cam's inequality [17]. Two Poisson processes with mean measures μ, μ' can be built from a common Poisson process of intensity $\mu \wedge \mu'$ and independent remainders, so their distance is at most $\|\mu - \mu'\|$. Finally, by (P2), $n r_n f(r_n s) = |s|^\alpha (1 + \rho(r_n s))$, and $\int_{-R}^{R} |s|^\alpha |\rho(r_n s)| ds \leq C(R r_n)^\epsilon R^{1+\alpha} \to 0$. □

Proof of Theorem 3.2. Process convergence. Fix M and $R \geq 4M$. For a finite configuration μ on $[-R, R]$ with a point in $[-R/2, R/2]$ and no point at zero, define $\Phi_R(\mu) \in C[-M, M]$ by $\Phi_R(\mu)(u) = \alpha[\sum_{s \in \mu, s \neq \tau_\mu(u)} \log|s-u| - \sum_{s \in \mu, s \neq \tau_\mu(0)} \log|s|]$, with $\tau_\mu(u)$ a point of μ nearest to u, and put $\Phi_R(\mu) = 0$ otherwise. For $|u| \leq M$, such a point lies within $3R/4$ of u, so nearest points are computed inside $[-R, R]$. Let $Y_{n,R} = \Phi_R(N_n|_{[-R,R]})$, $Y_R = \Phi_R(\mathcal{N}|_{[-R,R]})$ and $Y_n(u) = Q_n(ur_n) - Q_n(0)$. By Lemma 3.11, $Y_{n,R} \Rightarrow Y_R$ as $n \to \infty$. On the event that $\mathcal{N}$ has a point in $[-R/2, R/2]$, $\Psi_\alpha - Y_R = \alpha \sum_{|t|>R} \log|1 - u/t|$, which is at most $2M|\alpha| \sum_{|t|>R} |t|^{-1}$ in absolute value; hence $Y_R \to \Psi_\alpha$ almost surely in $C[-M, M]$ as $R \to \infty$. It remains to show that $\lim_R \limsup_n \Pr(\|Y_n - Y_{n,R}\|_\infty > \varepsilon) = 0$ for every $\varepsilon > 0$. Indeed, for a bounded Lipschitz function F on $C[-M, M]$, $|\mathbb{E}F(Y_n) - \mathbb{E}F(\Psi_\alpha)|$ is at most $\|F\|_{\mathrm{Lip}} \varepsilon + 2\|F\|_\infty \Pr(\|Y_n - Y_{n,R}\|_\infty > \varepsilon) + |\mathbb{E}F(Y_{n,R}) - \mathbb{E}F(Y_R)| + |\mathbb{E}F(Y_R) - \mathbb{E}F(\Psi_\alpha)|$, and the last two terms tend to zero as $n \to \infty$ and then $R \to \infty$.

On the event $G_{n,R}$ that N_n has a point in $[-R/2, R/2]$, which fails with probability at most $\exp(-cR^{1+\alpha})$ as in (E3), the omitted observations at ur_n and at zero are near points, and both near sums in Y_n have $N_n([-R,R]) - 1$ terms. By (P2), $\log f(r_n s) = \log A + \alpha \log r_n + \alpha \log|s| + \log(1 + \rho(r_n s))$, so the constants cancel and

$$Y_n(u) - Y_{n,R}(u) = \mathcal{E}_n(u) + F_{n,R}(u), \qquad |\mathcal{E}_n(u)| \le 4C_0 N_n([-R,R])(2Rr_n)^{\epsilon},$$

for $|u| \le M$ and large n, where $F_{n,R}(u) = \sum_{|t_i|>R}[\log f(X_i - ur_n) - \log f(X_i)]$. The error $\mathcal{E}_n$ tends to zero uniformly in probability since $\mathbb{E}N_n([-R,R]) \le CR^{1+\alpha}$. For the far field, $|X_i| > Rr_n \ge 2|u|r_n$, and Taylor's formula gives $F_{n,R}(u) = -ur_n S_1 + u^2 r_n^2 S_2(u)/2$ with $S_1 = \sum_{|X_i|>Rr_n} \psi(X_i)$ and $|S_2(u)| \le \sum_{|X_i|>Rr_n} \sup_{|\xi| \ge |X_i|/2} |\psi'(\xi)|$. Since $\psi f = f'$ is odd, $\mathbb{E}S_1 = n\int_{|x|>Rr_n} f' = 0$, and by (P3), (7) on $|x| \le 1$ and $\int_{|x|>1}(1 + x^{-2})f \le 2$ beyond,

$$r_n^2 \operatorname{Var} S_1 + r_n^2 \mathbb{E}|S_2| \le Cnr_n^2 \Big[\int_{Rr_n < |x| \le 1} (1 + x^{-2})|x|^{\alpha} dx + 2 \Big] \le C(nr_n^2 + R^{\alpha - 1}).$$

Thus $\mathbb{E}\sup_{|u|\le M} |F_{n,R}(u)| \le C(MR^{(\alpha-1)/2} + M^2 R^{\alpha-1}) + o(1)$, which proves part 1.

Argmax. Part 1 of Proposition 3.10 shows $\Pr(\sup_{M \le |u| \le M'} Y_n(u) \ge -1) \le C_b(M^{-b} + n^{-b})$ for $U_0 \le M < M'$. The set of $y \in C[-M', M']$ with $\sup_{M \le |u| \le M'} y(u) > -1$ is open, so by part 1 and the portmanteau theorem $\Pr(\sup_{M \le |u| \le M'} \Psi_\alpha > -1) \le \liminf_n \Pr(\sup_{M \le |u| \le M'} Y_n(u) > -1) \le C_b M^{-b}$, and letting $M' \to \infty$, $\Pr(\sup_{|u|>M} \Psi_\alpha > -1) \le C_b M^{-b}$. Hence almost surely $\sup_{|u| \ge M} \Psi_\alpha < 0$ for some M, so by Lemma 3.7 T_α is the almost surely unique maximizer, it is tight, and $\Pr(|T_\alpha| > M) \le C_b M^{-b}$. The argmax continuous mapping theorem [18, Theorem 3.2.2], with part 1, the tightness from Proposition 3.10 and the continuity of Ψ_α, gives $\widehat{\theta}_n / r_n \Rightarrow T_\alpha$.

Moments and tails. Part 3 is the last statement of Proposition 3.10. With $b = p + 1$ it gives $\sup_{n \ge n_b} \mathbb{E}|\widehat{\theta}_n / r_n|^{p+1/2} < \infty$, so the variables $|\widehat{\theta}_n / r_n|^p$ are uniformly integrable and their means converge to $\mathbb{E}|T_\alpha|^p$, which is finite by the tail bound for T_α.

Optimality. By Lemma 3.9, $H^2(f^{\otimes n}, f(\cdot - cr_n)^{\otimes n}) \le nH^2(f, f(\cdot - cr_n)) \le Cc^{1+\alpha}/A$, and the total variation distance is at most the Hellinger distance. For c small it is at most $1/2$, and the usual two-point argument gives part 4. □

Proof of Corollary 3.3. By Lemma 3.5, f_ν satisfies condition (P) with $A = A_{\nu,1}$, and $f_{\nu,\sigma}$ satisfies it with $A = A_{\nu,1}\sigma^{-2\nu}$ after rescaling the data by σ, under which $\widehat{\theta}_n$ is equivariant. The midpoint property is Lemma 3.6 with the convexity in Lemma 3.5. In the parameterization of Nitithumbundit and Chan the mixing variable has mean one and the conditional variance is s^2 times it, so $2\sigma^2 G = s^2 G/\nu$. □

3.7 Cusps: proof of Theorem 3.4

Proof of Theorem 3.4. Let $\nu > 1/2$, $\sigma = 1$, $\theta_0 = 0$, $f = f_\nu$, and $\beta = 2\nu - 1$; for $\nu \ge 1$ read $\beta = 1$ in the bounds below. Put $L_n(t) = \sum_i \log f(X_i - t)$ and $\omega(x) = \log(f(0)/f(x)) \ge 0$. Since f is even and bounded,

$$Q_n(t) = L_n(t) - \log f(d(t)), \qquad 0 \le \log f(0) - \log f(d(t)) = \omega(d(t)). \tag{10}$$

By Lemma 3.5, $|\log f(y) - \log f(z)| \le C|y - z|^\beta$ for $|y - z| \le 1$, so $\omega(x) \le C|x|^\beta$ for $|x| \le 1$ and $\omega(x) \le C(1 + |x|)$ in general. Define the rate function $w(h) = |h|^{1/2+\beta}$ if $\beta < 1/2$, $w(h) = |h|\sqrt{\log(1/|h|)}$ if $\beta = 1/2$, and $w(h) = |h|$ if $\beta > 1/2$, for $|h| \le 1/2$.

Step 1: Hellinger and L^2 bounds. For $|h| \le h_1$ small,

$$c\,w(h)^2 \le H^2(f, f(\cdot - h)), \qquad \mathbb{E} \sup_{|h'-h|\le \varrho} \big[\log f(Z-h') - \log f(Z-h)\big]^2 \le C\,w(\varrho)^2 \tag{11}$$

for $0 < \varrho \le h_1$. For the upper bound split at $|x-h| \le 2\varrho$, where the Hölder bound gives $C\varrho^{2\beta}$ on a set of probability $O(\varrho)$; at $2\varrho < |x-h| \le 1$, where the mean value theorem and Lemma 3.5 give $C\varrho^2|x-h|^{2\beta-2}$; and at $|x-h| > 1$, where ψ is bounded. The three parts sum to $C[\varrho^{1+2\beta} + \varrho^2 \int_{2\varrho}^1 x^{2\beta-2}dx + \varrho^2] \le Cw(\varrho)^2$. For the lower bound with $\beta < 1/2$, on $x \in [2h, 3h]$, $h > 0$, the expansion of Lemma 3.5 gives $f(x-h) - f(x) \ge b_\nu(x^\beta - (x-h)^\beta) - Ch^2 \ge ch^\beta$, so $\sqrt{f(x-h)} - \sqrt{f(x)} \ge ch^\beta$ and the integral over this interval is at least $ch^{1+2\beta}$. For $\beta = 1/2$ the same argument on $2h \le x \le x_1$, with $f(x-h) - f(x) \ge chx^{\beta-1}$, gives $ch^2 \int_{2h}^{x_1} x^{-1}dx$. For $\beta > 1/2$, $\sqrt{f}$ has a continuous nonzero derivative on an interval $[x_1, x_2] \subset (0, \infty)$, which gives ch^2.

Step 2: consistency and near maximization. By (10), $Q_n(\widehat{\theta}_n) \ge Q_n(t)$ implies

$$L_n(\widehat{\theta}_n) \ge L_n(t) - \omega(d(\widehat{\theta}_n)) \qquad \text{for every } t. \tag{12}$$

The maximizer lies in $[X_{(1)}, X_{(n)}]$, so $d(\widehat{\theta}_n) \le 2K_0 \log n$ and the slack is at most $C \log n$, except with probability $O(n^{-1})$ when K_0 is large. For $\delta \le |t| \le K_0 \log n$, $\mathbb{E}e^{[L_n(t)-L_n(0)]/2} \le e^{-nc_\delta/2}$ with $c_\delta = \inf_{|h|\ge\delta} H^2(f, f(\cdot - h)) > 0$, and L_n has modulus $Cn|t-t'|^\beta$; a grid of mesh $n^{-2/\beta}$ shows that $\sup_{\delta\le|t|\le K_0 \log n}[L_n(t) - L_n(0)] \le -c_\delta n/4$ except with probability $o(1)$. Hence $\Pr(|\widehat{\theta}_n| > \delta) \to 0$ for every $\delta > 0$. Given this, $d(\widehat{\theta}_n)$ is at most the largest spacing of the sample inside $[-2\delta, 2\delta]$, where $\underline{f} \ge f(3\delta) > 0$, and that spacing is $O_P((\log n)/n)$. Therefore $\omega(d(\widehat{\theta}_n)) = o_P(1)$, and (12) says that $\widehat{\theta}_n$ maximizes L_n up to $o_P(1)$.

Step 3: rate. Apply [18, Theorem 3.2.5] to $\mathbb{M}_n(\theta) = n^{-1}L_n(\theta)$ and $\mathbb{M}(\theta) = \mathbb{E}\log f(Z-\theta)$, with the distance $w(\theta)$. Here $\mathbb{M}(\theta) - \mathbb{M}(0)$ is minus a Kullback–Leibler divergence, which is at least the squared Hellinger distance, so $\mathbb{M}(\theta) - \mathbb{M}(0) \le -cw(\theta)^2$ by (11). The class $\mathcal{F}_\delta = \{\log f(\cdot - h) - \log f : w(h) \le \delta\}$ has L^2 norms at most $C\delta$ and a bounded envelope, and covering $\{w(h) \le \delta\}$ by intervals of length ϱ with $w(\varrho) = \varepsilon/C$ gives, by (11), $\log N_{[]}(\varepsilon, \mathcal{F}_\delta, L^2) \le C + C\log(\delta/\varepsilon)$ for $\varepsilon < \delta$; for $\beta = 1/2$ this uses $w^{-1}(\delta)/w^{-1}(\varepsilon) \le C(\delta/\varepsilon)\sqrt{1 + \log(\delta/\varepsilon)}$. The bracketing integral is therefore at most $C\delta$, and the maximal inequality [18, Lemma 3.4.2] gives $\mathbb{E}\sup_{\mathcal{F}_\delta} |\mathbb{G}_n| \le C(\delta + n^{-1/2})$. With $\phi_n(\delta) = \delta + n^{-1/2}$ the theorem yields $\sqrt{n}\,w(\widehat{\theta}_n) = O_P(1)$, using the near maximization of Step 2 and the consistency. Hence $\widehat{\theta}_n = O_P(s_n)$, where s_n is the normalization of the theorem with $\sigma = 1$, namely $(n\gamma_\nu^2)^{-1/(1+2\beta)}$, $(n \log n)^{-1/2}$ or $n^{-1/2}$ in the three cases.

Step 4: local limits. Let $\ell_u(x) = \log f(x - us_n) - \log f(x)$ and $\Lambda_n(u) = \sum_i \ell_u(X_i)$. Since $\|\ell_u\|_\infty \to 0$ uniformly on compacts and $\mathbb{E}e^{\ell_u(Z)} = 1$, $n\mathbb{E}\ell_u = -\frac{1}{2}n\mathbb{E}\ell_u^2 + o(1)$ uniformly on compacts. If $\beta < 1/2$, the expansion $\log f(x) = \log f(0) - (b_\nu/f(0))|x|^\beta + R(x)$ with $|R'(x)| \le C(|x|^{2\beta-1} + |x|)$ near zero and dominated convergence give $n\mathbb{E}\ell_u^2 = (b_\nu^2/f(0))C_\beta\, ns_n^{1+2\beta}|u|^{1+2\beta}(1 + o(1)) = |u|^{2H}(1 + o(1))$, with $C_\beta = \int(|y-1|^\beta - |y|^\beta)^2 dy$; its closed form follows from the Fourier transform $-2\Gamma(1+\beta)\sin(\pi\beta/2)|\xi|^{-1-\beta}$ of $|x|^\beta$ away from the origin and Plancherel. If $\beta = 1/2$, the region $|x| \le Lus_n$ contributes $O(L/\log n)$, and on $Lus_n < |x| < x_1$ the score satisfies $\psi^2 f = (b_\nu^2/(4f(0)))|x|^{-1}(1 + O(|x|^{1/2}))$, which gives $n\mathbb{E}\ell_u^2 \to u^2 b_\nu^2/(4f(0)) = u^2 I_*$ after letting $L \to \infty$, the second-order Taylor error contributing $O(1/L)$. If $\beta > 1/2$, $n\mathbb{E}\ell_u^2 \to u^2 I_\nu$ by dominated convergence. Polarization and translation invariance give the covariances, and the Lindeberg central limit theorem gives the finite-dimensional convergence of Λ_n to $W_H(u) - \frac{1}{2}|u|^{2H}$, respectively $u\Delta - \frac{1}{2}u^2 I$ with $\Delta \sim N(0, I)$. For tightness in $\ell^\infty[-K, K]$, the class $\{\sqrt{n}(\ell_u - \ell_v) : |u|, |v| \le K,\ |u - v| \le \eta\}$ has L^2 norms at most some ε_η, with $\varepsilon_\eta \to 0$ as $\eta \to 0$, by (11), sup norms $o(\sqrt{n})$, and polynomial bracketing numbers, so [18, Lemma 3.4.2] gives asymptotic equicontinuity of the centered process; the mean

is equicontinuous by Cauchy–Schwarz. By [18, Section 1.5], Λ_n converges in $\ell^\infty[-K,K]$ for every K.

Step 5: the leave-one-out maximizer. By (10), $|Q_n(us_n)-Q_n(0)-\Lambda_n(u)| \le 2\omega(\max_{|t|\le Ks_n} d(t))$, and the largest spacing near zero is $O_P((\log n)/n)$, so the leave-one-out process has the same limit. The limit $W_H(u)-\frac{1}{2}|u|^{2H}$ is a continuous Gaussian process with $\mathrm{Var}(W_H(u)-W_H(v)) = |u-v|^{2H} \neq 0$ for $u \neq v$, so its maximizer is almost surely unique by [19, Lemma 2.6]; it tends to $-\infty$ at infinity by self-similarity and the Borell–TIS inequality, so the maximizer is tight. In the other two cases the limit is the quadratic $u\Delta - \frac{1}{2}Iu^2$ with maximizer $\Delta/I \sim N(0,1/I)$. The argmax theorem [18, Theorem 3.2.2], with Step 3, completes the proof; rescaling restores σ. □

3.8 Numerical illustration

The scripts and their outputs are included with the article. Table 1 lists Monte Carlo quantiles of $|T_\alpha|$ for three shapes, from 3000 draws of $\mathcal{N}$. The functional (5) is evaluated exactly with all points of modulus at most $50W$; the points with $50W < |t| \le 2\cdot 10^6$ enter through the fourth-order expansion of $\log|1-m/t|$, and the maximum is taken over the midpoints with $|m| < W$, where $W = 400, 400, 200$. At most 0.2% of the maximizers lie within 10% of the edge of the window. For a strong pole the law is concentrated very close to the origin, where the intensity $|t|^{2\nu-1}$ piles up points, and it spreads out as $\nu \uparrow 1/2$.

ν	α	$q_{0.25}$	$q_{0.5}$	$q_{0.75}$	$q_{0.9}$	$\Pr(\lvert T_\alpha\rvert > 1)$	$\Pr(\lvert T_\alpha\rvert > 16)$
0.10	−0.8	$4.1\cdot 10^{-6}$	$7.3\cdot 10^{-5}$	$9.0\cdot 10^{-4}$	$8.9\cdot 10^{-3}$	0.002	0.000
0.25	−0.5	0.041	0.248	1.35	4.16	0.294	0.020
0.40	−0.2	1.57	5.58	16.2	35.9	0.807	0.253

Table 1: Monte Carlo quantiles q_p of $|T_\alpha|$, $\alpha = 2\nu - 1$, for the intensity $|t|^\alpha\, dt$ (3000 draws).

For the exact estimator, Lemma 3.6 reduces the maximization at $\nu \le 1/2$ to the $n-1$ midpoints of adjacent order statistics; for $\nu > 1/2$ we use the same candidates, since $\psi_\nu' > 0$ at all 201 points of a logarithmic grid in $[10^{-8}, 60]$ for every $\nu < 1$ tested in the verification scripts. At $n = 500$ we evaluate all of them; for larger n we evaluate the midpoints within $200r_n$ of θ_0 (within $30s_n$ for $\nu > 1/2$) together with the eight best midpoints outside this window under a near-field screening criterion. At $n = 2000$ the windowed and the exhaustive searches agree in all 180 samples tried (60 each at $\nu = 0.25, 0.4, 0.6$), and no observation beats the best midpoint. With 300 samples for each n, the two-sample Kolmogorov–Smirnov statistic between the laws of $|\widehat{\theta}_n - \theta_0|/r_n$ and $|T_\alpha|$ lies between 0.048 and 0.105 for $\nu = 0.25$ and $n = 500, \dots, 8000$ (p-values 0.46, 0.004, 0.33, 0.55, 0.08). For the symmetric gamma density with $\alpha = -0.5$ and $\alpha = -0.2$ it lies between 0.042 and 0.078 for n between 1000 and 16000 (p-values between 0.07 and 0.72), and the interquartile range of $(\widehat{\theta}_n - \theta_0)/r_n$ at $\alpha = -0.2$ is 10.8, 12.6 and 12.1 at $n = 1000$, 4000 and 16000, against 11.1 for $T_{-0.2}$. For the variance gamma with $\nu = 0.4$ the statistic is between 0.07 and 0.11 for $n \le 4000$ (p-values 0.005, 0.08, 0.003, 0.012), and the law of the estimator approaches the limit slowly from the concentrated side: by Lemma 3.5, the relative correction $|\rho(x)| \approx (B_\nu/A_{\nu,1})|x|^{1-2\nu} = 0.98|x|^{0.2}$ lies between 0.1 and 0.4 at the scales $r_n \le |x| \le 30r_n$ for these n, and it decreases only like $n^{-(1-2\nu)/(2\nu)} = n^{-1/4}$. The symmetric gamma density, with $\epsilon = 1$, shows the limit law at the same sample sizes.

Table 2 compares rate exponents. The effective exponent is minus the slope of $\log r^*(n)$ against $\log n$, where $nH^2(f_\nu, f_\nu(\cdot - r^*)) = 1$; it is the exponent displayed by the Hellinger scale over a finite range of n, and it differs from the limit exponent near $\nu = 1/2$ and $\nu = 3/4$, where logarithmic factors enter. For $\nu = 0.5, 0.6, 0.7$ the exponents fitted to the interquartile ranges of the exact

estimator agree with the effective exponents to within 0.035, and at $\nu = 0.25$ the fitted value 1.92 agrees with 2; the standard error of these fitted exponents is about 0.15 at 300 samples. At $\nu = 0.4$ the fitted value 1.36 is within this sampling error, and the law approaches the limit from the concentrated side, as described above. The exponents fitted by Nitithumbundit and Chan agree with the effective exponents for $\nu \le 0.4$. For $0.44 \le \nu \le 0.8$ their fitted exponents are larger than the effective exponents over their range $500 \le n \le 10000$ and than our fits for the exact estimator; for example 1.01 against 0.86 and 0.85 at $\nu = 0.5$, 0.90 against 0.66 and 0.69 at $\nu = 0.6$, and 0.76 against 0.56 and 0.58 at $\nu = 0.7$. Their estimates are produced by an expectation–conditional maximization iteration with a local search over the observations nearest the iterate [2, Sections 4.4–4.6]; the theorems and our computations concern the exact global maximizer. On the cusp range their fitted exponents lie below $1/(2\nu)$ for $\nu \ge 0.76$, which they note [2, Section 5.1]; Theorem 3.4 gives the exponents $1/(4\nu - 1)$ and $1/2$ there, both below $1/(2\nu)$.

ν	limit exponent	effective $n \le 4000$	exact estimator	effective $n \le 10^4$	fitted in [2]
0.25	2.000	1.999	1.92	1.999	2.08, 1.92
0.40	1.250	1.178	1.36	1.189	1.21
0.50	1 (with $\log n$)	0.849	0.85	0.860	1.01
0.60	0.714	0.657	0.69	0.660	0.90
0.70	0.556	0.565	0.58	0.563	0.76

Table 2: Rate exponents: the limit exponent of Theorems 2.1, 3.2 and 3.4; the Hellinger effective exponent over $500 \le n \le 4000$ and over $500 \le n \le 10^4$; the exponent fitted to the interquartile range of the exact estimator (300 samples at each of $n = 500, 1000, 2000, 4000$); and the exponent fitted in [2, Tables 1 and 2] (20000 samples, $n = 500, \dots, 10000$; at $\nu = 0.25$ their entries for $\nu = 0.24$ and 0.26).

4 Sharp location rates for multivariate variance-gamma laws

For fixed symmetric variance-gamma noise in dimension $d \ge 2$, location is regular above shape one and has a Poisson local experiment below shape one. At shape one the limit is Gaussian, with a rate involving $\log\log n$ in dimension two and $\log n$ in higher dimensions. We obtain the exact information matrix at both critical rates and construct an estimator attaining its inverse.

4.1 Model and phase diagram

Observe n independent copies of $X = \theta + Z$ for $\theta \in \mathbb{R}^d$, $d \ge 2$, and known shape $\nu > 0$. Normalize the characteristic function of Z to $(1 + |t|^2)^{-\nu}$. A fixed known positive definite scatter is included by an invertible linear transformation. Write f for the density of Z, I_d for the identity matrix, and ω_{d-1} for the surface area of the unit sphere in $\mathbb{R}^d$.

Theorem 4.1. *The minimax expected Euclidean error and root mean squared Euclidean error for θ both have order d_n, where*

$$d_n = \begin{cases} n^{-1/(2\nu)}, & 0 < \nu < 1, \\ (n \log\log n)^{-1/2}, & \nu = 1,\ d = 2, \\ (n \log n)^{-1/2}, & \nu = 1,\ d \ge 3, \\ n^{-1/2}, & \nu > 1. \end{cases}$$

For fixed θ_0 and every compact parameter set in u, the local experiments at $\theta_0 + d_n u$ converge in Le Cam distance. If $0 < \nu < 1$, the limit is the Poisson process on $\mathbb{R}^d$ with intensity

$$A_{\nu,d}|x-u|^{2\nu-d}\,dx, \qquad A_{\nu,d} = \frac{\Gamma(d/2-\nu)}{4^\nu \pi^{d/2}\Gamma(\nu)}.$$

If $\nu = 1$, the limit is a Gaussian shift of information

$$\mathcal{I}_d = \begin{cases} \frac{1}{2}I_2, & d = 2, \\ \dfrac{d-2}{2d}I_d, & d \geq 3. \end{cases}$$

There is an estimator with normalized centered limit $N(0, \mathcal{I}_d^{-1})$ and with normalized squared risk tending to $\operatorname{tr}(\mathcal{I}_d^{-1})$ in both critical cases. That trace is also the limiting normalized minimax squared risk. If $\nu > 1$, the Gaussian shift at the root-sample rate uses the ordinary location Fisher information.

In the symmetric parameterization of Nitithumbundit and Chan [3] with positive scatter Σ, the linear map from the normalization above is $(\Sigma/(2\nu))^{1/2}$. Thus at criticality the information in their coordinates is Σ^{-1} in dimension two and $(d-2)\Sigma^{-1}/d$ in higher dimensions, at the corresponding rates. Their one-dimensional numerical study concerns the shapes $0 < \nu \leq 1$. The sharp rates and the limits of the leave-one-out estimator in one dimension are given in Sections 2 and 3, and [12] gives the minimax rate at the logarithmic pole. The multivariate results here concern the experiment and attainable location rates with known shape and scatter. The power-pole limits of Ibragimov and Khas'minskii [14] in one-dimensional location and the Poisson models of Dachian [20] are related. Gupta, Lee and Price [21] use smoothed Fisher information for high-probability location bounds for known laws. The planar criticality here uses the annular information of a logarithmic density pole, which accumulates at a double-logarithmic rate.

4.2 Density and weighted differences

Let G be gamma with shape ν and rate one. Conditional on G, Z is centered normal with covariance $2GI_d$. With $r = |x|$ its density is

$$f(r) = \frac{1}{(4\pi)^{d/2}\Gamma(\nu)} \int_0^\infty t^{\nu-d/2-1} e^{-t-r^2/(4t)}\,dt.$$

For $\nu < d/2$, substitution $s = r^2/(4t)$ gives $f(r) \sim A_{\nu,d} r^{2\nu-d}$ and $f'(r) \sim (2\nu-d)A_{\nu,d} r^{2\nu-d-1}$, by differentiating first and using the same substitution. The first two derivative orders are bounded by $Cr^{2\nu-d-j}$, $j = 1, 2$, near zero. For $\nu = d/2$, split the integral at r^2 and at one to get $f(r) \sim a_d \log(1/r)$, where $a_d = 2/[(4\pi)^{d/2}\Gamma(d/2)]$. Differentiation and substitution give $f'(r) \sim -a_d/r$ and $|f''(r)| \leq Cr^{-2}$. For $\nu > d/2$, f tends to a positive finite value. The derivative formula then gives the bounds $O(r^{2\nu-d-1})$ for $\nu < d/2+1$, $O(r\log(1/r))$ at equality, and $O(r)$ above it. These bounds follow by splitting the derivative integral at r^2 and at one. The integral is the Bessel expression proportional to $r^{\nu-d/2}K_{\nu-d/2}(r)$, so its tail and differentiated tail estimates follow also from [13]. In particular the tails are exponential and the logarithmic derivative is bounded on radii bounded away from zero.

We record the shift costs in $H^2(f,g) = \int(\sqrt{f} - \sqrt{g})^2$. Let $D = |h|$ be the shift length and fix a small $r_0 > 0$. For a power pole with $0 < \nu \leq 1$ and $\nu < d/2$, the part $|x| \leq C_0 D$ costs

$O(D^{2\nu})$: bound by the sum of the masses of the two translates. For $C_0 D < |x| < r_0$, square-root differentiation and the preceding bounds give a cost at most

$$CD^2 \int_{C_0 D}^{r_0} r^{2\nu-3}\, dr.$$

On an outer ray beyond fixed radius the cost is $O(D^2)$, by the square-integrable derivative of $\sqrt{f}$ there. When $0 < \nu < 1$, these inequalities give $O(D^{2\nu})$ overall and $O(D^{2\nu} R^{2\nu-2}) + O(D^2)$ outside $|x| \le RD$, for $R > C_0$.

For $\nu = 1, d = 2$, the inner ball $|x| \le C_0 D$ has smaller shift cost $O(D^2/\log(1/D))$. To check this, the integral formula, or the K_0 series [13], gives $f(r) = a_2 \log(1/r) + O(1)$ and $a_2 = 1/(2\pi)$. Both densities on this ball are bounded below by $c \log(1/D)$. Set $x = Dy$ and use the square-root identity: the scaled squared density difference is bounded by a constant times one plus the squared logarithms of the distances to the two poles. These logarithms are square integrable on the bounded mark set. From $C_0 D$ to r_0 the derivative bound for $\sqrt{f}$ gives cost

$$CD^2 \int_{C_0 D}^{r_0} \frac{dr}{r \log(1/r)}.$$

These estimates show $O(D^2 \log\log(1/D))$ overall. For $\nu = 1, d \ge 3$ the previous power estimates give $O(D^2 \log(1/D))$. In all cases the bounds in this section can be restricted to a smaller radial annulus, with the corresponding endpoints in the integrals.

For $\nu > 1$, the location Fisher information is finite. Near a power pole the radial integrand is $O(r^{2\nu-3})$; near a logarithmic pole with $d > 2$ it is $O(r^{d-3}/\log(1/r))$. Near a positive finite density at zero, the derivative bounds above are square integrable. There is no boundary term in the weak derivative of $\sqrt{f}$ at zero: its boundary integral over radius r vanishes by the same local density bounds. Thus $\sqrt{f}$ has an L^2 gradient, and translations are differentiable in quadratic mean by the L^2 translation identity. The usual iid local asymptotic normality theorem then gives the positive Fisher limit [22, Theorem 7.2]. It also gives $H^2 = O(D^2)$; the sample mean attains the root-sample bound using the finite second moment.

Multiplying the shift costs by n and taking sufficiently small fixed multiples of d_n gives the two-point lower bounds for expected norm and root mean squared norm in Theorem 4.1. We prove the remaining upper bounds and experiments below.

4.3 The subcritical Poisson phase

Take $0 < \nu < 1$ and analyze the local model at $\theta_0 = 0$. In the mark $x = X/d_n$ the one-draw density multiplied by n converges in L^1 on bounded sets to $A_{\nu,d}|x - u|^{2\nu-d}$, by the near-zero asymptotic and domination by this integrable pole. In a fixed mark ball the binomial count and conditional marks therefore converge in total variation to the corresponding Poisson process. The omitted product squared Hellinger cost relative to $u = 0$ is $O(R^{2\nu-2}) + o(1)$ outside a mark ball of large radius R. The same bound holds for the omitted Poisson intensities by the square-root derivative of the power profile. Regenerate discarded observations from the normalized reference complement conditional on the retained count, and do the same with reference intensities for the Poisson marks. The complement has mass bounded away from zero, so normalization changes the Hellinger bound by at most a fixed factor. Sending R to infinity proves convergence of the finite experiments.

For an estimator, use one portion of the sample for a sample-mean pilot within a small fixed bandwidth of θ. The pilot has exponentially small failure probability at any fixed tolerance because the centered laws have exponential tails. On an independent portion count points within distance h of a proposed center z. For $|z - \theta| \le \epsilon h$ with fixed $\epsilon < 1/4$, the expected count fraction is at

least $ch^{2\nu}$. If $|z-\theta| \geq Bh$, the fraction is at most $C(B-1)^{2\nu-d}h^{2\nu}$ for large B, up to an $O(h^d)$ bound if the trial ball is far from the pole. Choose B large and h small to get a fixed gap between the two normalized probabilities. The probability of any ball is $O(h^{2\nu})$ uniformly, by the radial decrease and the pole estimate. Bernstein's inequality gives $O(\exp(-cnh^{2\nu}))$ failure probability for each candidate to estimate its normalized count fraction to within a fixed limit smaller than the gap.

Start at a small fixed bandwidth and use dyadic scales down to $h_f \asymp n^{-1/(2\nu)}$. At each level consider a deterministic lattice of spacing at most $\epsilon h/\sqrt{d}$ and examine the points within Kh of the previous center, where $K > 2B+2$. Choose one exceeding a threshold strictly between the near and far means, or retain the previous center if none does. Given all previous successes, the candidate set contains a near point and is contained in a deterministic set of bounded cardinality within $O(h)$ of the true location. Thus one may reuse the same data at every level and bound the probability of the first failure by $C\exp(-cnh^{2\nu})$ using an unconditional union bound over that deterministic set. If the first failure is at h, the total remaining movement is $O(h)$, giving error $O(h)$. The sum of squared errors from first failures is $O(\sum_{h \geq h_f} h^2 \exp(-cnh^{2\nu})) = O(h_f^2)$. On pilot failure the error is bounded by the pilot error plus a fixed constant, which contributes exponentially small squared risk by its finite fourth moment and the pilot failure bound. This proves the subcritical upper rate.

4.4 Critical experiments

Take $\nu = 1$, and put $\Lambda_n = \log\log n$ if $d = 2$ and $\Lambda_n = \log n$ if $d \geq 3$. The critical local displacement order is $\delta_n = (n\Lambda_n)^{-1/2}$. Set $\Delta = n^{-1/2}(\log n)^2$ and retain the annulus $\Delta \leq |x| \leq r_0$, with a single censoring symbol for all other values. We first show that this retains the local experiments. If $d \geq 3$, the omitted cost at $|x| < \Delta$ times n is

$$O(n\delta_n^2[1 + \log(\Delta/\delta_n)]) = o(1)$$

by the restricted shift bounds above. For $d = 2$ it is at most

$$Cn\delta_n^2 \left[\frac{1}{\log(1/\delta_n)} + \int_{C_0\delta_n}^{\Delta} \frac{dr}{r\log(1/r)}\right] = o(1).$$

Indeed the integral is the difference of double logarithms with ratio of the inner logarithms tending to one. Constants may depend on the fixed finite set of local parameters. The omitted tail cost is $O(n\delta_n^2) = o(1)$. The discarded region has reference mass bounded below by a fixed positive tail probability. Regenerating its normalized reference values thus proves asymptotic equivalence to censoring.

Let $\ell(x) = -\nabla \log f(x)$ be the location score, and define the matrix

$$J_\Delta = \int_{\Delta \leq |x| \leq r_0} \ell(x)\ell(x)^{\mathsf{T}} f(x)\, dx.$$

The angular average of the radial unit vector's outer product is I_d/d. In dimension two, the density and its radial derivative satisfy

$$f(r) \sim (2\pi)^{-1}\log(1/r), \qquad f'(r) \sim -(2\pi r)^{-1}.$$

They give $J_\Delta \sim \frac{1}{2}\log\log(1/\Delta) I_2$. For $d \geq 3$,

$$f(r) \sim \frac{r^{2-d}}{(d-2)\omega_{d-1}}, \qquad \frac{f'(r)}{f(r)} \sim -\frac{d-2}{r}.$$

The Gamma relation for ω_{d-1} gives the coefficient from $A_{1,d}$. Hence $J_\Delta \sim \frac{d-2}{d}\log(1/\Delta)I_d$. In both cases $J_\Delta/\Lambda_n \to \mathcal{I}_d$.

For clarity the likelihood expansion on the retained experiment is a triangular iid Gaussian expansion. On the annulus, shifted densities differ relatively from the reference by $O(\delta_n/\Delta) = o(1)$. Subtract the linear term $\delta_n u^\mathsf{T}\ell(x)$ from the relative difference. The resulting total weighted squared error is at most $Cn\delta_n^4\Delta^{-2} = o(1)$. On the relevant segment the radii are comparable to r. The Hessian of the density is $O(r^{-d})$ in the power case and $O(r^{-2})$ in the planar logarithmic case. Dividing its squared bound by the density and integrating on the annulus gives $O(\Delta^{-2})$ in both. The censoring atom has no linear term since the annulus is centered and the score is odd. Its quadratic remainder cost is bounded by the continuous cost using Cauchy–Schwarz and its positive mass. The exact likelihood increments have mean zero, uniformly vanish, and have limiting n-fold covariance $u^\mathsf{T}\mathcal{I}_d u'$. The multivariate triangular central limit theorem gives the linear Gaussian term; squared sums converge in probability to their covariance and the cubic remainder tends to zero by the uniform bound. The limit likelihood vector is that of the stated Gaussian shift. Its coordinates have mean one. Convergence in law and convergence of these means imply finite Le Cam convergence by truncating and quantizing the likelihood vector under the reference, then restoring the reference conditional within each quantization cell. Together with the censoring equivalence this establishes the critical local limits.

4.5 Attaining the critical information

Use a pilot mean $\widetilde{\theta}$ from $n_0 = \lfloor n/\log n \rfloor$ observations and let $n_1 = n - n_0$. Write $e = \theta - \widetilde{\theta}$. The pilot satisfies $\mathbb{E}|e|^4 = O(n_0^{-2})$ and $\Pr(|e| > \Delta/4) \le C\exp(-c(\log n)^3)$, by the exponential moment and a fixed-dimensional coordinate bound on sample means. Put $\ell_\Delta(x) = \ell(x)\mathbf{1}_{\{\Delta \le |x| \le r_0\}}$. Using the remaining sample define

$$\widehat{\theta} = \widetilde{\theta} + J_\Delta^{-1}\frac{1}{n_1}\sum_{i>n_0}\ell_\Delta(X_i - \widetilde{\theta}).$$

For $|e| \le \Delta/4$, its fitted score mean under the true parameter is

$$\int \ell_\Delta(x) f(x-e)\,dx = J_\Delta e + O(|e|^2/\Delta).$$

Oddness gives the zero-order term and the score definition gives the linear one. For the remainder, combine the Hessian estimates with $|\ell(x)| \le C/r$ on the annulus and integrate $C\int_\Delta^{r_0} r^{-2}\,dr$. Then the squared bias of the estimator after normalization by $\sqrt{n\Lambda_n}$ is bounded by

$$C\frac{n\,\mathbb{E}|e|^4}{\Lambda_n\Delta^2} = o(1).$$

The score covariance at $e = 0$ is J_Δ. When $|e|/\Delta \to 0$, it is asymptotic to the same matrix. In fact the reference-relative density error is $O(|e|/\Delta)$ on the score support, using the derivative bounds, and the squared mean is negligible by the preceding expansion. On the whole good pilot event the second moment is bounded by a fixed multiple of J_Δ. The normalized single-observation score envelope is $C/(\Delta\sqrt{n\Lambda_n}) = o(1)$. Conditionally on the pilot, the triangular central limit theorem therefore gives the asserted Gaussian limit and the conditional covariance yields the normalized mean squared risk. The probability of the bad pilot is superpolynomially small. The score is deterministically bounded by C/Δ and the pilot has bounded fourth moments of its centered error, so the bad event contributes $o(1)$ to normalized squared risk. This estimator is translation equivariant. All rate bounds hold uniformly in θ.

4.6 Compact parameter sets and the minimax constant

For compact local parameter sets, choose a finite net. The experiments on the net converge by the preceding proofs, with a Gaussian limit in the regular case as well. For two input parameters u, v with $s = |u - v| \le 1/2$, the product squared Hellinger bounds from Section 4.2 are $Cs^{2\nu}$ in the Poisson phase, $Cs^2(1 + \log(1/s))$ at criticality, and Cs^2 above it, uniformly for large n. At criticality this uses $D = d_n s$ inside the logarithms. The limits are also uniformly continuous in Hellinger distance, by the shifted intensity bound or the Gaussian affinity. A kernel that compares the restrictions to the net therefore compares every parameter to within the net error plus these two continuity moduli. Letting the mesh tend to zero yields compact-set convergence.

To see the exact critical minimax lower constant, fix a large ball in the local parameter scale and project the local decisions to that ball. Projection cannot increase squared loss for parameters there, and makes the loss bounded. The compact Gaussian limit thus lower bounds the original global minimax risk after normalization by $n\Lambda_n$. In the Gaussian shift, use a centered normal prior of covariance $\tau^2 I_d$ conditioned to the ball of radius R. For estimators valued in that ball, extending the integrated squared loss to the whole prior adds at most $\int_{|u|>R}(R+|u|)^2\, d\Pi_\tau(u)$, tending to zero as $R \to \infty$ for fixed τ. The whole-prior Gaussian Bayes risk is $\mathrm{tr}[(\mathcal{I}_d + \tau^{-2} I_d)^{-1}]$. Let $R \to \infty$, then $\tau \to \infty$. The resulting lower bound equals the upper constant from the risk of the estimator in Section 4.5.

5 Critical variance-gamma location with unknown shape, skew and scatter

At shape one, multivariate variance-gamma location has a faster rate than the other parameters. We show that the joint local experiment splits into independent Gaussian shifts for location and for shape, skew and scatter. The location normalization is $\sqrt{n \log\log n}$ in two dimensions and $\sqrt{n \log n}$ in higher dimensions. A cumulant pilot and an annular score estimator attain the same limiting location risk as in the experiment with the other parameters known.

5.1 Model and results

Let $d \ge 2$ be fixed. Observe independent copies of

$$X = \theta + \gamma G + \sqrt{G}\, S^{1/2} V, \qquad G \sim \mathrm{Gamma}(\nu, 1), \quad V \sim N(0, I_d),$$

where G, V are independent and S is real symmetric positive definite. Here $\eta = (\nu, \gamma, S)$ is unknown and θ is the location parameter and gives the pole position at shape one. Use ordinary Euclidean coordinates on the vector and independent matrix entries. Fix an interior base point (θ_0, η_0) with $\nu_0 = 1$, and put

$$\Lambda_n = \begin{cases} \log\log n, & d = 2, \\ \log n, & d \ge 3, \end{cases} \qquad \delta_n = (n\Lambda_n)^{-1/2}, \qquad \mathcal{K}(S) = \begin{cases} S^{-1}, & d = 2, \\ \dfrac{d-2}{d} S^{-1}, & d \ge 3. \end{cases}$$

Write f_η for the density with $\theta = 0$, $\ell_\eta = -\nabla_x \log f_\eta$ for the location score away from the pole, and $s_\eta = \nabla_\eta \log f_\eta$ for the ordinary nuisance score. All score identities are understood almost everywhere.

Theorem 5.1. *At $\theta_n = \theta_0 + \delta_n u$, $\eta_n = \eta_0 + h/\sqrt{n}$, the local experiments converge on every compact set of (u, h) in Le Cam distance to independent Gaussian shifts of information $\mathcal{K}(S_0)$ for u and*

$$\mathcal{J}(\eta_0) = \mathbb{E}_{\eta_0}[s_{\eta_0} s_{\eta_0}^{\mathsf{T}}]$$

for h. The nuisance information is finite and positive definite. In particular the location information is the same with the nuisance parameters unknown.

Theorem 5.2. *Fix compact brackets for θ, γ, S with the eigenvalues of S bounded away from zero. For each fixed $H < \infty$, there is an estimator using unknown shape within $|\nu - 1| \le H/\sqrt{n}$ and unknown γ, S that satisfies*

$$\sqrt{n\Lambda_n}(\widehat{\theta} - \theta) \implies N(0, \mathcal{K}(S_0)^{-1}) \quad \text{along every sequence with } S \to S_0.$$

Its normalized squared risk differs from $\operatorname{tr}(\mathcal{K}(S)^{-1})$ by $o(1)$ uniformly on these brackets. This is the sharp local minimax constant at every interior critical point.

Nitithumbundit and Chan [23, 3] study this normal mixture parameterization and construct likelihood algorithms for the skew multivariate model. Their simulations of the asymptotic location law and rate in [2, Section 5] use the univariate symmetric model with the other parameters fixed. Methods of moments for variance-gamma models are classical. Seneta [24] uses moment fitting in the skew univariate model. Fischer, Gaunt and Sarantsev [25] study their properties and modifications in the symmetric case. Here moments produce a preliminary estimate. The critical location information comes from observations near the pole, and the split experiments and attaining estimator show which information survives when the other parameters vary. The known symmetric phase diagram is derived in Section 4.

5.2 Density bounds near the pole

Let $L = (S/2)^{1/2}$, $y = L^{-1}x$, $\beta = L^{-1}\gamma$ and $r = |y|$. Completing the square in the normal density gives

$$f_\eta(x) = \frac{|\det L|^{-1} e^{\beta \cdot y/2}}{(4\pi)^{d/2}\Gamma(\nu)}\, B(\nu, a, r), \quad B(\nu, a, r) = \int_0^\infty t^{\nu - d/2 - 1} e^{-at - r^2/(4t)}\, dt, \quad a = 1 + |\beta|^2/4. \tag{13}$$

Uniformly on compact brackets for η near $\nu = 1$ and $0 < |x| \le r_0$ for small fixed r_0, this implies

$$|\nabla_x \log f_\eta| \le C/|x|, \qquad \frac{\|D_x^2 f_\eta\|}{f_\eta} \le C/|x|^2, \tag{14}$$

$$|s_\eta| \le C(1 + \log(1/|x|)), \qquad \|D_\eta \ell_\eta\| \le C(1 + \log(1/|x|))/|x|. \tag{15}$$

To check these bounds, inserting t^{-j} in the integral for B costs at most $C_j r^{-2j}$. For $t \ge r^2$ this follows directly. For $t < r^2$ put $z = r^2/(4t)$ and compare the resulting bound to the part of B from $r^2 \le t \le 2r^2$. A further $|\log t|$ costs at most $C(1 + \log(1/r))$ in the same bounds. For $r^2 \le t \le 1$ this is immediate and for $t > 1$ the exponential factor supplies the bound, using a fixed interval away from zero as a lower bound for B. Differentiation in a introduces t, harmless by splitting at one. Differentiating in ν introduces $\log t$. Spatial differentiation uses the $j = 1, 2$ bounds, while differentiation of $L^{-1}x$ in a parameter has size $O(|x|)$. Applying these estimates to (13), including its smooth prefactors, proves (14) and (15). Outside a fixed ball the log spatial derivative is uniformly bounded on the same parameter brackets, by the large-argument asymptotic of the equivalent Bessel kernel. The tails are uniformly exponential, also directly from the normal mixture moment-generating function.

At $\nu = 1$, remove the exponential tilt in (13) and use y coordinates. The radial part has the same leading coefficients and first logarithmic derivative as in the zero-skew density. Indeed the substitution $z = r^2/(4t)$ gives the power coefficient $1/[(d-2)\omega_{d-1}]$ for $d \ge 3$, independent of a. For

$d = 2$, the density is $(2\pi)^{-1}\log(1/r) + O(1)$ and its radial derivative is asymptotic to $-(2\pi r)^{-1}$. Splitting at r^2 and one gives the first expression with bounded remainder. Substitution in the differentiated integral gives the second. The tilt tends to one and adds a bounded term to the location score. Its multiplicative error $O(r)$ and the bounded score addition contribute $O(1)$ to the annular information, using integrability of $f_\eta/|x|$ at a critical pole. Thus, for $A_\Delta = \{x : \Delta \le |x| \le r_0\}$ and $\Delta = n^{-1/2}(\log n)^2$, at $\nu = 1$

$$\Lambda_n^{-1} J_\Delta(\eta) \;\longrightarrow\; \mathcal{K}(S), \qquad J_\Delta(\eta) = \int_{A_\Delta} \ell_\eta \ell_\eta^{\mathsf{T}} f_\eta. \tag{16}$$

The convergence is uniform on the brackets. To see the normalization, angular integration in the y coordinates gives $I_2/2$ times $\log\log(1/\Delta)$ in the plane, and $(d-2)I_d/d$ times $\log(1/\Delta)$ in higher dimensions. The radial cutoffs after the linear change lie between fixed multiples of the displayed cutoffs, which changes the leading terms by $o(\Lambda_n)$. Then conjugate by L^{-1} and use $\log(1/\Delta) \sim (\log n)/2$. In particular $f_\eta(1 + \log(1/|x|))/|x|$ is integrable near zero, uniformly at shape one. Multiplication by $|x|^{-\epsilon}$ for a fixed sufficiently small $\epsilon > 0$ preserves this property.

5.3 A moment pilot for the full parameter

A finite moment map is uniformly invertible on the stated brackets. In any coordinate, the marginal characteristic function after removing θ_j factors as

$$(1 - iat)^{-\nu}(1 + ibt)^{-\nu}, \qquad a - b = \gamma_j, \quad ab = S_{jj}/2, \quad a, b > 0.$$

To determine ν from the first coordinate, use its cumulants c_k of orders $k = 2, 3, 4$. They satisfy $c_k = (k-1)!\nu(a^k + (-b)^k)$. Set $\rho = (a^2 + b^2)^{1/2}$, $w = (a - b)/\rho$, $k_1 = c_3/(2c_2)$, and $k_2 = c_4/(6c_2)$. Then

$$\frac{k_1}{\sqrt{k_2}} = \frac{w(3 - w^2)}{\sqrt{2}\sqrt{1 + 2w^2 - w^4}}, \qquad k_2 = \frac{\rho^2}{2}(1 + 2w^2 - w^4).$$

The derivative of the first displayed function has strictly positive numerator proportional to $(1 - w^2)(3 - w^4)$. Since positive a, b stay in compact brackets, w stays a fixed distance from ± 1. Thus the inverse recovers w, ρ, then $\nu = c_2/\rho^2$, with uniform Lipschitz constants.

Given ν, put $B_1 = \mathrm{Cov}(X)/\nu = S + \gamma\gamma^{\mathsf{T}}$. For each coordinate the third cumulant divided by $\nu(B_{1,jj})^{3/2}$ equals $3v_j - v_j^3$ for $v_j = \gamma_j/\sqrt{B_{1,jj}}$. Its derivative is uniformly positive on the compact range away from ± 1. This recovers γ with a Lipschitz inverse, then $S = B_1 - \gamma\gamma^{\mathsf{T}}$ and $\theta = \mathbb{E}X - \nu\gamma$. All these quantities use only the mean, covariance, coordinate third cumulants and the first coordinate's fourth cumulant. They are smooth functions of the requisite raw moments on the relevant compact set.

Fit ν in a fixed compact positive bracket containing one in its interior. Let $n_0 = \lfloor n/a_n \rfloor$ with $a_n = \sqrt{\log\log n}$ for large n. Estimate the raw moments on the pilot sample, truncating their integrands to $|X| \le C_0 \log n$. The omitted expectations are $O(n^{-A_0})$ for any fixed A_0 upon choosing C_0 large enough. Use an approximate minimum-distance fit to the model moments over the compact brackets, to tolerance n^{-1}. The inverse Lipschitz inequalities, bounded variances and fourth moments of the empirical integrands give squared and fourth power pilot errors $O(n_0^{-1})$ and $O(n_0^{-2})$. At error threshold Δ/C for any fixed $C > 0$, the bounded integrand envelope is $O((\log n)^4)$ with $\Delta(\log n)^4 = o(1)$. Bernstein's inequality thus bounds the failure probability by $C\exp(-cn_0\Delta^2)$, which is superpolynomially small. These estimates hold uniformly on the brackets.

The nuisance model at fixed θ is differentiable in quadratic mean. Indeed the complete model (G, X) has square-integrable scores continuous in quadratic mean on compact interior brackets:

differentiate the gamma density in shape and the normal conditional density in mean and covariance. Gamma log moments and normal polynomial moments bound the resulting score norms uniformly. Marginalizing gives the usual conditional expectation of the complete score and preserves quadratic mean differentiability [22, Chapter 7]. This is s_η from (15). If a linear combination of its coordinates vanished, all the corresponding derivatives of the moment map would vanish, by differentiating under the complete model. The map has a smooth inverse in a neighborhood of the point, by the displayed formulas with nonzero denominators and derivatives. Hence that combination is zero only for the zero direction. This proves positive definiteness of $\mathcal{J}$.

5.4 The joint local experiment

Center at $\theta_0 = 0$. Fix finitely many (u, h) including $(0, 0)$, let $f_h = f_{\eta_0 + h/\sqrt{n}}$ and $f_0 = f_{\eta_0}$, and abbreviate $A = A_\Delta$. First replace $f_h(x - \delta_n u)$ off the annulus A by $f_h(x)$, and normalize the resulting density. The product laws change by $o(1)$ in total variation.

Here are details for the omitted region. At η_0 the planar pole at scale $D = |u|\delta_n$ has inner shift cost $O(D^2/\log(1/D))$ in the ball $|x| \le CD$. After $x = Dy$, the logarithmic density differences there have square integrable dependence on the two pole distances and a square-root denominator bounded below by $c\log(1/D)$. In the power case $d \ge 3$ the inner cost is $O(D^2)$ by the two ball masses. Between CD and Δ the cost is bounded by $CD^2 \int dr/(r\log(1/r))$ in the plane and $CD^2 \int dr/r$ in higher dimensions, by the square-root derivative and the asymptotics of (13). Both omitted costs times n tend to zero: $\log(\Delta/D) = O(\log\log n)$ for fixed nonzero u and the planar difference of inner double logarithms is $o(1)$. These bounds also hold at $\eta_0 + h/\sqrt{n}$ for the discarded inner ball in the required $o(1/n)$ sense. Indeed the restricted nuisance squared Hellinger cost between f_h and f_0, on a ball of radius 2Δ and likewise after the location shift, is bounded by

$$\frac{C}{n} \int_{|x| \le 3\Delta} f_0(x)(1 + \log(1/|x|))^2 |x|^{-\epsilon}\, dx = o(1/n)$$

using (15) along the short path. Combine by the triangle inequality for the square-root norms. Outside the fixed outer ball, the shift costs $O(D^2)$ uniformly in the nuisance by tail derivative bounds. Normalization of the spliced density changes the Hellinger bound by at most a fixed multiple, by the reverse triangle inequality for square-root norms. This proves the replacement for the finite experiment.

The replaced density relative to f_h is $(1 + k(x))/(1 + \mathbb{E}_h k)$, where

$$k(x) = \mathbf{1}_A(x) \left\{ \frac{f_h(x - \delta_n u)}{f_h(x)} - 1 \right\}, \qquad \mathbb{E}_h k = \int k f_h.$$

By (14), $|k| \le C\delta_n/\Delta = o(1)$. On the annulus the density ratios f_h/f_0 tend uniformly to one by (15). Taylor expansion, (14) and the critical density bounds give

$$n \int \left| k - \delta_n u^\mathsf{T} \ell_{\eta_0} \mathbf{1}_A \right|^2 f_0 \;\longrightarrow\; 0.$$

For the location Taylor remainder use a bound $Cn\delta_n^4 \Delta^{-2} \log(1/\Delta) = o(1)$ from integrating $f_0/|x|^4$ on the annulus. For the score replacement, (15) bounds the difference by $Cn^{-1/2}(1 + \log(1/|x|))/|x|$. The corresponding integral times $n\delta_n^2$ is $O(\delta_n^2(1 + \log(1/\Delta))^4) = o(1)$. Also $\mathbb{E}_h k = O(\delta_n)$. The difference between $\mathbb{E}_h k$ and $\mathbb{E}_0 k$, multiplied by n, is $O(\sqrt{n}\,\delta_n) = o(1)$ by (15), the bound $|k(x)| \le C\delta_n/|x|$ on A and the integrability following (16). It follows by expanding the log of the relative likelihood factor that its log contribution under f_0 is

$$\delta_n \sum_{i=1}^n u^\mathsf{T} [\ell_{\eta_0}(X_i) \mathbf{1}_A(X_i) - \mathbb{E}_0(\ell_{\eta_0} \mathbf{1}_A)] - \tfrac{1}{2} u^\mathsf{T} \mathcal{K}(S_0) u + o_{\Pr}(1).$$

For the quadratic term, the sum of the squares of k concentrates because of the vanishing envelope and bounded total squared mean. The normalization quadratic is $O(n\delta_n^2) = o(1)$. The information limit in (16) applies, since the truncated score mean is uniformly bounded.

Meanwhile the nuisance likelihood for f_h relative to f_0 has the regular $\sqrt{n}$ expansion by quadratic mean differentiability [22, Chapter 7]. The two central sequences have asymptotically zero cross covariance:

$$\left\| \sqrt{n}\, \delta_n \mathbb{E}_0[\ell_{\eta_0} \mathbf{1}_A s_{\eta_0}^{\mathsf{T}}] \right\| = O(\Lambda_n^{-1/2}),$$

using (15) and the stated radial integrability. The location rows have envelope $O(\delta_n/\Delta)$, and the nuisance rows meet Lindeberg by their finite second moment. Joint triangular normal convergence thus gives independent Gaussian shifts. The resulting likelihood coordinates have mean one. Truncation and quantization under the reference, followed by regeneration within the quantization cells, give finite Le Cam convergence.

To extend to compact sets, use finite nets and pairwise Hellinger continuity. At fixed location the nuisance product squared Hellinger distance is $O(|h-h'|^2)$, by smooth complete-data quadratic mean bounds uniformly on the brackets. At a fixed nuisance $\eta_0 + h/\sqrt{n}$, the annulus replacement just proved approximates the local locations uniformly for bounded u, h, with negligible error as $n \to \infty$. For uniformity on a bounded set of u, split at $C\delta_n$ with $C > 2 \sup |u|$. Within this ball, the bounded mark integrals give $O(\delta_n^2)$ in the power case and $O(\delta_n^2/\log(1/\delta_n))$ in the plane, uniformly for the allowed shifts. Between $C\delta_n$ and Δ use the same derivative integrals as before with $D^2 \le C^2\delta_n^2$. This gives the uniform $o(1)$ product bound. On the annulus the squared Hellinger cost for two splices at the same nuisance is at most $C|u-u'|^2/n$, since spatial score integration along the segment is bounded by $C\Lambda_n$, using density comparability under the bounded shifts and the critical information integral. Normalization obeys the same bound. The product limits are continuous. The triangle inequality through a nearest net point and contraction under Markov kernels prove compact-set convergence. This proves Theorem 5.1.

5.5 Estimating location with unknown nuisance

Let $\widetilde{\theta}, \widehat{\gamma}, \widehat{S}$ come from the moment pilot. In the score step set $\widehat{\eta} = (1, \widehat{\gamma}, \widehat{S})$; the unknown true shape may differ from one by $H/\sqrt{n}$. Thus the pilot nuisance error remains of order $O(n_0^{-1})$ in squared mean and its fixed-Δ failure probability is still superpolynomially small. Write $\ell_\Delta(x) = \ell_{\widehat{\eta}}(x)\mathbf{1}_A(x)$ and $m_\Delta = \int \ell_\Delta f_{\widehat{\eta}}$. Using the remaining $n_1 = n - n_0$ observations, set

$$\widehat{\theta} = \widetilde{\theta} + J_\Delta(\widehat{\eta})^{-1} \left\{ \frac{1}{n_1} \sum_{i>n_0} \ell_\Delta(X_i - \widetilde{\theta}) - m_\Delta \right\}.$$

In the skew model the subtraction of m_Δ is needed. It is uniformly bounded by radial integrability. For $e = \theta - \widetilde{\theta}$, on the good pilot event $|e| \le \Delta/4$ and $|\eta - \widehat{\eta}| \le \Delta/4$, the conditional centered score mean equals

$$J_\Delta(\widehat{\eta})e + O\left(|e|^2 \frac{\log(1/\Delta)}{\Delta} + |\eta - \widehat{\eta}| \right).$$

For the location expansion at the fitted nuisance, use the Hessian bound in (14), comparability of densities along the shift, and integrate $f_{\widehat{\eta}}/|x|^3$ on the annulus. This gives the first error bound. For the nuisance discrepancy use (15) on the same shifted annulus and integrate $f_{\widehat{\eta}}(1 + \log(1/|x|))/|x|$. That integral is uniformly bounded, giving the second error. By the pilot moment bounds and (16), the normalized squared bias tends to zero:

$$C \frac{n}{\Lambda_n} \left(n_0^{-2} \frac{\log^2(1/\Delta)}{\Delta^2} + n_0^{-1} \right) = o(1),$$

since $a_n = o(\Lambda_n)$.

On the score support the conditional true density divided by $f_{\widehat{\eta}}$ differs from one by at most a constant times $|e|/\Delta + |\eta - \widehat{\eta}| \log(1/\Delta)$. This tends to zero in probability and stays bounded on the good event. The score means have size $O(1) + o_{\Pr}(1)$ since $J_\Delta e = o_{\Pr}(1)$, so the covariance on the Λ_n scale tends to $\mathcal{K}(S)$. The conditional score rows have normalized envelope $O((\Delta\sqrt{n\Lambda_n})^{-1}) = o(1)$. The conditional triangular central limit theorem gives the stated normal limit, uniformly along convergent scatter sequences. The conditional score second moments are bounded by a fixed multiple of J_Δ on the good event, so the normalized variance converges also in expectation. On the bad event the score is still deterministically bounded by C/Δ, and the pilot lies in its compact brackets. Superpolynomially small probability makes this event's squared risk contribution negligible after normalization. All bounds are uniform. Finally fixing the nuisance gives the Gaussian local minimax lower constant by Theorem 5.1. Project decisions to fixed local balls to make squared loss bounded, then take diffuse normal priors truncated to expanding balls in the limit shift. The Gaussian Bayes risk tends to $\operatorname{tr}(\mathcal{K}(S_0)^{-1})$. The estimator therefore attains the sharp local constant also when the nuisance is unknown.

6 The sharp minimax risk at the variance-gamma logarithmic pole

For symmetric variance-gamma location at the logarithmic pole, the flat-prior posterior and the normalized leave-one-out likelihood converge to the same random Brownian probability measure. Their averages attain the sharp minimax squared-risk constant $16\zeta(3)\sigma^2/\pi^2$ at scale $(\log n)/n$. The posterior proof controls the integrated effect of the unbounded likelihood spikes. It also shows that averaging the actual leave-one-out likelihood improves the limiting risk of its maximizer.

6.1 Model and results

Observe independent $X_i = \theta + Z_i$, $i = 1, \dots, n$, $n \geq 2$, with Z_i of known density

$$f_\sigma(x) = \frac{1}{\pi\sigma} K_0(|x|/\sigma), \qquad \sigma > 0.$$

Write $j(t)$ for a sample index nearest to t, and set

$$L_n(t) = \prod_i f_\sigma(X_i - t), \qquad L_n^o(t) = \prod_{i \neq j(t)} f_\sigma(X_i - t).$$

The sample is distinct almost surely, so every pole of L_n is locally integrable. Both weights are integrable on $\mathbb{R}$ with finite moments of every fixed order, by exponential tails. Let $\widehat{\theta}_n$ and $\widehat{\theta}_n^o$ be the means of these weights after division by their integrals. The first is the flat-prior posterior mean. The second averages the same leave-one-out likelihood whose maximum is the leave-one-out estimator studied by Nitithumbundit and Chan [3] for this symmetric known-scale model.

Write $d_n = (\log n)/n$, $I = \pi/\sigma$, and let $\mathcal{B}$ be standard two-sided Brownian motion. Put

$$Z(u) = \exp\{\sqrt{I}\,\mathcal{B}(u) - I|u|/2\}, \qquad \Pi(du) = Z(u)\,du / \int_{\mathbb{R}} Z.$$

Theorem 6.1. *Under any θ, the normalized likelihood measures for L_n and L_n^o, both in the variable $u = (t - \theta)/d_n$, converge jointly in distribution to (Π, Π) in Wasserstein distance of every fixed order $p \geq 1$. Consequently both centered and normalized averages converge to $\int u\,\Pi(du)$, and their*

absolute moments of every fixed positive order converge. Their difference divided by d_n tends to zero in every fixed L^q, $q > 0$. Moreover,

$$\lim_{n\to\infty} d_n^{-2} \inf_{\widehat{t}} \sup_{\theta\in\mathbb{R}} \mathbb{E}_\theta(\widehat{t} - \theta)^2 = \frac{16\zeta(3)}{I^2}.$$

The true posterior mean is exactly minimax for each n. The leave-one-out average attains the displayed limit. Its limiting normalized risk is $8\zeta(3)/13$ times that of the leave-one-out maximizer.

Pitman [9] gives the best equivariant location estimator by averaging the likelihood, and Girshick and Savage [10] prove that Pitman estimators are minimax for squared error; Section 6.4 gives a direct proof here. Strasser [26] develops asymptotic minimax theory for Pitman estimates in translation invariant limit experiments. The canonical Brownian likelihood mean has squared moment $16\zeta(3)$ in unit information, as computed by Rubin and Song [8, 7] in the change-point setting. The ratio $8\zeta(3)/13$ is their asymptotic efficiency of the maximum likelihood estimator in that setting. The Brownian argmax has second moment 26, due to Terent'yev [6]; see also [7]. Section 2 proves that the unmodified variance-gamma leave-one-out maximizer has that canonical argmax limit and its associated risk. The general theorems of Ibragimov and Has'minskii on Bayes estimators [11, Chapter I, Section 10] give the limit law and moment convergence of posterior means from finite-dimensional convergence of the likelihood ratio and Hellinger bounds of the kind proved in Section 2; the global weighted integral control below gives the posterior half directly despite the infinite spikes of the likelihood. The new step is the same control for the leave-one-out likelihood, whose deletion changes with t, so that its criterion is not a likelihood ratio. This transfers the canonical risk to the actual variance-gamma likelihood averages. The rate $(\log n)/n$ itself is minimax by [12, Theorem 6.3].

6.2 Capped comparison

Reduce by translation and dilation to $\theta = 0, \sigma = 1$. Write $f = f_1$, $L = \log n$, and let $f_r(x) = f(|x| \vee r)$. For

$$C_r(t) = \sum_i \bigl[\log f_r(X_i - t) - \log f_r(X_i)\bigr]$$

we recall the estimates from the exact leave-one-out analysis of Section 2 and indicate the strengthening needed for weighted integrals. The density decreases as $|x|$ increases. Near zero,

$$f(x) = \pi^{-1} \log(1/|x|) + O(1), \qquad |f'(x)| \le C/|x|.$$

Its log derivative is bounded away from the pole, and its tails are exponential. In particular

$$\int_{|x|\le r} \{f(x) - f(r)\}\, dx \le Cr. \tag{17}$$

The normalized cap has product total variation difference $O(nr)$ from the true reference.

Fix positive constants b_0, b_1 and a weight order $p \ge 1$. Choose $D > b_0 + 5$ and a coarse cap $r_g = n^{-D}$. There is a fixed C_0 such that all observations are in $[-C_0 L, C_0 L]$ except with probability $O(n^{-b_0})$. Since $f \in L^2$, the probability of a pair gap below $2r_g$ is $O(n^2 r_g)$. The probability of an observation within r_g of zero is $O(n r_g L)$. The same union and mesh argument as in Section 2 gives, for any sufficiently large fixed B,

$$\sup_{BL^2/n \le |t| \le (C_0+1)L} C_{r_g}(t) \le -c_0 BL \tag{18}$$

with probability $1 - O(n^{-b_0})$. Here $c_0 > 0$ can be chosen independent of B. Indeed n times the single-observation squared Hellinger distance for the normalized coarse cap is at least cBL on this range. Under the normalized capped reference the half exponential moment is the affinity, so Markov's inequality at the threshold $-c_0BL$ gives an exponent proportional to $-BL$. The polynomial mesh has width $n^{-(D+3)}$, as the capped log likelihood is Lipschitz with constant Cn/r_g. This pays the mesh union for large B; the total variation difference $O(nr_g)$ transfers it. We include the observation range, pair-gap and zero-gap conditions in this coarse good event.

On that event the centered log leave-one-out criterion and C_{r_g} differ everywhere by a difference of capped deleted terms. On $|t| \le (C_0 + 1)L$ this is $O(L)$ with a constant independent of B. Thus (18) holds for the centered leave-one-out log criterion with a smaller c_0 by choosing B large. The uncapped original likelihood ratio equals $e^{C_{r_g}(t)}$ except on intervals of radius r_g around the observations. On such an interval it equals this cap times the factor $f(X_i - t)/f(r_g)$, since the intervals are disjoint and the reference at zero is uncapped. The excess of this multiplier after subtracting one has integral at most $Cr_g/f(r_g)$ in t, by (17). In the variable $u = t/d_n$ this integral is at most $Cr_g/(d_n f(r_g))$.

For either centered ratio, let $M = (C_0 + 1)L$. On the rest of each ray beyond M the ratio decays from its boundary value at least exponentially at rate $n - 1$ in t. Indeed $(\log K_0)'(x) \le -1$ for $x > 0$ by its integral as a mixture of e^{-xu}, $u \ge 1$. Along each outer ray the omitted index is fixed. Combining this fact with (18), the integrated spike bound, and $M/d_n = O(n)$, gives for either likelihood ratio $Z_n^*(u)$, true or leave-one-out,

$$\int_{|u|>BL} (1 + |u|^p) Z_n^*(u)\, du \le \exp\{-[c_1 B - (p+2)]L\} \tag{19}$$

on the coarse good event for all large n, for some $c_1 > 0$ independent of B. The factor n^{p+1} bounds the weighted integration length within the range. The n spike multipliers add at most $Cn^{p+1}r_g/(d_n f(r_g))$, which is smaller than the same bound before multiplication by the cap supremum. On the exterior rays the additional weighted integral is bounded by its boundary value times Cn^p, since $(n-1)d_n$ tends to infinity.

Both normalizing integrals have a deterministic polynomial lower bound on the coarse good event. Near $t = 0$ the coarse cap has centered log value at least -1 on an interval of width $cn^{-(D+3)}$, by the same Lipschitz bound. The true ratio dominates this cap because zero has no capped observation. On each nearest-point cell the log leave-one-out criterion uses $n - 1$ terms at distances at least r_g, since the pair gaps exceed $2r_g$. It is continuous across cell boundaries and hence has the same Lipschitz bound. Thus both normalizers in u are at least $cn^{-(D+3)}$. Together with (19) and sufficiently large B, this makes both scaled likelihood means at most $BL + 1$ in absolute value on the coarse good event.

6.3 Fine likelihood integrals and moments

On $[-BL, BL]$ in the u coordinate use the finer cap $r = n^{-1}L^{-K}$ with K large. By the near-zero squared-density bound, the probability of two points at distance at most $2r$ within the relevant t-window is $O(L^{4-K})$. The probability of a capped point at the reference zero is $O(L^{1-K})$. The number of points in that window enlarged by $2r$ has expectation $O(L^3)$. Fix a sufficiently large $Q > b_1 + 4$. Markov bounds this count by L^Q away from a set of probability $O(L^{-b_1})$. There is at least one point in the window with complement probability $O(e^{-cL^3})$. Choose $K > Q + p + b_1 + 10$ and combine these estimates. Then the cap and leave-one-out centered log criteria differ uniformly by $O(\log L/L)$, as the capped omitted term lies between $\log f(r)$ and $\log f(2BL^2/n)$. The true ratio differs only by the spike multipliers above, with each excess integral now bounded in u by

$Cr/(d_n f(r)) = O(L^{-K-2})$. On any fixed compact interval the capped log process converges in the continuous topology to $\sqrt{\pi}\,\mathcal{B}(u) - \pi|u|/2$ by the proof in Section 2. In particular the cap supremum there is tight. The spike excesses contribute $o_{\mathrm{Pr}}(1)$ to each weighted integral on the compact interval, by the cap supremum and $L^Q L^{-K-2} = o(1)$. The leave-one-out comparison also gives convergence of its compact integrals to the same ones.

We next control tails on the fine window. The Bernstein shell estimate from Section 2, applied from a radius R through the fine endpoint, gives

$$C_r(d_n u) \le -c_2|u| \quad (R \le |u| \le BL)$$

except with probability $Ce^{-c_3 R} + O(L^{-b_1})$. The constants $c_2, c_3 > 0$ are independent of R. In fact the cap drift under its normalized reference is at most $-c|u|$ on this range by Hellinger affinity. On a shell of size R, the centered increments have variance bounded by $C|u - v|$ and single-row envelope $C \log L/L$. With $O(\log L)$ dyadic levels, the Bernstein bound for the shell is less than $cR/2$ at probability cost $O(e^{-c'R})$; its linear envelope contribution is $o(R)$ uniformly for $R \le BL$. Pay the common grid interpolation error once and sum over shells. The capped and true reference sampling laws differ by $O(L^{-K})$, absorbed above.

The fine-window weighted tail integrals of both ratios are therefore bounded by $Ce^{-c_2 R/2}$ on that event. For the capped integral integrate $(1+|u|^p)e^{-c_2|u|}$. The uniform deleted-term comparison covers the leave-one-out case. In the true likelihood, use the supremum bound restricted to the fine tail, at most $(1 + (BL)^p)e^{-c_2 R}$ before multiplying by the sum CL^{Q-K-2} of the spike excess integrals. Thus, using (19) for the coarse rays, the full tail integrals tend to zero in probability as $R \to \infty$ after $n \to \infty$. The Brownian likelihood also has integrable weighted tails almost surely, since its exponent has negative linear drift and Brownian motion is sublinear at infinity. The compact convergence, followed by this tail bound and normalization, gives joint convergence of the random probability measures in W_p. For example compact test integrals converge jointly by the continuous cap limit and the fine comparison. The tail estimate along with compact denominators bounded away from zero in probability extends the convergence to the p-th moments, which yields the W_p topology.

For moment convergence of the two means we also need a denominator bound with an explicit tail probability. On $[-1, 1]$ in u the capped drift is bounded, by its Taylor estimate and the variance bound. The same Bernstein chaining on this fixed interval gives for any small fixed $\epsilon > 0$, for all sufficiently large $R \le BL$,

$$\inf_{|u| \le 1} C_r(d_n u) \ge -\epsilon R$$

except with probability $Ce^{-c_\epsilon R} + O(L^{-b_1})$. Indeed setting the concentration parameter proportional to R in the chain yields a square-root term $O(\sqrt{R})$ and a linear term $o(R)$. Since the reference has no capped point, the raw ratio dominates the cap there, and the leave-one-out difference is $o(1)$. This lower bounds both denominators by $ce^{-\epsilon R}$. Take $\epsilon < c_2/4$ and also $\epsilon < c_1/4$, and then take B large compared to $p + 2$ and D. For $R \le BL$, division of both fine and coarse tail bounds by the denominators gives a vanishing bound on the contribution of $|u| > R$ to their absolute means. Thus the probability that either mean exceeds $2R$ in absolute value is at most

$$Ce^{-c_4 R} + O(L^{-b_1}) + O(n^{-b_0}) \qquad (R \le BL)$$

for R above a fixed constant.

On the coarse good event the scaled means are bounded by $BL + 1$ as proved above. For bad events, the unscaled means lie between the sample extrema deterministically. Indeed if $t > x_{\max}$, reflect it to $2x_{\max} - t$. Each distance to a data point decreases weakly. The full likelihood increases

weakly, and so does the leave-one-out weight, since that weight is the product of the smallest $n-1$ factors. Pairing the integrals across $x_{\max}$ shows that either mean is at most $x_{\max}$. Reflection at $x_{\min}$ gives the other bound. Exponential sample tails imply $\mathbb{E}(\max_i |X_i|)^{2a} = O(L^{2a})$ for any fixed $a > 0$. Choose b_0, b_1 arbitrarily large for the power being integrated. Cauchy–Schwarz pays the coarse exceptional event, and the preceding exponential bound up to BL pays the other scaled moment tails. This establishes uniform integrability of every fixed power and proves convergence of the moments of both averages. The scaled difference tends to zero in probability by their common joint limit, and the same moment bounds give convergence in each fixed L^q.

6.4 Minimax identity and constant

For each fixed n, the ordinary likelihood mean is translation equivariant and has constant risk. To see the exact minimax identity without an asymptotic lower-bound approximation, write $w = (X_2 - X_1, \dots, X_n - X_1)$. At zero location let μ_w denote the conditional law of X_1 given w. Among equivariant estimates of the form $X_1 - a(w)$, the minimum conditional squared error is $\mathrm{Var}(\mu_w)$, attained by $a(w) = \int z\, \mu_w(dz)$. Changing variable in the flat-prior posterior integral shows that its mean is precisely this estimator.

For a minimax lower bound, put a uniform prior on $[-A, A]$ and fix $M < A$. In this model the marginal of w does not change. Given w, write the first observation as $x_1 = \theta + z$ with $z \sim \mu_w$ independent of the prior. The contribution for each x_1 to the integrated Bayes risk, with the common density factor $1/(2A)$ removed, is $\inf_a \int_{x_1-A}^{x_1+A} (z-a)^2\, \mu_w(dz)$. For $x_1 \in [-A+M, A-M]$, the integration interval contains $[-M, M]$. The Bayes risk is therefore at least

$$(1 - M/A)\mathbb{E}_w \inf_a \int_{-M}^{M} (z-a)^2\, \mu_w(dz).$$

First let $A \to \infty$, then $M \to \infty$. The restricted minima increase to the full conditional variance, by the conditional mass and moment formulas. This proves the exact minimax property of the true posterior mean.

By Brownian scaling, $I \int u\, \Pi(du)$ has the canonical Pitman law whose second moment is $16\zeta(3)$ [8, 7]. The moment convergence established above gives both limiting risks and, using the exact minimax identity, the sharp minimax constant. The corresponding maximizer has risk $26/I^2$ by Section 2. Rescaling the data gives $I = \pi/\sigma$ and completes Theorem 6.1.

7 Location inference as variance-gamma shape approaches one

When variance-gamma shape approaches one with sample size, location information changes on different shape scales in the plane and in higher dimensions. In dimensions at least three, deviations of order $1/\log n$ change the Gaussian information by a continuous factor. In the plane the transition from below occurs when the shape deficit times $\log n$ reaches $\log\log\log n$. We derive the local experiments and show that the location constants remain attainable with unknown shape, skew and scatter.

7.1 Results

In fixed dimension $d \geq 2$, observe n independent copies of

$$X = \theta + \gamma G + \sqrt{G}\, S^{1/2} V, \qquad G \sim \mathrm{Gamma}(\nu, 1), \quad V \sim N(0, I_d),$$

where G, V are independent and S is positive definite. Write $\eta = (\nu, \gamma, S)$. The reference sequence is $\theta_0, \eta_n = (\nu_n, \gamma_n, S_n)$ with $\nu_n \to 1, \gamma_n \to \gamma_0, S_n \to S_0$. Use compact parameter brackets, with eigenvalues of S uniformly bounded away from zero and the limiting parameter in the interior.

Theorem 7.1. *Define the normalization $d_n = (n\Lambda_n)^{-1/2}$ and matrix $\mathcal{K}$ in the following three cases.*

1. *For $d \geq 3$ and $(\nu_n - 1)\log n \to \tau \in \mathbb{R}$,*

$$\Lambda_n = \log n, \qquad \mathcal{K} = \frac{d-2}{d}\frac{1-e^{-\tau}}{\tau} S_0^{-1},$$

   with ratio one at $\tau = 0$.

2. *For $d = 2$, $\nu_n \leq 1$, and $\exp((1-\nu_n)\log n)/\log\log n \to \lambda \in [0, \infty)$,*

$$\Lambda_n = \log\log n, \qquad \mathcal{K} = (1+\lambda) S_0^{-1}.$$

   In particular $(1-\nu_n)\log n - \log\log\log n \to \kappa$ gives $\lambda = e^{\kappa}$.

3. *For $d = 2$ and $\nu_n \geq 1$, set*

$$\Lambda_n = \log \min\{\log n,\, 1/(\nu_n - 1)\}, \qquad \mathcal{K} = S_0^{-1},$$

   with $1/0 = \infty$.

On compact sets of u, h, the experiments at $\theta = \theta_0 + d_n u$, $\eta = \eta_n + h/\sqrt{n}$ converge in Le Cam distance to independent Gaussian shifts with location information $\mathcal{K}$ and ordinary nuisance information at $(1, \gamma_0, S_0)$. There is a common estimator using unknown η, independent of the choice among these regimes, with

$$d_n^{-1}(\widehat{\theta} - \theta) \Longrightarrow N(0, \mathcal{K}^{-1}), \qquad d_n^{-2}\mathbb{E}|\widehat{\theta} - \theta|^2 \longrightarrow \operatorname{tr}(\mathcal{K}^{-1}).$$

The risk limit is the sharp local asymptotic minimax squared-risk constant. The conclusions for the estimator hold as well under the stated bounded local perturbations.

Nitithumbundit and Chan [3] formulate the multivariate model and study location numerically in one dimension with shape fixed at each simulation setting. The fixed-shape experiments and the adaptation result at base shape one are given in Sections 4, 5. The phenomenon here is caused by the reference shape moving with n. Bakirov [27] studies fixed scalar location families with varying Hellinger behavior. We calculate the information of the moving multivariate density through its radial integral, including the planar transition where two terms of that integral have comparable sizes.

7.2 Annular information

Let $L = (S/2)^{1/2}$, $y = L^{-1}x$, $r = |y|$ and $\beta = L^{-1}\gamma$. The density at $\theta = 0$ is

$$f_\eta(x) = \frac{|\det L|^{-1} e^{\beta \cdot y/2}}{(4\pi)^{d/2}\Gamma(\nu)} \int_0^\infty s^{\nu - d/2 - 1} e^{-cs - r^2/(4s)}\, ds, \qquad c = 1 + |\beta|^2/4.$$

Let $\ell_\eta = -\nabla \log f_\eta$ and $s_\eta = \nabla_\eta \log f_\eta$. Near the pole and uniformly on compact brackets around $\nu = 1$,

$$|\ell_\eta| \leq C/|x|, \quad \|D_x^2 f_\eta\|/f_\eta \leq C/|x|^2, \quad |s_\eta| \leq C(1 + \log(1/|x|)), \quad \|D_\eta \ell_\eta\| \leq C(1 + \log(1/|x|))/|x|. \tag{20}$$

For example, each extra s^{-j} in the integral costs at most $C_j r^{-2j}$, by splitting at r^2 and using $r^2/(4s)$ below that cutoff. Each nuisance derivative in shape introduces a log of s, which costs at most $C(1+\log(1/r))$ by the same split and the exponential tail in s. The scatter derivative of y is $O(|y|)$. These estimates, detailed for the same parameter brackets in Section 5, give (20). They also imply $\|D_x \ell_\eta\| \le C/|x|^2$. On outer rays the log spatial derivative is uniformly bounded, by the Bessel tail representation. The density tails are uniformly exponential.

We use the deterministic cutoff

$$\Delta = n^{-1/2}(\log n)^2, \qquad A = \{x : \Delta \le |x| \le r_0\}, \qquad J_n(\eta) = \int_A \ell_\eta(x)\ell_\eta(x)^{\mathsf{T}} f_\eta(x)\, dx,$$

where r_0 is fixed and small. Uniformly for ν near one, the near-pole density is bounded by $C|x|^{2-d-\epsilon}(1+\log(1/|x|))$, for any fixed small $\epsilon > 0$ if the shape bracket is narrowed. In particular the weighted first score bounds in (20) are integrable at the pole, including their product $|\ell_\eta||s_\eta|$. The tilt $e^{\beta\cdot y/2}$ is $1+O(r)$, and adds a bounded term to the log gradient. Thus removing the tilt changes the annular information by $O(1)$, by these integrability estimates. The angular average in y coordinates is I_d/d. The radial cutoffs obtained from the ellipsoidal image of A differ by fixed multiplicative factors from Δ, r_0. In the asymptotics below such a change leaves the leading coefficient unchanged.

For $d \ge 3$, switching to $r^2/(4s)$ gives the leading radial density $A_{\nu,d} r^{2\nu-d}$ with $A_{\nu,d} = \Gamma(d/2-\nu)/(4^\nu \pi^{d/2}\Gamma(\nu))$. Its radial log derivative is asymptotic to $(2\nu-d)/r$. The convergence as $r \to 0$ is uniform for ν in a small fixed neighborhood of one, by dominated convergence in the transformed integrals, uniformly over c in compact brackets. The coefficient of the radial information integral, after angular integration, therefore tends to $(d-2)/d$ per whitened coordinate. For $s = \log(1/r)$, the remaining integral is $\int e^{-2(\nu_n-1)s}\, ds$ between a fixed lower limit and

$$T_n = \log(1/\Delta) = \tfrac{1}{2}\log n - 2\log\log n.$$

Its ratio to $\log n$ tends to $\frac{1}{2}(1-e^{-\tau})/\tau$. The bounded change of T_n under the ellipsoidal cut has $O(1)$ cost. Conjugation by L^{-1} gives the matrix in the first case of the theorem.

For $d = 2$, put $a = 1-\nu$ and $w = r^{-2a}$. The radial integral in the density has the uniform forms

$$B(a,r) = (w-1)/a + O(1+w), \tag{21}$$

$$-r\partial_r B(a,r) = 2\,4^a \Gamma(1+a) w[1+O(r^\kappa)] \tag{22}$$

for some fixed $\kappa > 0$ and all sufficiently small $|a|$. At $a = 0$ the quotient is $2\log(1/r)$. To prove (21), integrate s^{-a-1} between r^2 and one. The exponential factors change that integral by $O(1+w)$, using $|1-e^{-cs-r^2/(4s)}| \le C(s+r^2/s)$. Below r^2, substitution $r^2/(4s)$ gives $O(w)$; above one the bound is constant. The differentiated integral with the same substitution gives (22). Its remaining integrand is proportional to $z^a e^{-z} e^{-cr^2/(4z)}$. Splitting the difference from the gamma integral at $z = r$ gives a power remainder uniformly for $|a|$ small.

Writing $t = \log(1/r)$, the relative error in (21) is $O(|a|+1/t)$, because $|a|(1+w)/|w-1| \le |a|+1/t$. The density prefactor tends to $1/(4\pi)$ in y space. Consequently the leading radial information per whitened coordinate is the integral of $aw(t)^2/(w(t)-1)$, where $w(t) = e^{2at}$. If $a \ne 0$ its primitive is

$$\tfrac{1}{2}\{w(t) + \log|w(t)-1|\}.$$

At $a = 0$ it is $\frac{1}{2}\log t$. Indeed the angular factor is π and the squared radial derivative contributes $4w^2$. To justify retaining these leading terms, first start the integral at an arbitrarily large fixed t_0.

The relative error is then $O(1/t_0) + o(1)$ uniformly, while the discarded bounded interval has $O(1)$ information.

If $a_n \geq 0$ and the ratio defining λ is bounded, then $a_n \log n = O(1 + \log\log\log n)$. At T_n, $w(T_n)/\log\log n \to \lambda$. For $a_n > 0$, the logarithmic difference in the primitive is

$$\log\left(\frac{T_n}{t_0}\right) + O(1 + a_n T_n) \sim \log\log n,$$

by writing $w(t) - 1 = 2a_n t(e^{2a_n t} - 1)/(2a_n t)$. At $a_n = 0$ the same conclusion follows directly. Thus the leading whitened coefficient divided by $\log\log n$ is $(1 + \lambda)/2$.

If $a_n \leq 0$, put $b_n = -a_n$. The w difference is bounded. The logarithmic difference equals $\log\min(T_n, 1/b_n) + O(1)$, since $1 - e^{-2b_n t}$ is comparable to $\min(b_n t, 1)$. The coefficient divided by the specified Λ_n is $1/2$. Moving the cutoff by a bounded amount in t affects neither case: the primitive's derivative is bounded by $Cw(t)(a_+ + 1/t)$, with $a_+ = \max(a, 0)$, and this is $O(1)$ near T_n in the stated regimes. Finally conjugate by L^{-1}. Together these computations prove

$$J_n(\eta_n)/\Lambda_n \longrightarrow \mathcal{K}. \tag{23}$$

The same computations apply after any bounded root-n nuisance perturbation. In the last case, if ν_n is within $O(n^{-1/2})$ of one, either side has the same leading $\log\log n$ coefficient. Otherwise the clipping scale $\min(\log n, 1/b_n)$ has the same leading logarithm under the perturbation.

7.3 The local experiments

Center at $\theta_0 = 0$. We can replace a local location shift by zero outside A, preserving the nuisance density everywhere, at total variation cost $o(1)$ in the product. We first bound the omitted shift cost at η_n, uniformly for bounded u, with $D = d_n$. For $d \geq 3$, the inner ball $|x| \leq C_0 D$ costs $O(D^{2\nu_n}) = O(D^2)$ in squared Hellinger distance, by the two local masses. From $C_0 D$ to Δ the cost is bounded by $CD^2 \int_{C_0 D}^{\Delta} r^{2\nu_n - 3} dr$. The exponent factor $r^{2(\nu_n - 1)}$ is uniformly bounded there and the logarithmic interval has length $O(\log\log n)$. The product costs therefore tend to zero.

The inner ball in the plane requires using the difference of the two densities. Put $T_D = \log(1/D)$, $w_D = D^{-2a_n}$, and $b_D = (w_D - 1)/a_n$, with $b_D = 2T_D$ at zero. On $|x| \leq C_0 D$ both untilted radial density factors are bounded below by a positive constant times b_D, using (21) and monotonicity. Set $x = Dz$ and let ρ_1, ρ_2 be the whitened distances of z to the two shifted pole centers. By (22), the radial difference is bounded by

$$Cw_D(1 + \rho_1^{-2|a_n|} + \rho_2^{-2|a_n|})(1 + |\log\rho_1| + |\log\rho_2|).$$

This follows by integrating the radial derivative between the two radii. The squared integral of this bound after removal of w_D is uniformly bounded on the mark ball for $|a_n|$ small, for instance by Cauchy–Schwarz and integrability of a fixed small negative power times logarithms at either pole. The tilted factors change by relative $O(D)$ and contribute at most $O(D^2)$ times the inner mass, which is negligible in the product. The remaining squared Hellinger bound is

$$CD^2 w_D^2 / b_D.$$

For $a_n \leq 0$, $w_D \leq 1$ and $b_D \asymp \min(T_D, 1/|a_n|) \to \infty$. For $a_n \geq 0$ in the stated regime, $w_D = O(\Lambda_n)$ and $w_D^2/b_D \leq Cw_D(a_n + 1/T_D)$. In both cases multiplying the bound by n gives $o(1)$.

In the planar annulus $C_0 D \leq |x| \leq \Delta$, integrate the derivative of the square root against a displacement of size $O(D)$. In $t = \log(1/r)$ the squared derivative integral per coordinate is bounded

by $Cw(t)((a_n)_+ + 1/t)$, by (21) and (22). All t in this range differ from $\frac{1}{2}\log n$ by $O(\log\log n)$, and $w(t) \le C\Lambda_n$ for $a_n \ge 0$ and $w(t) \le 1$ for $a_n \le 0$. Hence the product cost is at most

$$C(\log\log n)((a_n)_+ + 1/\log n) = o(1).$$

This proves the omitted inner bounds in the plane. The outer bound for all dimensions is $O(nD^2) = o(1)$, by the log derivative bound and the tail probability.

Replacing η_n by $\eta_n + h/\sqrt{n}$ in the inner estimates preserves their $o(1)$ product order. Indeed the restricted squared Hellinger cost of this nuisance change, on a ball of radius 3Δ and on its small translate, is bounded by C/n times the integral of

$$|x|^{2-d-\epsilon}(1 + \log(1/|x|))^3$$

over the ball, by (20) along the nuisance segment. This integral tends to zero. The outer shift estimate is already uniform in the nuisance. Normalize the spliced single-observation density. The reverse triangle inequality for square-root norms shows that normalization costs at most a fixed multiple of the same Hellinger bound.

Now write $f_n = f_{\eta_n}$ and $f_h = f_{\eta_n + h/\sqrt{n}}$. The likelihood contribution of the splice relative to f_h is $(1 + k)/(1 + \mathbb{E}_h k)$, with $k = \mathbf{1}_A(f_h(\cdot - Du)/f_h - 1)$. Uniformly on the retained annulus, $|k| \le CD/\Delta = o(1)$. There and on slightly enlarged annuli the reference densities are bounded by

$$CM_n|x|^{2-d}, \qquad M_n = (1 + \log n)(1 + \log\log n).$$

In higher dimensions this follows from uniform power bounds and bounded $(\nu_n - 1)\log n$. In the plane it follows from (21), since $w = O(1 + \log\log n)$ when $a_n \ge 0$ and the density is $O(1 + \log n)$ when $a_n \le 0$. The same bound holds after bounded root-n nuisance changes.

As in the fixed-base splice in Section 5, Taylor expansion and (20) give

$$n \int |k - Du^{\mathsf{T}} \ell_{\eta_n} \mathbf{1}_A|^2 f_n = o(1).$$

Specifically, the squared shift remainder is bounded after integration by $CnD^4 M_n/\Delta^2 = o(1)$. The nuisance score substitution is bounded by $CD^2 M_n(1 + \log n)^3 = o(1)$, using the $O(n^{-1/2}(1 + \log(1/r))/r)$ score change. Also $\mathbb{E}_h k = O(D)$ and $n|\mathbb{E}_h k - \mathbb{E}_n k| = O(\sqrt{n}D) = o(1)$, by the integrable product of $1/r$ and the nuisance score. Together with (23), this identifies the location log likelihood as its centered annular score and the drift $-u^{\mathsf{T}}\mathcal{K}u/2$. The quadratic term concentrates by the vanishing envelope, and higher log terms vanish by the same bound.

The nuisance likelihood is regular $\sqrt{n}$-local asymptotic normality also along η_n. One way to verify the needed uniform quadratic mean expansion is to use the gamma-normal complete-data density. Its derivatives in η give a continuous L^2 square-root derivative on compact brackets, and the corresponding scores have uniformly bounded moments of order $2 + \epsilon$. The gamma shape score involves only $\log G$, and normal conditional scores involve V polynomials and at most a factor $\sqrt{G}$. Marginalizing preserves the expansion with the conditional-expectation score [22, Chapter 7] and Section 5. Its nuisance information tends to the ordinary positive definite information at the limiting point. The location and nuisance central sequences have cross covariance of order $O(\sqrt{n}D) = o(1)$, since $\ell_\eta s_\eta^{\mathsf{T}}$ is uniformly integrable near the pole. Joint triangular normal convergence gives the product shift. Finite likelihood vectors converge with mean-one limit coordinates, so truncation and quantization give finite Le Cam convergence.

The omission estimates were uniform for bounded parameters. On the annulus the spliced location squared Hellinger distance at fixed nuisance is at most $C|u - u'|^2/n$: integrate the shifted

square-root derivative along the segment and use annular information $O(\Lambda_n)$ on slightly enlarged cutoffs, as in Section 7.2. Normalization costs a fixed factor. Variation in nuisance at fixed location uses uniform complete-data quadratic mean bounds at the root-n scale. Thus a finite net in a compact set approximates the whole experiment in both directions, by these Hellinger moduli, the omitted $o(1)$ cost and continuity of the Gaussian product. This proves the compact convergence.

7.4 An attaining estimator

Use the cumulant pilot of Section 5 on one half of the data, recovering θ, ν, γ, S from the means, covariance, coordinate third cumulants and one fourth cumulant. The inverse is uniformly Lipschitz on the compact brackets. Estimate raw moments truncated at $C \log n$ and fit the model moments over the brackets as in that construction. Thus the pilot $(\widetilde{\theta}, \widehat{\eta})$ has squared and fourth power errors $O(n^{-1})$ and $O(n^{-2})$. The event $|\widetilde{\theta} - \theta| + |\widehat{\eta} - \eta| \le \Delta/4$ has superpolynomially small failure, by Bernstein with the bounded moment envelope $O((\log n)^4)$.

On the other half take the one-step estimate

$$\widetilde{\theta} + J_n(\widehat{\eta})^{-1} \left\{ \frac{1}{m} \sum_{\text{this half}} \ell_{\widehat{\eta}}(X_i - \widetilde{\theta}) \mathbf{1}_A(X_i - \widetilde{\theta}) - \int_A \ell_{\widehat{\eta}} f_{\widehat{\eta}} \right\}.$$

Here m is the number of observations on that half. Reverse the two halves, and average the two one-step estimates weighted by their score-sample sizes. This estimator uses the same deterministic annulus in every regime and never requires Λ_n as an input.

We give the L^2 linearization for the average. Write $e = \theta - \widetilde{\theta}$ for a given pilot. On its good event, the mean of the centered fitted score on the other half is $J_n(\widehat{\eta})e$ with error

$$O(|e|^2 M_n/\Delta + |\widehat{\eta} - \eta|).$$

For location, use the Hessian bound and integrate $f_{\widehat{\eta}}(x)/|x|^3$ on the annulus. For nuisance, integrate $f_{\widehat{\eta}}(x)(1 + \log(1/|x|))/|x|$, which is uniformly bounded. The resulting normalized squared bias is $O(M_n^2/(\Lambda_n \log^4 n) + 1/\Lambda_n) = o(1)$.

For the fluctuation, compare the fitted score including its shifted cutoff to the true annular score. On the overlap of their annuli, (20) gives the conditional squared $L^2(f_\eta)$ bound

$$C M_n \left[|e|^2/\Delta^2 + |\widehat{\eta} - \eta|^2 (1 + \log n)^3 \right].$$

The cutoff differences are confined to two shells of thickness $O(|e|)$. At the inner radius their squared contribution is bounded by $C M_n |e|/\Delta$ and at the outer radius by $C M_n |e|$. The expectation of all these bounds on the good event is $o(1)$, since $M_n = O(\log n \log\log n)$, $n\Delta^2 = (\log n)^4$, and $\sqrt{n}\Delta = (\log n)^2$. They also apply to the centered weights. Moreover $J_n(\widehat{\eta})$ and $J_n(\eta)$ differ by $o(\Lambda_n)$ uniformly on the good event: the nuisance score difference has squared bound $O(M_n \Delta^2 (1 + \log n)^3) = o(1)$ and the annular density ratio tends uniformly to one. The inverse matrices therefore differ by $o(1/\Lambda_n)$.

On the bad event the fitted score is bounded by C/Δ. The inverse matrix is bounded uniformly on the fixed parameter brackets, by the strictly positive score information on a fixed subannulus. Indeed in whitened coordinates the radial derivative is strictly negative, so opposing radial directions cannot both have zero score in a fixed direction after adding the tilt. Compactness gives a uniform positive lower eigenvalue. The pilot's compact bounds and its superpolynomial failure probability remove the bad-event contribution.

Conditioning each fold on its independent pilot now yields the full-sample representation

$$\widehat{\theta} - \theta = J_n(\eta)^{-1} \frac{1}{n} \sum_{i=1}^{n} \left[\ell_\eta(X_i - \theta) \mathbf{1}_A(X_i - \theta) - \int_A \ell_\eta f_\eta \right] + o_{L^2}(d_n).$$

This argument uses the two conditional variance bounds separately, so it does not require independent fold errors. The true centered score has covariance $J_n(\eta) - O(1)$ and a vanishing triangular envelope after scaling the likelihood rows by d_n. Equation (23) gives the Gaussian limit and the squared-risk limit of the displayed full-sample mean. The squared-risk lower constant follows by restricting the compact local experiment to fixed nuisance. Project decisions to local balls, transfer bounded squared loss to the Gaussian shift, and let truncated diffuse Gaussian priors fill expanding balls as in Section 4. Their Bayes risks tend to $\mathrm{tr}(\mathcal{K}^{-1})$. This proves Theorem 7.1.

Statements and Declarations

Funding. This research was self-funded.

Competing interests. The author owns Luminary Oddities and has no other relevant competing interests.

Data and code availability. The verification scripts for Section 3 and their outputs are included with this article.

AI assistance. OpenAI Codex assisted with derivations, proof development, source checks and independent internal review of the parts of editions 1–3. Anthropic Claude edited, consolidated and cross-referenced the parts, checked computations, and assisted with the derivations, proofs, computations and source checks of Section 3 in edition 4. Scott Schubr directed the research and is responsible for the content.

References

[1] K. Podgórski and J. Wallin, Maximizing leave-one-out likelihood for the location parameter of unbounded densities, Ann. Inst. Stat. Math. 67 (2015), 19–38. DOI.

[2] T. Nitithumbundit and J. S. K. Chan, Maximum leave-one-out likelihood estimation for location parameter of unbounded densities, arXiv:1602.01213 (2016). `https://arxiv.org/abs/1602.01213`.

[3] T. Nitithumbundit and J. S. K. Chan, Maximum leave-one-out likelihood method for the location parameter of variance gamma distribution with unbounded density, J. Stat. Comput. Simul. 93 (2023), 2642–2671. DOI.

[4] T. Polfeldt, The order of the minimum variance in a non-regular case, Ann. Math. Statist. 41 (1970), 667–672. DOI.

[5] T. Polfeldt, Minimum variance order when estimating the location of an irregularity in the density, Ann. Math. Statist. 41 (1970), 673–679. DOI.

[6] A. S. Terent'yev, Probability distribution of a time location of an absolute maximum at the output of a synchronized filter, *Radioengineering and Electronics* 13 (1968), no. 4, 652–657.

[7] S. Dachian, On limiting likelihood ratio processes of some change-point type statistical models, J. Statist. Plann. Inference 140 (2010), 2682–2692. Author manuscript.

[8] H. Rubin and K.-S. Song, Exact computation of the asymptotic efficiency of maximum likelihood estimators of a discontinuous signal in a Gaussian white noise, Ann. Statist. 23 (1995), 732–739. DOI.

[9] E. J. G. Pitman, The estimation of the location and scale parameters of a continuous population of any given form, Biometrika 30 (1939), 391–421. DOI.

[10] M. A. Girshick and L. J. Savage, Bayes and minimax estimates for quadratic loss functions, *Proc. Second Berkeley Symp. Math. Statist. Probab.*, University of California Press (1951), 53–73.

[11] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, Springer, New York (1981). https://doi.org/10.1007/978-1-4899-0027-2.

[12] S. Schubr, *Calibrating small noise from exteriors, frontiers, cusps and density singularities*, Luminary Research Archive, edition 1 (2026), Section 6, *Calibrating a noise endpoint from a logarithmic density singularity.* https://research.luminaryoddities.org/calibrating-small-noise-exteriors-frontiers-cusps-singularities/v/1, https://doi.org/10.5281/zenodo.24444424.

[13] NIST Digital Library of Mathematical Functions, Chapter 10, Sections 10.25–10.40. https://dlmf.nist.gov/10.

[14] I. A. Ibragimov and R. Z. Khas'minskii, Asymptotic behavior of statistical estimates of the shift parameter for samples with unbounded density, J. Soviet Math. 16 (1981), 1035–1041. Original paper.

[15] B. L. S. Prakasa Rao, Estimation of the location of the cusp of a continuous density, Ann. Math. Statist. 39 (1968), 76–87. DOI.

[16] G. Last and M. Penrose, *Lectures on the Poisson Process*, Cambridge University Press (2017). https://doi.org/10.1017/9781316104477.

[17] L. Le Cam, An approximation theorem for the Poisson binomial distribution, Pacific J. Math. 10 (1960), 1181–1197. DOI.

[18] A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer, New York (1996). https://doi.org/10.1007/978-1-4757-2545-2.

[19] J. Kim and D. Pollard, Cube root asymptotics, Ann. Statist. 18 (1990), 191–219. DOI.

[20] S. Dachian, Estimation of the location of a 0-type or ∞-type singularity by Poisson observations, Statistics 45 (2011), 509–523. Author manuscript.

[21] S. Gupta, J. C. H. Lee and E. Price, High-dimensional location estimation via norm concentration for subgamma vectors, Proc. Mach. Learn. Res. 202 (2023), 12132–12164. Proceedings.

[22] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press (1998). https://doi.org/10.1017/CBO9780511802256.

[23] T. Nitithumbundit and J. S. K. Chan, An ECM algorithm for skewed multivariate variance gamma distribution in normal mean-variance representation. Author manuscript.

[24] E. Seneta, Fitting the variance-gamma model to financial data, J. Appl. Probab. 41A (2004), 177–187. DOI.

[25] A. Fischer, R. E. Gaunt and A. Sarantsev, Modified method of moments for generalized Laplace distributions, Commun. Stat. Simul. Comput. 54 (2025), 2047–2064. Accepted author manuscript.

[26] H. Strasser, Local asymptotic minimax properties of Pitman estimates, Z. Wahrscheinlichkeitstheorie verw. Gebiete 60 (1982), 223–247. Author copy.

[27] N. K. Bakirov, The likelihood ratio for densities with singularities, J. Soviet Math. 34 (1986), 1403–1413. Original paper.